Guidelines for Preparing Economic Analyses

April 3, 2020

Review Copy Prepared for
EPA’s Science Advisory Board’s
Economic Guidelines Review Panel
# Table of Contents

- **Glossary** .................................................................................................................................................. i-8
- **Chapter 1 Introduction** ............................................................................................................................. 1-1
  - 1.1 Background ........................................................................................................................................... 1-1
  - 1.2 The Scope of the *Guidelines* ............................................................................................................... 1-1
  - 1.3 Economic Framework for Analysis ...................................................................................................... 1-2
    - 1.3.1 Assessing Economic Efficiency with Benefit-Cost Analysis (BCA) .............................................. 1-3
    - 1.3.2 Assessing Economic and Distributional Impacts .......................................................................... 1-4
- **Chapter 1 References** .................................................................................................................................. 1-5
- **Chapter 2 Executive Order and Statutory Requirements for Conducting Economic Analyses** ............. 2-1
  - 2.1 Executive Orders .................................................................................................................................. 2-3
    - 2.1.1 Executive Order 12866, "Regulatory Planning and Review" ......................................................... 2-3
    - 2.1.2 Executive Order 12898, "Federal Actions to Address Environmental Justice in Minority Populations and Low-Income Populations" ................................................................. 2-5
    - 2.1.3 Executive Order 13045, “Protection of Children from Environmental Health Risks and Safety Risks” ........................................................................................................................................ 2-5
    - 2.1.4 Executive Order 13132, “Federalism” ............................................................................................ 2-6
    - 2.1.5 Executive Order 13175, “Consultation and Coordination with Indian Tribal Governments”. 2-6
    - 2.1.6 Executive Order 13211, “Actions Concerning Regulations that Significantly Affect Energy Supply, Distribution, or Use” ................................................................................... 2-7
    - 2.1.7 Executive Order 13563, "Improving Regulation and Regulatory Review" ............................... 2-7
    - 2.1.8 Executive Order 13707, "Using Behavioral Science Insights to Better Serve the American People" ........................................................................................................................................ 2-7
    - 2.1.9 Executive Order 13771, "Reducing Regulation and Controlling Regulatory Costs" ............... 2-8
    - 2.1.10 Executive Order 13777, "Enforcing the Regulatory Reform Agenda" ...................................... 2-8
    - 2.1.11 Executive Order 13783, "Promoting Energy Independence and Economic Growth" (Sec. 1 & 2) ........................................................................................................................................ 2-9
    - 2.1.12 Executive Order 13891, "Promoting the Rule of Law Through Improved Agency Guidance Documents" ........................................................................................................................................ 2-9
2.2 Statutes ....................................................................................................................... 2-9
  2.2.1 Regulatory Flexibility Act of 1980 (RFA), as Amended by The Small Business Regulatory
  2.2.2 Unfunded Mandates Reform Act of 1995 (UMRA) (P.L. 104-4) ........................................ 2-10
  2.2.3 The Paperwork Reduction Act of 1995 (PRA) (44 U.S.C. 3501) ........................................ 2-10
  Chapter 2 References ....................................................................................................... 2-11

Chapter 3: Need for Regulatory Action and Evaluation of Policy Options .................................. 3-1
  3.1 The Statement of Need ................................................................................................. 3-1
    3.1.1 Problem Description ............................................................................................... 3-1
    3.1.2 Reasons for Market Failure .................................................................................... 3-2
    3.1.3 Need for Federal Action .......................................................................................... 3-4
  3.2 General Guidance on Policy Options to Evaluate ............................................................ 3-4
  Chapter 3 References ....................................................................................................... 3-7

Chapter 4: Regulatory and Non-Regulatory Approaches to Pollution Control ............................... 4-1
  4.1 Evaluating Environmental Policy .................................................................................... 4-1
    4.1.1 Economic Efficiency .............................................................................................. 4-1
    4.1.2 Cost-Effectiveness ............................................................................................... 4-2
  4.2 Traditional Command-and-Control or Prescriptive Regulation .......................................... 4-2
    4.2.1 Technology or Design Standards .......................................................................... 4-3
    4.2.2 Performance-based Standards .............................................................................. 4-4
  4.3 Market-Based Approaches ............................................................................................. 4-4
    4.3.1 Allowance Trading Systems ................................................................................... 4-5
    4.3.2 Emissions Tax .................................................................................................... 4-10
    4.3.3 Environmental Subsidies ..................................................................................... 4-10
    4.3.4 Tax-Subsidy Combinations ................................................................................. 4-12
  4.4 Hybrid and Other Approaches ......................................................................................... 4-13
    4.4.1 Combining Prescriptive and Market-Based Approaches .......................................... 4-13
    4.4.2 Information Disclosure ....................................................................................... 4-14
    4.4.3 Liability Rules ..................................................................................................... 4-15
    4.4.4 Behavioral Economics and "Nudge" Approaches ...................................................... 4-16
  4.5 Voluntary Initiatives ..................................................................................................... 4-18
<table>
<thead>
<tr>
<th>Chapter 4 References</th>
<th>4-25</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chapter 5: Setting the Foundation: Scope, Baseline, and Other Analytic Design</td>
<td>5-1</td>
</tr>
<tr>
<td>Considerations</td>
<td></td>
</tr>
<tr>
<td>5.1 Scope of Analysis</td>
<td>5-1</td>
</tr>
<tr>
<td>5.1.1 Standing</td>
<td>5-1</td>
</tr>
<tr>
<td>5.1.2 Market Effects</td>
<td>5-2</td>
</tr>
<tr>
<td>5.1.3 Externalities</td>
<td>5-3</td>
</tr>
<tr>
<td>5.2 Baseline</td>
<td>5-3</td>
</tr>
<tr>
<td>5.2.1 Baseline Definition</td>
<td>5-3</td>
</tr>
<tr>
<td>5.2.2 Guiding Principles of Baseline Specification</td>
<td>5-5</td>
</tr>
<tr>
<td>5.2.3 Multiple Baselines</td>
<td>5-8</td>
</tr>
<tr>
<td>5.3 Multiple Rules</td>
<td>5-8</td>
</tr>
<tr>
<td>5.3.1 Linked Rules</td>
<td>5-9</td>
</tr>
<tr>
<td>5.3.2 Unlinked Rules</td>
<td>5-9</td>
</tr>
<tr>
<td>5.3.3 Accounting for Benefits and Costs that Accrue Across Multiple Rules</td>
<td>5-11</td>
</tr>
<tr>
<td>5.4. Time Horizon of Analysis</td>
<td>5-12</td>
</tr>
<tr>
<td>5.5 Representing Economic Behavior</td>
<td>5-13</td>
</tr>
<tr>
<td>5.5.1 Behavioral Response</td>
<td>5-13</td>
</tr>
<tr>
<td>5.5.2 Potential for Cost Savings</td>
<td>5-14</td>
</tr>
<tr>
<td>5.5.3 Technological Change</td>
<td>5-16</td>
</tr>
<tr>
<td>5.5.4 Compliance</td>
<td>5-16</td>
</tr>
<tr>
<td>5.5.5 Voluntary Actions</td>
<td>5-18</td>
</tr>
<tr>
<td>5.5.6 Changes in Other Environmental Contaminants</td>
<td>5-18</td>
</tr>
<tr>
<td>Section</td>
<td>Page</td>
</tr>
<tr>
<td>------------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>5.6 Uncertainty</td>
<td>5-19</td>
</tr>
<tr>
<td>5.6.1 Performing Sensitivity Analysis</td>
<td>5-21</td>
</tr>
<tr>
<td>5.6.2 Other Considerations Related to Uncertainty and Risk</td>
<td>5-21</td>
</tr>
<tr>
<td>Chapter 5 References</td>
<td>5-22</td>
</tr>
<tr>
<td>Chapter 6: Discounting Future Benefits and Costs</td>
<td>6-1</td>
</tr>
<tr>
<td>6.1 Mechanics and Methods for Discounting</td>
<td>6-2</td>
</tr>
<tr>
<td>6.1.1 Net Present Value (NPV)</td>
<td>6-2</td>
</tr>
<tr>
<td>6.1.2 Annualized Values</td>
<td>6-4</td>
</tr>
<tr>
<td>6.1.3 Net Future Value (NFV)</td>
<td>6-5</td>
</tr>
<tr>
<td>6.1.4 Comparing the Methods</td>
<td>6-5</td>
</tr>
<tr>
<td>6.1.5 Sensitivity of Net Present Value Estimates to the Discount Rate</td>
<td>6-6</td>
</tr>
<tr>
<td>6.1.6 Issues in Discounting Applications</td>
<td>6-6</td>
</tr>
<tr>
<td>6.2 Background and Rationales for Social Discounting</td>
<td>6-9</td>
</tr>
<tr>
<td>6.2.1 Consumption Rate of Interest and Social Opportunity Cost of Capital</td>
<td>6-9</td>
</tr>
<tr>
<td>6.2.2 Social Rate of Time Preference as the Social Discount Rate</td>
<td>6-10</td>
</tr>
<tr>
<td>6.2.3 Social Opportunity Cost of Capital as the Social Discount Rate</td>
<td>6-11</td>
</tr>
<tr>
<td>6.2.4 Shadow Price of Capital Approach</td>
<td>6-13</td>
</tr>
<tr>
<td>6.2.5 Evaluating Alternative Social Discount Rate Estimates</td>
<td>6-15</td>
</tr>
<tr>
<td>6.3 Intergenerational Social Discounting</td>
<td>6-16</td>
</tr>
<tr>
<td>6.3.1 The Ramsey Framework in an Intergenerational Context</td>
<td>6-17</td>
</tr>
<tr>
<td>6.3.2 Efficiency and Intergenerational Equity</td>
<td>6-17</td>
</tr>
<tr>
<td>6.3.3 Declining Discount Rates</td>
<td>6-18</td>
</tr>
<tr>
<td>6.4 The Role of Private Discounting in Economic Analysis</td>
<td>6-21</td>
</tr>
<tr>
<td>6.4.1 Predicting private behaviors and choices</td>
<td>6-22</td>
</tr>
<tr>
<td>6.4.2 Treatment of interest payments</td>
<td>6-22</td>
</tr>
<tr>
<td>6.4.3 Selecting private discount rates</td>
<td>6-23</td>
</tr>
<tr>
<td>6.5 Recommendations and Guidance</td>
<td>6-23</td>
</tr>
<tr>
<td>Chapter 6 References</td>
<td>6-24</td>
</tr>
<tr>
<td>Chapter 7: Analyzing Benefits</td>
<td>7-1</td>
</tr>
<tr>
<td>7.1 The Benefits Analysis Process</td>
<td>7-2</td>
</tr>
<tr>
<td>7.2 Economic Value and Types of Benefits</td>
<td>7-10</td>
</tr>
<tr>
<td>7.2.1 Human Health Improvements</td>
<td>7-12</td>
</tr>
<tr>
<td>7.2.2 Ecological Benefits</td>
<td>7-17</td>
</tr>
<tr>
<td>7.2.3 Other Benefits</td>
<td>7-20</td>
</tr>
<tr>
<td>7.3 Economic Valuation Methods for Benefits Analysis</td>
<td>7-21</td>
</tr>
<tr>
<td>Section</td>
<td>Page</td>
</tr>
<tr>
<td>------------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>7.3.1 Revealed Preference Methods and the Cost of Illness</td>
<td>7-21</td>
</tr>
<tr>
<td>7.3.2 Stated Preference</td>
<td>7-36</td>
</tr>
<tr>
<td>7.3.3 Combining Revealed and Stated Preference Data</td>
<td>7-44</td>
</tr>
<tr>
<td>7.4 Benefit Transfer</td>
<td>7-45</td>
</tr>
<tr>
<td>7.5 Accommodating Non-Monetized Benefits</td>
<td>7-49</td>
</tr>
<tr>
<td>7.5.1 Qualitative Discussions</td>
<td>7-50</td>
</tr>
<tr>
<td>7.5.2 Alternative Analyses</td>
<td>7-50</td>
</tr>
<tr>
<td>Chapter 7 References</td>
<td>7-52</td>
</tr>
<tr>
<td>Chapter 8: Analyzing Costs</td>
<td>8-1</td>
</tr>
<tr>
<td>8.1 The Economics of Social Cost</td>
<td>8-1</td>
</tr>
<tr>
<td>8.1.1 Compliance Cost Approach</td>
<td>8-2</td>
</tr>
<tr>
<td>8.1.2 Partial Equilibrium Approach</td>
<td>8-3</td>
</tr>
<tr>
<td>8.1.3 General Equilibrium Approach</td>
<td>8-5</td>
</tr>
<tr>
<td>8.2 Estimating Social Cost</td>
<td>8-6</td>
</tr>
<tr>
<td>8.2.1 Compliance Cost Estimation</td>
<td>8-8</td>
</tr>
<tr>
<td>8.2.2 Measuring Social Cost</td>
<td>8-10</td>
</tr>
<tr>
<td>8.2.3 Evaluating Costs Over Time</td>
<td>8-11</td>
</tr>
<tr>
<td>8.2.4 Valuing Time</td>
<td>8-15</td>
</tr>
<tr>
<td>8.2.5 Compliance Assumptions</td>
<td>8-16</td>
</tr>
<tr>
<td>8.3 Models Used in Estimating the Costs of Environmental Regulation</td>
<td>8-16</td>
</tr>
<tr>
<td>8.3.1 Compliance Cost Models</td>
<td>8-19</td>
</tr>
<tr>
<td>8.3.2 Partial Equilibrium Models</td>
<td>8-21</td>
</tr>
<tr>
<td>8.3.3 Computable General Equilibrium Models</td>
<td>8-22</td>
</tr>
<tr>
<td>8.3.4 Other Input-Output Based Models</td>
<td>8-23</td>
</tr>
<tr>
<td>8.4 Modeling Decisions and Challenges</td>
<td>8-26</td>
</tr>
<tr>
<td>8.4.1 Aggregation</td>
<td>8-26</td>
</tr>
<tr>
<td>8.4.2 Choosing Between a Static and Dynamic Framework</td>
<td>8-27</td>
</tr>
<tr>
<td>8.4.3 Model Parameterization</td>
<td>8-28</td>
</tr>
<tr>
<td>8.4.4 Uncertainty</td>
<td>8-29</td>
</tr>
<tr>
<td>8.4.5 Potential for Double Counting</td>
<td>8-30</td>
</tr>
<tr>
<td>Chapter 8 References</td>
<td>8-31</td>
</tr>
</tbody>
</table>
Chapter 9: Economic Impacts ................................................................. 9-1
  9.1 Background .................................................................................. 9-1
  9.2 Statutes and Policies ................................................................. 9-2
  9.3 Connections between Economic Impacts and Frameworks of Distributional Effects ................................................. 9-3
  9.4 Analytic Components of an Economic Impact Analysis ......................................................................................... 9-6
    9.4.1 Baseline Profile ......................................................................... 9-6
    9.4.2 Screening Analysis .................................................................... 9-7
    9.4.3 In-Depth Examination ............................................................ 9-8
    9.4.4 Data ......................................................................................... 9-8
  9.5 Impact Categories ........................................................................ 9-11
    9.5.1 Impacts on Consumers ........................................................... 9-11
    9.5.2 Impacts on Producers and Factors of Production ...................... 9-13
    9.5.3 Impacts on Communities ........................................................ 9-25
    9.5.4 Impacts on Governments and Non-Profits ................................... 9-25
    9.5.5 Economy-Wide Impacts .......................................................... 9-28
    9.5.6 Impacts of Benefits ................................................................. 9-29
Chapter 9 References ........................................................................ 9-30

Chapter 10 Environmental Justice and Life Stage Considerations ..........10-1
  10.1 Executive Orders, Directives, and Policies .................................... 10-1
  10.2 Environmental Justice ................................................................. 10-3
    10.2.1 Background Literature ........................................................... 10-4
    10.2.2 Analyzing EJ Impacts in the Context of Regulatory Analysis .... 10-6
    10.2.3 Populations of Concern for EJ Analysis ..................................... 10-10
    10.2.4 Data Sources ........................................................................ 10-12
    10.2.5 Scope and Geographic Considerations ..................................... 10-13
    10.2.6 Comparison groups ............................................................... 10-16
    10.2.7 Measuring and estimating impacts ........................................... 10-16
  10.3 Children’s Environmental Health .................................................. 10-22
    10.3.1 Childhood as a Life Stage ....................................................... 10-22
    10.3.2 Analytical Considerations ....................................................... 10-23
    10.3.3 Intersection Between Environmental Justice and Children’s Health ............................................................. 10-24
  10.4 Other Distributional Considerations ............................................. 10-24
    10.4.1 Older adults .......................................................................... 10-24
    10.4.2 Intergenerational Impacts ....................................................... 10-24
Chapter 10 References ........................................................................ 10-25
<table>
<thead>
<tr>
<th>Chapter 11: Presentation of Analysis and Results</th>
<th>Appendix A: Economic Theory</th>
</tr>
</thead>
<tbody>
<tr>
<td>11.1 Presenting Results of Economic Analyses</td>
<td>A.1 Market Economy</td>
</tr>
<tr>
<td>11.1.1 Presenting the Results of Benefit-Cost Analyses</td>
<td>A.2 Reasons for Market or Institutional Failure</td>
</tr>
<tr>
<td>11.1.2 Presenting the Results of Cost-Effectiveness Analyses</td>
<td>A.3 Benefit-Cost Analysis</td>
</tr>
<tr>
<td>11.1.3 Presenting the Results of EIA and Distributional Analyses</td>
<td>A.4 Measuring Economic Impacts</td>
</tr>
<tr>
<td>11.1.4 Reporting the Effects of Uncertainty on Results of Economic Analyses</td>
<td>A.4.1 Elasticities</td>
</tr>
<tr>
<td>11.2 Communicating Data, Model Choices and Assumptions, and Related Uncertainty</td>
<td>A.4.2 Measuring the Welfare Effect of a Change in Environmental Goods</td>
</tr>
<tr>
<td>11.2.1 Data</td>
<td>A.4.3 Single Market, Multi-Market, and General Equilibrium Analysis</td>
</tr>
<tr>
<td>11.2.2 Model Choices and Assumptions</td>
<td>A.5 Optimal Level of Regulation</td>
</tr>
<tr>
<td>11.2.3 Addressing Uncertainty Driven by Assumptions and Model Choice</td>
<td>A.6 Conclusion</td>
</tr>
<tr>
<td>11.3 Use of Economic Analyses</td>
<td>Appendix A References</td>
</tr>
</tbody>
</table>

| Appendix B: Mortality Risk Valuation Estimates |
| B.1 Central Estimate of VSL |
| B.2 Other VSL Information |
| B.3 Benefit Transfer Considerations |
| B.4 Adjustments Associated with Risk Characteristics |
| B.5 Effects on WTP Associated with Demographic Characteristics |
| B.6 Conclusion |

Appendix B References
Glossary

**Annualized value**
An annualized value is a constant stream of benefits or costs. The annualized cost is the amount that a party would have to pay at the end of each period $t$ to add up to the same cost in present value terms as the stream of costs being annualized. Similarly, the annualized benefit is the amount that a party would accrue at the end of each period $t$ to add up to the same benefit in present value terms as the stream of benefits being annualized.

**Baseline**
A baseline describes an initial, status quo scenario that is used for comparison with one or more alternative scenarios. In typical economic analyses the baseline is defined as the best assessment of the way the world would evolve absent the proposed regulation or policy action.

**Benefit-cost analysis (BCA)**
A BCA evaluates the favorable effects of policy actions and the associated opportunity costs of those actions. It addresses the question of whether the benefits are sufficient for the gainers to potentially compensate the losers, leaving everyone at least as well off as before the policy. The calculation of net benefits helps ascertain the economic efficiency of a regulation.

**Benefit-cost ratio**
A benefit-cost ratio is the ratio of the net present value (NPV) of benefits associated with a project or proposal, relative to the NPV of the costs of the project or proposal. The ratio indicates the benefits expected for each dollar of costs. Note that this ratio is not an indicator of the magnitude of net benefits. Two projects with the same benefit-cost ratio can have vastly different estimates of benefits and costs.

**Cessation lag**
Cessation lag is the time between a reduction in exposure and the reduction in risk. See latency for a definition of a related but distinct concept.

**Command-and-control regulation**
Command-and-control regulation is a prescriptive regulation that stipulates how much pollution an individual source or plant is allowed to emit and/or what types of control equipment it must use to reduce pollution.

**Compliance cost**
A compliance cost is the private cost that a regulated entity incurs to comply with a regulation - for instance, through the installation and operation of pollution abatement equipment.

**Consumption rate of interest**
Consumption rate of interest is the rate at which individuals are willing to exchange consumption over time. Simplifying assumptions, such as the absence of taxes on investment returns, imply that the consumption rate of interest equals the market interest rate, which also equals the rate of return on private sector investments.

**Cost-effectiveness analysis (CEA)**
CEA examines the costs associated with obtaining an additional unit of an environmental outcome. It is designed to identify the least expensive way of achieving a given environmental quality target, or the way of achieving the greatest improvement in some environmental target for a given expenditure of resources.

**Distributional analysis**
Distributional analysis assesses changes in social welfare by examining the effects of a regulation across different subpopulations and entities.
Economic efficiency

Economic efficiency can be defined as the maximization of social welfare. Under the efficient level of production, there is no way to rearrange production or reallocate goods such that someone is better off without making someone else worse off in the process.

Economic impact analysis (EIA)

Economic Impact Analyses (EIAs) examine how compliance costs, transfers, and other policy outcomes are distributed across groups. EIAs describe and often quantify outcomes such as changes in employment, plant closures, or local government tax revenues that provide insight into the economic consequences of regulation.

Elasticity of demand

Elasticity of demand measures the relationship between changes in quantity demanded of a good and changes in its price. It is calculated as the percentage change in quantity demanded that occurs in response to a percentage change in price. As the price of a good rises, consumers will usually demand a lower quantity of that good. The greater the extent to which quantity demanded falls as price rises, the greater is the price elasticity of demand. Some goods for which consumers cannot easily find substitutes, such as gasoline, are considered price inelastic. Note that elasticity can differ between the short term and the long term. For example, if the price of gasoline rises, consumers will eventually find ways to conserve their use of the resource. Some of these ways, like finding a more fuel-efficient car, take time. Hence gasoline would be price inelastic in the short term and more price elastic in the long term.

Elasticity of supply

Elasticity of supply measures the relationship between changes in quantity supplied of a good and changes in its price. It is measured as the percentage change in quantity supplied that occurs in response to a percentage change in price. For many goods the quantity supplied can be increased over time by locating alternative sources, investing in an expansion of production capacity, or developing competitive products that can substitute. One might therefore expect that the price elasticity of supply will be greater in the long term than the short term for such a good, that is, that supply can adjust to price changes to a greater degree over a longer period of time.

Emissions tax

An emissions tax is a charge levied on each unit of pollution emitted.

Environmental justice

Environmental justice is the fair treatment and meaningful involvement of all people regardless of race, color, national origin, or income with respect to the development, implementation, and enforcement of environmental laws, regulations, and policies. Fair treatment means that no group of people, including racial, ethnic, or socioeconomic groups should bear a disproportionate share of the negative environmental consequences resulting from industrial, municipal, and commercial operations or the execution of federal, state, local, and tribal programs and policies. Meaningful involvement occurs when (1) potentially affected community members have an appropriate opportunity to participate in decisions about a proposed activity that will affect their environment and/or health; (2) the public’s contribution can influence the regulatory agency’s decision; (3) their concerns will be considered in the decision-making process; and (4) the decision makers seek out and facilitate the involvement of those potentially affected.¹

Expert elicitation

Expert elicitation is a formal, highly-structured and well-documented process for obtaining the judgments of multiple experts. Typically, an elicitation is conducted to evaluate uncertainty. This uncertainty could be associated with: the value of a parameter to be used in a model; the likelihood and frequency of various future events; or the relative merits of alternative models.

**Externality**
An externality occurs when the production or consumption decision of one party has an unintended negative (positive) impact on the profit or utility of a third party.

**Flow pollutant**
A flow pollutant is a pollutant for which the environment has some absorptive capacity. It does not accumulate in the environment as long as its emission rate does not exceed the absorptive capacity of the environment. Animal and human wastes are examples of flow pollutants.

**Hotspot**
A hotspot is a geographic area with a high level of pollution/contamination within a larger geographic area of low or “normal” environmental quality.

**Kaldor-Hicks criterion**
The Kaldor-Hicks criterion is really a combination of two criteria: the Kaldor criterion and the Hicks criterion. The Kaldor criterion states that an activity will contribute to Pareto optimality if the maximum amount the gainers are hypothetically prepared to pay is greater than the minimum amount that the losers are hypothetically prepared to accept. Under the Hicks criterion, an activity will contribute to Pareto optimality if the maximum amount the losers are hypothetically prepared to offer to the gainers in order to prevent the change is less than the minimum amount the gainers are hypothetically prepared to accept as a bribe to forgo the change. In other words, the Hicks compensation test is conducted from the losers’ point of view, while the Kaldor compensation test is conducted from the gainers’ point of view. The Kaldor-Hicks criterion is widely applied in welfare economics and managerial economics. It forms an underlying rationale for BCA.

**Latency**
Latency is the time between the increase in exposure to a pollutant and the increase in health risk. See cessation lag for a definition of a related but distinct concept.

**Marginal benefit**
The marginal benefit is the benefit received from an incremental increase in the consumption of a good or service. It is calculated as the increase in total benefit divided by the increase in consumption.

**Marginal cost**
The marginal cost is the change in total cost that results from a unit increase in output. It is calculated as the increase in total cost divided by the increase in output.

**Marginal social benefit**
The marginal social benefit is the marginal benefit received by the consumer of a good (marginal private benefit) plus the marginal benefit received by other members of society (external benefit).

**Marginal social cost**
The marginal social cost is the marginal cost incurred by the producer of a good (marginal private cost) plus the marginal cost imposed on other members of society (external cost).

**Market failure**
A market failure occurs when the allocation of goods and services by the free market is not economically efficient. The most common causes of market failure are externalities, market power, and inadequate or asymmetric information. Externalities are the most likely cause of market failure in an environmental context.

**Market-based incentives**
Market-based incentives include a wide variety of methods for environmental protection. Instruments such as taxes, fees, charges, and subsidies generally “price” pollution and leave decisions about the level of emissions to each source. Another
example is the market permit system, which sets the total quantity of emissions and then allows trading of permits among firms.

**Meta-analysis**

Meta-analysis is an umbrella term for a suite of techniques that synthesize the results of empirical research. This could include a simple ranking of results, a meta-analytic average or other central tendency estimate, or a multivariate regression.

**Net benefits**

Net benefits are calculated by subtracting total costs from total benefits.

**Net future value**

Net future value is similar to NPV, however, instead of discounting all future values back to the present, values are accumulated forward to some future time period — for example, to the end of the last year of a policy’s effects.

**Net present value (NPV)**

The NPV is calculated as the present value of a stream of current and future benefits minus the present value of a stream of current and future costs.

**Non-use value**

Non-use value is the value that an individual may derive from a good or resource without consuming it, as opposed to the value obtained from use of the resource. Non-use values can include *bequest value*, where an individual places a value on the availability of a resource to future generations; *existence value*, where an individual values the mere knowledge of the existence of a good or resource; and *paternalistic altruism*, where an individual places a value on others’ enjoyment of the resource.

**Opportunity cost**

Opportunity cost is the value of the next best alternative to a particular activity or resource. Opportunity cost need not be assessed in monetary terms. It can be assessed in terms of anything that is of value to the person or persons doing the assessing. For example, a grove of trees used to produce paper may have a next-best-alternative use as habitat for spotted owls. Assessing opportunity costs is fundamental to assessing the true cost of any course of action. In the case where there is no explicit accounting or monetary cost (price) attached to a course of action, ignoring opportunity costs could produce the illusion that the action’s benefits cost nothing at all. The unseen opportunity costs then become the implicit hidden costs of that course of action.

**Quality-adjusted life year (QALY)**

QALY is a composite measure used to convert different types of health effects into a common, integrated unit, incorporating both the quality and quantity of life lived in different health states. This metric is commonly used in medical arenas to make decisions about medical interventions.

**Shadow price of capital**

The shadow price of capital takes into account the social value of displacing private capital investments. For example, when a public project displaces private sector investments, the correct method for measuring the social costs and benefits requires an adjustment of the estimated costs (and perhaps benefits as well) prior to discounting using the consumption rate of interest. This adjustment factor is referred to as the “shadow price of capital.”

**Social benefits**

Benefits are the favorable effects society gains due to a policy or action. Economists define benefits by focusing on changes in individual well-being, referred to as welfare or utility. Willingness to pay (WTP) is the preferred measure of these changes as it theoretically provides a full accounting of individual preferences across trade-offs between income and the favorable effects.
Social cost
Social cost means the sum of all opportunity costs, or reductions in societal well-being, incurred as a result of the regulation or policy action. These opportunity costs consist of the value lost to society of all the goods and services that will not be produced and consumed if firms comply with the regulation and reallocate resources away from production activities and towards pollution abatement. To be complete, an estimate of social cost should include both the opportunity costs of current consumption that will be foregone as a result of the regulation, and also the losses that may result if the regulation reduces capital investment and thus future consumption.

Social opportunity cost of capital
Social opportunity cost of capital is the rate at which consumption in the next period is reduced because private investment is displaced by required investments from policy. This is the rate at which society can trade consumption over time due to productive capital.

Social rate of time preference
Social rate of time preference is the discount rate at which society is willing to trade consumption in one period (usually year) for consumption in the next period.

Social welfare function
A social welfare function establishes criteria under which efficiency and equity outcomes are transformed into a single metric, making them directly comparable. A potential output of such a function is a ranking of policy outcomes that have different aggregate levels and distributions of net benefits. A social welfare function can provide empirical evidence that a policy alternative yielding higher net benefits, but a less equitable distribution of wealth, ranks better or worse than a less efficient alternative with more egalitarian distributional consequences.

Stock pollutants
A stock pollutant is a pollutant for which the environment has little or no absorptive capacity, such as non-biodegradable plastic, heavy metals such as mercury, and radioactive waste. A stock pollutant accumulates through time.

Subsidy
A subsidy is a kind of financial assistance, such as a grant, tax break, or trade barrier, that is implemented in order to encourage certain behavior. For example, the government may directly pay polluters to reduce their pollution emissions.

Tax-subsidy
A tax-subsidy is any form of subsidy where the recipients receive the benefit through the tax system, usually through the income tax, profit tax, or consumption tax systems. Examples include tax deductions for workers in certain industries, accelerated depreciation for certain industries or types of equipment, or exemption from consumption tax (sales tax or value added tax).

Use value
Use value is the value that an individual may derive from consumption or use of a good or resource.

Value of statistical life (VSL)
VSL is a summary measure for the dollar value of small changes in mortality risk experienced by a large number of people. VSL estimates are derived from aggregated estimates of individual values for small changes in mortality risks. For example, if 10,000 individuals are each willing to pay $500 for a reduction in risk of 1/10,000, then the value of saving one statistical life equals $500 times 10,000 — or $5 million. Note that this does not mean that any single identifiable life is valued at this amount. Rather, the aggregate value of reducing a collection of small individual risks is, in this case, worth $5 million.
Value of statistical life year (VSLY)
The VSLY is an estimated dollar value for a year of statistical life. In practice this metric is typically derived by dividing a VSL estimate by remaining life expectancy or discounted remaining life expectancy. This approach usually assumes that each year of life over the life cycle has the same value.

Willingness to accept (WTA)
WTA is the amount of compensation an individual would be willing to take in exchange for giving up some good or service. In the case of an environmental policy, WTA is the least amount of money that an individual would accept to forego an environmental improvement (or endure an environmental decrement).

Willingness to pay (WTP)
WTP is the largest amount of money that an individual would pay to receive the benefits (or avoid the damages) resulting from a policy change, without being made worse off. In the case of an environmental policy, WTP is the maximum amount of money an individual would pay to obtain an improvement (or avoid a decrement) in an environmental effect of concern.
Chapter 1

Introduction

The Guidelines for Preparing Economic Analyses is part of the U.S. Environmental Protection Agency's (EPA's) commitment to improve the preparation and use of sound science in economic analysis to inform decision making. The primary purpose of this document is to define and describe best practices for economic analysis grounded in the economics literature. It also describes Executive Orders (EOs) and other documents that impose analytic requirements and provides detailed information on selected important topics for economic analyses.

1.1 Background

Thorough and careful economic analysis is an important component for informing and developing sound environmental policies. High quality economic analyses can greatly enhance the effectiveness of environmental policy decisions by providing policy makers and the public with data-driven information needed to systematically assess the consequences of various actions or options. An economic analysis of a rulemaking is a positive exercise, as opposed to a normative one, that provides information on the potential economic efficiency of policy alternatives and assesses the magnitude and distribution of an array of impacts through careful investigation. Economic analysis also serves as a mechanism for organizing information carefully, identifying the kinds of impacts associated with stated policy alternatives and projecting who will be affected. Ultimately, economic analysis based on sound science should lead to better-informed regulatory and policy decisions.

The Guidelines for Preparing Economic Analyses, hereafter Guidelines, focus on the conduct of economic analysis to inform policy decisions and to meet requirements described by related statutes, Executive Orders (EOs), and associated implementing guidance of those EOs. The document is intended to ensure high quality analyses and consistency in how these economic analyses are prepared, performed, and reported. In so doing, the Guidelines elevate the quality of information shaping environmental policy decisions and Agency-issued guidance. The Guidelines also describe an interactive policy analysis development process between analysts and decision makers; reviews and summarizes environmental economics theory and the practice of benefit-cost analysis; and emphasizes issues in practical applications.

1.2 The Scope of the Guidelines

The Guidelines apply to economic analyses conducted for environmental policies using both regulatory and non-regulatory management strategies (e.g., support for voluntary programs) as well as Agency-issued guidance. Separate EPA guidance documents exist for related analyses, such as risk assessments, which can be inputs to economic assessments. No attempt is made here to summarize such guidance materials. Instead, their existence and content are noted in the appropriate sections.

The Guidelines assume the reader has some background in microeconomics as applied to environmental and natural resource policies. To fully understand and apply the approaches and recommendations presented in the

---

2. It is important to note that economic analysis is but one component in the decision-making process. Depending on the statutory context, all or certain components of the economic analysis may not be used by or required for, the legal rationale for the regulation. Other factors that may influence decision makers include statutory requirements, health risks, distributional considerations, enforceability, technical feasibility, policy priorities and ethics.

3. Chapter 2 describes many of these statues, EOs, and the analytic and/or procedural requirements they impose, as well as associated guidance materials.
**Guidelines**, readers should be familiar with basic applied microeconomic analysis, the concepts and measurement of consumer and producer surplus, and the economic foundations of benefit-cost analysis. Appendix A provides a brief review of economic foundations and the Glossary defines selected key terms.

The **Guidelines** are designed to assist staff with the preparation of economic analyses but are not a rigid blueprint nor a detailed set of step-by-step directions for all economic analyses. The most productive and illuminating technical approaches for an analysis will depend on case-specific factors and will require professional judgment. The **Guidelines** are a summary of analytical methodologies, empirical techniques, best practices, and data sources that can assist in identifying and implementing those approaches.

Finally, it is important to note that while the **Guidelines** apply to all economic analysis the focus is on benefit cost analysis and economic impact analysis -- two mainstays of EPA’s economic analyses. Typically, these economic analyses are not independent from other analyses. Assessing the effects of environmental policy is an inherently complex process in which results from various disciplines are integrated and inform one another. Taken together, they are used to predict environmental and behavioral outcomes and their economic consequences.

### 1.3 Economic Framework for Analysis

Conceptually, the ideal economic framework for assessing policy actions is one of general equilibrium that defines the allocation of resources and interrelationships for an entire economy with all its diverse components (e.g., households, firms, government). Potential regulatory alternatives are then modeled as economic changes that move the economy from a state of equilibrium absent the regulation (the baseline) to a new state of equilibrium with the regulation in effect. The differences between the old and new states are measured as changes in prices, quantities of goods, services and factors produced and consumed, including environmental quality, as well as wealth, income, and other economic metrics. These measurements may then be used to characterize the net welfare change for each affected group to inform questions of efficiency and distribution, based on individuals’ expected changes in well-being.

Questions about efficiency focus on aggregate changes in welfare. Economists generally define benefits from environmental regulation as positive changes in well-being and costs as the opportunities foregone, or losses in individual welfare.\(^6\) To assess efficiency under this scenario, we add these changes in welfare measured in monetary terms across all affected individuals. In the ideal, general equilibrium framework, we can estimate and sum all benefits and costs; so, a policy is a movement toward efficiency if the sum is positive and a movement away from efficiency if the sum is negative. The policy that maximizes this sum, i.e., net benefits, is considered economically efficient.\(^5\)

Questions about how these benefits and costs are distributed across households and industry examine how specific groups are affected by the policy. The ideal framework would answer questions framed in terms of welfare changes for groups of individuals (e.g., is the policy welfare-improving for a specific group?) or in terms of specific economic factors (e.g., how much will prices change for some goods?). These assessments of distributional outcomes are often important, apart from analysis of benefits and costs (i.e., economic efficiency).

In practice, of course, capturing this idealized framework empirically can be difficult, if not impossible, due to data availability; in most cases it is not possible to monetize all benefits and costs. No single modeling tool allows us to answer all policy-relevant questions about efficiency and distribution.\(^6\) As a practical matter, most economic analyses assemble a set of models to address these issues separately; but, even then, not all effects can be monetized. If

---

\(^6\) Note that environmental deregulation often results in disbenefits and cost-savings

\(^5\) Appendix A provides a conceptual overview of the economic theory of welfare changes and benefit-cost analysis.

\(^6\) As discussed in Chapter 8, computable general equilibrium models (CGE) capture most, or all, modeled market benefits and costs, but may not include non-market benefits. In practice, CGE models may be unable to analyze relatively small sectors of the economy. See Chapter 8, Section 4.6.
limitations are appropriately described, however, it is still informative to present comparisons of benefits and costs that can be monetized and qualitatively characterized, as well as evaluations of effects on specific groups.

As detailed more fully in Chapter 2, today economic analysis of benefits, costs, and distributional impacts are required by Executive Order (E.O.) 12866 for economically significant rules. Although E.O. 12291 in 1981 was the first to require an economic assessment of significant regulatory actions in a regulatory impact analysis (RIA), these analyses were not as extensive as the economic analyses required now by E.O. 12866. A complete economic analysis today, though it may still at times be labeled as an RIA, consists of a benefit-cost analysis and any related cost-effectiveness analyses and assessments of economic and distributional impacts. The Office of Management and Budget (OMB) has a useful checklist (shown in Text Box 1.1) for all components of an economic analysis conducted under E.O. 12866 (OMB, 2010).

### 1.3.1 Assessing Economic Efficiency with Benefit-Cost Analysis (BCA)

Benefit-cost analyses assess economic efficiency using the Potential Pareto criterion: is it theoretically possible for those who gain from the policy to fully compensate those who lose, and remain better off? When the answer to this question is 'yes', then net benefits, (benefits minus costs), are positive and the policy is a movement toward economic efficiency.

While conceptually identical, benefits and costs are often evaluated separately due to practical considerations. The benefits of reduced pollution are often attributable to changes in outcomes not exchanged in markets, such as improvements in public health. In contrast, the costs generally are measured through changes in outcomes that are exchanged in markets, such as pollution control equipment. As a result, different techniques are used to estimate benefits and costs.

Social benefits analyses evaluate the welfare gains individuals are expected to experience as the result of a policy. Once the changes in pollution levels or other environmental effects resulting from a policy are predicted, these changes are translated into health outcomes or other outcomes of interest using information provided by risk assessment and other disciplines. Benefits analyses then apply a variety of economic methodologies to estimate the value of these anticipated health improvements and other types of environmental benefits. Chapter 7 provides details on methods for estimating social benefits. Within a benefits assessment, pollution exposure may increase for some (e.g., emissions of a pollutant other than the one being regulated may increase or when the policy is deregulatory). Such costs may be presented as negative benefits and may be described as disbenefits or foregone benefits provided that the analysis is internally consistent.

Social cost analyses evaluate the welfare losses experienced by individuals as a result of environmental policies. In most instances, these costs are measured by higher prices for goods and services for consumers and lower earnings for producers and factors of production. Sometimes one modeling effort can be used to estimate both social costs and inputs for benefits analyses, such as predicted changes in pollution from regulated sources. Chapter 8 provides detailed information on methods for estimating social costs. As with benefits, some costs in a cost analysis may decrease due to the regulation (e.g., profits may increase for certain related entities or when the action is deregulatory). These outcomes may be presented as negative costs and may be described as avoided costs, again, provided that the analysis is internally consistent. Ultimately, from the perspective of economic theory, the treatment of disbenefits and avoided costs in the analysis is primarily a communications issue and should not affect efficiency analysis and whether net benefits are positive or negative.

---

7 More details about the checklist can be found at https://www.whitehouse.gov/sites/whitehouse.gov/files/omb/inforeg/inforeg/regpol/RIA_Checklist.pdf.

8 Appendix A describes the underlying economic theory in greater detail.
1.3.2 Assessing Economic and Distributinal Impacts

The assumptions and modeling framework developed for the BCA often do not include or allow for detailed examination of impacts on specific groups. Understanding the nature and magnitude of policy impacts and who will gain or lose from a regulation can be important to policy evaluation, and this requires analyses to supplement BCA.

EPA addresses economic and distributional impacts of environmental policy through two sets of analyses:

- **Economic Impact Analyses (EIAs)** provide insight into how compliance costs, transfers, and other policy outcomes are distributed across groups. EIAs describe and often quantify outcomes such as changes in employment, plant closures, or local government tax revenues that provide insight into the economic consequences of regulation. Economic impacts may fall on groups such as industry sectors, small businesses, state governments, consumers, or workers, that may benefit or be harmed by a policy. Chapter 9 provides information on analyzing economic impacts.

Adapted from OMB’s Agency Checklist: Regulatory Impact Analysis (2009).
• Other analyses evaluate the distribution of changes in environmental risks or health outcomes due to regulation from environmental justice (i.e., on minority, low-income, or indigenous populations), life stage (i.e., on children, the elderly), and intergenerational perspectives. Consideration of costs may also be relevant in such analyses. Chapter 10 provides information on how to analyze impacts from these perspectives.

1.4 Principles for Conducting Economic Analysis

While many specific aspects of an economic analysis will vary depending on the purpose, area of focus, available data, and needed level of detail for the analysis, there are core principles that apply to all analyses. These principles draw in part from, and are consistent with, those described in OMB Circular A-4 (OMB 2003).

• **Economic analyses should be based on sound economics and science.** Economic analyses should be grounded in well-established economic methods, theory, and principles. The effects considered in BCA, for example, should follow from economic principles and are independent of what is considered in legal or policy analyses, or what may be defined by science policy in other disciplines. Economic analysis should also be flexible enough to incorporate new information and advances in theory and the practice of economics. Economic analyses often rely upon or draw from the tools and results of other scientific analyses. These analyses should also be grounded in the principles, theories, and methods appropriate to their discipline.

• **Economic analyses should be objective and avoid bias.** The goal of the economic analysis is to provide objective information about the consequences of policy decisions. Professional judgments and assumptions are generally required for economic analyses, but these judgments and assumptions should not be based on the preferences of the analyst or policy maker. Economic analyses should seek to capture the expected behavioral responses of households, firms, and governments to incentives and options created by the actual requirements of the regulation or other context being analyzed as accurately as possible. Analyses should not be framed or performed in a manner to obtain predetermined results, nor should judgments or assumptions be made to favor one conclusion over another. For instance, sensitivity analysis can be used to explore a range of possible outcomes but should examine both higher and lower values rather than only one or the other.

• **Economic analyses should be transparent and replicable.** Economic analysis requires choices about data sources, methods, models, and assumptions. The reasons for these choices should be presented explicitly and clearly, along with appropriate justification. Economic analysis should also explicitly acknowledge and characterize important uncertainties in the analysis, state the judgments and decisions associated with these uncertainties, and should identify the implications of these choices. Specific references should be made to all data sources and models, and publicly available data and models should be used to the maximum extent possible. The analysis should provide enough information for readers to see clearly how final empirical estimates and conclusions were reached.

**Key best practices covered in the Guidelines**

• Key best practices that apply to all or most economic analyses are also covered in these Guidelines. These are listed below along with the chapter in which they are covered:

• Economic analyses produced by the Agency should adhere to directives from applicable statutes and executive orders (Chapter 2).

• Analyses should describe the economic basis for the policy action and evaluate multiple options to arrive at the most desirable decision (Chapter 3).

• Economics and economic analysis can also inform the consequences of different regulatory designs under consideration, identifying those that are likely to be most cost-effective (Chapter 4).

• The economic impact and consequences of policy must be evaluated relative to some alternative setting, generally one without the policy action. This alternative setting is called the analytic baseline. Specifying
baseline can sometimes be challenging, but it is essential for sound and informative economic analysis. The scope of the analysis should also be clearly defined (Chapter 5).

- The economic effects of policies usually take place over time periods of several years, and consistent application of discounting is necessary to make these effects comparable (Chapter 6).

- Analysis of benefits and costs should be grounded in sound, well-established economic principles and approaches, should capture all relevant outcomes to the extent possible, and should incorporate advances in the field where warranted (Chapter 7 and Chapter 8).

- Analysis of the distribution of impacts associated with policy decisions should adhere to the same high standards of an economic analysis, should start with the same baselines as the economic analysis, and should provide a balanced accounting of who gains and who loses as a result the policy action (Chapter 9 and 10).

- Finally, an economic analysis must be clearly and effectively communicated for it to be valuable for decision-making (Chapter 11).

Chapter 1 References


Chapter 2

Executive Order and Statutory Requirements for Conducting Economic Analyses

Federal agencies are subject to executive orders (EOs) and statutes that direct them to conduct specific types of economic analyses. Many of these directives are potentially relevant for all EPA programs, while others target individual programs. This chapter identifies directives for conducting economic analyses that may apply to all EPA programs (see Table 2.1 - Overview of Executive Orders and Statutes). Although not discussed here, analysts should carefully consider the relevant program-specific statutory requirements when designing and conducting economic analyses, recognizing that these requirements may mandate specific economic analyses.

The scope of the requirements for economic analyses in these directives varies substantially. In some cases, the language in a statute or EO may limit its applicability to only those regulatory actions that exceed a specified threshold in significance or impact. To determine whether a regulatory action meets such a threshold and is covered by the statute or EO, the Agency may need to conduct a brief economic analysis. Covered regulatory actions may be subject to additional requirements, such as:

- economic analysis (e.g., analysis of benefits and costs as required by EO 12866, "Regulatory Planning and Review"),
- procedural steps (e.g., consultation with affected state and local governments as required by EO 13132, "Federalism"), or
- a combination of both an economic analysis and procedural steps.

This chapter identifies the thresholds that trigger an economic analysis or additional procedural requirements for a regulatory action, summarizes the general requirements for economic analyses contained in selected statutes and EOs contingent on the thresholds, and provides further direction for analysts seeking guidance on compliance with the statute or EO. It also provides references to applicable OMB and EPA guidelines for each EO or statute discussed. For further information about the type and scope of analysis required, the program's Office of General Counsel (OGC) attorney is a

---

9. Specific numeric or other detailed criteria identified in EO or statute.

10. Note that for some statutes and EOs, requirements for proposed regulatory actions may vary slightly from the requirements for final regulatory actions.
Table 2.1 - Overview of Executive Orders and Statutes

<table>
<thead>
<tr>
<th>Executive Order/Statute</th>
<th>Economic Threshold¹</th>
<th>Guidance/Information Available</th>
</tr>
</thead>
<tbody>
<tr>
<td>E.O. 12898, Federal Actions to Address Environmental Justice in Minority Populations</td>
<td>General</td>
<td>EPA</td>
</tr>
<tr>
<td>and Low-Income Populations (1994)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>E.O. 13045, Protection of Children from Environmental Health Risks and Safety Risks</td>
<td>Specific</td>
<td>EPA</td>
</tr>
<tr>
<td>(1997)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>E.O. 13175, Consultation and Coordination with Indian Tribal Governments (2000)</td>
<td>General</td>
<td>EPA, OMB</td>
</tr>
<tr>
<td>E.O. 13211, Actions Concerning Regulations that Significantly Affect Energy Supply,</td>
<td>Specific</td>
<td>OMB</td>
</tr>
<tr>
<td>Distribution, or Use (2001)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>E.O. 13563, Improving Regulation and Regulatory Review (2011)</td>
<td>Specific</td>
<td>OMB</td>
</tr>
<tr>
<td>E.O. 13707, Using Behavioral Science Insights to Better Serve the American People</td>
<td>General</td>
<td>White House Memo</td>
</tr>
<tr>
<td>(2015)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>E.O. 13771, Reducing Regulation and Controlling Regulatory Costs (2017)</td>
<td>Specific</td>
<td>OMB</td>
</tr>
<tr>
<td>E.O. 13777, Enforcing the Regulatory Reform Agenda (2017)</td>
<td>General</td>
<td>OMB</td>
</tr>
<tr>
<td>E.O. 13891, Promoting the Rule of Law Through Improved Agency Guidance Documents</td>
<td>Specific</td>
<td>OMB</td>
</tr>
<tr>
<td>(2019)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Regulatory Flexibility Act (RFA), as Amended by the Small Business Regulatory</td>
<td>Specific</td>
<td>EPA</td>
</tr>
<tr>
<td>Enforcement Fairness Act of 1996 (SBREFA)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unfunded Mandates Reform Act of 1995 (UMRA)</td>
<td>Specific</td>
<td>EPA, OMB</td>
</tr>
<tr>
<td>Paperwork Reduction Act of 1995 (PRA)</td>
<td>Specific</td>
<td>EPA, OMB</td>
</tr>
</tbody>
</table>

¹ For OGC's reference guide on cross-cutting statutory and executive order reviews that may apply to rulemakings, see U.S. EPA 2003c (http://intranet.epa.gov/ogc/memoranda/checklist703.pdf) and 2005b (http://intranet.epa.gov/ogc/memoranda/desktopprefguide.pdf) (accessed 05/02/19).

This chapter does not address provisions of the statutes and EOs that do not require economic analysis.
2.1 Executive Orders

2.1.1 Executive Order 12866,12 "Regulatory Planning and Review"

Threshold: Significant regulatory actions as defined by the Executive Order. A “significant regulatory action” is defined by Section 3(f)(1)-(4) as any regulatory action that is likely to result in a rule that may:

- Have an annual effect on the economy of $100 million or more or adversely affect in a material way the economy, a sector of the economy, productivity, competition, jobs, the environment, public health or safety, or State, local, or tribal governments or communities;
- Create a serious inconsistency or otherwise interfere with an action taken or planned by another agency;
- Materially alter the budgetary impact of entitlements, grants, user fees, or loan programs or the rights and obligations of recipients thereof; or
- Raise novel legal or policy issues arising out of legal mandates, the President’s priorities, or the principles set forth in this Executive order.

Meeting one or more of the threshold criteria trigger the classification of a regulatory action as “significant;” a regulatory action that meets the first criteria is generally defined as “economically significant.” While the determination of economic significance is multi-faceted, it is most often triggered by the $100 million threshold. This threshold is interpreted by OMB as being based on the annual costs, benefits, or transfers of the proposed or finalized option in any one year. EO 12866 does not distinguish between regulatory and deregulatory actions.13 The word "or" is important: $100 million in annual benefits, or costs, or transfers is sufficient to meet the threshold.14 For example, suppose Congress passes a new law that requires the EPA to collect user fees from an industry that manufactures chemicals. The user fees will be used to defray the costs associated with an existing obligation for the EPA to conduct risk evaluations of new chemicals. Previously, the EPA’s costs to conduct these evaluations were provided by Congress through its annual congressional appropriation. This new rule requires the EPA to recoup these costs from industry. Assume that the fees to be collected from industry total $120 million per year. In this case, no new burden is being placed on society. The $120 million is simply a transfer of payments from businesses to government;15 however, because the transfer is more than $100 million annually, this action is economically significant. While the threshold for economic significance is important, the level of analysis is somewhat of a continuum; OMB clarifies, “The level of detail in the analysis can vary with the expected effects of the rule...”16

OMB does not adjust the $100 million threshold for inflation.17 As such, nominal values have been used in practice, implying that as inflation increases, the threshold becomes more stringent. Although most economic analyses report

---

12 EO 13563, “Improving Regulation and Regulatory Review,” issued in January 2011, supplements and reaffirms the provisions of EO 12866. It emphasizes the importance of reducing regulatory costs and burdens and maintaining flexibility and freedom of choice. To achieve these goals, it encourages:
- public participation,
- integration across federal agencies to promote simplification and harmonization of regulatory action,
- innovation in regulatory approaches, and
- consideration of alternative regulatory approaches.
It highlights the importance of scientific integrity, and retrospective analyses of existing rules. Finally, it states, “Our regulatory system must protect public health, welfare, safety, and our environment while promoting economic growth, innovation, competitiveness, and job creation.”

13 See EO 13771 “Reducing Regulation and Controlling Regulatory Costs” for more information on deregulatory actions.


16 OMB (2011b) page 5, Question 8.

17 OMB (2011b) page 2, Question 1.
costs and benefits using a base year other than the year the rule is issued,\textsuperscript{18} OMB has applied the economic significance trigger using current dollars.

A rule can also be deemed to meet the threshold of economic significance based on careful consideration of the phrase: “adversely affect in a material way.” There are no hard-and-fast rules for interpreting this criterion. OMB offers an example: a regulation that would (1) impose $98 million in first-year costs for pollution control equipment, with lower annual costs thereafter, (2) disproportionately and adversely affect a small sector of the economy, and (3) threaten to create significant job loss would be considered economically significant.\textsuperscript{19}

**Requirements contingent on threshold:** For all significant regulatory actions, a statement of the need for the proposed action and an assessment of potential benefits and costs are required (Section 6(a)(3)(B)). The requirements for the analysis of benefits and costs increase in complexity and detail for *economically* significant rules (i.e., those that fall under the definition in the first bullet above). For these rules, the EO requires that in addition to assessing potential costs and benefits, agencies must include the underlying analysis informing that assessment, quantify benefits and costs to the extent feasible, assess the benefits and costs of potentially effective and reasonably feasible alternative approaches, and provide the underlying analysis of that alternatives assessment (Section 6(a)(3)(C)).\textsuperscript{20} OMB’s *Circular A-4* (discussed below) states that analysts should generally analyze at least three options: the preferred option; a more stringent option; and a less stringent one.\textsuperscript{21}

**Guidance:** OMB’s *Circular A-4* (2003) provides guidance to federal agencies on the development of regulatory analysis of *economically* significant rules as required by EO 12866. More specifically, *Circular A-4* is intended to define good regulatory analysis and standardize the way benefits and costs of federal regulatory actions are measured and reported. Parts of *Circular A-4* guidance are standardized. For example, agencies are asked to provide a clear executive summary of their central conclusions, including a prominent standardized accounting statement, with one or more tables summarizing costs and benefits (both quantitative and qualitative), and transfers, at both 3% and 7% discount rates.\textsuperscript{22} In other respects, OMB has stated that “The level of detail in the analysis can vary with the expected effects of the rule; you should use more rigorous analytical approaches, and more comprehensive sensitivity analysis, for rules with especially large consequences.”\textsuperscript{23} To help clarify the requirements of EO 12866 and the guidance in A-4, OMB has also issued supplemental references on regulatory analyses for agencies.\textsuperscript{24}

\textsuperscript{18} Circular A-4 states that all costs and benefits should be reported in 2001 dollars but most Economic Analyses report results in a more recent base year. OMB states that you should use the GDP deflator to convert dollars to a different year. The Annual OMB Report to Congress on Benefits and Costs of Federal Regulations began to report estimates in both 2001 and 2010 dollars in the 2014 report. See OMB (2003) and OMB (2015).

\textsuperscript{19} OMB (2011b), page 1, Question 1.

\textsuperscript{20} EO 13422 and amended EO 12866 formerly required analysts to “identify in writing the specific market failure (such as externalities, market power, lack of information) or other specific problem” and extended the BCA requirement to “significant” guidance documents. Although EO 13497, issued in January 2009, revoked EO 13422 together with any “orders, rules, regulations, guidelines, or policies” enforcing it, a subsequent memo issued by then Director of OMB Peter R. Orszag offering guidance on the implementation of the new EO indicated that “significant policy and guidance documents…remain subject to OIRA’s review.”


\textsuperscript{22} See Chapter 11 of this document, Presentation of Analysis and Results, for agency guidance on presenting economic analysis results.

\textsuperscript{23} OMB (2011b) page 5, Question 8.

\textsuperscript{24} The supplemental OMB references are:

- 2010 Agency Checklist for RIAs (OMB 2010b)
- 2011 FAQ on regulatory analysis (OMB 2011b)
  https://www.whitehouse.gov/sites/whitehouse.gov/files/omb/assets/OMB/circulars/a004/a-4_FAQ.pdf
- 2011 "Primer" on RIAs per Circular A-4 (OMB 2011d)

2 - 4
The Guidelines provide more in-depth Agency guidance, building on the OMB’s guidance with a focus on approaches and methods that are relevant to environmental regulations. Chapters 3 through 8 of this document provide more detailed guidance for meeting the EO 12866 benefit-cost analysis requirements, consistent with provisions in OMB’s Circular A-4. Chapters 9 and 10 provide guidance on addressing distributional effects of environmental regulations, with a focus on economic impact analysis examining compliance costs effects (e.g., profitability, employment, prices) in Chapter 9 and on environmental justice and life stage considerations in Chapter 10.25

2.1.2 Executive Order 12898, “Federal Actions to Address Environmental Justice in Minority Populations and Low-Income Populations”

**Threshold:** No specific threshold; Agencies are required to “…identify and address disproportionately high and adverse human health or environmental effects of its programs, policies, and activities [including rulemaking] on minority populations and low-income populations…”

**Requirements contingent on threshold:** No specific analytical requirements in the EO. However, President Clinton issued a memorandum to accompany EO 12898 directing federal agencies to analyze environmental effects, including human health, economic, and social effects, of federal actions when such analysis is required under the National Environmental Policy Act (NEPA).

**Guidance:** The EPA’s Technical Guidance for Assessing Environmental Justice in Regulatory Analysis (U.S. EPA 2016) is designed to help EPA analysts understand how to evaluate potential EJ concerns associated with EPA regulatory actions. The Agency also has guidance for considering environmental justice in the Action Development Process (U.S. EPA 2015) which provides direction on when EJ should be considered during rulemaking. The EPA and the Council on Environmental Quality (CEQ) have prepared guidance for addressing environmental justice concerns in the context of National Environmental Policy Act (NEPA) requirements (U.S. EPA 1998a; U.S. CEQ 1997). These materials provide guidance on key terms in the EO. Chapter 10 of this document addresses environmental justice analysis.

2.1.3 Executive Order 13045, “Protection of Children from Environmental Health Risks and Safety Risks”

**Threshold:** Economically significant regulatory actions as described by EO 12866 that involve environmental health risk or safety risk that an agency has reason to believe may disproportionately affect children.

**Requirements contingent on threshold:** An evaluation of the health or safety effects of the planned regulation on children, as well as an explanation of why the planned regulation is preferable to other potentially effective and reasonably feasible alternatives the agency is considering.

**Guidance:** The EPA has prepared guidance to assist EPA staff on the implementation of EO 13045 (U.S. EPA 2006). The EPA’s Children’s Health Valuation Handbook (U.S. EPA 2003b) discusses special issues related to estimation of the value of health risk reductions to children. The Office of Children’s Health Protection also provides online information with links to resource materials on guidance and tools.26 Guidance in Chapter 10 of this document addresses distributional analyses focused on children.

---

25 In its Statement of Regulatory Philosophy, EO 12866 states that agencies should consider the distributional and equity effects of a rule (Section 1(a)).

2.1.4 Executive Order 13132, “Federalism”

Threshold: Rules that have “federalism implications” that either impose substantial compliance costs on state and local governments or preempt state or local law. According to EPA policy, rules are considered to impose substantial compliance costs if:

- the action is likely to result in the expenditure by state and local governments, in the aggregate, of $25 million or more in any one year; or
- the action is likely to result in expenditures by small governments that equal or exceed 1 percent of their annual revenues.²⁷

Exception: An action that imposes substantial compliance costs (meets the $25 million threshold or the 1 percent test) does not have a federalism implication if: (1) the action is expressly required by statute (without any discretion by the EPA); or (2) there are federal funds available to cover the compliance costs.

Requirements contingent on threshold: Actions with federalism implications require pre-proposal consultation with elected state/local officials or their representative national organizations. Rules must include a Federalism Summary Impact Statement in the preamble, and a signed Federalism Certification from the Agency’s designated official should be provided to OMB along with any written communications that the EPA received from state or local officials.

Guidance: Specific guidance on EO 13132 can be found in the internal EPA document Guidance on Executive Order 13132: Federalism (U.S. EPA 2008c).

2.1.5 Executive Order 13175, “Consultation and Coordination with Indian Tribal Governments”

Threshold: Regulations that have substantial direct effects on one or more Indian tribes, on the relationship between the federal government and tribes, or on the distribution of power and responsibilities between the federal government and tribes and that: (1) impose substantial direct compliance costs on Indian tribal governments that are not required by statute, or (2) preempt tribal law.

Requirements contingent on threshold: To the extent practicable and permitted by law, the Agency must either provide the funds necessary to pay the tribal governments’ direct compliance costs, if applicable, or prior to the formal promulgation of the regulation, consult with tribal officials early in the process of developing the proposed regulation and include in the preamble of the regulation a Tribal Summary Impact Statement. The Statement must include a description of the extent of the Agency's prior consultation with tribal governments; a summary of the nature of the tribe's concerns and the agency's position supporting the need to issue the regulation; and a statement of the extent to which the concerns of tribal governments have been met.

Guidance: OMB issued Guidance for Implementing EO 13175 in 2010, to provide direction for compliance and documentation.²⁸ The EPA issued Policy on Consultation and Coordination with Indian Tribes (2011) to establish national guidelines and institutional controls for consultation across the EPA. This policy states, "EPA’s policy is to consult on a government-to-government basis with federally recognized tribal governments when EPA actions and decisions may affect tribal interests" [emphasis added].²⁹ Chapter 10 of this document addresses distributional analyses focusing on minority, low income, and indigenous populations.


2.1.6 Executive Order 13211, “Actions Concerning Regulations that Significantly Affect Energy Supply, Distribution, or Use”

**Threshold:** Rules that are significant regulatory actions under EO 12866 and that are likely to have significant adverse effects on the supply, distribution, or use of energy.

**Requirements contingent on threshold:** Submission of a detailed Statement of Energy Effects to OMB. The Statement of Energy Effects must address any expected adverse effects on energy supply, distribution, or use, the reasonable alternatives to the action, and the expected effects of such alternatives on energy supply, distribution, and use.

**Guidance:** OMB has guidance for implementing EO 13211.\(^{30}\)

2.1.7 Executive Order 13563, "Improving Regulation and Regulatory Review"

**Threshold:** Significant regulatory actions under EO 12866 (reaffirms EO 12866 and adds additional requirements).\(^{31}\)

**Requirements contingent on threshold:** Among other requirements, agencies must use best available techniques to quantify costs and benefits, give the public meaningful opportunity to comment online, include relevant scientific and technical findings in the rulemaking docket, consider the combined effects of their regulations on particular sectors and industries, and promote coordination across agencies. Agencies are required to develop plans for retrospective review of significant rules.

**Guidance:** OMB issued implementation guidance in three memos: M-11-10 February 2, 2011, M-11-19 April 25, 2011, M-11-25 June 14, 2011.\(^{32}\)

2.1.8 Executive Order 13707, "Using Behavioral Science Insights to Better Serve the American People"

**Threshold:** No specific threshold; the EO encourages agencies to "identify policies, programs, and operations where applying behavioral science insights may yield substantial improvements in public welfare, program outcomes, and program cost effectiveness..."

**Requirements contingent on threshold:** Agencies are encouraged to use behavioral science insights when designing policies and specifically when determining access to programs, presenting Information to the public, structuring choices within programs, and designing incentives.

**Guidance:** The White House Social and Behavioral Sciences Team issued implementation guidance in a memo on September 15, 2016.\(^{33}\) Chapter 4 of this document includes a discussion of behavioral responses.


\(^{32}\) EO 13563 and OMB's implementation guidance (OMB 2011a; OMB 2011c; and OMB 2011e) is located at: [https://www.whitehouse.gov/omb/information-regulatory-affairs/regulatory-matters/#eo13563](https://www.whitehouse.gov/omb/information-regulatory-affairs/regulatory-matters/#eo13563) (accessed March 21, 2019).

2.1.9 Executive Order 13771, "Reducing Regulation and Controlling Regulatory Costs"

**Threshold:** OMB Guidance defines an EO 13771 regulatory action as "a significant regulatory action as defined in Section 3(f) of EO 12866 that has been finalized and that imposes total costs greater than zero," and defines an EO 13771 deregulatory action as "an action that has been finalized and has total costs less than zero."  

**Requirements contingent on threshold:** For every new EO 13771 regulatory action proposed, the Agency must identify at least two prior regulations to be repealed. The Agency must also offset any incremental costs associated with the new regulation in order to meet a cost allowance set by OMB each fiscal year. Agencies are directed to calculate the present value (as of 2016) of costs for EO 13771 regulatory actions and cost savings for EO 13771 deregulatory actions over the full duration of the expected effects of the actions using both 3% and 7% end-of-period discount rates.  

**Guidance:** OMB issued guidance on implementing the EO on April 5, 2017 in the form of Questions and Answers for agencies as well as interim guidance for implementing the EO on February 2, 2017. The guidance notes that EO 12866 remains the primary governing EO regarding regulatory planning and review.

OMB's guidance also notes that agencies may proceed with significant regulatory actions that need to be finalized in order to comply with an imminent statutory or judicial deadline even if they are not able to identify offsetting regulatory actions by the time of issuance but must subsequently identify other regulations to be repealed to meet the requirements of the EO.

2.1.10 Executive Order 13777, "Enforcing the Regulatory Reform Agenda"

**Threshold:** No specific threshold; Agency’s Regulatory Reform Task Force is charged with making recommendations to agency head on repeal, replacement or modification of existing rules.

**Requirements contingent on threshold:** To make recommendations, the Task Force is to evaluate existing regulations to identify those that, among other things, "(i) eliminate jobs, or inhibit job creation; (ii) are outdated, unnecessary, or ineffective;" or "(iii) impose costs that exceed benefits."

**Guidance:** OMB issued a guidance memo on regulatory reform accountability on April 28, 2017. The memo states that "agencies should establish and report other meaningful performance indicators and goals for the purpose of evaluating and improving the net benefits of their respective regulatory programs (i.e., all of the existing regulations in place that address a specific regulatory objective)." See also Chapter 9 of this document for a discussion of analysis of economic impacts.

---

37 OMB (2017a), page 5.
2.1.1 Executive Order 13783, "Promoting Energy Independence and Economic Growth" (Sec. 1 & 2)

Threshold: All existing regulations, orders, guidance documents, policies, and any other similar agency actions (collectively, agency actions) that potentially burden the development or use of domestically produced energy resources, with particular attention to oil, natural gas, coal, and nuclear energy resources.

Requirements contingent on threshold: Agencies are to develop and submit a review plan to OMB and CEQ, conduct the required review, and issue a report which "shall include specific recommendations that, to the extent permitted by law, could alleviate or eliminate aspects of agency actions that burden domestic energy production."

Guidance: OMB issued guidance for Section 2 of the EO which covers review of actions that "potentially burden the safe, efficient development of domestic energy resources." The review should include "any quantitative analysis (e.g., costs, lost production) the agency plans to perform” and the report should include "preliminary estimates by agency action of the costs and cost savings, increased production, or other beneficial effects, that may be achieved by implementing each recommended action" using the guidance for EO 13771 and Circular A-4.

2.1.12 Executive Order 13891, "Promoting the Rule of Law Through Improved Agency Guidance Documents"

Threshold: According to OMB, "[a]n analysis is required for any guidance document that may bring about $100 million in benefits, costs, or transfer impacts in at least one year (i.e., in one consecutive twelve-month period), or that otherwise qualifies as economically significant under Executive Order 12866.” In determining whether a guidance document is significant, agencies should provide at least the same level of analysis that would be required for a major determination under the Congressional Review Act.

Requirements contingent on threshold: Agencies are to conduct a Regulatory Impact Analysis for economically significant guidance documents that is consistent with the analysis that would be conducted for an economically significant rulemaking. In addition, agencies are to explain how the guidance document maximizes net benefits and any alternatives considered.

Guidance: OMB issued a guidance memo on October 31, 2019. The RIA for a significant guidance document should generally follow the principles of Circular A-4, although there may be some differences in estimating behavior change due to the non-binding nature of guidance and in considering baseline considerations. The memo also discusses the definition of guidance document, waivers and exemptions.

2.2 Statutes

2.2.1 Regulatory Flexibility Act of 1980 (RFA), as Amended by The Small Business Regulatory Enforcement Fairness Act of 1996 (SBREFA) (5 U.S.C. 601-612)

Threshold: Regulations that may have a “significant economic impact on a substantial number of small entities,” (SISNOSE), including small businesses, governments and non-profit organizations. The RFA does not define the terms significant or substantial.

Requirements contingent on threshold: For rules that are expected to have a SISNOSE, agencies are required to prepare an initial regulatory flexibility analysis (IRFA) and a final regulatory flexibility analysis (FRFA) examining potential adverse economic impacts on small entities and complying with a number of procedural requirements to

---

solicit and consider flexible regulatory options that minimize adverse economic impacts on small entities and address significant issues raised in public comments. The IRFA and FRFA are published with the proposed and final rules, respectively.

Guidance: The EPA has issued specific guidance for complying with RFA/SBREFA requirements in the *EPA Final Guidance for EPA Rulewriters: Regulatory Flexibility Act as amended by the Small Business Regulatory Enforcement Fairness Act.* See also Chapter 9 of this document on economic impact analysis.

The guidance offers approaches for determining whether a specific rule may have a SISNOSE but provides flexibility to use alternative methods or reach different conclusions where appropriate in the context of a specific rule.

### 2.2.2 Unfunded Mandates Reform Act of 1995 (UMRA) (P.L. 104-4)

**Threshold one (Sections 202 and 205 of UMRA):** Regulatory actions that include federal mandates “that may result in the expenditure by State, local, and tribal governments, in the aggregate, or by the private sector, of $100 million or more (adjusted annually for inflation) in any one year.” An action contains a federal mandate if it imposes an enforceable duty on state, local or tribal governments or the private sector.

**Requirements contingent on threshold one:** Section 202 of UMRA requires preparation of a written statement that includes the legal authority for the action; a BCA; a distributional analysis; estimates of macroeconomic impacts; a description of an agency’s pre-proposal consultation with elected representatives of the affected state, local, or tribal governments; and a summary of concerns raised and how they were addressed. Section 205 of UMRA requires an agency to consider a reasonable number of regulatory alternatives and select the least costly, most cost-effective, or least burdensome alternative, or to publish with the final rule an explanation of why such alternative was not chosen.

Per OMB’s Circular A-4, the analytical requirements under EO 12866 are similar to the analytical requirements under Sections 202 and 205 of UMRA, and thus the same analysis may permit compliance with both analytical requirements.

**Threshold two (Section 203 of UMRA):** Regulatory requirements that might “significantly” or “uniquely” affect small governments. Small governments include governments of cities, counties, towns, townships, villages, school districts, or special districts with a population of less than fifty thousand.

**Requirements contingent on threshold two:** Agencies must solicit involvement from, and conduct outreach to, potentially affected elected officers of small governments (or their designated employees) during development and implementation.

Guidance: The EPA has issued *Interim Guidance on the Unfunded Mandates Reform Act of 1995,* (1995b), and OMB issued a memo on *Guidance for Implementing Title II of S.1* that provides general guidance on complying with requirements contingent on each of the two thresholds under UMRA.

### 2.2.3 The Paperwork Reduction Act of 1995 (PRA) (44 U.S.C. 3501)

**Threshold:** Any action that includes record-keeping, reporting, or disclosure requirements or other information collection activities calling for answers to questions seeking the same information imposed upon or posed to ten or more persons, other than federal agency employees. 

---


42 Note that the threshold in this case is “adjusted annually for inflation” as opposed to the threshold under EO 12866.


44 Note that Section1320.3(c)(4)(ii) states that “any collection of information addressed to all or a substantial majority of an industry is presumed to...
Requirements contingent on threshold: The Agency must submit an information collection request (ICR) to OMB for review and approval and meet other procedural requirements including public notice. The ICR should: (1) describe the information to be collected, (2) give the reason the information is needed, and (3) estimate the time and cost for the public to answer the request.

Guidance: Both guidance and templates for completing an ICR and associated Federal Register (FR) notices can be found on the EPA’s intranet site, “ICR Center.”

Chapter 2 References


---

involve ten or more persons.” However, OMB guidance on this issue indicates that if agencies have evidence showing that this presumption is incorrect in a specific situation (i.e., fewer than 10 persons would be surveyed), the agency may proceed with the collection without seeking OMB approval. Agencies must be prepared to provide this evidence to OMB on request and abide by OMB’s determination as to whether the collection of information ultimately requires OMB approval (OMB 1999).


Chapter 3

Need for Regulatory Action and Evaluation of Policy Options

A clear statement of need for regulatory action describing the problem to be addressed by the policy and a detailed evaluation of policy options are both essential components of an economic analysis. The statement of need should include a description of the market failure, an explanation as to why the market and other institutions have failed to correct the problem, and a justification for federal action to address it.

The economic analysis should consider and evaluate multiple policy options that address the problem. This is true for analyses of both proposed rules and final rules, even when the Agency has settled on a specific option. When identifying policy options, the analysis should describe any statutory or judicial requirements that must be considered. The options should include those permissible under the relevant statutory authority and may include those that are unavailable but have other advantages. The options may also differ in their levels of stringency or they may represent entirely different regulatory approaches. Detailing the possible options under consideration is a necessary step in establishing why the preferred option is the appropriate choice.

3.1 The Statement of Need

Consistent with EO 12866 and OMB (2003), each economic analysis should include a statement of need that provides (1) a clear description of the problem being addressed, (2) the reasons for and significance of the market failure causing this problem, and (3) the compelling need for federal government intervention in the market to correct the problem. This statement sets the stage for the subsequent benefit-cost analysis and allows one to judge whether the problem is being adequately addressed by the policy.

3.1.1 Problem Description

The statement of need should begin with a brief review of the problem or public need that is to be addressed by the policy. While not always the case, the compelling public need for EPA regulations is generally to address an environmental problem. In this case, the following considerations are often relevant:

- The primary environmental contaminants causing the problem and their concentration;
- The media through which exposures or damages take place;
- Private and public sector sources responsible for creating the problem;

---

EO 12866 states that “Federal agencies should promulgate only such regulations as are required by law, are necessary to interpret the law, or are made necessary by compelling need, such as material failures of private markets to protect or improve the health and safety of the public, the environment, or the well-being of the American people…” (emphasis added). The Office of Management and Budget’s guidance for how to comply with EO 12866, Circular A-4 (OMB 2003), provides recommendations to federal agencies on the development of economic analyses supporting regulatory actions. OMB (2003, pg. 2) states that “a statement of the need for the proposed action” is a “key element” of a regulatory analysis, and that “an agency must demonstrate that the proposed action is necessary”.

---
• Human exposures involved and the health effects due to those exposures;
• Non-human resources affected and the resulting outcome;
• Expected evolution of the environmental problem over time, absent additional regulation;
• Available and potential abatement and mitigation techniques and technologies;
• The amount and/or proportion (or both) of the environmental problem likely to be corrected by federal action;
• Any existing state, local and other federal activities that partially or fully address the problem.

3.1.2 Reasons for Market Failure

A regulation can be promulgated for a number of social purposes including improving government function, increasing distributional equity, or promoting privacy and personal freedom. For pollution problems, the social purpose is to correct a “market failure.” A market failure occurs when the allocation of goods and services by the free market is not economically efficient. After defining the problem, the statement of need should examine the reasons why the market and other public and private sector institutions have failed to correct it. This identification is an important component of policy development because the underlying failure itself often suggests the most appropriate remedy for the problem (see Chapter 4). While other social purposes, including improving government function, increasing distributional equity, or promoting privacy and personal freedom, may be enhanced by the regulation, correcting market failures is often the driver behind environmental policy.

A market failure occurs when the allocation of goods and services by the free market is not economically efficient. The most common causes of market failure are externalities, market power, and inadequate or asymmetric information. Externalities are the most likely cause of market failure in an environmental context. Technically, externalities occur when non-monetary variables chosen by one individual enter the utility or production function of another. Put another way, externalities occur when the market does not account for the effect of one party’s activities on another party’s well-being. Consider for example, a factory that produces smoke as a by-product of manufacturing that in turn affects individuals living downwind. The factory does not weigh the costs of its actions on the downwind community when making production decisions. Although the factory imposes an externality on the downwind community, the mere existence of an externality is not enough to justify a regulation. Under certain conditions, namely the ability to bargain, availability of full information, and presence of low transactions costs, externalities can be internalized by the free market (Coase 1960). Text Box 3.1 describes this Coase solution in more detail.

It is important to differentiate externalities from other external effects when an individual or firm is affected by the behavior of others. For example, a negative outcome caused by another individual is not an externality if the affected individual rationally and willingly accepted the risk of that outcome through a private transaction between them. This may occur when a worker accepts a job with a greater risk of injury in exchange for a higher wage. Under the conditions of complete and perfect markets with full information, the stipulations of the transaction reflect and incorporate the expected risk such that there is no externality associated with increased risk of injury. Similarly, external effects that function through the price system (e.g., higher prices faced by certain consumers because of rising demand) or zero-sum transfers from one person to another (e.g. through taxes or redistribution of consumer and producer surplus) are not externalities by definition and do not constitute a market failure. For example, if person A outbids person B in an auction, B may be made worse off than had they won the auction but was unwilling to pay the higher bid. This is a result

---

47 If the social purpose of a regulation is not to address a market failure (e.g., to improve Agency processes or solely to define a statutory term), then the statement of need still should include a description of the problem being addressed and an explanation of why government action is necessary to address this problem.

48 For further discussion of market failure, types of market failures, and externalities see Scitovsky (1954), Mishan (1969), Baumol and Oates (1988), Cornes and Sandler (1996), Hanley et al. (2019), Peman et al. (2003), Tietenberg and Lewis (2014). OMB (2003) also describes different categories of market failure as well as other reasons for regulation. Section A.2 of these Guidelines provides further discussion of externalities.
Text Box 3.1 - Coase Solution

Government intervention for the control of environmental externalities may not be necessary if parties can work out an agreement between themselves. Coase (1960) outlined conditions under which transaction costs are low enough that a private agreement between affected parties might result in the attainment of a welfare-maximizing level of pollution without government intervention. First, property rights must be fully and clearly defined and transferable. In situations where the resource in question is not "owned" by anyone, there is no ability to negotiate, and the offending party can "free ride," or continue to pollute, without facing the costs of its behavior.

When property rights have been allocated, a welfare-maximizing solution can be reached regardless of which party is assigned the property rights, although the distribution of the gains from bargaining will differ. Take for example a farm whose pesticide application to its crops pollutes the well water of nearby homeowners. If property rights of the watershed are assigned to the homeowners, and information is available to them about potential damages from the pollution, then the farm may negotiate with the homeowners about its continued use of the pesticide. Potential payments from the farm to the homeowners agreed upon through such negotiations need not be in the form of cash but could be payments in kind. If property rights of the watershed are given to the farm, then the homeowners could negotiate to pay the farm to stop applying the pesticide.

The effectiveness of such agreements is contingent on meeting additional conditions: bargaining must be possible, damages must be known, and transaction costs must be low. These conditions are more likely to be met when there are only a small number of individuals involved. If either party is unwilling to negotiate or faces high transaction costs, then no private agreement will be reached. Asymmetric information or bargaining power can also hinder a socially optimal solution. Going back to the example, consider a case where there are many farms in the watershed using the pesticide on their crops, and it may be difficult to identify the relative contribution of each farm’s effluent on damages experienced by the homeowner. Clearly homeowners would have more difficulty in negotiating an agreement with many farms than they would in negotiating with a single farm.

of the price system working to ensure scarce resources go to those willing to pay the most for them, and as such does not result in an inefficient allocation of resources.\(^{49}\)

When left unaddressed, however, externalities prevent the market from achieving economic efficiency and reduce economic welfare. This can occur in the presence of high transaction costs that make it difficult for injured parties to ensure that polluters internalize the cost of damages through bargaining, legal action, or other means. High transaction costs may result when activities that pose environmental risks are difficult to link to the resulting damages because they occur over long periods of time or occur in a different location than where the pollution originates.\(^{50}\)

However, even the presence of an unaddressed externality is not enough to justify a regulation; what is required is a compelling need for government intervention at any level of government (federal, state, or local). That is, there must be some form of evidence that government intervention can improve economic welfare.\(^{51}\) For instance, government regulation may not be warranted if the benefits of regulation do not justify the costs. Circumstances where this may occur include when a regulation designed to reduce a negative externality (e.g., direct emission controls) exacerbates pre-existing distortions or market imperfections. In this case, government intervention could make things worse. Even if an externality warrants government intervention, it may not warrant direct, prescriptive regulation. Some externalities may be addressed more efficiently through other means such as providing information, requiring firms to carry insurance, defining legal liability, or assigning property rights. The nature of the externality may determine the best approach for government action (See Chapter 4).

---

\(^{49}\) External effects operating through the price system are sometimes referred to as pecuniary externalities.

\(^{50}\) The concept of an externality is closely tied to the concept of a public good, which is a good that either can be used simultaneously by many (i.e., nonrival) or that is difficult to prevent others from using (i.e., nonexcludable). The environment is a classic example of a public good.

\(^{51}\) Lusk (2013) provides a useful 9-point checklist for externalities that require prescriptive regulation.
3.1.3 Need for Federal Action

The final component of the statement of need for the regulatory action is an evaluation and explanation of why a federal remedy is preferable to actions by private and other public-sector entities, such as the judicial system or state and local governments. Federal involvement is often required for environmental problems that cross jurisdictional boundaries (e.g., when pollution in one state affects the population of another). In some cases, federal involvement is mandated by statute or Executive Order as described in Chapter 2. The basis for federal regulation could be justified by comparing its expected performance to realistic alternatives that rely on other institutional arrangements. This component of the statement of need for regulatory action, justifying federal regulation, should verify that the proposed policy action is necessary, within the jurisdiction of the relevant statutory authorities, and yields results that will be preferable to no action. Finally, the statement of need should identify those aspects of the regulation necessitated by statutory requirements and those that are discretionary.

3.2 General Guidance on Policy Options to Evaluate

Following the statement of need, the economic analysis should identify and describe in detail all policy options or potential regulatory alternatives that were considered. This includes clearly explaining which options were selected for emphasis and further analysis and why other important options were not. Since the benefit cost analysis informs the public, stakeholders, and Congress and other decision-makers of the effects of the policy (OMB 2003), assessing a robust set of policy options is important.

The identification of policy options should describe any statutory or judicial requirements that must be considered when designing the regulation, how these requirements may influence the options considered, and how the preferred option satisfies them. For example, the description should identify any economic considerations (e.g., costs incurred by regulated entities) that must or may be used to shape the form and stringency of the regulation. The analysis should identify those options that are more efficient or cost-effective even if the regulatory approaches may be prohibited by statutory or judicial requirements (OMB 2003). For example, the Supreme Court has held that the Clean Air Act requires that National Ambient Air Quality Standards be set based on health or welfare considerations only; the Act bars EPA from considering the costs of implementing them when setting the standards.

At a minimum, the economic analysis should fully assess and present three options for consideration: the preferred option; a more stringent option; and a less stringent one. The incremental benefits and costs for each option, as well as other important criteria (e.g. distributional consequences), should be compared across the options. Measuring the

---

52 As discussed in Chapter 2, EO 13132, “Federalism”, describes principles of federalism and identifies requirements for federal preemption of state or local law. Also, there is a robust economics literature on the pros and cons of regulating environmental quality at different jurisdictional levels that may be informative when determining whether federal regulation is appropriate as a substitute or complement to state or local regulation (e.g., Oates, 2002). See also Circular A-4 (OMB 2003) on Showing that Regulation at the Federal Level Is the Best Way to Solve the Problem.

53 Often consideration of different regulatory options is required or encouraged by statute (e.g., different stringencies of emissions standards). Any qualitative or quantitative analysis that supports these considerations should be summarized in the benefit-cost analysis, even if estimates of the benefits and costs of those options were not produced.


55 An exception may occur if the preferred option is at or near the limit of technical feasibility, in which case the analysis might not need to examine a more stringent option. However, it is possible that even if abatement of an environmental contaminant using on-site controls is technically infeasible, the value of the good or activity whose production creates the contaminant may be less than the harm the contaminant causes. In such circumstances a more stringent option that shifts production away from the good or activity should be evaluated.

56 In the course of developing a regulation, the decision maker may choose the more or less stringent option after weighing the results of the analysis. Doing so demonstrates the usefulness of the analysis. In this circumstance, the analysis should include an additional option to satisfy this guidance if time allows. If there is insufficient time to evaluate an additional option, the other two options should still be presented and the analysis should explain why the central option was not selected.

57 OMB's Primer on Regulatory Impact Analysis (OMB n.d.) provides similar guidance stating that "at a minimum, agencies should compare, with their preferred option, a more stringent and less stringent alternative, and assess the benefits and costs of the three possibilities, with careful consideration of which achieves the greatest net benefits."
incremental benefits and costs of successively more stringent regulatory options provides a clear indication of the most economically efficient option, provided important benefits and costs can be quantified.

Assessing at least three options applies in any circumstance. It is not adequate to solely evaluate the preferred option even for a final rule that establishes the option to be promulgated. Similarly, in cases where the form of the rule is dictated by statute, presenting multiple options is still necessary even though the Agency may have no discretion in the form of the rule even at the proposal stage. Because the regulatory analysis is meant to inform the public about the anticipated effects of the Agency’s final action and the options not pursued, it is imperative that the analyst fully assess multiple options.

The analysis should consider whether alternatives to federal regulation may suitably address the market failure. Alternatives may include using existing product liability rules to encourage firms to internalize the costs of the environmental damages, or the potential for state or local regulation. Even in cases when these options may not be available, the economic analysis should discuss the statutory requirement limitations and, if possible, estimate the opportunity cost of not being allowed to pursue them. There is no prohibition against analyzing these options as a way of estimating the opportunity cost of the limitation.

When a rule includes several distinct regulatory provisions, the benefits and costs of each provision should be analyzed both separately and jointly (i.e., as a package of provisions). Doing so may yield insights as to when certain regulatory requirements are not necessary or otherwise undesirable among others. Jointly analyzing provisions becomes more complicated when the existence of one provision affects the benefits or costs arising from another. Even so, it is still possible to evaluate each specific provision by estimating the net benefits of a regulatory option with and without that provision.

Ultimately, the number and choice of options to evaluate is a matter of judgment, but the analysis should strive for a balance between thoroughness and analytic capacity. Realistically, analyzing all possible combinations of provisions is impractical if the number is large and interactions between provisions are common. Generally, some options can be eliminated through a preliminary and less rigorous analysis, leaving a more manageable number to be evaluated in the formal benefit-cost analysis. For a proposed rule, it may be especially useful to provide economic analysis that illuminates important tradeoffs associated with key specific aspects of the rule on which the Agency is soliciting comment.

The analysis should carefully describe the policy design being evaluated and, when the costs or benefits vary substantially with alternative policy designs, assess alternative design options. The policy design includes the core regulatory approach as well as key features of its implementation and structure. Prescriptive regulation (e.g., technology, design or performance standards) is common in federal regulation. Performance standards, which specify the allowable emissions limit but not the way regulated entities must achieve that limit, are generally less costly than standards that dictate technologies or techniques. Economic analyses in some cases may include assessments of policy designs that currently are not statutorily allowed to highlight potential tradeoffs between the required approach and other more desirable approaches (for example, more flexible market-based approaches such as emissions taxes and allowance trading systems that may be prohibited).

In many cases, aspects of the market failure can help identify which types of regulatory approaches to consider. A key principle in the design of environmental regulations is that the regulatory structure and incentives should align with the environmental objective. A practical example is that, if the effect of emissions on human health depends on the proximity to the emission, then the optimal regulation should more stringently control emissions from emitters that are closer to population centers. Another example is that regulations should control emissions rather than the amount of

---

58 Sections 4.2 through 4.5 of these Guidelines provide a detailed description of different regulatory approaches.

59 Section 4.6 provides a detailed discussion of considerations for selecting among different regulatory approaches.
input associated with the emissions, if emissions monitoring costs are not too high relative to the costs of monitoring input use.

Variations in policy design features other than regulatory stringency and regulatory approach may help identify viable alternative regulatory approaches. While regulatory stringency (e.g., the level of a pollution standard) and the regulatory approach are important dimensions in establishing policy alternatives, options varying other policy design features, both alone and in combination, should also be considered (OMB 2003). In so doing, the analysis may identify approaches that increase net benefits or reduce the impact to certain entities. These features include but are not limited to:

- Compliance dates: Delaying when the regulation takes effect could reduce costs by allowing the regulated entities additional planning time which can be weighed against a possible reduction in benefits.
- Enforcement methods: Alternatives include regular on-site inspections, random monitoring, periodic reporting, and noncompliance penalties, which may have different costs and efficacy.
- Requirements for different-sized firms or facilities: In some cases, small firms or facilities may face proportionately higher compliance costs, especially if there are large fixed compliance costs. When a market-based approach cannot be used, varying the regulatory stringency or pollution requirement by firm size may increase economic efficiency.
- Requirements for different geographic regions: Differentiating requirements by region may be desirable if there is significant regional variation in pollution reduction benefits or the costs of compliance.

It is important to present both the total benefits and costs of each alternative and the incremental benefits and costs between the alternatives. Reporting the total benefits and costs for all policy options is important because any options where the benefits exceed the costs is an improvement in efficiency relative to future conditions absent the policy according to the potential Pareto principle. By this standard, selecting any option with positive net benefits would improve societal welfare. However, the most economically efficient option is the one that produces the largest increase in net benefits. While the option with the highest net benefits is obvious from the presentation of total benefits and costs, presenting the incremental benefits and costs of each option compared to the next less-stringent alternative helps to indicate if there is an even more economically efficient option other than those being considered. In general, economic efficiency is maximized (i.e., net benefits are highest) when incremental benefits are equal to incremental costs.

Carefully detailing the sources of the benefits and costs of a rule, rather than looking only at its total net benefits, may help identify policy options to consider. Some actions produce benefits from reductions in environmental contaminants other than those related to the statutory objective of the regulation. When the benefits associated with reductions in these other contaminants are a large share of total benefits, or net-benefits would be negative without them, the analysis should describe other options that directly regulate those contaminants. Furthermore, an analysis of a policy option in which the other contaminant(s) are regulated directly, either separately or simultaneously with the regulation being analyzed, may be warranted. Correspondingly, there may be costs from increases in environmental contaminants

---

60 The potential Pareto principle, or the compensation principle, states that economic welfare is improved by an action if the benefits of the action outweigh the costs (provided both benefits and costs can be measured accurately) because the gainers (those who benefit) could, theoretically, compensate the losers (those who bear the costs) and still be better off. Section A.3 of these Guidelines provides a further description of the potential Pareto principle.

61 The preferred option should also be reasonably robust to alternative baseline conditions. See the discussion of uncertainty in Chapter 5.

62 The statutory objective of the regulation is the specific objective of the statutory provision under which the regulation is promulgated.

63 All benefits and costs that result from a policy change, including from decrease in contaminants subject to the regulation and consequent increases or decreases in other contaminants, should be accounted for in a benefit-cost analysis. Determining whether an action may increase economic efficiency also requires accounting for all benefits and costs of an action. Furthermore, Executive Order 12866 and OMB’s Circular A-4 (2003) consistently affirm that all benefits and costs should be assessed in benefit-cost analyses of regulatory actions. Chapter 5 provides further discussion and guidance on how to treat in an economic analysis those benefits from environmental contaminants other than those related to the statutory objective.
other than those related to the statutory objective of the regulation that occur as a result of the regulatory action. As noted, analysis of additional options to mitigate these effects may be warranted if they are large.

Chapter 3 References


---

64 Such costs attributable to increases in other pollutants (and other environmental contaminants) should be accounted for even if future regulation might reduce them.
Chapter 4

Regulatory and Non-Regulatory Approaches to Pollution Control

This chapter describes several regulatory and non-regulatory approaches used in environmental policy making. The goals of this chapter are to introduce several important analytic terms, concepts, and approaches; to describe the conceptual foundations of each approach; and to provide additional references for those interested in a more in-depth discussion. This chapter covers the following four general approaches that a government agency can take to environmental policy making: (1) command-and-control regulation; (2) market-based incentives; (3) hybrid and other approaches; and (4) voluntary initiatives. While command-and-control regulations have been a commonly used method of environmental regulation in the United States, the EPA also employs other approaches. Market-based incentives and hybrid approaches offer the regulated community an opportunity to meet standards with increased flexibility and lower costs compared to many command-and-control regulations, while voluntary initiatives may allow environmental improvements in areas not traditionally regulated by the EPA. The chapter also includes a discussion of criteria used to evaluate the effectiveness of regulatory and non-regulatory approaches to pollution control.

4.1 Evaluating Environmental Policy

Once federal action is deemed necessary to address an environmental problem, policy makers have various options at their disposal to influence pollution levels. In deciding which approach to implement, policy makers must be cognizant of the constraints and limitations of each approach in addressing specific environmental problems. It is important to account for how political and information constraints, imperfect competition, or pre-existing market distortions interact with various policy options. Even if one approach is appealing from a social welfare perspective, it may be inconsistent with statutory requirements or generate additional concerns when considered along with other existing regulations. While any policy option under consideration must balance cost considerations with other important policy goals (including benefits), economic efficiency and cost-effectiveness are two economic concepts useful for framing the discussion and comparing these options.

4.1.1 Economic Efficiency

Economic efficiency can be defined as the maximization of social welfare. An efficient market is one that allows society to maximize the net present value (NPV) of benefits: the difference between a stream of social benefits and social costs over time. Under the efficient level of production, there is no way to rearrange production or reallocate goods such that someone is better off without making someone else worse off in the process. The efficient level of production occurs without government intervention in a market with no externalities or other market failures. Government intervention may be justified on economic efficiency grounds when a market failure or externality exists, in which case the

Baumol and Oates (1988), particularly Chapters 10-14; Kolstad (2010); Tietenberg and Lewis (2018); Phaneuf and Requate (2016) and Field and Field (2021) are useful references on the economic foundations of many of the approaches presented here.
government may attempt to determine the socially optimal point of production accounting for such externalities. Said differently, government analysts may evaluate which of the various policy approaches under consideration maximizes the benefits of reducing environmental damages, net the resulting costs.66

The socially optimal level of pollution is determined by reducing emissions until the benefit of abating one more unit (i.e., the marginal abatement benefit) — measured as a reduction in damages — is equal to the cost of abating one more unit (i.e., the marginal abatement cost).67 This is the level of pollution that maximizes the present value of net social benefits. In the simplest case, when each polluter chooses the level of emissions according to this decision rule, an efficient aggregate level of emissions is achieved such that the cost of abating one more unit of pollution is equal across all polluters. Any other level of emissions would result in a lower level of net benefits. It is also possible to evaluate policies based on their dynamic efficiency, the degree to which net benefits are equalized across periods in present value terms.

The reality of environmental decision-making is that Agency analysts are rarely in the position to select the economically efficient level of pollution. This is the case when legislation sets the level of abatement, or when it directs the Agency to set the level of abatement based on factors other than marginal benefits and/or costs. The scope of the standard may also influence its overall economic efficiency. For instance, the EPA may only have authority to set standards for a subset of an industry or must base standards in each subsector on different criteria. In cases where the Agency has some say in the stringency of a policy, its degree of flexibility in determining the approach varies by statute, which may also have efficiency implications. This may limit its ability to use certain policy instruments or consider certain technologies. It is also important to keep in mind analytic constraints. In cases where it is difficult to quantify benefits, cost-effectiveness may be the most defensible analytic framework.

4.1.2 Cost-Effectiveness

A policy is considered cost-effective when marginal abatement costs are equal across all polluters. The efficiency of a policy option differs from its cost-effectiveness. A policy is cost-effective if it meets a given goal at least cost, but cost-effectiveness does not encompass an evaluation of whether that goal has been set appropriately to maximize social welfare. All efficient policies are cost-effective, but it is not necessarily true that all cost-effective policies are efficient. For instance, a policy that mandates a certain emissions limit but allows firms to meet the limit in any way they choose - including paying other firms to reduce emissions - may be cost effective, but net benefits may not be maximized if the emissions limit was set either too high or too low.

4.2 Traditional Command-and-Control or Prescriptive Regulation

Many environmental regulations in the United States are prescriptive in nature (and are often referred to as command-and-control regulations).68 A prescriptive regulation can be defined as a policy that stipulates how much pollution an individual source or plant is allowed to emit and/or what types of control equipment it must use to reduce pollution. Such regulations are sometimes defined in terms of a source-level emissions rate. Despite the introduction of potentially more cost-effective methods for regulating emissions, this type of regulation is still commonly used and is sometimes

---

66 The terms market failure and externality are discussed in Chapter 3. Briefly, a market failure is a condition where the allocation of goods and services by a market is not efficient. A market failure may be due to externalities, market power, and/or inadequate or asymmetric information. An externality is a cost or benefit resulting from an action that is borne or received by parties not directly participating in the action. See Appendix A for further discussion of efficiency.

67 A certain amount of pollution may be efficient due to the economic concept of diminishing returns. For many pollutants, marginal social benefits decrease with each additional unit of abatement, while marginal social costs increase. Thus, it only makes sense to abate until the point where marginal benefits and marginal costs are just equal. Any abatement beyond that point will incur more additional costs than benefits. The efficient level of pollution will be zero only when the marginal benefits ≥ marginal costs for the first unit of emissions (e.g., new products containing asbestos).

68 Goulder and Parry (2008) refer to these as “direct regulatory instruments” because they feel that “command-and-control” has a “somewhat negative connotation.” Ellerman (2006) refers to them as prescriptive regulations.
statutorily required. It is almost always available as a “backstop” if other approaches do not achieve desired pollution limits.

Even when a prescriptive regulation is defined in terms of an emissions rate, it does not directly control the aggregate emission level. In such cases, aggregate emissions will depend on the number of polluters and the output of each polluter. As either production or market size increase, so will aggregate emissions. Even when the standard is defined in terms of an emission level per polluting source, aggregate emissions will still be a function of the total number of polluters.

When abatement costs are similar across regulated sources, a source-level standard may be reasonably cost-effective. However, when abatement costs vary substantially across polluters, reallocating abatement activities so that some polluters abate more than others could lead to substantial cost savings. For example, if reallocation were possible through a non-prescriptive, market-oriented approach, a polluter facing relatively high abatement costs could continue to emit at its current level but would have to pay an emissions tax or purchase allowances, while a polluter with relatively low abatement costs could reduce its emissions, allowing it to avoid the tax or sell its allowances (see Section 4.3 for more discussion of these approaches).69

Note that regulators can account for some variability in costs by allowing prescriptive regulations to vary according to size of the polluting entity, production processes, geographic location, or other factors. Beyond this, however, a prescriptive regulation usually does not allow for reallocation of abatement activities to take place — each entity is still expected to achieve a specified emissions rate or use certain abatement technologies. Thus, while pollution may be reduced to the desired level, it is often accomplished at a higher cost under a prescriptive approach.70

"Grandfathering" is a practice in which older polluters are exempted from new prescriptive regulations or are subjected to a less stringent standard than newer polluters. Grandfathering creates a bias against constructing new facilities and investing in new pollution control technology or production processes. As a result, grandfathered older facilities with higher emission rates tend to remain active longer than they would if the same emissions standard applied to all polluters (e.g., Helfand 1991; and Stavins 2006).

The most stringent form of prescriptive regulation is one in which the standard specifies zero allowable source-level emissions. For instance, the EPA has completely banned or phased out the use or production of chlorofluorocarbons (CFCs) and certain pesticides. This approach to regulation is potentially useful in cases where the level of pollution that maximizes social welfare is at or near zero. For cases where the optimal level of pollution is at or near zero, the literature also indicates that market-based incentives can sometimes be useful as a transition instrument for the phasing-out of a chemical or pollutant (e.g., Sterner and Coria 2012; and Kahn 2005).

Two types of prescriptive regulations exist: technology or design standards; and performance-based standards.

### 4.2.1 Technology or Design Standards

A technology or design standard, mandates the specific control technologies or production processes that an individual pollution source must use. This type of standard constrains firm behavior by mandating how a source must reduce pollution, regardless of whether such an action is cost-effective. Technology standards may be particularly useful in cases where the costs of emissions monitoring are high but determining whether a specific technology or production process has been put in place (and is operating properly) to meet a standard is relatively easy. However, since these types of standards specify the abatement technology required to reduce emissions, sources do not have an incentive to invest in

---

69 It is important to note that while emissions trading can equalize marginal abatement costs across polluters, there is no assurance that the change in marginal exposure or risk for the individuals in different locations will be equal. See Section 4.3.1 for more discussion.

70 See Tietenberg and Lewis (2014) for a discussion of empirical studies that examine the cost-effectiveness of prescriptive air pollution regulations. Of the ten studies included, eight found that prescriptive regulations cost at least 78 percent more than the most cost-effective strategy.
more cost-effective methods of abatement or to explore new and innovative abatement strategies that are not permitted by regulation.

4.2.2 Performance-based Standards

A performance-based standard requires that polluters meet a source-level emissions standard, but allows a polluter to choose among available methods to comply with the standard. At times, the available methods are constrained by additional criteria specified in a regulation. Performance-based standards that are technology-based do not specify a particular technology, but rather consider what is possible for available and affordable technology to achieve when establishing a limit on emissions.  

In the case of a performance-based standard, the level of flexibility depends on whether the standard specifies an emission level or an emission rate (i.e., emissions per unit of output or input). A standard that specifies an emission level allows a source to choose to implement an appropriate technology, change its input mix, or reduce output to meet the standard. An emission rate may be more restrictive depending on how it is defined. If the emissions rate is defined per unit of output, then it does not allow a source to meet the standard through a reduction in output. If the standard is defined as an average emissions rate over a certain time period, then the source may reduce output to meet the standard.

While performance-based standards encourage firms to meet the standard at lower cost, they generally do not provide incentives to reduce pollution beyond what is required to reach compliance. Also, because permitting authority is often delegated to the States, approval of a technology in one state does not ensure its use is allowed in another. For both of these reasons, there is little incentive for regulated firms to develop new, less expensive, and potentially superior technologies (See Swift, 2000; and Johnstone, et al., 2010 for more discussion).

4.3 Market-Based Approaches

Market-based approaches implemented by government to correct for an externality create an incentive for the private sector to incorporate pollution abatement into production or consumption decisions and to innovate to search for the least costly method of abatement. Market-based approaches can differ from more traditional regulatory methods in terms of economic efficiency (or cost-effectiveness) and the distribution of benefits and costs. Many market-based approaches minimize abatement costs, an objective that often is not achieved under command-and-control based approaches. Because market-based approaches do not mandate that each polluter meet a given emissions standard, they typically allow firms more flexibility than more traditional regulations and capitalize on the heterogeneity of abatement costs across polluters to reduce aggregate pollution efficiently. Environmental economists generally favor market-based policies because they tend to be less costly, they place lower information burden on the regulator, and they provide incentives for technological advances.

Market-based approaches can pose implementation challenges as well. Market-based policies require effective approaches to measure and monitor emissions. Quantifying and valuing marginal pollution damages can be difficult but is necessary to design socially optimal policies that balance marginal damages with marginal costs. In addition, measuring and forcing polluters to pay for emissions may lead to illegal dumping. Other considerations when contemplating the use of market-based policies include the distribution of compliance costs across firms or households over space and time.
compared with other regulatory approaches, political incentives to make the policy too lax, and the collection of revenues and distribution of economic rents that result from such programs.

Four classic market-based approaches are discussed in this section:

- Allowance trading systems;
- Emissions taxes;
- Environmental subsidies; and
- Tax-subsidy combinations.73

While operationally different (e.g., taxes and subsidies are price-based while allowance trading systems are quantity-based), these market-based instruments put similar incentives in place. This is particularly true of emissions taxes and cap-and-trade systems, which can be designed to achieve the same goal at equivalent cost. The sections that follow discuss each of these market-based approaches in turn.

4.3.1 Allowance Trading Systems

Several forms of emissions trading exist, including cap-and-trade systems and project-based trading systems. The common element across these programs is that sources can trade credits, offsets, or allowances so that those with opportunities to reduce emissions at lower costs have an incentive to do so. Emission-rate trading systems, a hybrid approach between tradable allowances and command-and-control, is discussed in section 4.4.1.3.

4.3.1.1 Cap-and-Trade Systems

In a cap-and-trade system, the government sets the level of aggregate emissions, allowances are distributed to polluters, and a market is established in which allowances may be bought or sold. An allowance is a right to emit one unit of pollution; polluters must own an allowance for each unit emitted. The price of emission allowances is determined by supply and demand in the market and can vary over time. Because different polluters incur different private abatement costs to control emissions, they are willing to pay different amounts for allowances. Therefore, a cap-and-trade system allows polluters who face high marginal abatement costs to purchase allowances from polluters with low marginal abatement costs, instead of installing expensive pollution control equipment or using more costly inputs. Cap-and-trade systems also differ from command-and-control regulations in that they aim to limit aggregate emissions over a compliance period rather than establish an emissions rate.

For a uniformly mixed pollutant where marginal damages are identical for all sources and in all locations, if the cap is set at the efficient level then the equilibrium price of allowances adjusts so that it equals the marginal external damages from a unit of pollution. This equivalency implies that any externality associated with emissions is completely internalized by the firm. For polluters with marginal abatement costs greater than the allowance price, the cheapest option is to purchase allowances and continue to emit. For polluters with marginal abatement costs less than the allowance price, the cheapest option is to reduce emissions and forego purchasing allowances (or to sell any allowances that they own at the market price). As long as the price of allowances differs from individual firms’ marginal abatement costs, firms will continue to buy or sell them. Trading will occur until marginal abatement costs equalize across all firms, ensuring a cost-effective allocation of pollution abatement.74 If the allowance price is lower than the marginal damages from pollution, this implies that the cap is set at an inefficiently high level.

---


74 Schmalansee and Stavins (2017) provide an overview of major emission trading programs over time and lessons learned with regard to implementation, system design, and performance.
When the government sells allowances at auction, the revenue generated represents a transfer from the purchasers to the government. The revenue generated by the allowance auction affords the government the opportunity to reduce pre-existing market distortions, to alter the distributional consequences of the policy, or to invest in other social priorities. Allowance auctions can be designed in a variety of ways. Typically, allowances are purchased through a bidding process that reveals' buyers' willingness-to-pay, with allowances going to be highest bidder.

The government could also decide to allocate allowances to polluters for free according to a specified rule. This represents a transfer from the government to polluting firms, some of which may find that the value of allowances received exceeds the firm's aggregate abatement costs (i.e., rents). The way in which allowance allocations occur can also affect firm entry and exit decisions. For example, when allowances are allocated based on historical emissions, some old, dirty plants may continue to operate to qualify for allowances.

The distribution of rents under cap-and-trade systems should be considered when comparing these systems with more traditional regulatory approaches. If the allowances are auctioned or otherwise sold to polluters, the distributional consequences will be similar to those from emissions taxes. If allowances are distributed for free to polluters, distributional consequences will depend on the allocation mechanism (e.g., historical output or inputs), on who receives the allowances, and on the ability of the recipients to pass their opportunity costs onto their customers. If new entrants must obtain allowances from existing polluters or through auction, then the policy maker should also consider barrier-to-entry effects. Differing treatment applied to new versus existing polluters can affect the eventual distribution of revenues, expenses, and rents within the economy.

Additional considerations in designing an effective cap-and-trade system include the number of market participants, transaction costs, banking, and consequences for noncompliance. The United States' experience suggests that a market characterized by low transaction costs and being “thick” with many buyers and sellers is critical if pollution is to be reduced at the lowest cost. This is because small numbers of potential traders in a market make competitive behavior unlikely, and fewer trading opportunities result in lower cost savings. Likewise, the number of trades that occur could be significantly hindered by burdensome requirements that increase the transaction costs associated with each trade.

Banking introduces increased flexibility into a trading system by allowing polluters to save unused allowances for future use. A firm may reduce emissions below the allowance level earlier (resulting in environmental benefits earlier than what would have occurred without banking), and bank remaining allowances to cover excess emissions or sell to another polluter at a later time. In this way, polluters that face greater uncertainty regarding future emissions or that expect increased regulatory stringency can bank allowances to offset potentially higher future marginal abatement costs.

For a cap-and-trade system to be effective, it is important to reliably measure and monitor emissions and establish predictable consequences for noncompliance. At the end of the compliance period, emissions at each source are compared to the allowances held by that source. If a source has fewer allowances than the monitored emission levels, it is in noncompliance and the source must provide allowances to cover its environmental obligation and pay a penalty.

Cap-and-trade systems for non-uniformly mixed pollutants have the potential to create temporal or spatial spikes or "hotspots" — areas in which the pollution level has the potential to increase as a result of allowance trading. While one potential solution to this problem is to adjust trading ratios (i.e., the rate at which allowances from one source can be

---

75 Economic rents are any payment to the owner of capital or a resource above what it would cost to induce them to engage in a certain behavior. In the context of a cap-and-trade market, these rents occur because firms are given something of value that can be bought and sold in the market (i.e., allowances) for free.

76 Text Box 4.3, presented later in this chapter, provides an example where thin markets resulted in few trades.

77 The U.S. Acid Rain Trading Program has high levels of compliance and requires fewer than 50 EPA staff to administer because penalties are automatically levied for each ton of excess emissions. Most of these Individuals certified and audited monitoring equipment and data (Napolitano, et al., 2007).
traded to another) to equalize the impact of particular polluters on overall environmental quality, determining the appropriate adjustments to these ratios can be costly and difficult. Another possible solution is zone-based trading.

Two reviews of the literature (Burtraw et al. 2005 and Harrington et al. 2004) find little evidence of spatial or temporal spikes in pollution resulting from the use of market-based approaches. In fact, market-based approaches have led to smoothing of emissions across space in some cases. These results come primarily from studies of the SO\textsubscript{2} and NO\textsubscript{x} trading programs (see Text Box 4.1). If the market-based policy is not carefully designed, the results may not transfer to other pollutants that have more localized effects.

Even if the cap is set at an efficient level when the system is created, changing conditions over time can result in inefficient levels of pollution control. For instance, the incentive to innovate means that the marginal abatement cost curve may shift downward over time as cheaper abatement options are introduced. If innovation causes the cost of pollution control to fall, the marginal cost of further decreasing pollution levels could drop well below the marginal benefit. Establishing a price floor below which allowances are removed from the market is one approach to dynamically adjusting the cap over time. Similarly, a price ceiling above which additional allowances are introduced to the market can be used to ensure that marginal costs do not rise too far above marginal benefits (Fell, et al. 2012).

### 4.3.1.2 Project-Based Trading Systems

Offsets and bubbles (sometimes known as “project-based” trading systems) allow restricted forms of emissions trading across or within sources to allow sources greater flexibility in complying with emission limits or facility-level permits.\textsuperscript{78}

An offset allows a new polluter to negotiate with an existing source to secure a reduction in the latter’s emissions. A bubble allows a facility to consider all sources of emissions of a specific pollutant within the facility to achieve an overall target level of emissions or environmental improvement. Offsets, which entail cross-firm emissions trading, have been historically hindered by high administrative and transaction costs due to the case-by-case negotiation to convert a technology or emission rate limit into tradable emissions per unit of time, to establish a baseline, and to determine the number of offsets generated or required (U.S. EPA 2001a). Regulators can improve the efficiency of offsets by allowing third parties, who are not themselves polluters, to participate in the market. Offsets have also been included in cap-and-trade programs for greenhouse gas emissions such as the Clean Development Mechanism of the Kyoto Protocol. Such systems allow entities covered under the cap to purchase offsets for emission reductions or sequestration from firms in industries or locations not covered under the program, further increasing the flexibility and reducing the costs of meeting the aggregate GHG emissions target.

---

\textsuperscript{78} See Bennear and Coglianese (2012) for an evaluation of how these types of flexibilities have worked in the United States.
In 1995, Title IV of the 1990 Clean Air Act Amendments established a cap-and-trade system for SO2 emissions to address the problem of acid rain. The 263 highest SO2-emitting units of 110 electricity-generating units (EGUs) were selected to participate in the first phase of the trading program. Emissions of SO2 in 1995 were limited to 8.7 million tons for facilities in Phase I. Of the EGUs that participated in Phase I, most were coal-fired units located east of the Mississippi River. Allowances were allocated to units on a historical basis, after which they could use the allowances, sell them to other units, or “bank” the allowances for use in subsequent years. Phase I EGUs were required to install continual emission monitoring systems, which allowed the government to easily monitor and enforce emission restrictions in accordance with the allowances. The second phase of the program, initiated in 2000, imposed a national SO2 emissions cap of 10 million tons and brought almost all SO2 generating units into the system.

Evaluations of the first phase of implementation suggest that the SO2 trading system significantly reduced emissions. Compliance costs were estimated to be between 15 and 90 percent lower than an equally stringent command-and-control alternative. The success of the program continued into the second phase. Chan, et al. (2018) estimated Phase II annual cost savings at several hundred million dollars compared to a simulated uniform performance standard.

Schmalensee and Stavins (2013) reported that emissions declined by 36 percent between 1990 and 2004, even as coal-fired electricity generation increased. One reason for such large emission reductions was the ability to bank allowances for future use. In addition, incentives to innovate continued to reduce abatement costs over time (e.g. Bellas and Lange 2011, Frey 2013). Railroad deregulation and investment by utilities in mining and infrastructure also played a role by making low-sulfur coal cheaper. That said, researchers have observed that there was less inter-firm trading than expected, which meant that marginal abatement costs were not equalized across plants (e.g., Swift 2001, Swinton 2004). Estimates of the SO2 allowance program’s annual benefits range from $59-116 billion with estimated annual costs of $0.5 to $2 billion (in 2000$) (Schmalensee and Stavins 2013).

Congress did not grant the EPA the authority to adjust the cap in response to new information on either the costs or benefits of reducing emissions. For this reason, the EPA pursued additional reductions in SO2 emissions via more traditional regulatory approaches, which restricted the ability of sources to trade and reduced allowance prices to zero by 2012 (Schmalensee and Stavins 2013).

4.3.2 Emissions Tax

Emissions taxes are imposed per unit of pollution emitted. Under an emissions tax, the polluter will abate emissions up to the point where the additional cost of abating one more unit of pollution is equal to the tax. The tax will result in an efficient outcome if it is set equal to the external damage caused by the last unit of pollution emitted.\textsuperscript{79}

For example, suppose that emissions of a toxic substance are subject to a tax based on the damages the emissions cause. To avoid the emissions tax, polluters find the cheapest way to reduce pollution. This may involve a reduction in output (and in the extreme, exiting the industry), a change in inputs, the installation of pollution control equipment, or a process change that prevents the creation of pollution. Polluters decide individually how much to control their emissions, based on the costs of control and the magnitude of the tax. The polluting firm reduces emissions to the point where the cost of reducing one more unit of emissions is just equal to the tax per unit of emissions. For any remaining emissions, the polluter prefers to pay the tax rather than to abate further. The government collects revenue that it may use to address other environmental problems, reduce other distorting taxes, or redistribute to finance other public services.\textsuperscript{80} While difficult to implement in the case of a non-uniformly mixed pollutant, policy makers can more closely approximate the ambient impact of emissions by incorporating adjustment factors for fluctuations in marginal damages across time, geographic area, or populations affected.\textsuperscript{81}

User or product charges are a variation on emissions taxes that are occasionally utilized in the United States. These charges may be imposed on users of publicly operated facilities or on intermediate or final products whose use or disposal harms the environment. User or product charges may be effective approximations of an emissions tax for those cases in which the product is closely related to environmental damage. User charges have been imposed on firms that discharge waste to municipal wastewater treatment facilities and on non-hazardous solid wastes disposed of in publicly operated landfills. Product charges have been imposed on products that release CFCs into the atmosphere, that utilize more gasoline (such as cars), or require more fertilizer. In practice, both user and product charges are sometimes set at a level only sufficient to recover the \textit{private costs} of operating the public system, rather than being set at a level to reduce pollution to the socially optimal level.

Emissions taxes should lead to outcomes similar to those from allowance trading systems when both are designed to achieve the same level of emissions. Rather than specifying the total quantity of emissions, taxes, fees, and charges specify the effective “price” of emitting pollutants. However, these two types of policy instruments differ in their usefulness when there is uncertainty about the costs or benefits of abatement. Section 4.6.4 discusses instrument choice under uncertainty.

4.3.3 Environmental Subsidies

Subsidies paid by the government to firms or consumers for reductions in pollution create the same abatement incentives as emissions taxes or charges. If the government subsidizes the use of a cleaner fuel or the purchase of a particular control technology, economic theory predicts that firms will switch from the dirtier fuel or install the control technology to reduce emissions up to the point where the additional private costs of control are equal to the subsidy. This type of subsidy is fundamentally different from the many subsidies already in existence in industries such as oil and gas, forestry, and agriculture, which exist for other reasons apart from environmental quality, and therefore can exacerbate existing environmental externalities. Environmental subsidies may be particularly useful when the government lacks the legal authority to impose an emissions tax or allowance trading system that restricts firms' right to pollute.

\textsuperscript{79} These taxes are called “Pigovian” after the economist, Arthur Pigou, who first formalized them. See Pigou (1932).

\textsuperscript{80} For more information on how the government can use revenues from taxes to offset distortions created by other taxes, see Goulder (2013) and McKibbin, et al. (2015).

\textsuperscript{81} See Fullerton, Leicester, and Smith (2010) for a discussion of the advantages and disadvantages of emissions taxes.
Unlike an emissions tax, an environmental subsidy lowers a firm’s total and average costs of production, encouraging both the continued operation of existing polluters that would otherwise exit the market, and the entry into the market by new firms that might otherwise face a barrier to entry. Given the potential entrance of new firms under an environmental subsidy, the net result may be a decrease in emissions from individual polluters but a proportionally smaller decrease (or even an increase) in the overall amount. For this reason, environmental subsidies and emissions taxes may not have the same aggregate social costs or result in the same degree of pollution control. An environmental subsidy also differs from an emissions tax because it requires government expenditure (vs. generating government revenue). Analysts should always consider the opportunity costs associated with collecting and spending public funds. It is possible to minimize the entry and exit of firms resulting from subsidies by redefining the subsidy as a partial repayment of verified abatement costs, instead of defining it as a per-unit payment for emissions reductions relative to a baseline. Under this definition, the subsidy only relates to abatement costs incurred and does not shift the total or average cost curves, thereby leaving the entry and exit decisions of firms unaffected. Defining the subsidy in this way also minimizes strategic behavior because no baseline must be specified. Government funding for research and development of new pollution control technologies is another form of subsidy.

Cost sharing, in which the government partially covers the private costs of selected actions to a firm or consumer, is another type of subsidy. For example, if the government wishes to encourage investment in specific pollution control technologies, the subsidy may take the form of reduced interest rates, accelerated depreciation, direct capital grants, and loan assistance or guarantees for investments. Cost-sharing policies targeted towards specific technologies may not induce broader changes in private behavior. In particular, such subsidies may encourage investment in pollution control equipment, rather than encouraging other changes in operating practices such as recycling and reuse, process redesign, or input substitution, which may not require such costly capital investments. However, in conjunction with direct controls, pollution taxes, or other regulatory mechanisms, cost sharing may influence the nature of private responses and the distribution of the cost burden. As is the case with emissions taxes, subsidy rates also can be adjusted to account for both spatial and temporal variability.

A government “buy-back” constitutes another type of subsidy. Under this system, the government offers a payment for the return of an older, high-polluting product, or a rebate on a new, cleaner substitute if the older model is turned in. For example, the EPA has funded changeout programs to encourage the replacement of old wood stoves with EPA-certified gas, electric, or wood appliances that substantially reduce indoor air pollution (US EPA 2014). Buy-back programs also exist to promote the scrapping of old, high-emission vehicles. In 2009, the U.S. government offered rebates to consumers trading in old, inefficient, but still drivable vehicles for new, fuel efficient vehicles to stimulate auto sales during a recession through a program called "Cash for Clunkers." An EPA grant program funds buy-back programs at the state and local levels to reduce diesel emissions.

The cost-effectiveness of buy-back programs and other subsidies depends on the degree to which they motivate behavior that would not have already occurred without the subsidy (an effect called "additionality"). In the Cash for Clunkers program, researchers estimated that most of the funds were received by consumers who would have purchased a vehicle in 2009 regardless, though the program did induce sales of more fuel-efficient vehicles than would have been purchased without the subsidy (Li et al. 2013). Similar to allowance trading systems, auctions can be incorporated into subsidy programs to incentive participants to reveal their opportunity costs and avoid payments in excess of this amount. In these programs, sometimes referred to as conservation or reverse auctions, subsidies are awarded to the lowest bidder (de Vries and Hanley 2016).

---

82 See Sterner and Coria (2012) and Goulder and Parry (2008) for a more in-depth discussion of how subsidies work and for numerous examples of subsidy programs in the United States and other countries.

83 Strategic behavior is a problem common to any instrument or regulation that measures emissions relative to a baseline. In cases where a firm or consumer may potentially receive funds from the government, they may attempt to make the current state look worse than reality, in order to receive credit for large improvements. If firms or consumers are responsible for paying for emissions above a given level, they may try to influence the establishment of that level upward in order to pay less in fines or taxes.
Environmental subsidies in the United States have been used to encourage proper waste management and recycling by local governments and businesses; the use of alternative fuel vehicles by public bus companies, consumers, and businesses; and conservation of wetlands and other undeveloped areas by property owners using cost-sharing measures. While most of these subsidies are not defined per unit of emissions abated, they can be effective when they induce behavioral changes that are closely related to environmental improvements.

### 4.3.4 Tax-Subsidy Combinations

Emissions taxes and environmental subsidies can also be combined to achieve the same level of abatement as achieved when the two instruments are used separately. One example of this type of instrument is referred to as a deposit-refund system in which the deposit operates as a tax and the refund serves as a partially offsetting subsidy. As with the other market instruments already discussed, a deposit-refund system creates economic incentives to return a product for reuse or proper disposal, or to use a particular input in production, provided that the deposit exceeds the private cost of returning the product or switching inputs.

Under a deposit-refund system, firms or consumers pay an upfront deposit on the production or use of certain goods. A refund is then provided if firms or consumers demonstrate that they used a cleaner form of production than is typical or engaged in proper disposal. In the case where a deposit-refund system encourages firms to use less-polluting inputs, a deposit on output induces the firm to reduce its use of all inputs, both clean and dirty (i.e., the output effect). The refund provides the firm with an incentive to switch to a specific input such as a cleaner fuel (i.e., the input substitution effect). The main advantages and disadvantages of deposit-refund systems are discussed in Walls (2013) and Fullerton and Wolverton (2001, 2005).

Deposit-refund systems have been designed to encourage consumers to reduce litter and increase recycling. The most prominent examples are deposit-refunds for plastic and glass bottles, lead acid batteries, toner cartridges and motor oil. Other countries have implemented deposit-refund systems on a wider range of products and behaviors, including to reduce the sulfur content of fuels (Sweden) and product packaging (Germany). Tax-subsidy combinations have also been discussed in the literature as a means of controlling pollution from motor vehicles, nonpoint source water pollution, cadmium, mercury, and the removal of carbon from the atmosphere.

A tax and subsidy combination functions best when there is a direct relationship between use of a product and emissions. For instance, a tax on the production or use of hydrochlorofluorocarbons (HCFCs) combined with a refund for HCFC recycled or collected in a closed system is a good proxy for an HCFC emissions tax.

The main advantage of a combined tax and subsidy is that both parts apply to a market transaction. Because the taxed and subsidized items are easily observable in the market, this type of economic instrument may be particularly appealing when it is difficult to measure emissions or to control illegal dumping. In addition, polluters have an incentive to reveal accurate information on abatement activity to qualify for the subsidy. Because firms have access to better information than the government does, they can measure and report their actions with greater precision and at a potentially lower cost.

Disadvantages of the combined tax-subsidy system may include potentially high implementation and administrative costs, and the political incentive to set the tax too low to induce proper behavior (a danger with any tax). Policy makers may adjust an emissions tax to account for temporal variation in marginal environmental damages, but a tax on output sold in the market cannot be matched temporally or spatially to emissions during production. In addition, when emissions are easily and accurately monitored (e.g., SO2 from power plants), other market incentives may be more appropriate. If a production process has many different inputs with different contributions to environmental damages, then it is necessary to tax the inputs at different rates to achieve efficiency. Likewise, if firms are heterogeneous and

---

84 See Fullerton and West (2010), Walls (2013), and Sterner and Coria (2012) for detailed descriptions of these and other examples of tax-subsidy combinations. The OECD Database on Policy Instruments for the Environment also provides numerous examples at: https://pinedatabase.oecd.org/.

4-12
In addition to the instruments discussed above, several other approaches have been used alone or in combination. This section discusses the following approaches:

- Combining prescriptive and market-based approaches;
- Liability rules;
- Information disclosure;
- Behavioral economics and "nudge" approaches.

### 4.4.1 Combining Prescriptive and Market-Based Approaches

In practice, some policies combine aspects of prescriptive and market-based policies. As such, they may not represent the most economically efficient approach. The cost of the policy is likely to be greater than what would be achieved using a pure market-based approach. Nevertheless, such approaches are appealing to policy makers because they often combine the certainty associated with a given standard or technology with the flexibility of allowing firms to comply at least cost. Safety-valve systems and tradable performance standards are two such hybrid approaches.

### 4.4.1.3 Safety-Valve Systems

Emissions taxes restrict costs by allowing polluters to pay a tax on the amount they emit rather than undertake excessively expensive abatement. Taxes, however, do not set a limit on the quantity of emissions and leave open the possibility that pollution may be excessively high. Some researchers suggest a policy that limits both costs and pollution, referred to as a “safety-valve” approach to regulation, which combines quantity and pricing mechanisms (See Roberts and Spence, 1976; and Spence and Weitzman, 1978). In the case of a prescriptive standard and tax combination, an emissions standard is imposed on all polluters, but polluters can pay a unit tax for emissions in excess of the standard. Safety-valve systems can also be entirely market-based, combining a cap-and-trade policy that sets an aggregate level of allowable emissions across all polluters with an emissions tax if allowance prices go above or below a certain level. Safety valves can be implemented as a symmetric approach that sets a price floor, or minimum price for allowances, in addition to a price ceiling (Burtraw, Palmer, and Kahn 2010).

This policy combination has several attractive features. First, it protects against excessively damaging pollution levels. Combining approaches allows for more certainty in the expected environmental and health effects of the policy than would occur with a pricing approach alone. Second, high abatement cost polluters can defray costs by paying the...
emissions fee instead of cleaning up. This feature preserves the flexibility of emissions taxes: overall abatement costs are lower because polluters with low abatement costs reduce pollution while polluters with high abatement costs pay taxes.

4.4.1.3 Tradable Performance Standards

Rather than establish an emissions cap, the regulatory authority under a tradable performance standard establishes a performance standard or emissions rate. Sources that perform better than the standard can earn credits and sell them to sources that perform worse. A credit allows a source to emit one unit of a pollutant in excess of what would normally be permitted (e.g., reducing emissions below a baseline or existing emissions cap).88

In rate-based trading systems, sources able to reduce their emissions rate at low cost have an incentive to do so since they can sell the resulting credits to those sources facing higher costs of abatement. However, emissions may increase under these programs if sources increase their utilization or if new sources enter the market. Therefore, the regulating authority may need to periodically impose new rate standards to achieve and maintain the desired emission target, which in turn may lead to uncertainty in the long term for the regulated sources. Rate-based trading programs have been used in the United States to phase out lead in gasoline (Newell and Rogers 2006, Schmalensee and Stavins 2017) and to control emissions of light-duty and heavy-duty vehicles (U.S. EPA and Department of Transportation 2012; U.S. EPA and Department of Transportation 2016).

In addition, the EPA's Renewable Fuel Standard and state Renewable Portfolio Standards both set standards requiring the use of particular technologies (bio-based transportation fuels and renewable electricity sources, respectively), but they incorporate tradable credits so that firms can meet the overall standard at least cost. While a standard and pricing approach does not necessarily ensure the maximization of social welfare, it can lead to the most cost-effective method of technology adoption or pollution abatement.

4.4.2 Information Disclosure

Disclosure of environmental information is required by certain EPA regulations.89 Disclosure requirements attempt to minimize inefficiencies in regulation associated with asymmetric information, such as when a firm has more and better information on what and how much it pollutes than is available to the government or the public. By collecting and making such information publicly available, firms, government agencies, and consumers can become better informed about the environmental and human health consequences of their production and consumption decisions. In some cases, the availability of this information may also encourage more environmentally benign activities and discourage environmentally detrimental ones. For example, warning labels on hazardous substances that describe risks or safe-handling procedures may encourage consumers to take greater precautions or switch to less damaging substitutes. Similarly, a community with information on a nearby firm’s pollution activity may exert pressure on the firm to reduce emissions, even if formal regulations or monitoring and enforcement are weak or nonexistent.90

Requirements for information disclosure need not be tied explicitly to an emissions standard; however, such requirements are consistent with a standard-based approach because the information provided allows a community to easily understand the level of emissions and the polluters’ level of compliance with existing standards or expectations. As is the case with market-based instruments, polluters still have the flexibility to respond to community pressure by reducing emissions in the cheapest way possible.

88 Ellerman, et al. (2003) use the terms credit and allowance interchangeably.
89 See OMB (2010b) for guidance issued to regulatory agencies on the use of information disclosure and simplification in the regulatory process. Information disclosure can also be an important component of non-regulatory EPA programs.
90 For more information on how information disclosure may help to resolve market failures, see Pargal and Wheeler (1996), Tietenberg (1998), Tietenberg and Wheeler (2001), and Brouhle and Khanna (2007).
The use of information disclosure or labeling rules has other advantages. When expensive emissions monitoring is required to collect such information, reporting requirements that switch the burden of proof for monitoring and reporting from the government to the firm might result in lower costs, because firms are often in a better position to monitor their own emissions. However, random inspections may be needed to ensure that monitoring equipment functions properly and that firms report results accurately.

While information disclosure has advantages, it is important to keep several caveats in mind when considering this method of environmental regulation. Information disclosure alone does not typically result in a socially efficient level of pollution when externalities are present. As discussed previously in Chapter 3, certain conditions are necessary for a private agreement between affected parties to lead to an efficient level of pollution as outlined by Coase (1960). These include low transaction costs and the possibility of bargaining, which are more likely when a small number of individuals are involved. In addition, the amount of pressure a community exerts on an emitting plant may be related to socioeconomic status. Poorer, less-educated populations tend to exert far less pressure than communities with richer, well-educated populations (see Hamilton, 1993; Arora and Cason, 1999; and Earnhart, 2004). The effect that public pressure has on behavior may also vary by firm and depend on factors such as the firm’s market power and societal reputation. Finally, consumers may not understand the health risks associated with pollution exposure and do not always act to further their own best interests even when complete information is available. As discussed in section 4.4.4, the behavioral economics literature has documented some examples in the latter category.

EPA-led information disclosure efforts include the Toxics Release Inventory (TRI) and the mandatory reporting of greenhouse gases (GHG). Both the TRI and the GHG Reporting Program require firms to provide the government and public with information on pollution at each plant, on an annual basis, if emissions exceed a threshold. There are also consumer-based information programs targeting the risks of specific toxic substances, the level of contamination in drinking water, the dangers of pesticides, and air quality index forecasts for more than 300 cities. There is some evidence in the literature regarding the impact of TRI reporting on firm value: the most polluting firms experience small declines in stock prices on the day TRI emission reports are released to the public. Firms that experienced the largest drop in their stock prices also reduced their reported emissions by the greatest quantity in subsequent years (Hamilton, 1995).91

4.4.3 Liability Rules

Liability rules are legal tools that individuals (or the government) can use to force polluters to pay for environmental damages after they occur. These instruments serve two main purposes: (1) to create an economic incentive for firms to incorporate careful environmental management and the potential cost of environmental damages into their decision-making processes; and (2) to compensate harmed individuals when damages occur. These rules are used to guide compensation decisions when the court rules in favor of the victim. Liability rules can serve as an incentive to polluters. To the extent that polluters are aware that they will be held liable before the polluting event occurs, they may minimize involvement in activities that inflict damages on others. In designing a liability rule, it is important to evaluate whether damages depend only on the amount of care taken on the part of the polluter or also on the level of output; and whether damages are only determined by polluter actions or are also dependent on the behavior of the harmed individuals. For instance, if harmed individuals do not demonstrate some standard of care to avoid damages, the polluter may not be held liable for the full amount. If damages depend on these other factors in addition to polluter actions, then the liability rule should be designed to provide adequate incentives to address these other factors.

While a liability rule can be constructed to mimic an efficient market solution in certain cases, there are reasons to expect that this efficiency may not be achieved. First, payments need not reflect the social damages. The amount that polluters are required to pay after damages have occurred is dependent on the legal system and may be limited by an

---

91 Konar and Cohen (1997); Khanna, Quimio, and Bojilova (1998); Bui and Mayer (2003); Banzhaf and Walsh (2008); and Mastromonaco (2015) also have investigated how the TRI has affected firm behavior, stock market valuation, and housing markets.
inability to prove the full extent of damages or by the ability of the firm to pay. Second, liability rules can generate relatively large costs, both in terms of assessing the environmental damage caused and the resources used to take legal action. Segerson (1995), and Alberini and Austin (2001) discuss different types of liability rules and the efficiency properties of each.

Liability rules are most useful in cases where damages requiring compensation are expected to be stochastic (e.g., accidental releases), and where monitoring firm compliance with regulatory procedures is difficult. Finally, strict liability rules can create disincentives for the redevelopment of contaminated land because newly involved firms become liable for past contamination (Jenkins, Kopits, and Simpson 2009). Depending on the likely effectiveness of liability rules to provide incentives to firms to avoid damages, they can be thought of as either an alternative or a complement to other regulatory approaches.

Strict liability and negligence are two types of liability rules relevant to polluters. Under strict liability, polluters are held responsible for all health and environmental damage caused by their pollution, regardless of actions taken to prevent the damages. Under negligence, polluters are liable only if they do not exhibit “due standard of care.” Regulations that impose strict liability on polluters may reduce the transactions costs of legal actions brought by affected parties. This may induce polluters to alter their behavior and expend resources to reduce their probability of being required to reimburse other parties for pollution damages. For example, they may reduce pollution, dispose of waste products more safely, install pollution control devices, reduce output, or invest in added legal counsel.

Liability rules have been used in the remediation of contaminated sites under the Comprehensive Environmental Response, Compensation and Liability Act (CERCLA), also known as Superfund, and under the Corrective Action provisions of the Resource Conservation and Recovery Act (RCRA). The Small Business Liability Relief and Brownfields Revitalization Act eased some of CERCLA’s liability provisions to encourage the redevelopment of potentially contaminated industrial sites, known as brownfields.

4.4.4 Behavioral Economics and "Nudge" Approaches

The neoclassical economics paradigm that has helped inform the design of market-based and other policy instruments makes several simplifications about human behavior—for instance, that people are rational, well-informed, self-interested, and disciplined. While these may be reasonable assumptions in many contexts, they do not always hold in the real world. Behavioral economics is a subfield at the intersection of economics and psychology that examines departures from the neoclassical or standard economics model. Such behavioral anomalies include cognitive limitations, altruism, inequality aversion, procrastination, status quo bias, and loss aversion, among others. Behavior that is altruistic, short-sighted, or inattentive may have important implications for the way environmental policies are designed and enforced.

Insights from behavioral economics present the opportunity to design policies that “nudge” people to make choices that improve their well-being. Nudges have been proposed as an approach to encourage socially beneficial actions by making small changes to the context in which people make decisions. Thaler and Sunstein (2008) define a nudge as “any aspect of the choice architecture that alters people’s behavior in a predictable way without forbidding any options or significantly changing their economic incentives,” elaborating that, “the intervention must be easy and cheap to avoid. Nudges are not mandates.”

While market-based policies are typically designed to correct externalities, nudges may be especially relevant in situations where the market under-provides environmental quality due to lack of information, cognitive limitations, procrastination, or other behavioral anomalies. Inattentive or impatient behavior may help explain some consumers’ reluctance to invest in energy-saving appliances or fuel-efficient cars that cost more upfront but save money in the long

---

92 Loss aversion occurs when individuals facing risky choices place greater weight on losses compared to gains of an equivalent value. Empirical research suggests that many people tend to give losses double the weight of gains (Kahneman and Tversky 1979, Tversky and Kahneman 1992). Loss aversion can contribute to status quo bias, which describes a preference for avoiding any change from the current situation.
run. Altruism and social norms may lead people to purchase eco-labeled products even absent regulation or price signals.

In contrast to the use of information disclosure alone as a policy instrument, nudges emphasize the visual design, timing, delivery method, and other aspects of the way information is presented to make it more salient and useful to the public. Other general strategies that have been used as nudges include default rules that require individuals or firms to opt out of a program instead of opting in, moral suasion or pro-social messages that appeal to a sense of altruism or fairness, ordering choices to put the most beneficial option first, and the use of social norms that tap into individuals’ desire to match or outperform their peers. Examples of nudges outside the realm of environmental policy include automatic enrollment of employees into retirement savings plans (Madrian and Shea 2001) and rearranging cafeterias to make healthy foods more convenient or eye-catching (Hanks et al. 2012).

There are many other potential applications of behavioral economics to environmental policy. For example, research has shown that providing residential consumers with real-time information about electricity consumption and prices can reduce electricity use, which can lead to decreased pollution from fossil-powered electricity generation. Signals conveyed visually, such as with a “glowing orb” that changes color to reflect changes in prices or demand, have been shown to be particularly effective. Residential consumers who received reports comparing their own consumption of water or electricity to that of their neighbors also reduced their resource consumption, with effects ranging from about two to five percent (Allcott 2011b, Ferraro and Price 2013). See Text Box 4.2 for a description of EPA examples.

Behavioral economics insights can be used in tandem with other policy instruments. The implementation of plastic bag taxes provides one example. Standard economic models predict that individual consumers will respond similarly to market incentives regardless of whether they are presented as a tax on damaging activities or a subsidy for beneficial activities. However, research has found that consumers faced with a fee for disposable bags cut their bag use by more than 40 percent, but no change occurred in response to a subsidy for reusable bag use (Homonoff 2018). This result is consistent with loss aversion and suggests that consumer responsiveness to market-based policies can depend on how

---

93 Executive Order 13707, “Using Behavioral Insights to Better Serve the American People” (Sept. 15, 2015) encouraged federal agencies to consider behavioral science strategies with particular attention to 1) access to programs, 2) presentation of information to the public, 3) the structure of choices within programs, and 4) the design of financial and non-financial incentives.

94 Shogren and Taylor (2008), Shogren, Parkhurst, and Banerjee (2010), and Croson and Treich (2014) provide in-depth discussions of the intersection between behavioral economics and environmental economics.

95 Allcott (2011a) and Jessoe and Rapson (2014) focused on real-time electricity pricing, while Houde et al. (2013) examined the effect of real-time electricity consumption information.
the incentives are framed. That said, nudges do not always complement financial incentives; in some cases, one approach can “crowd out” the other.\(^96\)

Nudges that are effective in one situation are not always transferable to different contexts. For example, the residential electricity consumption reports mentioned above led to larger reductions in electricity use for high-user households and for environmentalists, while they have been less effective for other households (Allcott 2015). This observation highlights the importance of using rigorous empirical approaches such as randomized controlled trials to test the effectiveness of new nudges before adopting them on a wide scale (List and Metcalfe 2014, Allcott and Mullainathan 2010, Hahn and Metcalfe 2016).

### 4.5 Voluntary Initiatives

The EPA has pursued a number of non-regulatory approaches that rely on voluntary initiatives to reduce emissions and other environmental hazards. These programs are usually not intended as substitutes for regulation, but instead act as complements to existing regulation. Many of the EPA’s voluntary programs encourage polluting entities to go beyond what is mandated by existing regulation. Other voluntary programs address environmental quality in areas that policy makers expect may be regulated in the future but are currently not regulated.\(^97\)

The foundation for these voluntary initiatives rests on a “pollution prevention” approach to environmental management choices. In the Pollution Prevention Act of 1990, Congress established a national policy that:

- Pollution should be prevented or reduced at the source whenever feasible;
- Pollution that cannot be prevented should be recycled in an environmentally safe manner whenever feasible;
- Pollution that cannot be prevented or recycled should be treated in an environmentally safe manner whenever feasible; and

---

\(^96\) Preliminary research on electricity consumers in both Ecuador and the United Kingdom found that combining social comparison messages with financial incentives yielded no improvement in conservation compared to using either approach alone (Dolan and Metcalfe 2013, Pellerano et al. 2017).

\(^97\) While this chapter only discusses government-led voluntary initiatives at the EPA, other government agencies, industry, non-profits, and international organizations have also organized voluntary initiatives designed to address environmental issues. These initiatives are beyond the scope of this chapter, which limits itself to a brief description of policy options available to the EPA.
Disposal or other release into the environment should be employed as a last resort and should be conducted in an environmentally safe manner.

The EPA typically designs its voluntary programs through consultation with affected industries or consumers. In many cases, voluntary programs facilitate problem solving between the EPA and industry because information on practices that reduce pollutants and waste at the source are shared through the consultative process. Voluntary programs also frequently encourage peer education and resource sharing among program participants.

Partner organizations have included small and large businesses, citizen groups, state and local governments, universities, and trade associations. Voluntary programs tend to have either broad environmental objectives targeting a variety of firms from different industries or focus on more specific environmental problems relevant to a single industrial sector. These programs strive to provide participating firms with targeted and effective technological expertise and assistance.\(^98\)

Voluntary programs can use the following four general methods to achieve environmental improvements: (1) require firms or facilities to set specific environmental goals; (2) promote firm environmental awareness and encourage process change; (3) publicly recognize firm participation; and (4) use labeling to identify environmentally responsible products. These methods are not mutually exclusive, and many voluntary programs use a combination of methods.

Goal setting is a common method used in the design of voluntary programs. Implementation-based goals are typically EPA-specified, program-wide targets designed to provide a consistent objective across firms. Target-based goals are usually qualitative and process-oriented so that firms may individually set a unique target. The EPA's Green Power Partnership specifies implementation-based goals to promote the use of renewable energy. These goals vary with annual electricity use; larger users have more stringent goals. The EPA's WasteWise program is an example with a target-based goal.

Programs designed to promote environmental awareness and process change within firms often involve implementing a system to evaluate firms’ ongoing operations and to provide information on new technologies. Examples of this approach include the SmartWay program, which encourages fuel efficient trucking and logistics, and the Green Suppliers Network program, which provides technical reviews and suggestions on how to eliminate waste from production processes. These programs may also promote or recognize use of third-party industry standards for products and materials.

Voluntary programs that publicly recognize firm participation are designed to provide green consumers and investors with new information that may alter their consumption and investment patterns in favor of cleaner firms. Firms may also use their environmental achievements to differentiate their products from competitors’ products.\(^99\) These information and firm differentiation effects are the intent of the Green Power Partnership and the WasteWise program.

Product labeling can be applied to either intermediate inputs in a production process or to a final good. Labels on intermediate goods encourage firms to purchase environmentally responsible inputs. Labels on final goods allow consumers to identify goods produced using a relatively clean production process. For example, products that use chemicals that are less harmful to human health and the environment may be eligible for the Safer Choice (formerly called Design for the Environment) labels. The WaterSense program provides a label for independently certified water-efficient products. Section 4.4.4 and Text Box 4.2 discuss how labeling can be made more effective by using behavioral economics concepts.

\(^98\) See Prakash and Potoski (2012), Borck and Coglianese (2009), Brouhle, Griffiths, and Wolverton (2005), and Lyon and Maxwell (2007) for discussions of how voluntary programs work and how they are used in U.S. environmental policy making.

\(^99\) See Konar and Cohen (2001); Videras and Alberini (2000); Brouhle, Griffiths, and Wolverton (2005); Morgenstern and Pizer (2007); and Borck and Coglianese (2011) for more information on the main arguments for why firms participate in voluntary programs.
The economic literature evaluating the efficacy of voluntary programs is decidedly mixed. The majority of empirical studies have focused on a few large, multi-sector programs such as 33/50, Green Lights, and ENERGY STAR. For these programs, there is some evidence of success in reducing participant emissions. However, studies generally fail to account for reductions would have occurred absent the program. Comparing emissions relative to pre-program levels rather than a true business-as-usual counterfactual may overstate these reductions. The potential for beneficial information or technology spillovers from program participants to other firms in the target industry also makes it more difficult to measure a program’s impact (Lyon and Maxwell 2007). Researchers have been less successful in demonstrating that voluntary programs have led to greater emission reductions than would have occurred without the program in place. One thread of literature points to the positive impact of a regulatory threat on voluntary program effectiveness. When the threat of regulation is weak, abatement levels are lower. However, when the threat of regulation is strong, levels achieved are closer to those under optimal regulatory action.100

4.6 Selecting the Appropriate Policy Approach

Selection of the most appropriate policy approach depends on a wide variety of factors, including: 101

- The type of market failure being addressed;
- The specific nature of the environmental problem;
- The type of pollutant information that is available and observable;
- The degree of uncertainty surrounding costs and benefits;
- Concerns regarding market competitiveness;
- Monitoring and enforcement issues;
- Potential for exacerbating economy-wide distortions; and
- The goals of policy makers.

Note that the policy ultimately chosen will also depend on statutory and other legal limitations on what approaches may be used.

4.6.1 The Type of Market Failure

There are two main types of market failure that are commonly addressed through environmental policy. The first, externality, occurs when firms or consumers fail to integrate into their decision making the impact of their own production or consumption decisions on entities external to themselves. The second type of market failure, asymmetric information, occurs when firms or consumers are unable to make optimal decisions due to lack of information on available abatement technologies, emission levels, or associated risks. Market-based or hybrid instruments that incorporate the marginal external damages of a unit of pollution into a firm or consumer’s cost function address the first type of market failure. Information disclosure, labeling, or nudges are often suggested when the second type of market failure occurs. Information disclosure alone may be sufficient if private firms and individuals will act to address an environmental problem once information has been disseminated. If their actions are also affected by behavioral factors such as inattention, impatience, and social norms, coupling information with nudges could be a more effective approach. Multiple policy instruments can also be optimal when an environmental problem is caused by multiple market failures, or when exogenous constraints limit use of the first-best policy (Bennear and Stavins 2007).


101 Helpful references that discuss aspects to consider when comparing among different approaches include Hahn and Stavins (1992), OECD (1994a, 1994b), Portney and Stavins (2000), and Sterner and Conia (2012).
4.6.2 The Nature of the Environmental Problem

The use of a specific approach is often motivated by the nature of the environmental problem. Do emissions derive from a point source or a nonpoint source? Which media are affected (e.g., air, water, land)? Do the pollutants persist in the environment (e.g., heavy metals or dioxins in soil) or dissipate rapidly (e.g. volatile organic compounds)? Are emissions uniformly mixed or do they vary by location? Does pollution originate from stationary or mobile sources? Point sources, which emit at identifiable and specific locations, are much easier to control than diffuse and often numerous nonpoint sources, and therefore are often responsive to a wide variety of instruments. Although nonpoint sources typically are not regulated by the EPA, the pollution emitted from a nonpoint source is. Clearly, this makes the monitoring and control of nonpoint source emissions challenging (See Text Box 4.3). In instances where both point and nonpoint sources contribute to a pollution problem, a case can be made for a tax-subsidy combination or an allowance trading system. Under these alternatives, emissions from point sources might be taxed while nonpoint source controls are subsidized.

Flow pollutants that dissipate quickly are responsive to a wide variety of market and hybrid instruments for emissions control. In contrast, stock pollutants that persist in the environment may require strict limits to prevent bioaccumulation or detrimental health effects at small doses, making direct regulation a potentially more appealing approach. If these limits are not close to zero, then instrument options include a standard-and-pricing approach or an allowance trading approach that defines trading ratios to ensure that emission standards are not violated at any given source. These same instruments are appealing when pollutants are not uniformly mixed across space. In the case of non-uniformly mixed emissions, it is important to account for differences in baseline pollution levels and in emissions across more and less polluted areas.

Stationary sources of pollution are easier to identify and control through a variety of market instruments than are mobile sources. Highly mobile sources are usually numerous, each emitting a relatively small amount of pollution. Emissions therefore vary by location, and damages can vary by time of day or season. For example, health impacts associated with vehicle emissions may be substantially larger during rush hour because roads are congested, and cars spend time idling or in stop-and-go traffic. Differential pricing of resources used by these mobile sources (such as higher tolls on roads or greater subsidies to public transportation during rush hour) is a potentially useful tool.

4.6.3 The Type of Pollutant Information that is Available and Observable

The selection of approach may depend on the available data. Is the level of pollutant observable or measurable? Or will the level need to be imputed based on inputs and technology used? Are the sources heterogeneous? Does the pollutant vary across time and space? Are information technologies available to improve data collection? When the pollutant concentration can be directly and easily measured then it is possible to directly regulate the level of the pollutant. But if monitoring costs are high, it may be easier to target a related input or require a technology known to reduce pollutants by a certain amount. The pollutant levels can be imputed based on regulation placed on the input or the technology used.

The link between pollution and heterogeneous sources is often difficult and costly to determine, and costs may increase if the pollutant levels vary over time. Uniform policies are often used for the sake of simplicity. However, information technologies such as continuous emissions monitoring (CEM) equipment or geographical information systems (GIS) can

---

102 For a detailed discussion of how the nature of the environmental problem affects instrument choice, see Kahn (2005), Goulder and Parry 2008, Parry and Williams (1999), Tietenberg and Lewis (2014), and Sterner and Coria (2012).
Text Box 4.3 - Water Quality Trading of Nonpoint Sources

In 2003, the EPA issued a “Water Quality Trading Policy” (U.S. EPA 2003d) that encourages states and tribes to develop and implement voluntary water-quality trading to control nutrients and sediments in areas where it is possible to achieve these reductions at lower costs. A 2019 memo announced additional flexibilities available to states and tribes to further facilitate the uptake of water quality trading, particularly between point and non-point sources. The memo cites the increased availability of effective non-point emission reducing technologies and practices and enhanced monitoring capabilities as reasons to modernize the 2003 policy (U.S. EPA 2019).

Under the Clean Water Act states are required to establish Total Maximum Daily Loads (TMDL) of pollutants for impaired water bodies. The TMDL is not a regulation and does not establish an enforceable cap on discharges to the watershed, but it does provide a method for allocating pollutant discharges among point and nonpoint sources. Point sources are regulated under the Clean Water Act by the EPA and, as such, are required to hold National Pollutant Discharge Elimination System (NPDES) permits that limit discharges. Where a TMDL exists, the point source NPDES discharge limit is informed by the TMDL allocation. Nonpoint sources are not regulated under the Clean Water Act. However, many water bodies are still threatened by pollution from these sources. Nutrients and sediment from urban and agricultural runoff have led to water quality problems that limit recreational uses of rivers, lakes, and streams; create hypoxia in the Gulf of Mexico and other coastal waters; and decrease fish populations in the Chesapeake Bay and other areas.

To account for various uncertainties and differences associated with nonpoint source pollution, trading ratios are often applied. These ratios account for the differential effects resulting from a variety of factors, which may include:

- location of the sources in the watershed relative to the downstream area of concern;
- distance between the allowance buyer and seller;
- uncertainty about nonpoint source reductions;
- equivalency of different forms of the same pollutant discharged by the trading partners; and
- additional water quality improvements above and beyond those required by regulation.

Trading can also be used as a tool to allow continued growth in production, while providing nonpoint sources with an incentive to reduce pollution through participation in the market. To the extent that it is cheaper for a nonpoint source to reduce pollution than to forgo revenues earned from the sale of any unused credits to point sources, economic theory predicts that the nonpoint source will choose to emit less pollution.

As of 2014, EPA had identified 19 nutrient trading programs in 11 states, with the majority of trades occurring in just three states, Connecticut, Pennsylvania, and Virginia (GAO 2017). Trading has been limited in many of these programs for several reasons. First, as previously mentioned there is no enforceable cap on discharges that applies to both point and nonpoint sources within a watershed. Reductions by nonpoint sources are voluntary absent state-level mandates. Point-source dischargers often explore trading as a way to expand production while meeting the requirements of their individual permits, but there is no general signal in the market to do so. Second, these are often thin markets (i.e., markets with few trades). The way that the market is designed or trading ratios are established can make it difficult or expensive for an entity to identify and complete a trade. Third, while Best Management Practices (BMPs) are typically used to define a pollution reduction credit from a nonpoint source, uncertain or changing climatic conditions, river flow, and stream conditions make it difficult to measure the effect of a BMP on downstream water quality. Such uncertainty also makes measuring and enforcing a pollution reduction from a nonpoint source difficult. Fourth, encouraging nonpoint source involvement in trading is challenging. Finally, it is difficult to define appropriate trading ratios between point and nonpoint sources (Morgan and Wolverton 2008; U.S. EPA 2008).

be used to link sources to pollutant levels. In these cases, policies that make use of this information may be used and often can reduce costs. As technology improves or more data become available, analysts should consider reassessing the regulation design.\textsuperscript{103}

\textsuperscript{103} For more information see Xabadia, Goetz, and Zilberman, 2008.
4.6.4 Uncertainty in Abatement Costs or Benefits

When abatement costs and benefits are certain, price-based instruments (e.g., emissions taxes) and quantity-based instruments (e.g., cap-and-trade) are theoretically equivalent and can be designed to achieve the same outcomes. However, this result may not hold when there is uncertainty about the benefits and costs of pollution control, or when marginal benefits and costs change substantially with the stringency of the pollution control target. If uncertainty associated with the abatement costs exists but damages do not change much with additional pollution, then policy makers can limit costs by using a price instrument without having much impact on the benefits of the policy. If, on the other hand, there is more uncertainty associated with the benefits of controlling pollution and policy makers wish to guard against high environmental damages, a quantity instrument is preferable. In this way, the policy maker can avoid potentially costly or damaging mistakes. The policy maker should also be aware of any discontinuities or threshold values above which sudden or large changes in damages or costs could occur in response to a small increase in the required abatement level (Pindyck 2007). Hybrid approaches that combine features of price and quantity instruments can also address uncertainty (Pizer 2002; see Section 4.4.1.3).

4.6.5 Market Competitiveness

Market power is the ability of a firm to raise the price of a good or service above the marginal cost of production. This situation results in lower output than would otherwise occur in a competitive market. Market power is a type of market failure because the allocation of goods and services is not efficient, absent other externalities or distortions, which results in a loss in economic welfare. Policy instruments that cause firms to restrict output (e.g., an emissions tax) may create additional inefficiencies in sectors where firms have some degree of market power. A combination of market-based instruments may work more effectively than a single instrument in this instance. To the extent that cost burdens are differentiated, the use of certain market-based instruments may cause a change in market structure that favors existing firms by creating barriers of entry and allowing existing firms a certain amount of control over price. Cap-and-trade systems that set aside a certain number of allowances for new firms, for instance, may guard against such barriers.

4.6.6 Monitoring and Enforcement Issues

Policy approaches differ in the degree of effort required to monitor and enforce the desired emissions level. When it is difficult to monitor or estimate emissions, attempts to restrict or tax the actions of polluters could fail due to the risk of widespread noncompliance (e.g., illegal dumping to avoid the tax) and costly enforcement. It may be easier to monitor and enforce regulations on a smaller number of “upstream” sources (e.g., oil refineries) rather than a larger set of “downstream” sources (e.g., gasoline consumers) (Mansur 2012). Mandating the use of specific technologies can sometimes reduce monitoring and enforcement burdens, as noted in section 4.2.1. For example, CEM equipment installed at power plants can measure pollution releases directly. The EPA’s SO2 allowance trading program required CEM for large sources (see Text Box 4.1).

4.6.7 Potential for Economy-Wide Distortions

Analysts should consider the potential distortionary effects of any policy option considered. Even if a policy is deemed relatively efficient on its own, it may interact with pre-existing environmental, economic, or agricultural policies (e.g., product standards, non-environmental taxes or subsidies, trade barriers) in non-intuitive ways that can exacerbate distortions in the economy and result in unintended environmental consequences. Instruments that include a revenue-raising component, such as auctioned allowances or taxes, may allow for opportunities to reduce other taxes and fees

---

104 See Weitzman (1974) for the classic paper on the ways in which uncertainty (also referred to as lack of information) affects instrument choice.

105 Baumol and Oates (1988) illustrate this point using a simple diagram.

106 Sigman (2012) presents a theoretical model showing that compliance need not decrease when regulations are broadened beyond industries with low cost monitoring to include those where monitoring costs are higher but abatement costs are lower.
and the associated inefficiencies. See Chapter 8 and Appendix A for a more detailed discussion of economy-wide distortions.

### 4.6.8 The Goals of the Policy Maker

Finally, the goals of policy makers may influence the instrument selected to regulate pollution. Each considered instrument may have different distributional and equity implications for both costs and benefits; these implications should be accounted for when deciding among instruments. For example, policy makers may wish to ensure clean-up of future pollution by firms. Policy makers may consider using insurance and financial assurance mechanisms to supplement existing standards and rules when there is a significant risk that sources of future pollution might be incapable of financing the required pollution control or damage mitigation method. In addition, the degree to which policy makers want to allow the market to determine environmental outcomes may influence the choice of instrument. The quantity of marketable allowances issued, for example, sets the total level of pollution control, but the market determines which polluters reduce emissions. On the other hand, taxes let the market determine the total level of control.

### 4.7 Measuring the Effectiveness of Regulatory Approaches or Voluntary Initiatives

Several policy criteria should be considered when evaluating the success of regulatory or non-regulatory approaches. While a formal analysis may not be required when considering the implementation of a non-regulatory approach, these factors are still important to consider. It is unlikely that any one policy will dominate on all of these factors (Harrington et al. 2004, and Goulder and Parry 2008).

In determining the effectiveness of a policy approach, policy makers should consider the following factors and questions:

- **Environmental Effectiveness**: Does the policy instrument accomplish a measurable environmental goal? Does the policy instrument result in environmental improvements or emission reductions? Does the approach induce firms to reduce emissions by greater amounts than they would have in the absence of the policy?
- **Economic Efficiency**: How closely does the approach approximate the most efficient outcome? Does the policy instrument reach the environmental goal at the lowest possible cost to firms and consumers?
- **Distributional or Equity Impacts**: Does the policy instrument have distributional or equity implications, when considering both costs and benefits?
- **Reductions in Administrative, Monitoring, and Enforcement Costs**: Does the government benefit from reductions in costs? How large are these cost savings compared to those afforded by other forms of regulation?
- **Inducement of Innovation**: Does the policy instrument lead to innovation in abatement techniques that decrease the cost of compliance with environmental regulations over time?

These evaluation criteria are relevant to several of the chapters in this guidance document: Chapter 7 covers approaches for analyzing social benefits. Chapter 8 offers guidance on how to measure social costs. Chapters 9 and 10 describe analyses that may help inform distributional and equity issues, respectively.

---

107 For useful references on the issues concerning the uses of revenues from pollution charges (e.g., applying environmental tax revenues so as to reduce other taxes and fees in the economy) and ways to analyze these policies, see Bovenberg and Goulder (1996), Goulder (2013), Jorgenson, et al. (2013), and McKibbin, et al. (2015).
Chapter 4 References


Chapter 5
Setting the Foundation: Scope, Baseline, and Other Analytic Design Considerations

Analysts must make choices regarding the design of the analysis of a regulation, including the appropriate scope of a benefit-cost analysis (BCA), how to specify the baseline, how to account for behavioral and technological change, what to assume about regulatory compliance, and how to address analytic uncertainty, among others. Identifying key issues or questions surrounding these decisions early in the regulatory development process is important because they can have a profound impact on analytic outcomes. This chapter provides an overview of a broad set of issues related to analytic design. Subsequent chapters on benefits (Chapter 7), costs (Chapter 8), economic impacts (Chapter 9), and environmental justice and other distributional analyses (Chapter 10) delve into these considerations in more depth as they apply in those specific contexts.

5.1 Scope of Analysis

Several early analytic decisions determine the scope or breadth of a BCA of a regulation. Specifically, analysts must consider: whose costs and benefits to count in a regulatory analysis, and the types of markets and non-market effects that should be evaluated, including those that cannot be quantified, to adequately capture the costs and benefits of the regulatory options under consideration.

Analysts should consider all the potential benefits and costs of the regulatory action to avoid potentially misleading conclusions regarding the net benefits and relative rankings of the analyzed regulatory options. In practice, however, not all changes in economic welfare can be quantified and monetized due to constraints in available tools, data, and resources. Therefore, the results of a BCA should be interpreted with care, evidence for welfare effects that cannot be quantified and monetized should be described, and any analytic limitations and omissions should be explicitly documented and discussed. While this section provides guidance on the scope of a BCA, Chapters 9 and 10 provide guidance on the scope of economic impact and distributional analysis.

5.1.1 Standing

One of the first scoping questions an analyst must answer when conducting BCA is: who has economic “standing,” or put another way, whose gains and losses should be counted in the analysis? The most inclusive answer is all persons who may be affected by the policy regardless of where (or when) they live. However, for domestic policy making standing is typically limited to the national level in order to maximize the welfare of residents. Consistent with this interpretation,

---

108 E.O. 12866 and OMB’s Circular A-4 (2003) require and affirm that all benefits and costs that result from a policy change should be considered in a BCA.

109 While section 5.1 focuses on the scope of a BCA, the same set of issues applies broadly to economic impact and distributional analysis.

110 Regulations typically only apply to a nation’s own residents who have consented to adhere to the same set of rules and values for collective decision making. In addition, most domestic policies are expected to have relatively negligible effects on other countries (Gayer and Viscusi, 2016; Kopp et al. 1997, Whittington et al. 1986).
OMB guidance states that analysts should "focus on the benefits and costs that accrue to citizens and residents of the United States" (OMB 2003).

Limiting standing to citizens and residents of the United States can be complicated to operationalize in practical terms (e.g., how should multi-national firms with plants in the United States but shareholders elsewhere be treated?). Analysts should ensure that its application is supported by the available data and that standing is consistently applied when estimating costs and benefits; in other words, if a group has standing for estimating costs, they should also have standing for benefit estimation.

Ultimately, who has economic standing is a policy decision. However, because it has important implications for the scope of the analysis it should be determined early in the process. When evaluating impacts beyond national borders, OMB recommends that those effects are reported separately.\textsuperscript{111}

### 5.1.2 Market Effects

Another analytic scoping question is identifying which markets will be affected by the regulation. Understanding the ways in which a regulation may affect different markets will help inform the analytic approach taken (see Chapters 7 and 8 for more discussion). Ideally, the analyst should capture all costs (and benefits) of a regulation. In practice, this may not always be feasible due to limitations in available data, methodologies, or resources. When prioritizing which costs (and benefits) to include, it is important to consider the effect of the regulation on related markets.

A "distorted" market is one in which factors such as pre-existing taxes, externalities, regulations, or imperfectly competitive markets, moves consumers or firms away from what would occur under perfect competition.\textsuperscript{112} In the absence of market distortions, it is not necessary to consider impacts in related markets. This is because, while a policy may have effects on these other markets, market-clearing conditions ensure that they are effectively canceled out from an aggregate welfare perspective (Farrow and Rose 2018, Just et al. 2004). However, in reality, every market is distorted to some degree.

Effects in related markets are important to consider when there are both pre-existing distortions in these markets, and there are significant cross-price effects between the regulated sector and these other economic sectors (Harberger 1964; Boardman et al. 2011; U.S. EPA 2017). Related markets may include upstream suppliers of major inputs to the regulated sector, downstream producers who use the regulated sector’s output as an input, producers of substitute or complimentary products, as well as consumers. Given that all markets are distorted to some degree, a key question for the analyst to consider then is, when is it reasonable to assume away these effects (e.g. Hahn and Hird 1990)? Evidence suggests that effects outside of the regulated sector, and therefore changes in welfare, may be substantial even with a relatively small sector-specific regulation (Marten, et al. 2019, Goulder and Williams 2003). The presence of a distortion alone, however, may not warrant a broader analytic approach, particularly if the value of information from accounting for its effect on costs and benefits is relatively small. The analyst should take special care to justify her choice of which markets to explicitly analyze as part of the regulatory analysis and identify key assumptions and limitations underlying this choice.\textsuperscript{113,114}

\textsuperscript{111} For discussion of when the effects of US policy on non-residents might be relevant in domestic BCA, see Viscusi, et al. (1988); Cropper, et al. (1994); and Gayer and Viscusi (2016).

\textsuperscript{112} Perfectly competitive markets are characterized by the following conditions: all economic agents have complete information; there are no barriers to entry or exit; firms have constant returns to scale; there are no taxes, subsidies or policies that create a wedge between the price suppliers receive for a good and the price consumers pay for it..

\textsuperscript{113} Analysts should also keep in mind that even in cases where effects in other sectors contribute little to the overall social cost or benefits of the policy, they may have important distributional consequences that warrant a broader analytic treatment than one that focuses solely on the directly regulated market. See Chapters 9 and 10 for more discussion.

\textsuperscript{114} Choosing the model that is most appropriate for capturing the key impacts of a policy is sometimes referred to as "horses for courses." Just as the best horse for a race depends on the features of the course, the best economic model(s) to evaluate the benefits and costs of a regulation depend on...
5.1.3 Externalities

BCA should aim to evaluate all benefits and costs resulting from the regulation, which includes effects from any changes in environmental contaminants or other externalities.\textsuperscript{115} If some of these effects cannot be quantified, they should be evaluated qualitatively (including a discussion of their potential magnitude).

The analysis should estimate the welfare effects arising from the primary statutory objective of the regulation as well as other welfare effects. These other welfare effects could include both favorable or adverse impacts on societal welfare. Analogous to how a regulation’s interactions with existing distortions in other markets (e.g., pre-existing taxes) could lead to additional social costs, a regulation could ameliorate or exacerbate other pre-existing externalities. Changes in other environmental contaminants may arise from the compliance methods of regulated sources. For example, the use of an abatement technology by regulated sources to reduce a pollutant into one medium (e.g. air) may change the emissions of another pollutant into the same medium (e.g., from the same smokestack) or cause changes in emissions of pollutants into another medium (e.g., water). Changes in other environmental contaminants may also occur as a result of market interactions induced by the regulation. For example, more stringent vehicle emissions standards can lead to reductions in upstream refinery emissions. Section 5.5.6 discusses the importance of ensuring that projected changes in contaminants are consistent with expected market behavior and consider other regulations. This guidance also applies to expected changes in externalities beyond those associated with environmental contaminants. For example, changes in vehicle emissions standards may reduce the marginal cost of driving due to greater fuel efficiency and lead to an increase in vehicle miles traveled that affect road safety, congestion, and other transport related externalities. Those welfare effects should also be evaluated and, if data and resources allow, limited resources on refining the estimated for the BCA of a regulation.

When presenting the results of the benefit cost analysis, clearly distinguishing between benefits that arise from the statutory objective of the regulation and other welfare effects of the regulation, when it is possible to do so, provides transparency. For example, in a BCA of a regulation promulgated under a Clean Air Act provision whose objective is reducing hazardous air pollutants (HAPs), the air pollution benefits resulting from reductions in HAP emissions should be clearly distinguished from other welfare effects resulting from the expected compliance strategies of regulated entities.\textsuperscript{116} However, when calculating net benefits all welfare effects should be included, as it is the total willingness to pay for all changes induced by a regulation that determines economic efficiency.

5.2 Baseline

Clearly establishing the baseline of an economic analysis is a critical step for accurately evaluating benefits and costs. Because a BCA considers the impact of a policy or regulation in relation to the baseline, its specification can have a profound influence on the outcome of that analysis. The level of detail presented in the baseline specification is also an important determinant of the type of analysis that can be conducted when evaluating regulatory options. For these reasons, careful thought in specifying the baseline is crucial.

5.2.1 Baseline Definition

The baseline is defined as the best assessment of the way world would evolve absent the proposed regulation. It is the primary point of comparison for assessing the effects of the regulatory options under consideration. Specifically, the BCA models two states of the world: the expected state without the regulation (the baseline scenario) and the expected

---

\textsuperscript{115} These distortionary effects are among those discussed in Section 5.1.2 as the presence of an externality represents a deviation from perfect and complete markets, which may be ameliorated or exacerbated by behavioral changes induced by a regulation. They differ from marketed goods in that the welfare effects due to changes in an externality are not reflected in market prices.

\textsuperscript{116} This means, for example, that if the air pollution reduction also reduces harmful deposition of the pollutant into the water, the benefits from reducing water pollution should be distinguished from the benefits arising from the reduction of the pollutant in the air.
state with the proposed regulation option(s) in effect (the policy scenario(s)). The impact for each policy scenario is measured by examining the differences in net benefits between the scenarios.

The baseline scenario describes the expected future extent of the environmental problem and level of environmental contaminants along with the affected markets and exposed population in the absence of the proposed regulation. While the policy scenario is described in a similar fashion to the baseline, it reflects different environmental and/or market conditions as well as the expected compliance activities to comply with the policy.

Figure 5.1 illustrates the overall structure of a BCA to further demonstrate the difference between the baseline and a policy scenario, although there may be multiple policy scenarios in practice. An economic analysis begins with a description of the state of the world in the current period as a foundation, before any analytic scenarios are constructed. The current period, or current state of the world, includes a description of the environmental problem as well as other variables such as the level of environmental contaminants; the number and characteristics of the affected markets, firms, consumers, and state and local governments; the consumption and production of affected goods; characteristics of the exposed population; and existing federal, state, and local regulations that may affect the environmental problem. Based on the description of the current state of the world, the next step is to develop a projection of the future state of the world without the regulation, which is referred to as the baseline. This step is done by characterizing how economic and environmental conditions are expected to change over time. Changes may occur in demographics, the pace and direction of technology, energy and/or other prices, sector-specific economic activity, consumer behavior, and other related policies and programs that are already in place. The baseline should reflect likely outcomes, not an extreme scenario. The policy scenario is evaluated in a similar fashion, but the economic and environmental conditions reflect the future state of the world with the regulation in place. The two scenarios are then compared.

It is important to note that the comparison of the world with the policy to the world without the policy is distinct, and quite different, from a comparison of the state of the world before the action to the state of the world after the action. In other words, the baseline is a future scenario without the regulatory program under consideration; it is not a scenario assuming no change from current conditions. The economy and other factors (e.g., baseline health risks) may change over the period of analysis even in the absence of regulation, so a proper baseline should incorporate assumptions about the changes in the economy that may affect relevant benefits and costs.

In most cases, future economic and environmental conditions in the baseline are expected to have changed solely in response to factors unrelated to the regulation under consideration. On occasion this may not be the case. For example, a regulation under consideration may extend the compliance period of an existing regulation. In this case, the baseline specification might incorporate the expiration of the existing program. However, changes between the baseline and policy scenario should be solely attributable to the introduction of the regulation. The economic and environmental characteristics specified in the baseline should be used in the policy scenario unless the policy scenario is anticipated to change those characteristics. This is what makes the baseline the relevant point of comparison for the policy. In general, the construction of the baseline needs to be balanced to equally identify factors that may meaningfully affect both benefits and costs. For example, the analyst should not assume a great deal of technological innovation in one sector (e.g., the pollution abatement sector) and ignore potential technology improvements in other sectors.

The final step in an analysis, as illustrated in Figure 5.1, is to use the information from the baseline and policy scenarios as a basis for estimating the benefits, costs, economic impacts and distributional impacts of the regulatory option(s) under consideration. The damages from exposure to environmental contaminant levels in the baseline and policy scenarios can be valued using a variety of economic techniques (Chapter 7: Analyzing Benefits). The value of the change in damages in the policy scenario are the benefits of the policy. The new compliance activities and other effects identified in the policy scenario can be used to quantify the costs of the policy (Chapter 8: Analyzing Costs).
Figure 5.1: Structure of a Benefit-Cost Analysis

Current Period
- Current environmental problem/level of environmental contaminants
- Number and characteristics of affected markets, firms, consumers, and state/local governments
- Consumption and production of affected goods
- Characteristics of exposed population
- Federal/state/local regulations that have bearing on the environmental problem

Future Period
Baseline: without the regulation
- Expected extent of future environmental problem/level of environmental contaminants
- Number and characteristics of affected markets, firms, consumers, and state/local governments
- Consumption and production of affected goods
- Characteristics of exposed population
- Anticipated federal/state/local regulations that have bearing on the environmental problem
- Damages from environmental contaminants on exposed population
- Valuation of damages

Future Period
Policy Scenario: with the regulation
- Expected new compliance activities
- Expected new environmental conditions/level of environmental contaminants
- Possibly new market configurations
- Number and characteristics of affected markets, firms, consumers, and state/local governments
- Consumption and production of affected goods
- Characteristics of exposed population
- Anticipated federal/state/local regulations that have bearing on the environmental problem
- Damages from environmental contaminants on exposed population
- Cost of new compliance activities
- Valuation of damages

5.2.2 Guiding Principles of Baseline Specification

- Benefits = (Baseline valuation of damage) - (Valuation of damages w/policy)
- Costs = (Policy cost of controlling environmental contaminants)
- Net benefit = Benefits - Costs
- Economic impacts = (Baseline market changes) - (Market changes w/policy)
- Distributional impacts = (Baseline changes in exposures) - (Changes in exposures w/policy)
In specifying the baseline, analysts should employ the following guiding principles:

1. Clearly specify the environmental problem that the regulation addresses and the regulatory approach being considered in the statement of need;
2. Identify all required variables for the analysis;
3. Clearly specify the current and future state of relevant economic and regulatory variables;
4. Concentrate on the components of the analysis that have the greatest influence on the results;
5. Clearly identify all assumptions made in specifying the baseline conditions;
6. Detail all aspects of the baseline specification that are uncertain; and
7. Use the baseline assumptions consistently throughout the analysis of a regulation.

Though these principles exhibit a general common-sense approach to baseline specification, the analyst is advised to provide statements on each of these points. Failure to do so may result in a confusing presentation, inefficient use of time and resources, and misinterpretation of the economic results.

**Clearly specify the environmental problem that the regulation addresses and the regulatory approach being considered in the statement of need.** As discussed in detail in Chapter 3, the analysis should begin with a statement of need that identifies: (1) the problem being addressed; (2) the current regulatory environment; (3) the method by which the problem will be addressed; and (4) the regulated and other affected parties. There should also be a discussion of how the specific regulatory or non-regulatory approaches being evaluated were chosen.\(^{117}\) This statement of need will help clarify the appropriate baseline to be used.

In general, the most appropriate baseline will assume no change in behavior to comply with the regulation being analyzed or existing regulations; but in some cases, a different baseline may be considered. For example, if an industry is certain to be regulated by some other means (e.g., by court order or state action) but that regulation has not yet been implemented, then the baseline should include this regulation. Also, it is common practice to assume full compliance with existing regulatory requirements in the baseline even if there is noncompliance, although a separate sensitivity analysis assuming less-than-full compliance may be considered to determine the implication of this assumption. (See Section 5.5.4 and 5.6.1 for more discussion of this issue.)

**Identify all required variables for the analysis.** To ensure that the baseline scenario can be compared to the policy scenario, there should be a clear understanding of the path from economic behavior to environmental changes to impacts on humans or ecosystems. The models, parameters, and variables required for the baseline analysis should be chosen so that they can inform all subsequent analyses. Differences between the baseline and policy scenario may include changes in use or production of toxic substances, changes in production processes and costs, changes in pollutant emissions and ambient concentrations, and incidence rates for adverse health and environmental outcomes associated with exposure to pollutants. This does not mean that the analyst must identify all the variables that could possibly change, but the analyst should recognize all relevant variables needed to compare the baseline scenario to the policy scenario. At a minimum, the analyst should identify the variables that are expected to have the largest impact on cost and benefit across options.

**Clearly specify the current and future state of relevant economic and regulatory variables.** Future baseline trajectories of certain types of economic variables such as energy prices, the level and growth of economic activity, and population growth, may be important for modeling the effects of a regulation. Even small changes in the rate of economic growth may, over time, result in considerable differences in emissions and control costs. Assuming no change in the baseline

---

\(^{117}\) See Chapter 4 for a description of various regulatory and non-regulatory approaches.
economic activity may then lead to incorrect results. Likewise, assumptions about the future growth and age distribution of the population affected by a regulation are important for predicting the number of individuals exposed or even the magnitude of aggregate damages. Other variables, such as broad trends in consumer spending patterns and technological growth, are also important for modeling the effects of a regulation but are more difficult to estimate. In these cases, the analyst should specify the baseline levels and changes over time and explicitly discuss all assumptions.

If environmental and other policies or programs influence baseline conditions, they should also be accounted for in the baseline. For example, changes in farm subsidy programs may influence future pesticide use. Accounting for the way existing regulations affect compliance behavior and economic and environmental outcomes of a new regulation assures that the BCA properly accounts for the cumulative effects of regulations. In an ideal analysis, all potential direct and indirect influences on baseline conditions (and on the costs and benefits of regulatory alternatives) would be examined and estimated. However, it is up to the analyst to determine if these indirect influences are important enough to incorporate into the regulatory analysis. If indirect influences are known but are not considered significant enough to be included in the quantitative analysis, they can be discussed qualitatively. On the other hand, there may be value in quantifying small or insignificant impacts of indirect influences if there is a prior expectation that they are important.

Concentrate on the components of the analysis that have the greatest influence on the results. The analyst should concentrate analytic efforts on components (e.g., assumptions, data, models) of the baseline that are most important to the analysis, taking into consideration factors such as the time given to complete the analysis, the person-hours available, the cost of conducting the analysis, and the availability of models and data. If several components of the baseline are uncertain, the analyst should concentrate on components that have the greatest influence on the results of the analysis and can be refined through additional analysis or data collection. Analysts should pay special attention to the components that will be used to calculate costs and benefits and those that are important in the evaluation and selection of a policy option.

Clearly identify all assumptions made in specifying the baseline conditions. The analyst should explain key assumptions in detail, including those related to changes in consumer and producer behavior, and how these trends may be affected by the regulatory options. Analysts may observe trends in economic activity or pollution control technologies that occur for reasons unrelated to environmental regulation. For example, as a consumer’s income increases over time, the demand for different commodities may change. Demand for some commodities may grow at rates faster than the rate of change in income, while demand for other goods may decrease. Where these trends are highly uncertain or are expected to have significant influence on the evaluation of regulatory alternatives, the analyst should clearly explain and identify the assumptions used in the analysis, with the goal of laying out the assumptions clearly enough so that other analysts (with access to the appropriate models) would be able to replicate the baseline specification.

Detail all aspects of the baseline specification that are uncertain. Because the analyst does not have perfect foresight, baseline conditions cannot be characterized with certainty. To the extent possible, estimates of current values should be based on actual data and estimates of future values should be based on clearly specified models and assumptions. Where reliable projections of future economic activity and demographics are available, this information should be used and adequately referenced. In general, uncertainties underlying the baseline conditions should be treated in the same way as other types of uncertainties in the analysis.

It is also important to detail information that was not included in the analysis due to scientific uncertainty. For example, a regulated pollutant may have a suspected health or ecological effect but the science behind this connection may be too uncertain to quantify. In this case, the effects generally are not quantified in the analysis, but why the effects were excluded should be discussed — especially if the expected magnitude is such that it could significantly affect the net benefit calculation. Analysts should also explain how scientific uncertainty affects model choice and parameter values.

For example, if the regulated industry is in significant decline, or is rapidly moving overseas, this information should be accounted for in the baseline. In such cases, incremental costs to the regulated community (and corresponding benefits from the regulation) are likely to be less than if the targeted industry were stable or growing.
Important aspects of the analysis which are not included in the baseline due to scientific uncertainty should be included in dedicated uncertainty section(s) of the analysis (See Section 5.6 below). Significant uncertainty in important variables may require the construction of alternative baselines (discussed below). While sensitivity analysis is usually a better choice, multiple baselines may provide insights when evaluating different policy options.

Use the baseline assumptions consistently for all analyses of a regulation. The economic and environmental characteristics used in the baseline should be consistent with those used in the policy case. For example, the calculation of both costs and benefits should draw upon estimates derived using the same underlying assumptions about future economic and environmental conditions. If the benefits and costs are derived using multiple economic and environmental models, then the baseline conditions applied in those models should be compared to ensure that they are consistent. Likewise, when comparing and ranking alternative regulatory options, comparison to the same baseline should be used for all options under consideration.

In some cases, a secondary analysis may be required. For example, it may be useful to single out a sector for more detailed analysis, or a follow-on analysis might be needed to assess impacts on a specific set of households based on their socioeconomic characteristics, region or sector. In this case, it may not be possible to specify a baseline for the secondary analysis that is fully consistent with the primary analysis, but the analyst should endeavor to make them as similar as possible. The analyst also should explicitly describe the differences between the two baselines and any uncertainty associated with the secondary baseline.

5.2.3 Multiple Baselines

In most cases, a single, well-defined baseline is generally all that is needed as a point of comparison with the policy. However, there are a few situations where it may be informative to compare the policy options to more than one baseline. Multiple baseline scenarios are needed when it is difficult to identify a single, reasonably reliable description of the world in the absence of the proposed regulation. For instance, if the current level of compliance with existing regulations is not known and may substantially influence the net benefits, then it may be necessary to compare the policy scenario to both a full compliance baseline (the standard assumption) as well as a partial compliance baseline. Also, if the impact of other rules currently under consideration fundamentally affects the analysis of the rule being analyzed, then multiple scenarios with and without these rules in the baseline may be necessary. For example, for the 2019 rule to repeal the 2015 rule defining “Waters of the United States,” the degree to which states would continue to regulate their waters at the 2015 standard was uncertain. Since the states' decisions dramatically affected the avoided costs and forgone benefits of the repeal, multiple baselines were used to illustrate the range of potential impacts (U.S. EPA 2019).

The decision to include multiple baselines should not be taken lightly, since it may result in a complex set of modeling choices and analytic findings. Multiple baselines increase the possibility of erroneous comparisons of costs and benefits across them if the modeling choices and results are not interpreted and communicated clearly. The number of baselines should be limited to as few as possible that cover the key dimensions of the analysis and any phenomena in the baseline that are uncertain. Each baseline-to-policy comparison should be internally consistent in its definition and use of baseline assumptions.

5.3 Multiple Rules

Although regulations that have been finalized clearly belong in the baseline of a proposed rule, the baseline specification may be complicated if regulations other than the one being promulgated are under consideration or nearing completion. It is important to consider how these other regulations affect market conditions and the degree to which

---

119 In the less common case in which more than one baseline scenario is modeled, the analyst must avoid the mistake of combining analytic results obtained from different baseline scenarios. To limit confusion on this point, if multiple baseline scenarios are included in an analysis, the presentation of economic information should clearly describe and refer to the specific baseline scenario being used.
they might influence the costs or the benefits associated with the policy of interest. This is true not only for multiple rules promulgated by the EPA, but for rules passed by other federal, state, and local agencies. In addition to agencies that regulate environmental behavior, other agencies that regulate consumer and industrial behavior (e.g., Occupational Safety and Health Administration (OSHA), Department of Transportation (DOT), and Department of Energy (DOE)) develop rules that may impact some of the same entities as EPA regulations.

5.3.1 Linked Rules

When rules affect the same industry or when multiple rules are needed to achieve a policy objective, it may be possible to analyze the multiple rules together. For example, EPA may issue a rule covering both the effluent limitation guidelines (ELGs) for an industry, providing the technical requirements for pollution discharge, and requirements under the National Pollution Discharge Elimination System (NPDES), providing details of the permitting system for the industry (e.g., U.S. EPA 2002c). Since the ELGs and NPDES work together to achieve one objective, it makes sense to analyze them together. In some cases, linked rules may affect the same industry but have different enabling statutes. For example, in 1997, EPA issued a single rule for the pulp and paper industry covering the National Emission Standards for Hazardous Air Pollutants under the Clean Air Act and the Effluent Limitations Guidelines for Pretreatment and New Source Performance Standards under the Clean Water Act (U.S. EPA 1997c).

The best approach for linked rules is to include all of them in the same analysis. Analyzing multiple rules as if they were one rule simplifies the baseline specification, comparing them to the world in which none of the linked rules are in place. When statutory requirements and judicial deadlines complicate promulgating multiple rules as one, coordination between rulemaking groups is still possible. The sharing of data, models, and joint decisions on analytic approaches may make a unified baseline possible so that the total costs and benefits resulting from the package of policies can be assessed in a way that avoids omissions or double counting.

5.3.2 Unlinked Rules

In some cases, it is not feasible to analyze a collection of rules being developed at the same time in a single analysis. This may be true for rules originating from different program offices or different regulatory agencies, or when the timing of the various rules is not clear. In this case, each rule should be analyzed separately, but the order in which the rules are being analyzed should be stated explicitly. If two rules are issued in sequence but some of the costs of complying with the second rule are incurred in the process of complying with the first rule, then these costs should be included in the baseline and should not be considered as costs of the second rule. Only the incremental benefits and costs should be included in the second rule. For example, in 2005, the baseline of the Clean Air Mercury Rule (CAMR) included mercury emission reductions from the previously issued Clean Air Interstate Rule (CAIR) (See Text Box 5.1).

The practical assumption commonly made when rules cannot be evaluated together is to consider the actual or statutory timing of the rules and use this to establish the sequence in which they are analyzed. However, this may not always be possible. For example, a rule may be phased in over time, complicating the analysis of a new rule going into effect during that same period. For this case, the baseline for the new rule should include the timing of each stage of the phased rule and its resulting environmental, health, and economic changes.

In the absence of some orderly sequence of events that allows the attribution of changes in behavior to a unique regulatory source, there may be no clear way to allocate the costs and benefits of a package of policies being developed at the same time to each individual regulation. By implication, there is no theoretically correct order for conducting a sequential analysis of multiple policies that are promulgated simultaneously. The only solution in this case is to make a
Because the cost and benefit estimates of one regulation may be affected by those of others, it is important to carefully consider if they should be incorporated into the baseline. As a rule, it is important to be transparent and use objective reasoning when deciding to account for other regulations in a baseline. Transparency requires that the analyst clearly state all assumptions. Objective reasoning requires that the analyst not engage in speculation. If there is uncertainty about an anticipated rule, then two baselines, one with the anticipated rule and one without, might be considered. If only one baseline is considered due to time or resource constraints, then it should be constructed using only final rules and, in some cases, imminent rules that are expected with a high degree of certainty in the absence of EPA action. General guidelines to follow are given below.

**All final rules, including those that have not fully taken effect, should be included:** The analysis should assume firms will comply with already promulgated rules. For example, on March 15, 2005, the EPA promulgated the Clean Air Mercury Rule (CAMR) to reduce mercury emissions from coal-fired power plants (U.S. EPA 2005b). Five days earlier, on March 10, 2005, the EPA finalized the Clean Air Interstate Rule (CAIR) (U.S. EPA 2005a). While the primary purpose of CAIR was to reduce sulfur dioxide (SO₂) and nitrogen oxides (NOₓ), the control technologies necessary to achieve these reductions also lowered mercury emissions. Because the final CAIR rule had been issued, the analysis for CAMR assumed that the mercury reduction from CAIR was in the baseline. This meant that the estimated incremental reduction in mercury from CAMR was much smaller than if CAIR had not been included in the baseline.

**Including imminent final rules may be appropriate if the impacts are known with a high degree of certainty:** If another (final) rule is imminent and will take effect prior to the effective date of the new rule under consideration, then the imminent rule should be included in the baseline but only if its requirements and impacts are known with a high degree of certainty. The analyst should not speculate that another rule will be implemented. In addition, the analyst should be clear as to what assumptions have been made to include the imminent rule in the baseline.

**Proposed rules should not be in the primary baseline:** While a proposed rule signals the intent to issue a final rule and the agency maintains a schedule to do so, there is no guarantee that the final rule will be issued or that it will follow the planned schedule. Even if the agency does issue a final rule, it may differ significantly from the proposed rule, which means that the assumptions embedded in a baseline using a proposed rule will not accurately reflect the likely future effects of the final rule. An alternative baseline for a proposed rule may have another proposed rule in it, however, if the two rules are expected to be finalized in the same sequence and the existence of the first rule may influence the benefits and costs of the second substantially.

**Future regulatory actions of other jurisdictions should be considered carefully:** Actions by state and local governments and even international organizations can affect the costs and benefits of federal rules, particularly if they are regulating the same sector or pollutant. In this case, the analyst must use professional judgment to determine what would happen in the baseline (i.e. in the absence of EPA action) and how the regulatory response of other jurisdictions may change in the policy scenario.

State regulations that have been finalized should be included in the baseline. The more difficult case occurs when a state has a legal obligation to implement a regulation but either has not done so or is in the process of doing so. For example, the EPA occasionally issues rules establishing numeric water quality standards for some states when the states themselves have not done so. One might argue that the state regulation should be in the baseline since they had the legal obligation to issue the criteria, but this is not the case. The EPA’s justification for action is that it assumes the state will not act. In this example, only if the state would issue the water quality standard in the absence of EPA action can a reasonable case be made for including the state action in the baseline.

Compliance with a finalized international agreement cannot simply be assumed in the baseline, especially if some EPA action (such as codifying the international standard) is required for it to become effective. The costs and benefits associated with any behavioral response by firms to the EPA action should be part of the policy scenario. In the case where firms will meet the international standard on their own, even without EPA action, then the compliance with the standard can be included in the baseline; but establishing that this behavioral response will occur requires justification.

reasonable assumption and clearly explain it, detailing which rules are included in the baseline (see Text Box 5.1). If the impact of other rules on the costs and benefits of the rule under consideration is small, then this may be all that is necessary; it may not be worth additional time and resources to reconcile the baseline of rules being developed at the same time. On the other hand, when the impact on the costs and benefits is large or if the number of overlapping rules is small, then a sensitivity analyses can be included to test the implications of including or omitting other regulations.
Under this sensitivity analysis, it may also be possible to use the overlapping nature of the regulations to allow for some regulatory flexibility in compliance dates and regulatory requirements.

5.3.3 Accounting for Benefits and Costs that Accrue Across Multiple Rules

When EPA targets the same contaminants or industries through a sequence of regulations, the benefits and costs of these actions are additive. To ensure consistency in regulatory accounting, these Guidelines recommend that regulatory analyses fulfill an “adding-up condition” when comparing a single large regulation to multiple smaller regulations that imply the same requirements for the same set of entities. The adding-up condition means that the sum of the estimated benefits (and costs) from a set of small regulations analyzed separately should be the same as the benefits (and costs) from the same actions evaluated jointly in a single regulation. Benefits and costs from previous rules should be included in the baseline so that they are not double counted in a new regulation.

The adding-up condition was originally proposed in the context of contingent valuation studies (Diamond and Hausman 1994; Kling and Phaneuf 2018) and has been applied to valuation of water quality improvements (Newbold et al. 2018). If analysts do not impose an adding-up condition and fail to account for improved environmental quality in the baseline when valuing incremental improvements from successive regulations, then inconsistent results could arise if people value marginal improvements more when the environmental good is scarce.

In some cases, environmental regulations yield relatively small changes in health or the environment that may not be noticeable to the public until multiple regulations have achieved a large aggregate improvement. Just as it is important to account for small average costs imposed by regulations—which can be economically significant when aggregated over a sufficiently large population—it is conceptually correct to account for small improvements in public health and the environment. For instance, the EPA’s Science Advisory Board (U.S. EPA 1998) noted that, “small effects distributed across a large population exert large total health effects,” and recommended that the Agency quantify changes in IQ resulting from regulations that reduce lead exposure, including changes of less than a single IQ point on average per child.

Some benefits only occur after a threshold has been reached. However, a specific benefits threshold may not be met with a single rule. In such cases, it is reasonable to account for the benefits of making progress toward a goal, even if the threshold is not met in the rule under consideration. Otherwise, if the benefits are associated only with the rule that actually passes the threshold, it may be impossible to justify the previous rules that made incremental progress.

For example, the EPA has calculated the benefits associated with improving river miles for various designated uses (e.g., swimming, fishing, and boating) in several rules. In each case, some river miles were improved for the designated use, while other miles were improved, but not enough to change their designated use. Analyses of earlier rules claimed benefits only if a river mile changed its designation, implicitly giving a value of zero to partially improved river miles. More recent analyses have included estimates of the partial benefits from incremental improvements toward the threshold. Either approach can be used; but, accounting for the benefits of partial gains provides useful information to decision makers and the public and allows the Agency to justify incremental progress to a threshold. Once partial gains have been valued in one rule, then subsequent rules cannot claim full credit for crossing the threshold. Doing so would double count those benefits.

In the special case when new data or methods make estimates of benefits or costs for earlier rules obsolete, the analyst should develop a baseline based on the new Information and discuss all changes made since the previous regulatory analysis.
5.4. Time Horizon of Analysis

The time horizon of analysis is the period over which the baseline and policy scenarios are compared. The time horizon is defined by the starting and ending points.\(^\text{120}\) A guiding principle is that the time horizon should be chosen to capture all of the benefits and cost for the policy alternatives analyzed, subject to available resources.\(^\text{121}\) This principle is consistent with the requirement that a BCA sufficiently reflects the welfare outcomes of those affected by the policy.

The appropriate time horizon will depend on the economic and legal conditions unique to the regulatory context under consideration. In many cases, the time span of the physical effects that drive the benefit estimates and the economic lifetime of any pollution control investments will be key factors in its determination. Legal conditions that affect the time horizon of analysis include the timing of compliance dates. While selecting the appropriate time horizon is challenging, the analysis should identify the time horizon chosen and explain why it is expected to capture all benefits and costs. It should also identify the extent to which the sign of net benefits or the ranking of policy alternatives by their magnitude of net benefits may be sensitive to the choice of time horizon.\(^\text{122}\)

The starting point for the analysis should be based on when conditions between the baseline and policy scenarios diverge, and thus benefits and costs of the regulation begin to be realized. Two possible choices for the starting point are when an enforceable regulatory requirement becomes effective or when the final rule is promulgated. These dates are convenient starting points because they are clearly defined under administrative procedures and represent specific deadlines. However, the starting point should precede when the regulatory requirements become effective if firms or households are expected to make anticipatory investments or other behavioral changes after the rule is finalized and leading up to the effective date. In these circumstances, an earlier starting point should be used.\(^\text{123}\)

The duration of costs and benefits resulting from a policy generally determines the appropriate ending point for the analysis. In theory, the longer the time horizon, the more likely that the analysis will capture all the major benefits and costs. However, other analytical considerations, such as the relative uncertainty in projecting out-year conditions, may also need to be weighed. Forecasts of economic, demographic, and technological trends are required over the entire time horizon of the analysis. Because long term forecasts are less reliable than near term forecasts, the analyst should balance the advantages of capturing important effects against the disadvantages of decreased reliability of forecasts further out in time.

The ending point may differ for assessing costs and for assessing benefits when their accrual does not coincide. For example, the human health benefits of a policy to reduce leachate from landfills may not occur for many years after the cost of compliance is incurred either because decreases in groundwater contamination take time or because even after contamination is reduced some health improvements do not manifest immediately. In other contexts, while control costs are incurred upfront, changes in pollution may lead to health and ecological benefits that continue to accrue over time. While the choice of ending points for costs and benefits may differ, analysts should still ensure consistent

---

\(^{120}\) The time horizon for analysis may also be called the time frame or time period of analysis.

\(^{121}\) Chapter 6 provides a formal method of identifying the ending point of the time horizon analysis. A symmetric method may be used to identify the starting point. In addition, Chapters 7 and 8 also provide detailed guidance on selecting the time horizon of the analysis for benefits and costs, respectively.

\(^{122}\) To compare the benefits and costs of a proposed policy, the analyst should estimate the present discounted values of the total costs and benefits attributable to the policy over the time horizon of analysis. Chapter 6 provides guidance on how to discount benefits and costs.

\(^{123}\) In most circumstances a starting point that precedes final rule promulgation is unnecessary, but an earlier starting point may be desirable if significant behavioral changes were made in anticipation of the final rule. Two possible starting points that precede promulgation of the final rule and are clearly defined legal milestones are when authorizing legislation was signed into law and when EPA formally proposed the rule. However, when using a starting point that precedes regulatory requirements, it is important for the analysis to identify which behaviors occurred specifically because of the anticipated rule versus those that happened for other reasons. This will likely be difficult to do.
accounting of benefits and costs. All of the costs from activities that lead to quantified benefits should be accounted for in the analysis and vice versa.

Generally, the analysis should account for costs at least until the end of the economic lifetime of any pollution control methods adopted for compliance with the regulation.\textsuperscript{124} Any costs incurred will then correspond with the total abatement, and therefore benefits, achieved by these pollution control methods. Similarly, the length of the cost analysis should capture any turnover in markets for regulated goods (e.g., vehicles) and the length of time those goods are in use. This guidance may be challenging to implement in an analytic framework that captures the possibility that additional regulated sources will appear in the future. Again, the analysts should weigh the value of additional benefit and cost information gleaned from a longer time horizon of analysis against uncertainty about future economic conditions.

Some statutory provisions have schedules for when regulations need to be reviewed, and an ending point corresponding to this review date may be a tempting choice. However, care should be taken when using regulatory or statutory deadlines to determine the ending point of the time horizon of analysis. For example, these provisions may not envision the regulation being loosened but only tightened, and therefore the requirements under consideration are expected to persist over time, at least at the promulgated level of stringency, potentially yielding additional benefits and costs.\textsuperscript{125} Again, a time horizon that reflects the span over which the baseline diverges from the policy case and accounts for all of the benefits and costs is appropriate even if the period extends beyond the scheduled review.

In certain circumstances where benefits and costs are not expected to notably change over time, it may be analytically convenient to estimate benefits and costs over a shorter time period (e.g., year) that is representative of a longer time horizon of analysis (e.g., decade). In other cases, it may be analytically challenging to estimate benefits and costs over the entire time horizon and benefits and/or costs are estimated for only a few representative periods. In these cases, the analysis should still identify the entire time horizon over which the representative periods of analysis are applicable and discuss any limitations or uncertainty introduced by this approach. In addition, the representative periods of analysis should be chosen such that they adequately identify the relative net benefits of the various options under consideration.

5.5 Representing Economic Behavior

To measure the benefits and costs of a regulation, it is important to clearly characterize the behavior of firms and households in both the baseline and the policy scenarios. In particular, assumptions regarding how firms and households engage in technological change, comply with regulations, participate in voluntary actions, and affect levels of other contaminants in the baseline and policy scenarios can also influence costs and benefits.

5.5.1 Behavioral Response

Predicting firm and household responses to regulation requires some underlying model of economic behavior. These Guidelines recommend that analysts assume behavior consistent with utility or profit maximization unless there is evidence supporting other behavioral assumptions (see Section 5.5.2).

When modeling the response to regulation, it is important to capture how regulated firms may choose to comply with new requirements, for instance by changing production practices, output, location, or even exiting the industry. Likewise, it is important to capture households' responses, such as changing the products they buy, where they choose to live, or the types and frequency of their averting behaviors (e.g., purchasing bottled water or staying inside on bad air

\textsuperscript{124} The economic lifetime is the length of time a piece of equipment is expected to be operational before it is worn out and needs to be replaced.

\textsuperscript{125} Furthermore, if there is a some credible reason to assume that the regulation will be loosened in the future then this possibility should be clearly acknowledged in the analysis and the compliance choices of regulated sources should reflect this possibility (e.g., regulated sources would be more likely to adopt easily reversible compliance strategies if they thought the regulation may be loosened in the future). Another reason to evaluate the benefits and costs of the rule beyond the statutory review date is if the rule currently under consideration is expected to be accounted for in the baseline of any analysis with a time frame beyond the statutory review date, including the rulemaking subsequent to the statutory review.
quality days). These household and firm responses also may result in changes in market prices, which could further influence economic behavior. Behavioral response to the regulation may also precede compliance dates, which can make it difficult to disentangle how much of the behavior is attributable to the regulation.

Depending on the types of behavioral responses that are anticipated, the analyst will need to identify and select the most appropriate economic and environmental model(s) for the regulatory analysis. Any modeled changes in behavior should be supported by empirical estimates of demand, cross-price, and income elasticities. Uncertainty analysis is useful for examining the sensitivity of benefit and cost estimates to different elasticity assumptions. See Text Box 5.2 regarding other considerations when selecting models for estimation of costs and/or benefits.

5.5.2 Potential for Cost Savings

If firms and households behave in ways consistent with profit and utility maximization, they will adopt available cost-effective technologies or practices absent regulation. Even if they are not in widespread use when a new regulation is being developed, cost-effective technologies could be adopted under baseline conditions in the future as information about their effectiveness spreads. When households and firms voluntarily undertake these changes without the regulation, the regulatory action cannot be credited with any private cost savings resulting from their adoption. In cases where a regulation is estimated to result in net private cost savings, it is important to provide evidence of why these cost-saving measures would not already be undertaken in the baseline.

When evidence to explain this phenomenon is not available, analysts should consider whether the finding of private cost savings is defensible and whether other costs are not being accounted for. For instance, a regulation may impose "hidden" costs that are not easily quantified in a standard engineering cost model but still represent welfare losses for firms or households that offset cost savings. Lower operating expenditures from a new technology required by a regulation might be offset by increases in other costs if the new technology breaks down more frequently, requires special training to operate, or has other undesirable features. If data are available on such costs, analysts should include them in the analysis.

In some cases, evidence may suggest that firms or households do not adopt cost-saving measures because of market failures (e.g., informational asymmetries or transaction costs). If the regulation addresses these market failures, it could lead to net private cost savings. In these instances, analysts should provide a clear description and evidence of the market failure and how the new action addresses it.

The economics literature has also documented specific instances in which households or firms act in ways that appear to run counter to their self-interest (Shogren and Taylor 2008; Shogren et al. 2010; Croson and Treich 2014). However, research also indicates that market experience can eliminate behavior that is inconsistent with profit-maximization in certain settings (List 2003, List 2011). If estimated net private cost savings could be due to widespread suboptimal behavior, analysts should provide empirical evidence specific to the affected market. In the absence of such evidence, analysts should assume rational profit- or utility-maximizing behavior by firms and households in the primary analysis, which would eliminate the possibility of estimating net private cost savings as a result of regulation. Sensitivity analysis can be used to consider other behavioral assumptions if warranted.

——

Demand elasticities show how the quantity of a product purchased changes as its price changes, all else equal. Cross-price elasticities show how a change in the price of one good can result in a change in the price of another good (either a substitute or a complement), thereby altering the quantity purchased. Income elasticity allows a modeler to forecast how much more of a good or service consumers will buy when their income increases. See Appendix A for more information about elasticities.

5-14
"Models are constructed to provide the simplest analysis possible that allows us to understand the issue at hand…. The real world is typically much more complex than the models we postulate. That doesn’t invalidate the model, but rather by stripping away extraneous details, the model is a lens for focusing our attention on specific aspects of the real world that we wish to understand" (McAfee and Lewis 2009). Models used to inform EPA decision-making should also be reliable, transparent, defensible, and useful (U.S. EPA 2009). Uncertainty tends to be higher when a model is either exceedingly simple (e.g., because it misses key interactions or feedbacks) or increasingly complex (e.g., due to data requirements). EPA guidance advises analysts to seek balance: "the optimal choice generally is a model that is no more complicated than necessary to inform the regulatory decision" (U.S. EPA, 2009).

When selecting models for use in regulatory analysis, analysts should evaluate the following:

**Is the model based on sound science?** Prior to use in regulatory analysis, the model should be subject to credible and objective peer review to ensure that it is consistent with scientific and economic theory and based on the best available data and empirical evidence. Many of the questions that follow can also be put to peer reviewers to evaluate the particulars of a specific model and/or appropriateness of the model within a specific policy context.

**Is the model fit for purpose?** It is important for analysts to identify the best model(s) for the analysis and thoroughly explain why it is applicable given the features and expected effects of the rule. A model may be based on sound science and still not be appropriate to evaluate the circumstances of interest.

**Is the model supported by the best available data?** Data quality and resolution may limit the ability to use some models in a regulatory context. For this reason, it is important to identify what data are available or can be collected to adequately parameterize the model (U.S. EPA, 2009). Analysts should base assumptions and calibration/estimation of key parameters that are peer reviewed.

**Does the model reasonably approximate the system or market(s) of interest?** A model should capture the most salient details of the policy and the system or markets affected. A model selected to evaluate a regulation should be no more complicated than is necessary to inform decision making. If model capabilities add complexity without substantially improving performance, the more transparent option is to eliminate them (NRC 2007).

**Is the model transparent?** In addition to model tractability, it is important that comprehensive documentation of all aspects of the model be publicly available, including details about model structure, key assumptions, sources and values of key parameters, and limitations. When possible, models and their underlying data should be publicly available.

**Can key assumptions or parameter values that warrant scrutiny be evaluated within the model?** Analysts should use sensitivity analysis to explore the robustness of results to key input values, specifications, or assumptions, particularly when the literature is inconclusive regarding the most defensible approach or estimates. Sensitivity analysis may be application specific: parameters that may matter little in one context may be key drivers of results in other contexts (U.S. EPA 2009).

Conducting uncertainty analysis is also important, as it "investigates the effects of lack of knowledge and other potential sources of error in the model" (U.S. EPA 2009). Sensitivity and uncertainty analysis inform users about the confidence that can be placed in model results. In some cases, analysts also may need to rely on multiple models. Section 5.5 provides detailed guidance regarding when sensitivity and uncertainty analyses are appropriate.

**What are the key limitations of the model?** Every model has its strengths and weaknesses. It is important that decision makers and stakeholders understand a model's limitations. What does the model capture? What is not captured or only captured with large bounds of uncertainty? These should be communicated and documented within the analysis in a way that is easy for a non-technical audience to interpret and understand.

It is also important for analysts to make consistent assumptions about firm and consumer behavior under the baseline and policy scenarios unless there is reason to believe the regulation will change underlying behavioral patterns. For example, the economics literature has found mixed evidence on whether car buyers fully account for future gasoline
expenses when choosing fuel economy. A fuel economy standard could reduce the impact of undervaluation of fuel economy on consumer decisions, but if such behavior occurs in the baseline, it is likely to persist regardless of regulatory requirements. Section 4.4.4 offers more discussion about possible insights from behavioral economics for policy design.

5.5.3 Technological Change

It is important to capture future changes in production techniques or pollution control that may influence the baseline, costs and benefits of regulatory alternatives. Technological change can be thought of as having at least two components: genuinely new technological innovation, such as the development and adoption of a new pollution control method; and learning effects, in which experience leads to cost savings through improvements in operations, capability, or similar factors.

While technological innovation in the regulated sector can reduce the cost of compliance, technological innovation can also affect the costs in other sectors and/or the benefits of the regulation. For example, the cost of phasing out ozone-depleting substances has declined over time due to technological improvements in substitutes. However, innovation in mitigating factors, such as improvements in skin cancer treatments and efficacy of sunscreen lotions have also occurred. Further, the analysis should include the costs associated with research and development (R&D), including the potential to crowd out other investments that would have occurred absent the regulation, to correctly value cost-reducing technological innovation, but only if the costs are induced by the regulation. Distinguishing R&D induced by the regulation from changes in other investment decisions is sometimes difficult.

While innovation will occur in the baseline and policy scenarios, rates across scenarios may differ because regulation may lead to innovation to reduce the cost of compliance. It is not advisable to assume a constant, generic rate of technological progress, even if the rate is small, simply because the continuous compounding of this rate over time can lead to implausible rates of technological innovation. In cases where small changes in technology could dramatically affect the costs and benefits, or where technological change is reasonably anticipated, the analyst should consider exploring these effects in a sensitivity analysis. This might include probabilities associated with specific technological changes or adoption rates of a new technology, or it may be an analysis of the rate required to alter the policy decision. Such an analysis should show the policy significance of emerging technologies that have already been accepted, or are, at a minimum, in development or reasonably anticipated.

In some cases, there also may be empirical evidence of reductions in costs as firms accumulate experience in production or abatement over time. Before incorporating learning effects, the analyst should carefully examine the existing evidence for relevance to the specific context at hand. Estimated learning effects can vary according to many factors, including already accumulated experience with a technology, industry, and the length of the time period considered. Also, because estimates of learning effects are based on doubling of cumulative production, including learning effects will have a greater influence on analyses with longer time horizons. See Chapter 8 for further discussion.

5.5.4 Compliance

One aspect of analytic design that can be complex is what to assume about the extent of compliance with current and future environmental regulations. Assumptions about compliance in both the baseline and policy cases can significantly affect the results of the analysis and should be clearly described. In most cases, a baseline and policy scenario that assumes full compliance should be analyzed. When an industry has not been regulated before, data will not typically be available to gauge the likelihood of compliance with a new rule, but compliance should be expected.

---

127 Recent studies suggest that consumers account for about fifty to one hundred percent of future gasoline expenses in their vehicle purchase decisions (Alcott and Wozny 2014; Busse et al. 2013; Sallee et al. 2016).
There are many proposed mechanisms by which environmental regulation could cause technological change. One mechanism is by induced innovation: the induced innovation hypothesis states that as the relative prices of factors of production change, the relative rate of innovation for the more expensive factor will also increase. This idea is well accepted; for example, Newell et al. (1999) found that a considerable amount of the increase in energy efficiency over the last few decades has been caused by the increase in the relative price of energy over that time.

A similar idea has also been described (somewhat less formally) as the “Porter Hypothesis” (Porter and van der Linde 1995, and Heyes and Liston-Heyes 1999). Jaffe and Palmer (1997) delineate three versions of the hypothesis: weak, narrow, and strong.

The weak version of the hypothesis assumes that an environmental regulation will stimulate innovation, but it does not predict the magnitude of these innovations or the resulting cost savings. This version of the hypothesis is very similar to the induced innovation hypothesis. The narrow version of the hypothesis predicts that flexible regulation (e.g., incentive-based) will induce more innovation than inflexible regulation and vice versa. There is empirical evidence that this is the case (Kerr and Newell 2003, and Popp 2003). Analysts may be able to estimate the rate of change of innovation under the weak or narrow version of the hypothesis, or under induced innovation. Note, however, that these types of innovation may crowd out other forms of innovation. By raising the cost of pollution, the regulation makes it profitable to find cheaper compliance strategies, but finding these strategies also has its own opportunity cost (e.g., firms use their engineers, scientists and other experts to develop more cost-effective compliance strategies instead of developing some other technology).

The strong version of the Porter Hypothesis predicts cost savings from environmental regulation under the assumption that firms do not maximize cost savings without pressure to do so. While anecdotal evidence of this phenomenon may exist, the available economic literature has found no statistical evidence supporting it as a general claim (Jaffe et al. 1995; Palmer et al. 1995; Jaffe and Palmer 1997; and Brännlund and Lundgren 2009). For the strong version to be true, it requires special assumptions and an environmental regulation combined with other market imperfections that are difficult to generalize. Thus, analysts should not assume cost savings from a regulation based on the strong version of the Porter Hypothesis.

**Text Box 5.3 - Technological Change, Induced Innovation, and the Porter Hypothesis**

When there are significant compliance issues with an existing regulation, an assumption of under-compliance in the baseline for a new regulation may be warranted when supported by data from monitors, inspections, or enforcement actions. Analysts may establish a “current practice” baseline incorporating data on actual compliance rates rather than assume full compliance. Current practice baselines are particularly useful for regulations intended to address compliance problems with existing policies. Assuming a full-compliance baseline that disregards under-compliant behavior could obscure the value of these types of regulations.\(^\text{128}\) If the policy being evaluated is not designed to address the underlying reason for non-compliance, then under-compliance data may be applicable to the policy case as well as the baseline.

If under-compliance is assumed either in the baseline or in the policy case, then identifying the reason for non-compliance is important and could affect the sign of the regulation's net benefits. For example, non-compliance could occur selectively where compliance costs are high. If compliance is not systematically correlated with costs, then the compliance assumption is less likely to change the sign of the regulation's net benefits.

When analyzing new requirements for an industry subject to existing regulations, it is important to carefully specify the assumptions about baseline compliance to avoid double counting benefits and costs from the same set of actions across multiple regulations. Assuming full compliance with existing regulations in the baseline makes it easier for analysts to focus on the incremental effects of the new regulatory action without double counting. If there is evidence of under-compliance in the baseline, analysts should consider whether the regulation is structured to reduce the compliance problem or whether the problem is likely to persist in the policy case. If it will persist and this behavior is not captured, the net benefits of a regulation will not be estimated correctly. For example, if analysts repeatedly factor under-compliance into the baselines for a sequence of emissions tightening rules but assume that entities will fully comply

---

\(^{128}\) For instance, banning lead from gasoline was precipitated, in part, by the noncompliance of consumers. When consumers put leaded gasoline in vehicles that required non-leaded fuel, this resulted in increased vehicle emissions (U.S. EPA 1985).
under the policy case, inconsistent results will arise. Summing the benefits and costs from the sequence of rules will overstate the benefits and costs because each rule takes credit for a portion of the same actions.

Conversely, there may be cases in which firms over-comply with regulations. Over-compliance in the policy scenario should be assumed in limited circumstances. As with under-compliance, it is important to identify the reason for over-compliance and assure it is consistent with expected behavior. The analysis should not typically assume that a regulation will motivate abatement greater than what is legally required. However, over-compliance may occur for reasons such as to reduce the risk of non-compliance or because least-cost compliance methods achieve greater reductions than required (e.g., shifting to a different process that does not pollute rather than installing abatement equipment). In such cases, the benefits and costs of over-compliance in the policy case should be accounted for. If more additional regulations are considered later on, current practices can be used to define baseline conditions for the new regulation unless these practices are expected to change.

To summarize, analysts should assume full compliance with regulations unless there is strong evidence to support an alternate assumption. Whenever scenarios other than full compliance are included in regulatory analysis, the analyst should discuss the sensitivity of the results to the compliance rate assumption.

### 5.5.5 Voluntary Actions

Occasionally, polluting industries adopt voluntary measures to reduce emissions. Firms or sectors can undertake such actions independently, or they might participate in formal, government-sponsored programs. Such voluntary measures are adopted for a variety of reasons, including to improve public relations, to avoid regulatory controls, to reduce other legal risks, or to access incentives associated with joining a formal program. When this is the case, it is important to account for these actions in the baseline for new regulations and to be explicit about the assumptions of firms’ future actions. If participation in voluntary programs was motivated by the threat of the regulation, then a new regulation could affect future participation in these programs.

Typically, voluntary emission reductions that are expected to occur without a new regulation may be included in the baseline consistent with the guidance on over-compliance above. This is not always possible, however, as voluntary actions are often difficult to measure (Brouhle, Griffiths, and Wolverton 2005). Sensitivity analysis could shed light on the importance of assumptions about voluntary emission reductions under the baseline if this is a significant source of uncertainty.

### 5.5.6 Changes in Other Environmental Contaminants

Decreases or increases in environmental contaminants that are not the subject of the regulation may occur for a variety of reasons that the analyst should consider. Projections of changes in the levels of all environmental contaminants should be consistent with expected economic behavior. These changes should be based on expected outcomes of least cost compliance, existing economic relationships, and continued compliance with existing regulations. The analysis should take a balanced approach to identifying increases and decreases in other contaminants that may be affected by the regulation.\(^{129}\)

As discussed in Section 5.1.3, changes in other environmental contaminants may result from the compliance approaches used by regulated entities. For example, the use of an abatement technology to reduce one air pollutant may simultaneously reduce other air pollutants from the same source, and/or it could change the emissions of another pollutant into a different medium (e.g., water). It is also possible for changes in other environmental contaminants to

\(^{129}\) The benefits from changes in environmental contaminants other than those related to the statutory objective of the regulation have sometimes been called “co-benefits”, and these contaminants sometimes called “co-pollutants”. However, these terms are imprecise and have been applied inconsistently in past practice, and as such should be avoided (unless these terms are used explicitly in statutes). Similarly, benefits from changes in environmental contaminants other than those related to the statutory objective of the regulation are sometimes referred to as “ancillary benefits” (or “ancillary costs”). These terms should be used cautiously in an analysis because they may be interpreted as having legal or policy meaning that is unintended.
occur as a result of market interactions. For example, a regulation may cause consumers or firms to substitute away from one commodity towards another, whose increased production may be associated with additional emissions of an environmental contaminant as well as the costs of abating it. Other examples include when a regulation induces beneficial reuse of a waste product and thereby reduces production and the associated emissions and costs of producing an input the waste product replaces; a controlled pollutant might be a precursor to multiple secondary pollutants.

Care should be taken when estimating changes in other contaminants to ensure they are consistent with expected market behavior. For example, consider an abatement technology that may potentially reduce emissions of multiple pollutants. The analyst should consider whether the technology will actually achieve similar reductions in all of the pollutants in new applications as it had in previous applications, or if the regulated entities will tailor the technology to control the regulated pollutant to reduce costs.

In estimating the welfare effects of changes in environmental contaminants other than those related to the statutory objective of the regulation, analysts should also consider the implications of existing pollution control regulations on behavior. For example, consider the case where a regulation on one pollutant leads to installations of a technology that reduces a second pollutant, and that second pollutant is subject to an allowance trading program with a cap that is economically binding (i.e., there is a positive allowance price). In this case, the regulation may not ultimately lead to reductions in the second pollutant. Instead, reductions in the second pollutant at regulated entities that install the new technology may be offset by reductions in abatement activities by entities subject only to the cap. As in this example, to the extent that any new regulation affects the cost of complying with an existing regulation, these changes in cost should be accounted for in the analysis.

If a regulation is expected to increase environmental contaminants not subject to the regulation, they should be accounted for in a BCA even if an anticipated future regulation is expected to mitigate them. This guidance follows directly from the specification of the baseline discussed in Section 5.2. It is important to account for these changes for completeness, such that the sum of the benefits and costs of rules evaluated in sequence should sum to the costs and benefits of the rules if evaluated collectively.

Finally, as discussed in Chapter 3, if the regulation is expected to induce large benefits from changes in environmental contaminant(s) beyond those arising from the primary statutory objective of the regulation, an analysis of a policy option where those contaminant(s) are regulated, either separately or simultaneously with the contaminants that are the primary statutory objective of the regulation, it may be useful to determine whether there are more economically efficient or appropriate ways of obtaining these unrelated benefits.

### 5.6 Uncertainty

Uncertainty is inherent in BCAs, particularly when estimating and valuing environmental benefits for which there are no existing markets. The primary issue is often not how to reduce uncertainty, but how to account for it and present useful conclusions to inform policy decisions. Most analyses specify baseline and policy scenarios based on the expected or most plausible outcomes. However, point estimates alone cannot provide policy makers with information about whether these estimates are robust to alternate assumptions nor can they convey the full range of potential outcomes. Treatment of uncertainty is an essential component of analysis that enhances the communication process between analysts and policy makers.

---

130 There may still be benefits (or negative benefits) from changes in the timing and location of emissions of these environmental contaminants even if the cap continues to bind. Chapter 4 describes how allowance trading programs work.

131 Stemming from definitions given in Knight (1921) economists have often distinguished risk and uncertainty according to how well one can characterize the probabilities associated with potential outcomes. Risk applies to situations or circumstances in which a probability distribution is known or assumed, while uncertainty applies to cases where knowledge of probabilities is absent. However, these definitions are not always adhered to in economics. Also, note that the economic definitions for these terms may differ from those used in other disciplines.
The guiding principles for assessing and describing uncertainty in analysis are transparency and clarity of presentation. Although the extent to which uncertainty is treated and presented will vary according to the specific needs of the analysis, some general minimum requirements apply to most BCAs. In assessing and presenting uncertainty, analysis should:

- present outcomes or conclusions based on expected or most plausible values;
- provide descriptions of all known key assumptions, biases, and omissions;
- perform sensitivity analysis on key assumptions;
- sensitivity analysis should examine both higher and lower values rather than only one or the other;
- justify the assumptions used in the sensitivity analysis; and
- make full use of available probability distributions of key parameters that drive benefit or cost estimates.

Sensitivity analysis on key assumptions may be all that is needed for an uncertainty analysis, or it may be only the initial assessment. Statistical confidence intervals and probability distributions, if available, are used to describe the statistical uncertainty associated with specific variables as well as sets of variables and to provide a more complete characterization of uncertainty. The outcome of the initial assessment may be sufficient to understand the influence of key parameters on outcomes and to inform the policy decisions. If, however, the implications of uncertainty are not adequately captured in the initial assessment then a more sophisticated analysis should be undertaken when the data allow. The need for additional analysis should be clearly stated, along with a description of the methods used for assessing uncertainty.

Probabilistic methods such as Monte Carlo analysis can be particularly useful because they explicitly characterize analytical uncertainty and variability. Where probability distributions of relevant input assumptions are available and can be feasibly and credibly combined, BCAs should characterize how the probability distributions of the relevant input assumptions would on net affect the resulting distribution of benefit and cost estimates. In this case the analysis should consider sources of uncertainty jointly rather than singly.

However, probabilistic methods can be challenging to implement when data needed to characterize distributions are limited. In the absence of data to specify distributions for specific parameter values, it is more transparent and defensible to use simpler sensitivity analysis. Note that for rules with annual economic effects of $1 billion or more, OMB Circular A-4 requires a formal quantitative uncertainty analysis that provides some estimate of the probability distribution of benefits and costs.

The analysis should make clear that the statistical uncertainty captured by the Monte Carlo or other probabilistic analysis generally does not account for model uncertainty, the degree to which mathematical models represent real-world systems. For example, when quantifying changes in a specific health effect from a reduction in an environmental contaminant, the statistical uncertainty analysis assumes that a particular dose-response model is the "true" model; that is we are 100 percent certain that there is a causal relationship, and that the dose-response function used in the analysis is the truth. There are some approaches to incorporating model uncertainty in probabilistic analyses, such as model averaging. More often, model uncertainty (including uncertainty over whether an environmental contaminant causes a specific type of health impact) will need to be captured and described independent of the statistical uncertainty analysis. When possible, alternative model specifications that are supported by or consistent with underlying biological, engineering or economic evidence or theory should be used to illustrate the consequences of assuming a different model.

---

132 Jaffe and Stavins (2007) provide a useful overview of probabilistic analysis of uncertainty in regulatory analysis.
133 Moral-Benito (2015) provides an overview of model averaging in economics.
5.6.1 Performing Sensitivity Analysis

Sensitivity analysis is a systematic method for describing how net benefit estimates or other outputs of the analysis change with assumptions about input parameters. Some basic principles for sensitivity analysis include:

- **Identify key parameters.** For most applied analyses, a full sensitivity analysis that includes every variable is not feasible. Instead the sensitivity analysis will often need to be limited to those input parameters considered to be key or particularly important, a determination informed by the range of possible values for input parameters and each one’s functional relationship to the output of analysis. The analyst should specify a plausible range of values for each key variable and describe the rationale for the range of values tested.

- **Vary these key parameters.** The most common approach is a partial sensitivity analysis that estimates the change in net benefits (for a BCA) or other economic outcome while varying a single parameter, leaving other parameters at their base value. A more complete analysis will present the marginal changes in the economic outcome as the input parameter takes on progressively higher or lower values. When an input has known or reasonably determined maximum and minimum values, it can be informative to investigate if outcomes are robust to these alternative input values.

Varying two parameters simultaneously can often provide a richer picture of the implications of base values and the robustness of the analysis but can be more difficult to communicate effectively. Analysts should consider using graphs to present these combined sensitivity analyses by plotting one parameter on the x-axis, the economic outcome on the y-axis, and treating the second parameter as a shift variable. Results of the sensitivity analysis should be presented clearly and accompanied with descriptive text.

- **Identify switch points.** Switch points are defined as those conditions under which the economic analysis would recommend a different policy decision. For BCA, the switch point would typically be the input parameter value where estimated net benefits changes sign. Switch point values for key input parameters can be very informative. For instance, they can be compared to the available literature to assess whether the values are plausible or well outside known distributions or observations. While switch points are not tests of confidence in the statistical sense, they can help provide decision-makers with an understanding of how robust the analytic conclusions are.

- **Assess the need for more detailed analysis.** Finally, sensitivity analyses may be used as a screening tool to determine where more extensive effort may be needed. For example, the plausible range of values for an influential uncertain parameter may be narrowed with further research or data gathering, which can be used to better characterize the parameter’s uncertainty. If several parameters independently have a large influence on the results of the analysis when they are varied, then a more sophisticated treatment of uncertainty that allows for joint consideration of their effects may be necessary.

5.6.2 Other Considerations Related to Uncertainty and Risk

There are additional issues related to uncertainty that may also merit consideration, including how individuals affected by environmental policies may perceive or respond to risk information, and how they evaluate policies with irreversible decisions when new information may become available.

Lay and expert risk perceptions: Lay perceptions of risk may differ significantly from scientific assessments of the same risk. An extensive literature has developed on the topic. When the analysis contains many highly uncertain variables, presentation may be facilitated by noting the uncertainty of each in footnotes and carrying through the central analysis using best point estimates. Because individuals respond according to their own risk perceptions, it is important for the analyst to be attentive to situations where there is an obvious divergence in these two measures. In such cases, analysts should clearly state the basis for the economic value

---

134 When the analysis contains many highly uncertain variables, presentation may be facilitated by noting the uncertainty of each in footnotes and carrying through the central analysis using best point estimates.

135 For a general overview see Renner et al. 2015.
estimates used or developed in their analysis. Analysts should also consider providing information to the public that may reduce differences between lay and expert perceptions of risk and may allay public concerns.

**Provision of information:** Some policy actions focus on providing information to individuals on risks to health and welfare. If this information allows them to make better decisions that improve household welfare, there is an economic benefit to providing this information. When this is the case, revealed preference approaches can make new information appear to have a net negative effect on household welfare because households may undertake new (and costly) activities in response. For example, information on drinking water quality may lead consumers to buy and use costly filtration systems at home, which could be misconstrued to mean that providing the information diminished consumer welfare. An appropriate framework for evaluating the benefits of information provision under these circumstances is to assess the costs of sub-optimal household decisions under the less complete information.\(^{136}\) Analysts should carefully consider these issues when they evaluate policies that focus on information provision.

**Quasi-option value:** Some environmental policies involve irreversible decisions that must be made in the face of uncertainty. If information that reduces this uncertainty can be expected to develop over time, then there is a positive "quasi-option" value to waiting until this information is available.\(^{137}\) In this case, the value originates from possessing the option to hold off on making the decision until uncertainties are resolved and an analysis can show the potential costs of making a decision without this new information. Generally, it is difficult to quantitatively include quasi-option values in an analysis of environmental policy, but the concept is useful and may be highlighted qualitatively if circumstances warrant.

---

### Chapter 5 References


---

\(^{136}\) Foster and Just (1989) describes this approach more fully, demonstrating that compensating surplus is an appropriate measure of willingness-to-pay under these conditions. The authors illustrate this with an empirical application to food safety.


U.S. EPA. 2005a, Rule to Reduce Interstate Transport of Fine Particulate Matter and Ozone (Clean Air Interstate Rule); Revisions to Acid Rain Program; Revisions to the NOX SIP Call; Final Rule, Federal Register. 70(91):25162-25405.


Chapter 6
Discounting Future Benefits and Costs

Discounting allows for economically consistent comparisons of benefits and costs that occur in different time periods. In practice, it is accomplished by multiplying changes in future consumption (including market and non-market goods and services) by a discount factor. Discounting reflects that (1) people prefer consumption today over consumption in the future, and (2) invested capital is productive and provides greater consumption in the future. Properly applied, discounting can tell us how much future benefits and costs are worth today.

Social discounting is the main type of discounting discussed in this chapter. This is discounting from the broad society-as-a-whole point of view embodied in benefit-cost analysis (BCA). Private discounting, on the other hand, is discounting from the specific, limited perspective of private individuals or firms. This distinction is important to maintain. Using a private discount rate instead of a social discount rate to evaluate the effects of a policy on societal well-being may incorrectly represent benefits and costs of regulatory alternatives, and in some cases may be significant enough to change the sign of net benefits.

This chapter addresses both discounting over the relatively short term, called intragenerational discounting, as well as discounting over much longer time horizons, or intergenerational discounting. Intragenerational, sometimes called conventional, discounting applies to contexts that may have decades-long time frames, but where the private planning horizon is within the lifetime of current generations. Intergenerational discounting addresses very long time horizons in which the effects being discounted will impact generations to come. To some extent, this distinction is a convenience as there is no discrete point at which one moves from one context to another. However, the relative importance of specific considerations, such as uncertainty and equity, can change as the time horizon lengthens, which can have implications for the appropriate discount rate to be used.

This chapter focuses on the most important discounting issues for applied policy analysis beginning with practical, basic mechanics and methods for discounting. It then turns to the theory and foundational logic for discounting and the different approaches to estimate discount rates. The sensitivity of the discount rate to the choice of discounting approach is discussed throughout this chapter. Other important issues include the economic rationale and framework for discounting, the Ramsey discounting framework, prescriptive vs. descriptive approaches to discount rate selection, private vs. social discounting, and declining discount rates. The chapter closes with recommendations for BCA. Foremost among these are:
• the full streams of benefits and costs over time should be shown without discounting;
• the full streams of benefits and costs should be appropriately discounted when calculating net-benefits;
• analysts should follow OMB's Circular A-4 guidelines (OMB 2003) and present the results using both a 3 percent and a 7 percent discount rate; and
• the consideration of other discount rates may be warranted for policies with long time horizons.\textsuperscript{138}

6.1 Mechanics and Methods for Discounting

The most common methods for discounting involve estimating either net present values or annualized values. \textsuperscript{139} An alternative method is to estimate a net future value. Net present value, annualization, and net future value are different ways to express and compare the costs and benefits of a policy in a consistent manner. These three methods will be discussed in turn below.

6.1.1 Net Present Value (NPV)

The net present value (NPV) of a stream of benefits and costs in the future is the value that those benefits and costs provide to society today. The NPV at time 0 of a projected stream of current and future benefits and costs is calculated by multiplying the benefits and costs in each year by a time-dependent weight, or discount factor, $d$, and adding all of the weighted values. This can be done by discounting the benefits and subtracting the discounted costs, or by discounting the net benefits over all $n$ years as shown in the following equation:

\[ NPV = B_0 + d_1B_1 + ... + d_{n-1}B_{n-1} + d_nB_n \]

\[ - C_0 - d_1C_1 - ... - d_{n-1}C_{n-1} + d_nC_n \]

\[ = NB_0 + d_1NB_1 + d_1NB_2 + ... + d_{n-1}NB_{n-1} + d_nNB_n \]

\[ = NB_0 + d_1NB_1 + d_1NB_2 + ... + d_{n-1}NB_{n-1} + d_nNB_n \]

where

$B_t$ are the benefits in year $t$,

$C_t$ are the costs in year $t$, and

$NB_t$ are net benefits, the net difference between benefits and costs ($B_t - C_t$) in year $t$.

Alternatively, NPV can be calculated by estimating the present value of costs and the present value (PV) of benefits separately and then the PV of costs from the PV of benefits:

\[ NPV = (B_0 + \sum_{t=1}^{n} d_tB_t) - (C_0 + \sum_{t=1}^{n} d_tC_t) \] \textsuperscript{(2)}

The discounting weights, $d_t$, are given by:

\[ d_t = \frac{1}{(1+r)^t} \] \textsuperscript{(3)}

\textsuperscript{138} This chapter summarizes some key aspects of the core literature on social discounting, but it is not a detailed review of the vast and varied social discounting literature. Excellent sources for additional information are: Lind (1982a, 1982b; 1990; 1994), Lyon (1990, 1994), Pearce and Turner (1990), Pearce and Ulph (1994), Arrow et al. (1996), Portney and Weyant (1999), Frederick et al. (2002), Moore et al. (2004), Spackman (2004), Groom et al. (2005), Cairns (2006), Zerbe and Burgess (2011a), Moore, et al. (2013a), and Harberger and Jenkins (2015).

\textsuperscript{139} Note that discounting is distinct from inflation, although observed nominal market rates of return reflect expected inflation. While most of the discussion in this chapter focuses on real discount rates and values, benefits and costs should also be adjusted for inflation when relevant.
where $r$ is the discount rate and $t$ is the year.

### 6.1.1.1 Beginning-of-year vs. end-of-year discounting

In the NPV equation, $B_0$, $C_0$, and $NB_0$ are the benefits, costs, and net benefits incurred $r$ immediately (when $t=0$) so they are not multiplied by a discount factor. This makes sense when time is continuous, but what is “immediate” becomes less clear when time, $t$, is an entire year. For example, if a rule is finalized at the beginning of a year and costs and benefits will be realized throughout that year, are these values “immediate” or should they be discounted one period? If costs and benefits incurred throughout the year are considered immediate, then they would be $B_0$ and $C_0$ in equation (1) above. This is known as beginning-of-year discounting because all intra-annual effects in the current year are treated as if they occur at the beginning of the year, when $t=0$. The alternative is to treat all intra-year effects in the current year as if they occur at the end of the year, when $t=0$, and discount them back one period. Effects in the next year would then be discounted back two periods. This is known as end-of-year discounting. The choice between beginning- or end-of-year discounting does not generally have a large quantitative effect on the analysis. Whichever approach is adopted should be explicitly stated and should be applied to both benefits and costs so that the analysis is internally consistent.

### 6.1.1.2 Time periods of less than one year

When estimating the NPV, it is important to explicitly state how time periods are designated and when, within each time period, costs and benefits accrue. Typically, time periods are in years, but alternative time periods can be justified if costs or benefits accrue at irregular or non-annual intervals. To correctly discount intra-year effects, the annual discount rate, $r$, must be adjusted to an “effective rate” $\tilde{r}$, which produces the same result as the annual discount rate if compounded for one year. The effective discount rate for any non-annual period is

$$\tilde{r}_t = (1 + r)^{1/(\text{# of periods})} - 1$$

(4)

For example, if the annual discount rate is 7% and costs are incurred on a quarterly basis (i.e., there are four periods in a year), then the effective quarterly discount rate, $\tilde{r}$, is approximately 1.7%. The formula for discounting weights, $d_t$, given above, can now be used with this effective rate, but $t$ is measured in quarters rather than years.

While the discounting formula can be adjusted to account for intra-annual discounting periods, it may not be necessary unless very exact values are required. The NPV generated by an intra-annual effective discount rate, $\tilde{r}$, will be between the NPVs using beginning-of the year discounting and the NPV using end-of-the-year discounting using the annual discount rate, $r$. These NPVs don't usually differ by much in a typical economic analysis.

### 6.1.1.3 Continuous discounting

Costs and benefits may also be discounted on a continual basis during the year. In this case, benefits or costs occurring at the end of a future year (or period) $t$ are discounted by the weight:

$$d_t = e^{-\tilde{r}t}$$

(5)

Where $e$ is the irrational number, which when rounded to three decimal places is 2.783 and is the base of the natural logarithm. This is a commonly used expression in economics and finance. Furthermore, continuous discounting provides a convenient way to represent a discount weight for some theoretical economic concepts related to discounting. As with intra-annual discounting discussed above, an effective discount rate, $\tilde{r}$, should be used to produce the same result as the annual discount rate. The effective discount rate for continuous discounting is

$$\tilde{r}_t = \ln (1 + r)$$

(6)

In this case, $t=1$ represents one year, but the discounting weight is assumed to be applied to every moment, continuously throughout the year.
### 6.1.2 Annualized Values

An annualized value is an illustrative cost or benefit which, if incurred every year over the entire length of the analysis, would produce the *same net present value (NPV)* as the original time-varying stream of costs, benefits, or net benefits. In some cases, annualized values are easier to understand than NPV.

Because the annualized value is constructed to generate the same net present value as the actual stream of values, comparing the annualized values is equivalent to comparing the net present values. That is, one can use either the NPV or the annualized values to determine whether benefits exceed costs or to determine which option produces the highest net benefits. As with NPV, benefits and costs may be annualized separately and then compared, or the stream of net benefits can be annualized.

The formulas below illustrate the estimation of annualized costs; the formulas are identical for benefits.\(^{(140)}\) The exact equation for annualizing depends on whether there are any immediate costs (i.e., any costs at time zero, \(t=0\)).

**Annualized costs when there is no cost at \(t=0\) (e.g., no \(C_0\) in equation (1))** are estimated using the equation:

\[
\text{Annualized Cost} = \frac{PVC \times (1+r)^n}{(1+r)^{n+1}-1}
\]  

(7)

where

- *Annualized Cost* = annualized cost accrued at the end of each of \(n\) years;
- *PVC* = present value of costs (estimated as in equation (1), above);
- \(r\) = the discount rate per year; and
- \(n\) = the timeframe of the annualization.

**Annualized costs when there is initial cost at \(t=0\)** are estimated using a slightly different equation:

\[
\text{Annualized Cost} = \frac{PVC \times (1+r)^n}{(1+r)^n-1}
\]  

(8)

Note that the numerator expression is the same in both equations, although the PVC is calculated differently depending upon whether there are costs at \(t=0\). The only difference is the “\(n+1\)” and “\(n\)” terms in the denominator of (7) and (8).

Some important caveats are associated with the use of annualized values. First, they are generally illustrative in nature; the annualized value is not the actual value that will manifest every year. Second, the annualized value changes with the timeframe of the annualization. This means that the annualized value will be *different* for each value of \(n\), even for the same discount rate, \(r\). The longer the timeframe assumed for the annualization, the lower the annualized value.

One special case of equation (7) (the annualization formula when there is no cost at \(t=0\)) is when \(n=\infty\). In this case, the annualized cost is simply

\[
\text{Annualized Cost} = PVC \times r
\]  

(9)

For example, suppose an action permanently eliminates the use of an environmental amenity (e.g., a wetland) and the estimated present value of that amenity is $1 million at a discount rate of 3\%. The cost of this policy is the lost value of the amenity in perpetuity -- the period of the analysis is effectively infinity. The annualized cost of that policy (that is, the

\(^{(140)}\) Variants of these formulas may be common in specific contexts. See, for example, the Equivalent Uniform Annual Cost approach in the EPA’s Air Pollution Control Cost Manual (U.S. EPA 2017).
cost that if lost every year, forever, would be equivalent to $1 million in present value today) is $1 million * 3% = $30,000 per year.

The corollary to equation (9) is

\[ PV_C = \frac{\text{Annualized Cost}}{r} \]  \hspace{1cm} (10)

Thus, if an environmental amenity is estimated to be worth $30,000 per year, the present value of that amenity using a 3% discount rate is $1 million, assuming that the amenity provides benefits into perpetuity.

### 6.1.3 Net Future Value (NFV)

Instead of discounting all future values to the present using the NPV, it is possible to estimate the stream of values from the perspective of some future year, for example, at the end of the last year of the policy’s effects, \( n \). This would be the net future value (NFV). This might be particularly useful when conducting a retrospective analysis.

The net future value for net benefits (\( NB_t \)) is estimated using the following equation:

\[ NFV = a_0NB_0 + a_1NB_1 + a_2NB_2 + \ldots + a_{n-1}NB_{n-1} + NB_n \]  \hspace{1cm} (11)

Where, as before, \( NB_t \) are net benefits, \((B_t - C_t)\), in year \( t \). This formula can also be used for either benefits or costs alone.

In the NFV equation, the *accumulation* weights, \( a_t \), are different from the discounting weights in equation (3) for NPV, and are given by:

\[ a_t = (1 + r)^{(n-t)} \]  \hspace{1cm} (12)

where \( r \) is the annual discount rate. The net future value for year \( n \) can be expressed in relation to the net present value for \( t=0 \), as follows:

\[ NPV = \frac{NFV}{(1+r)^n} \]  \hspace{1cm} (13)

The NFV can be modified for intra-annual values by using an effective discount rate as described in the NPV section above. It can also be calculated assuming continuous accumulation using the effective discount rate in equation (6) and accumulation weights:

\[ a_t = e^{\bar{r}t} \]  \hspace{1cm} (14)

The only difference between equation (14) and equation (5) is the use of \( \bar{r} \) rather than \(-\bar{r}\) in the exponent.

### 6.1.4 Comparing the Methods

NPV represents the value of a stream of costs and benefits from some point in time (often the present moment) going forward. NFV represents the value of the stream of costs and benefits at some future time. Annualization is the calculation of a constant, annual value for costs and benefits that would produce the same NPV as the actual stream of costs and benefits.

Depending on the circumstances or application of the analysis, one method might have certain advantages over the others. Discounting to the present to get a NPV is likely to be the most informative for the standard economic analysis of a policy that will generate future benefits and costs. NFV may be more appropriate for evaluating the cumulative impacts of regulation or when conducting a retrospective analysis. Annualized values may be used in conjunction with the NPV as a means of communicating the result or comparing options when the costs or benefits are highly variable over time. It is important to remember, however, that annualized values are sensitive to the annualization period -- the
annualized value will be lower the longer the annualization period -- so analysts should be aware of potentially different annualization periods when comparing annualized values from one analysis to those from another.\textsuperscript{141}

The choice of discount rate affects the values generated by these discounting methods. For a given stream of net benefits, the NPV will be lower with higher discount rates, the NFV will be higher with higher discount rates, and the annualized value may be either higher or lower depending on the time which impacts occur and the length of time over which the values are annualized. The important point is that the ranking among regulatory alternatives is unchanged across these three methods for any given discount rate.

\subsection*{6.1.5 Sensitivity of Net Present Value Estimates to the Discount Rate}

Both the size and sign of NPV can be sensitive to the choice of discount rate when there is a significant difference in the timing of costs and benefits. This is the case for policies that require large initial outlays or have long delays before benefits are realized, as do many EPA policies. Text Box 6.1 illustrates how discount rates affects NPV.

In other cases, the discount rate is not likely to affect the sign of the NPV estimate. Specifically, the NPV will not be affected by the discount rate when:

- All effects occur in the same period. In this case, discounting may be unnecessary or superfluous because net benefits are positive or negative regardless of the discount rate used.
- Costs and benefits of a policy occur consistently over the period of the analysis and their relative values do not change over time.

In these cases, whether the NPV is positive does not depend on the discount rate, but the discount rate can still affect how the present value compares to another policy.

\subsection*{6.1.6 Issues in Discounting Applications}

There are several important analytic components that need to be considered when discounting costs and benefits.

\subsubsection*{6.1.6.1 Consistent use of the discount rate}

It is important that the same discount rate be used for both benefits and costs, as the discount rate reflects society's intertemporal preferences for trading off consumption over time. This allows for a consistent comparison of results across policies and prevents the discount rate from being used as a tool to justify a preordained policy. A high discount rate reduces the weight given to costs and benefits in the future and minimizes their impact on the NPV, whereas a low discount rate weights future impacts more heavily and increases their impact on the NPV. Therefore, almost any policy can be justified by using a sufficiently low discount rate for benefits and a sufficiently high rate for costs. The inverse is also true: almost any policy can be rejected by using a high discount rate for benefits and a low rate for costs.

\subsubsection*{6.1.6.2 Future value of environmental effects and uncertainty}

There are two issues that are sometimes confounded with social discounting and the choice of social discount rate, but should be treated separately: the value of environmental impacts may change over time, and future benefits and costs may be uncertain. While these issues are important, they both should be addressed separately in the economic analysis rather than adjusting the discount rate to account for them.\textsuperscript{142}

\textsuperscript{141} This is important when aggregating the cost-savings across multiply regulations to comply with Executive Order 13771.

\textsuperscript{142} See, for example, Moore et al. 2017.
Text Box 6.1 - Potential Effects of Discounting

Suppose the benefits of a given program occur 30 years in the future and are valued (in real terms) at $5 billion at that time. The rate at which the $5 billion future benefits is discounted can dramatically alter the economic assessment of the policy: $5 billion 30 years in the future discounted at 1 percent is worth $3.71 billion in the present, at 3 percent it is worth $2.06 billion, at 7 percent it is worth $657 million, and at 10 percent it is worth only $287 million. In this case, changing the discount rate from 1 percent to 10 percent generates more than an order of magnitude of difference in the present value of benefits. Longer time horizons will produce even more dramatic effects of discounting on a policy’s NPV. After 100 years, the present value of $5 billion is $260 million at 3 percent and only $5.8 million at 7 percent. (See Section 6.3 on intergenerational discounting). Particularly in the case where costs are incurred in the present and therefore are not affected by the discount rate, it is easy to see that the choice of the discount rate can determine whether a policy has positive or negative net benefits.

The future value of environmental effects (i.e., the “current price” in future years) depends on many factors, including the availability of substitutes and the level of wealth in the future. For example, the relative price of environmental goods in the future may rise if those environmental goods are expected to become scarcer over time. These changes in relative prices should be applied to future effects and the associated values discounted, but the discount rate itself should not adjusted to incorporate a change in relative prices. (An exception would be when using a decreasing social discount rate using a Ramsey Framework, Section 6.3.3.)

Uncertainty or riskiness of future benefits and costs from a policy should also not be incorporated into the social discount rate. While it is technically possible to adjust the discount rate to account for uncertainty, doing so may hide important assumptions and information from decision makers. Uncertainty about future values should be treated separately from discounting.

However, uncertainty about the discount rate itself is different from uncertainty of future benefits and costs and can affect discounting. Social discounting the using the ‘Ramsey’ Framework (Section 6.2.2) reflects: (1) the amount of time between the present and the point at which in consumption impacts occur; (2) the rate at which consumption is expected to change over time in the absence of the policy; (3) the rate at which the marginal value of consumption diminishes with increased consumption; and (4) the rate at which the future utility from consumption is discounted with time. Changes in these components or uncertainty about them can lead to a discount rate that changes over time (Section 6.3.3). But for many analyses, and particularly for intragenerational discounting, a fixed discount rate without considering uncertainty in the discount rate itself may be sufficient.

6.1.6.3 Placing Effects in Time

Placing effects properly in time is essential for all calculations involving discounting. Analyses should account for implementation schedules and the resulting changes in emissions or environmental quality, including possible changes in behavior that occur between the announcement of policy and compliance deadlines. Additionally, a lag time may occur between changes in environmental quality and the corresponding change in welfare. It is the change in welfare which defines economic value, and not the change in environmental quality itself. The EPA’s Science Advisory Board addressed this issue in 2001 for the Arsenic Rule (U.S. EPA 2001). If exposure to arsenic in drinking water is reduced, the number of cancer case is expected to decline over time to a lower, steady-state level. How fast this reduction in risk occurs depends on the "cessation-lag" following reduction in exposure. Enumerating the full time path of welfare changes is essential for proper valuation and BCA.143

---

143 It is inappropriate to characterize the net-benefits or effects of the regulation by only reporting benefits or costs over any one period (e.g., year). For example, capital outlays for compliance may be required the year the rule comes into effect, but not afterward, with benefits realized in following years. Presenting only the effects of the rule in the first year would misrepresent the rules net benefits and full consequences.
6.1.6.4 Length of the Analysis

As described in Chapter 5 (Section 5.2), while there is little theoretical guidance on the appropriate time horizon of economic analyses, a guiding principle is that the time horizon should be sufficient to capture major welfare effects from policy alternatives. This principle is consistent with the underlying requirement that BCA reflect the welfare outcomes of those affected by the policy. Previously, $n$ was defined as the final period of the policy’s future impacts but is possible that it is impractical to model all future years. A solution may be to consider that the time horizon, $T \leq n$, of an analysis should be chosen such that:

$$\sum_{t=T}^{\infty} (B_t - C_t) \frac{1}{(1+r)^t} \leq \varepsilon,$$

(15)

where $\varepsilon$ is a tolerable estimation error for the NPV of the policy. That is, the time horizon should be long enough that the net benefits for all future years (beyond the time horizon) are expected to be negligible when discounted to the present. In practice, however, it is not always obvious when this will occur because it may be unclear whether or when the policy will be changed by policy makers, whether or when the policy will become obsolete or “non-binding” due to exogenous technological changes, how long the capital investments or displacements caused by the policy will persist, etc.

As a practical matter, reasonable alternatives for the time span of the analysis may be based on assumptions regarding:

- The expected life of capital investments required by or expected from the policy (e.g., when the emissions of flow pollutants are affected);
- The point in time at which benefits and costs are negligible for the indefinite future;
- Statutory or other requirements for the policy or the analysis; and/or
- The extent to which benefits and costs are allocated to generations.

The time horizon choice for the analysis should be explained and well-documented, and the analysis should highlight the extent to which the sign of net benefits or the relative rankings of policy alternatives are sensitive to the choice of time horizon.

6.1.6.5 Discounting Non-Monetized Effects

A common criticism of discounting for environmental policies is that health impacts such as “lives saved” or physical impacts such as “improved water quality” are not like money flows. They cannot be deposited in a bank and withdrawn later after earning interest. This criticism does not appreciate that the valuation approaches are designed to estimate the amount of money that is as valuable to individuals as the environmental or health effects being examined. If all environmental and health impacts have been appropriately valued (monetized), we can proceed to discount those money-equivalent flows just as we would discount real money flows over time.

However, beneficial effects cannot always be monetized. In this case, the undiscounted stream of the non-monetized effects should be presented as they occur over time. As a general matter, these non-monetized effects should also still be discounted in benefit-cost analysis and cost-effectiveness analysis if they are aggregated over time. This is because they are assumed to hold some value, albeit unspecified, and, discounting assumes that individuals prefer the benefit of that value today over the future. This is usual practice in cost-effectiveness analysis where monetized costs and non-monetized effectiveness measures are both discounted. OMB Circular A-4 (OMB 2003) recommends discounting non-monetized health effects for cost-effectiveness.

For some effects, however, the (unknown) marginal value of a change in the non-monetized effect might be dependent upon the level and timing of that change. That is, marginal values are not constant. For example, suppose there are annual emissions thresholds below which environmental effects are negligible, but above which lead to major environmental damages. The economic value of emissions depends upon whether those emissions are above or below
this threshold and discounting these economic values would be appropriate. If we lack these values, however, and discount the effects themselves, we are treating all changes as if they had the same value. Here it would be preferable to only display the undiscounted stream of non-monetized effects with an appropriate justification and explanation.

6.2 Background and Rationales for Social Discounting

The goal of social discounting is to compare benefits and costs that occur at different times based on the rate at which society is willing to make such trade-offs. The analytical and ethical foundation of the social discounting literature rests on the traditional test of a potential Pareto improvement in social welfare; whereby those who, on net, benefit from a policy could potentially compensate those who, on net, experience costs, such that everyone is at least as well off as they were before (see Chapter 1 and Appendix A). This framework casts the consequences of government policies in terms of individuals contemplating changes in their own consumption over time. Trade-offs (benefits vs. costs) in this context reflect the preferences of those affected by the policy, and the time dimension of those trade-offs should reflect the intertemporal preferences of those affected. Thus, social discounting should seek to mimic the discounting practices of the affected individuals. Simultaneously, social discounting must reflect social tradeoffs in consumption over time, which may differ from tradeoffs from a private, individual perspective.

The literature on discounting often uses a variety of terms and frameworks to describe identical or very similar key concepts. For the purposes of the Guidelines, we focus on the following terminology and fundamental concepts for defining a social discount rate:

- **The social rate of time preference** is the discount rate at which society is willing to trade consumption in one period (usually year) for consumption in the next period.

- **Consumption rate of interest** is the rate at which individuals are is willing to trade consumption in one period (usually year) for consumption in the next period. This rate reflects the individual’s rate of time preference and, following the potential Pareto principle, the social rate of time preference should be based on this individual rate.

- **The social opportunity cost of capital** is the rate at which consumption in the next period is reduced because private investment is displaced by required investments from policy. This is the rate at which society can trade consumption over time due to productive capital. Social discounting should account for future consumption losses from displaced investment.

- **Market interest rates** are what we observe in the markets for loanable funds. We observe several real market rates which, to varying extents and accounting for tax distortions, can be taken as estimates for the individual rates of time preferences and the social opportunity cost of capital needed for social discounting.

Social discounting is largely concerned with the relationships among these concepts and how they are measured.

6.2.1 Consumption Rate of Interest and Social Opportunity Cost of Capital

If capital markets were perfect and complete with no distortions or uncertainties, the market interest rate would equal both the consumption rate of interest and the social opportunity cost of capital: it reflects both how individuals value present vs. future consumption and how productive capital can be transformed into future consumption. Following the potential Pareto principle and valuing future costs and benefits in the same way as the affected individuals, this market rate would be the appropriate social discount rate.

However, such perfect and complete markets do not exist. Private sector returns are taxed (often at multiple levels), capital markets are not perfect, and capital investments often involve private (and not necessarily social) risks. These factors cause us a divergence in the consumption rate of interest and the social opportunity cost of capital. That is, there

144 The term consumption is broadly defined to include both the use of both private and public goods and services by households in BCA, and includes the intergenerational nature of this change in consumption.
is now a divergence between the rates at which individuals and society can trade consumption over time. These are two different rates once we consider the role of taxes. Text Box 6.2 illustrates how these rates can differ.

A large body of economic literature analyzes the implications for social discounting of divergences between the consumption rate of interest and the social opportunity cost of capital. Most of this literature is based on the evaluation of public projects, but many of the insights still apply to regulatory BCA, and the dominant approaches from the literature are briefly outlined here. More complete recent reviews can be found in Spackman (2004), Zerbe and Burgess (2011b), Moore et al. (2013a, 2013b), and Harberger and Jenkins (2015). Section 6.2.2 discusses social discounting using the consumption rate of interest as the social rate of time preference, whereas Sections 6.2.3 and 6.2.4 discuss methods for discounting changes in investments.

### 6.2.2 Social Rate of Time Preference as the Social Discount Rate

If costs and benefits can be represented as changes in consumption profiles over time, then discounting should be based on the rate at which society is willing to postpone consumption today for consumption in the future. Thus, the rate at which society is willing to trade current for future consumption, or the social rate of time preference, is the appropriate discounting concept for evaluating public policy decisions.

The social rate of time preference differs from individual rates of time preference. An individual rate of time preference includes factors such as the probability of death, whereas society can be presumed to have a longer planning horizon. Additionally, individuals routinely are observed to have several different types of savings, each possibly yielding different returns, while simultaneously borrowing at different market interest rates. For these and other reasons, the social rate of time preference is not directly observable and may not equal any particular market interest rate. Generally, there are two primary approaches to deriving the social rate of time preference.

#### 6.2.2.1 Estimating a Social Rate of Time Preference Using Risk-Free Assets

One common approach to estimate the social rate of time preference is to use the market rate of interest from long-term, risk-free assets such as government bonds. The rationale behind this approach is that this market rate reflects how individuals discount future consumption, and government should value policy-related consumption changes as individuals do. In this approach, the social discount rate should equal the consumption rate of interest found in the market.

In principle, estimates of the consumption rate of interest could be based on after-tax interest rates consumers face for either saving (i.e., lending) or borrowing. Because individuals have different marginal tax brackets, different levels of assets, and different opportunities to borrow and invest, the type of market interest rate that best reflects the consumption rate of interest will differ among individuals. However, the fact that, on net, individuals generally accumulate assets over their working lives suggests that the after-tax returns on savings instruments widely available to the public will provide a reasonable estimate of the consumption rate of interest for society.

The historical rate of return on long-term government bonds, after-tax and in real terms, is a useful measure because it is relatively risk-free, which maintains the distinction between risk and social discounting described in section 6.1.6. Also, because these are long-term instruments, they provide more information on how individuals value future benefits over time frames which are more relevant for policy analysis.
6.2.2.2 Estimating a Social Rate of Time Preference Using the ‘Ramsey’ Framework

A second option is to construct the consumption rate of interest as the social rate of time preference in a framework attributed to Ramsey (1928), which explicitly reflects: (1) preferences for utility in one period relative to utility in a later period; and (2) the value of additional consumption as income changes. These factors are combined in the equation:

\[ r = \rho + \eta g \]  

(16)

where

\( r \) = the consumption rate of interest that can be used as the social rate of time preference;

\( \rho \) = the pure rate of time preference;

\( \eta \) = the elasticity of marginal utility with respect to consumption; and

\( g \) = the consumption growth rate.

The pure rate of time preference, \( \rho \), is the rate at which the representative agent discounts utility in future periods due to a preference for utility sooner rather than later. The elasticity of marginal utility with respect to consumption, \( \eta \), defines the rate at which the wellbeing from an additional dollar of consumption declines with total level of consumption. The consumption growth rate, \( g \), defines how consumption is expected to grow over time, for example, it may be expected to increase because incomes are expected to increase over time. Estimating a social rate of time preference in this framework requires information on each of these arguments, and while \( t \eta \) and \( g \), can be derived from data, \( \rho \) is unobservable and must be assumed.145 A more detailed discussion of the Ramsey equation can be found in Text Box 6.3 and a discussion of using the Ramsey framework to guide intergenerational discounting can be found in section 6.3.1.

6.2.3 Social Opportunity Cost of Capital as the Social Discount Rate

The social opportunity cost of capital recognizes that funds for government projects or funds required to meet government regulations may have an opportunity cost in terms of foregone investments and therefore, future consumption. If a regulation displaces private investments, society would lose the total returns from those forgone investments, including the tax revenues generated. In these cases, ignoring such capital displacements and using a social discount rate equal to the consumption rate of interest, which is post-tax, does not capture the fact that society loses the higher total returns (pre-tax) on forgone investments.

145 The Science Advisory Board (SAB) defined discounting based on a Ramsey equation as the “demand-side” approach, noting that the value judgments required for the pure rate of time preference make it an inherently subjective concept (U.S. EPA 2004c).
The Ramsey discounting framework provides an intuitive approach to thinking about, and potentially calibrating, the social discount rate. The Ramsey framework can be derived by considering a representative individual with utility in period $t$ denoted by $u(c_t)$, where $c_t$ denotes consumption. The agent is assumed to make choices to maximize lifetime welfare $\int_0^T e^{\rho t} u(c_t) dt$, where $\rho$ is the pure rate of time preference, or the rate at which the agent discounts utility. In this case, the minimum rate of return, $r$, the representative agent would require for a one period investment that cost one dollar in consumption would satisfy:

$$\frac{du}{dc_t} = e^{-\rho} \frac{du}{dc_{t+1}}$$ (1)

This states that the agent would require the benefits of the investment in terms of increased utility from the extra consumption in period $t+1$ discounted back at the pure rate of time preference, $\rho$, to be equal to the costs of the investment in terms of the forgone utility in period $t$. Here, the rate $r$ defines the additional benefits beyond getting back the initial investment that would be required of investment for the agent to be just as well off as before. Therefore, the rate $r$ also represents the consumption discount rate appropriate for comparing a future change in consumption with a change in consumption in the current period.

The relationship shown above can be used to solve for the consumption discount rate under certain assumptions. First, assume that the utility function has the commonly assumed iso-elastic functional form, such that:

$$u(c_t) = \frac{c_t^{1-\eta}}{1-\eta}$$

where $\eta$ is the (inverse) elasticity of marginal utility. Second, for simplicity, assume that consumption grows over time at a constant rate $g$. Substituting these assumptions into the equation in (1) and taking the natural log of both sides yields the definition for $r$, where:

$$r = \rho + \eta g$$

This definition highlights two reasons that future changes in consumption should be discounted. First is a general preference by individuals for utility sooner rather than later, as captured by the pure rate of time preference, $\rho$, which measures the rate at which individuals discount their own utility over time (taking a positive view of the optimal growth framework). The second reason is that a marginal change in consumption in the future may not have the same value as a marginal change in consumption today, represented by the term, $\eta g$. For example, if baseline consumption is expected to increase over time as income increases this will cause the marginal utility of consumption to decrease, implying that a future change in consumption will be valued less than a contemporaneous change.

As shown by Ramsey (1928), in an economy with no taxes, market failures, or other distortions, the social discount rate as defined in equation (2) would be expected to equal the market interest rate. The market interest rate, in turn, would be equal to the social rate of return on private investments and the consumption rate of interest. However, pre-existing distortions and market failures cause these rates to diverge in practice.

Private capital investments might be displaced if, for example, public projects are financed with government debt or regulated firms cannot pass through capital expenses to households, and the supply of investment capital is relatively fixed. The resulting demand pressure in the investment market will tend to raise market interest rates and reduce private investments that would otherwise have been made.\footnote{Another justification for using the social opportunity cost of capital argues that the government should not invest (or compel investment through its policies) in any project that offers a rate of return less than the social rate of return on private investments. While it is true that social welfare will be improved if the government invests in projects that have higher values rather than lower ones, it does not follow that rates of return offered by these alternative projects define the level of the social discount rate. If individuals discount future benefits using the consumption rate of interest, the correct}
cost of capital as the social discount rate in a BCA for an environmental policy requires that foregone investments fully crowd out private investments.\textsuperscript{147}

In principle, the social opportunity cost of capital can be estimated by a pre-tax, marginal, risk-free rate of return on private investments, but this rate is not observed in the marketplace. As a result, these values are sometimes derived by using National Accounts data to estimate rates of return on reproducible capital (e.g., Zerbe and Burgess, 2011b; Harberger and Jenkins, 2015), though there are some differences in the exact accounts included and their relative weights across these analyses. In practice, average returns that are likely to be higher than the marginal returns are typically observed, given that firms will make the most profitable investments first. In fact, it is not clear how to estimate marginal returns. Observed rates also reflect some unknown risk premium faced in the private sector, which causes them to be higher than a risk-free rate.

\subsection*{6.2.4 Shadow Price of Capital Approach}

As noted above, because capital markets are taxed and suffer from other market distortions, the consumption rate of interest and the social opportunity cost of capital are not equal. This means that costs and benefits that affect consumption should be discounted at a different rate than costs and benefits that affect investment. The shadow price of capital approach adjusts the costs and benefits that affect investment into equivalent consumption impacts (i.e., their shadow values) that reflect the social value of altered private investments.\textsuperscript{148} All impacts -- costs and benefits that affect consumption and the shadow costs and benefits that affect investment -- are then discounted using the social rate of time preference that represents how society trades and values consumption over time.\textsuperscript{149} Many sources recognize this method as the preferred analytic approach to social discounting for public projects and policies.\textsuperscript{150}

The shadow price, or social value, of private capital investment captures the fact that a unit of private capital produces a stream of social returns at a rate greater than that at which individuals discount them, due to distortions in the capital market noted earlier. If the social discount rate is the consumption rate of interest, then the social value of a $1 private capital investment will be greater than $1. This is because a capital investment produces a rate of return for its owners equal to the consumption rate of interest (which is post-tax), plus a stream of tax revenues for the government (generally considered to be used for consumption). Text Box 6.4 illustrates this idea of the shadow price of capital.

When compliance with environmental policies displaces private capital investments, e.g., machinery and equipment the shadow price of capital approach suggests first adjusting any capital-displacing project or policy cost upward by the shadow price of capital, and then discounting all costs and benefits using a social discount rate equal to the consumption rate of interest. The most complete frameworks for the shadow price of capital also note that while the costs of regulation might displace private capital, the benefits could encourage additional private investments in capital. In

\begin{footnotesize}
\begin{itemize}
\item \textsuperscript{147} The term “crowding out” refers to how total private investment in the economy is reduced due to new investment in response to the environmental policy. That is, how new Investment in response to the policy displaces investment that would have occurred without the policy. An environmental policy has fully crowded out private investment if private investment is reduced by the full amount of investment required by that policy.

\item \textsuperscript{148} A “shadow price” can be viewed as a good’s true opportunity cost, which may not equal the market price. Adjusting the cost and benefits of investment to reflect their consumption equivalent impact is, basically, reporting their shadow values. Lind (1982a) remains the seminal source for this approach in the social discounting literature.

\item \textsuperscript{149} Because the consumption rate of interest is often used as a proxy for the social rate of time preference, this method is sometimes known as the “consumption rate of interest – shadow price of capital” approach. However, as Lind (1982b) notes, what is really needed is the social rate of time preference, so more general terminology is used. Discounting based on the shadow price of capital is referred to as a “supply side” approach by the EPA’s SAB (U.S. EPA 2004c).

\item \textsuperscript{150} See OMB Circular A-4 (2003), Freeman (2003), and the report of the EPA’s Advisory Council on Clean Air Compliance Analysis (U.S. EPA 2004c).
\end{itemize}
\end{footnotesize}
Text Box 6.4 – Example of Estimating and Applying the Shadow Price of Capital

Suppose that the pre-tax rate of return on private investments (i.e., the social opportunity cost of capital) is 5 percent and the post-tax consumption rate of interest is 3 percent. A $1 private investment under these conditions will produce a stream of private consumption of $.03 per year, and tax revenues of $.02 per year. Further assume that the net-of-tax earnings from these investments are consumed in each period but the $1 investment exists in perpetuity.

The net present value of the perpetual stream of constant income is simply the stream of income divided by the discount rate (Equation (9)). Thus, the present value of the stream of $.03 per year in private, pre-tax consumption at the 3 percent consumption rate of interest is $0.03/$0.03 = $1. This is the present value of the benefit to individuals. The present value of the $0.02 per year stream of tax revenues at 3 percent is $0.02/0.03 = $0.67. This is the present value of the benefit to taxpayers (via the government). The social value of this $1 private investment – the shadow price of capital – is thus $1.67, which is substantially greater than the $1 private value that individuals place on it.

Generally, the simplest form of the shadow price of capital is:

$$\text{Shadow price of capital} = \frac{\text{opportunity cost of capital}}{\text{consumption rate of interest}}.$$

As can be seen with this example and form, the shadow price of capital is 5%/3%, which is about 1.67.

To apply this shadow price of capital estimate to an actual public project, we need additional information about how the investment is financed (i.e., through debt or through taxes) and how total investment and consumption in the economy are affected by each type of financing. Assume the following:

- The $1 current cost of a public project is financed 75 percent with government debt and 25 percent with current taxes.
- The increase in government debt fully displaces private investments dollar for dollar.
- The increase in taxes reduces individuals’ current consumption also on a dollar-for-dollar basis.
- This project produces a benefit 40 years from now that is estimated to be worth $5 at that time in the future.

The shadow price of capital approach is implemented in the following steps:

1. Separate the costs that displace capital investment from the costs that displace consumption
   - The $0.75 of the costs financed through debt displace investment
   - The $0.25 of the costs financed through taxes displace consumption
2. Apply the shadow price of capital (1.67 from the example above) to the $0.75 of costs that displace private investment. This yields $1.25.
3. Add to this the remaining current cost ($0.25) that displaces current consumption, which is not adjusted for the shadow price of capital.
   - The total social cost is therefore $1.50.
4. Calculate the net present value of benefits using the consumption rate of interest.
   - $5 in 40 years discounted at 3 percent is approximately $1.53.
5. Calculate net benefits by comparing the present value of social benefits and social costs.
   - $1.53 - $1.50 = $0.03

principle, a full analysis of shadow price of capital adjustments would treat costs and benefits symmetrically in this sense.

Policies analyzed in a general equilibrium framework (Chapter 8) will implicitly apply a shadow price of capital approach when household savings are endogenous. In the case of partial equilibrium analysis with fixed savings rates (including foreign investment flows), additional steps are necessary to apply the shadow price of capital approach. The first step is
to determine whether private investment flows will be altered by a policy. Next, the altered private investment flows (positive and negative) are multiplied by the shadow price of capital to convert them into consumption-equivalent units.

All flows of consumption and consumption equivalents are then discounted using the consumption rate of interest. A simple illustration of this method applied to the costs of a public project is shown in Text Box 6.4.151

6.2.4.1 Estimating the Shadow Price of Capital

The shadow price of capital approach is data intensive. It requires, among other things, estimates of the social rate of time preference, the social opportunity cost of capital, and the extent to which regulatory costs displace private capital investment and benefits stimulate it. While the first two components can be estimated as described earlier, information on how regulation affects capital formation is more difficult to obtain, making the approach difficult to implement.152

How policies affect capital investment depends largely on whether the economy is assumed to be open or closed to trade and capital flows, and on the magnitude of the policy intervention relative to the flow of investment capital from abroad. Some argue that early analyses implicitly assumed that capital flows into the nation were either nonexistent or very insensitive to market interest rates, known as the “closed economy” assumption.153 If, however, international capital flows are quite large and are sensitive to market interest rate changes (the “open economy” assumption) then total investment in private capital is likely to be less sensitive to regulatory policy interventions and there is little, if any, crowding out.154 If there is no crowding out of private investment, then no adjustments using the shadow price of capital are necessary; benefits and costs should be discounted using the consumption rate of interest alone.

The literature is not conclusive on the degree of crowding out. There is little detailed empirical evidence of relationship between the nature and size of projects and capital displacement. Ultimately, this makes it difficult to implement the shadow price of capital approach outside of a general equilibrium framework.

6.2.5 Evaluating Alternative Social Discount Rate Estimates

The empirical literature for choosing a social discount rate focuses largely on estimating the consumption rate of interest at which individuals trade off consumption through time with reasonable certainty. Historical real rates of return have expanded this portfolio to include other bonds, stocks, and even housing. This generally raises the range of rates slightly. It should be noted that these rates are realized rates of return, not anticipated, and they are somewhat sensitive to the choice of time period and the class of assets considered.155

Other economists have constructed a social discount rate by estimating the individual arguments in the Ramsey equation. These estimates necessarily require judgments about the pure rate of time preference. Moore et al. (2013a) and Boardman et al. (2006) estimate the social discount rate to be 3.5 percent under this approach. The Ramsey

---

151 An alternative approach for addressing the divergence between the higher social rate of return on private investments and lower consumption rate of interest is to set the social discount rate equal to a weighted average of the two. The weights would equal the proportions of project financing that displace private investment and consumption respectively. This approach has enjoyed considerable popularity over the years, but it is technically incorrect and can produce NPV results substantially different from the shadow price of capital approach. For an example of these potential differences see Spackman 2004.

152 Depending on the magnitudes of the various factors, shadow prices from about 1 to infinity can result (Lyon 1990). Lyon (1990) and Moore et al. (2004) contain excellent reviews of how to calculate the shadow price of capital and possible settings for the various parameters that determine its magnitude. Boardman, et al (2011) contains a textbook explanation as well as empirical examples.

153 See Lind (1990) for this revision of the shadow price of capital approach.

154 See, for example, Warnock and Warnock (2009).

155 Ibbotson and Sinquefield (1984 and annual updates) provide historical rates of return for various assets and for different holding periods.
equation has been used more frequently in the context of intergenerational discounting, which is addressed in the next section.

The use of the social opportunity cost of capital as the social discount rate requires a situation where investment is crowded out dollar-for-dollar by the investment costs of environmental policies. This is an unlikely outcome, but it can be useful for sensitivity analysis and special cases. Estimates of the social opportunity cost of capital typically range from 4.5 percent to 8 percent depending upon the type of data used.\textsuperscript{156}

The utility of the shadow price of capital approach hinges on the magnitude of altered capital flows from the environmental policy. If the policy will substantially displace private investment, then a shadow price of capital adjustment is necessary before discounting consumption and consumption equivalents using the consumption rate of interest. The literature does not provide clear guidance on the likelihood of this displacement, but it has been suggested that if a policy is relatively small and capital markets fit an “open economy” model, there is probably little displaced investment.\textsuperscript{157} Changes in yearly U.S. government borrowing during the past several decades have been in the many billions of dollars. It may be reasonable to conclude that EPA programs and policies costing a fraction of these amounts are not likely to result in significant crowding out of U.S. private investments. Primarily for these reasons, some argue that for most environmental regulations it is sufficient to discount costs and benefits with an estimate of the consumption rate of interest with some sensitivity analysis.\textsuperscript{158}

\section*{6.3 Intergenerational Social Discounting}

Policies designed to address long-term environmental problems such as global climate change, radioactive waste disposal, groundwater pollution, or biodiversity present unique challenges because they can involve significant economic effects across generations. Often, costs are imposed mainly on the current generation to achieve benefits that will accrue primarily to unborn, future generations. Discounting in this context is generally referred to as intergenerational discounting.

This section presents a discussion of the main issues associated with intergenerational social discounting using the Ramsey discounting framework as a convenient structure for considering how the “conventional” discounting procedures might need to be modified for policy analysis with very long, multi-generational time horizons. This discussion presents alternative modeling approaches to estimate the term structure, or the sequence of discount rates over time, along with important caveats when using these approaches.

Intergenerational discounting is complicated by at least three factors: (1) the “investment horizon” is longer than what is reflected in observed market interest rates representative of intertemporal consumption tradeoffs made by the current generation; (2) intergenerational investment horizons involve greater uncertainty than intragenerational time horizons; and (3) future generations without a voice in the current policy process are affected. These complications limit the utility of using observed market rates to evaluate long-term public investments. The leading alternative is to use model--based approaches to forecast a discount rate representative of expected household preferences, which often suggest the use of rates lower than those currently observed in the marketplace. This holds regardless of whether the estimated rates are measured in private capital or consumption terms, especially when uncertainty over the future state of the world is taken into consideration.

\textsuperscript{156} OMB (2003) recommends a real, pre-tax opportunity cost of capital of 7 percent. Harberger and Jenkins (2015) estimate an average rate of 8 percent for “advanced countries”. Zerbe and Burgess (2011b) estimate a rate of 6 to 8 percent, and Moore, et al. (2013b) estimate a rate of approximately 5 percent using the same model but with different inputs. Similar to the approach taken by OMB (2003), the CEA (2017) estimated real rates of return to capital to be around 7 percent based on National Accounts data but noted that approach may be subject to measurement error leading to an overestimate.

\textsuperscript{157} Lind (1990) first suggested this.

\textsuperscript{158} See Lesser and Zerbe (1994) and Moore et al. (2004).
The problem of comparing benefits borne by future generations to costs experienced by the current generation involves both economic and ethical questions. Therefore, the normative choice of how a decision maker should weigh the welfare of present and future generations, along with the preferences of the current generations regarding future generations, cannot be made on economic grounds alone. Nevertheless, economics offers important insights concerning intergenerational discounting, the implications and consequences of alternative discounting methods, and the systematic consideration of uncertainty.

### 6.3.1 The Ramsey Framework in an Intergenerational Context

The Ramsey framework introduced in Section 6.2.2 is one of the most commonly used approaches for modeling consumption discount rates.\(^{159}\) It is based on basic economic theory and provides an intuitive organizing framework to think about consumption discount rates over long time horizons. If per capita consumption grows over time — as it has at least since the industrial revolution (Valdés 1999) — then future generations will be richer than the current generation. Due to the diminishing marginal utility of consumption, increases in consumption will be valued less in future periods than they are today. In a growing economy, changes in future consumption would be given a lower weight (i.e., discounted at a positive rate) than changes in present consumption in the Ramsey framework, even setting aside discounting due to the pure rate of time preference, \(\rho\).

This framework can be viewed in positive terms as a description (or first-order approximation) of how the economy works in practice. It can also be viewed in normative terms to define how individuals should optimally consume and reinvest economic output over time. As a result, the individual parameters of the Ramsey equation can be specified using two approaches: the descriptive (or positive) approach and the prescriptive (or normative) approach.

- **The descriptive (positive) approach** attempts to calibrate the parameters of the Ramsey equation by using estimates from observed behavior. The resulting consumption discount rate reflects society's observed preferences for trading off consumption over time as well as the best available information on the future growth rate of consumption. Advocates of the descriptive approach generally call for inferring the discount rate from market rates of return “because of a lack of justification for choosing a social welfare function that is any different than what decision makers [individuals] actually use” (Arrow et al. 1996). However, this can be difficult to do in practice.

- **The prescriptive, (normative) approach** is based on defining a social welfare function that formalizes the normative judgments that the decisionmaker wants to explicitly incorporate into the policy evaluation. In the case of the Ramsey equation, parameters would then be chosen to match these desired normative judgements.\(^{160,161}\) The main argument against the prescriptive approach is that it may not be consistent with the preferences for inter-temporal tradeoffs revealed by individuals through their market behavior.

While the Ramsey framework is commonly used and is based on an intuitive description of the general problem of trading off current and future consumption, it has limitations. Arrow et al. (1996) contains detailed discussion of descriptive and prescriptive approaches to discounting over long time horizons, including examples of rates that emerge under various assumptions about components of the Ramsey equation.

### 6.3.2 Efficiency and Intergenerational Equity

A principal concern with policies that span long time horizons is that future generations affected by the policy are not yet alive. Therefore, they cannot participate in the decision-making process and their preferences are uncertain. These

---


\(^{160}\) Arrow et al. (1996).

\(^{161}\) For instance, there has been a long debate, starting with Ramsey himself, on whether the pure rate of time preference, which shows a general preference for consumption by the current as opposed to future generations, should be greater than zero when evaluating public policy decisions.
are not always severe problems for practical policy analysis. Many policies impose relatively modest costs and benefits, or have costs and benefits that begin immediately or occur in the not-too-distant future. In most cases, it suffices to assume future generations will have preferences much like those of present generations. However, for policies where the costs and benefits are large and distributed asymmetrically over large expanses of time, the choice of discount rate may involve both efficiency and ethical considerations. Based on these considerations (along with uncertainty discussed in the next section) OMB in Circular A-4 advises analysts, “if your rule will have important intergenerational benefits or costs you might consider a further sensitivity analysis using a lower but positive discount rate” in addition to the approaches discussed above.

6.3.2.1 Efficiency considerations
As discussed in Chapter 1 and Appendix A, the BCA efficiency test is grounded in the notion of a potential Pareto improvement, whereby those who benefit from a policy on net could potentially compensate those who experience costs on net, such that everyone is at least as well off as they were before. The potential for this compensation to occur across generations hinges on the rate of interest at which society can transfer wealth across long time horizons. The choice of social discount rate therefore, contains an implicit assumption about whether, and at what price, the distribution of wealth across generations could be adjusted to compensate those who bear costs, on net. Some have argued that in the U.S. context, the federal government’s borrowing rate is a good candidate for this rate while others have argued that practical difficulties associated with implementing intergenerational transfers suggest that the Kaldor-Hicks potential compensation test is limited in its ability to assess policies affecting multiple generations.162,163 Still others argue that the discount rate should be below market rates to correct for market distortions and uncertainties or inefficiencies in intergenerational transfers of wealth (the role of uncertainty is discussed in more detail below).164

6.3.2.2 Equity considerations
Because future generations cannot participate in decisions made by current generations, social discounting may raise ethical issues regarding intertemporal distribution of wealth. This concern does not suggest forgoing the use of a positive discount rate, but has led to suggestions that the discount rate used in intergenerational contexts should be below market rates to ensure that generations are treated equally based on ethical principles (e.g., Arrow et al. 1996, Portney and Weyant 1999).165 One interpretation of this idea is to forgo discounting the utility of future generations by setting the pure rate of time preference in the Ramsey framework to zero. These suggestions are for using a prescriptive (i.e., normative) approach for discounting.

6.3.3 Declining Discount Rates
Theoretical and empirical support is growing for discount rates that decline over time for intergenerational discounting (Arrow et al., 2014). That is, the appropriate rate to use in discounting effects in year 101 to year 100 will be lower than the appropriate rate to use in discounting effects in year 2 to year 1. Multiple rationales support a declining discount rate, most notably slowing consumption growth rates and uncertainty about economic growth.

6.3.3.1 Rationales for declining discount rates
A slowing of consumption growth rates leads to declining discounting, as is clear from the Ramsey framework. Using a constant discount rate in BCA is technically correct only if the rate of economic growth per capita will remain fixed over the time horizon of the analysis. In principle, any set of known changes to income growth, the elasticity of marginal

162 See Lind (1990) and a summary by Freeman (2003).
163 For more information and theoretical foundations of the Kaldor-Hicks test for potential Pareto improvements see Appendix A
164 Arrow et al. (1996); Weitzman (1998).
165 Another issue is that there are no market rates for intergenerational time periods.
utility of consumption, or the pure rate of time preference will lead to a discount rate that changes accordingly. If economic growth per capita is changing over time, then the discount rate will also fluctuate. In particular, an assumption that the growth rate is declining systematically over time (perhaps to reflect some physical resource limits), will lead to a declining discount rate. This is the approach taken in some models of climate change.\textsuperscript{166}

Uncertainty about future consumption growth can also lead to a declining discount rate. The longer time horizon in an intergenerational policy context implies greater uncertainty about the investment environment and economic growth over time, and a greater potential for environmental feedbacks to economic growth (and consumption and welfare). These feedbacks further increase uncertainty when attempting to estimate the social discount rate. This additional uncertainty implies effective discount rates lower than those based on observed average market interest rates\textsuperscript{167} (Weitzman 1998, 2001; Newell and Pizer 2003; Arrow et al., 2013; Cropper et al., 2014).\textsuperscript{168}

The effect of uncertainty on discount rates is a result of the fact that discounting is a non-linear operation, such that the average of discount factor (i.e., $E[e^{-rt}]$) is not equal to the discount factor calculated at the average discount rate (i.e., $e^{-E[r]}$). As an alternative to estimating the average discount factor, one can calculate the certainty equivalent discount rate schedule, which is the discount rate schedule that yields the same discount factor in any time period as the average of discount factor across the possible discount rates. Uncertainty about future consumption growth will cause this certainty equivalent discount rate schedule to decline over time as the potential for low discount rates will increasingly dominate the expected NPV calculations for benefits and costs far in the future (Weitzman 1998). Text Box 6.5 provides a simple example to highlight how declining discount rates arise in this fashion.\textsuperscript{169}

6.3.3.2 Approaches to estimate declining discount rates

Declining discount rate schedules can be derived from specifications of the Ramsey formula or from historically estimated stochastic models of interest rates.

If there is uncertainty in the rate of consumption growth, then the standard Ramsey formula may need to be adjusted. Incorporating uncertainty in consumption growth results in a third precautionary term being subtracted from the Ramsey formula to account for the potential of low growth futures (Gollier 2002; Arrow et al., 2014). If the shocks to consumption growth are independent and identically distributed, then the precautionary term will cause the discount rate to be lower but not decline. However, if the shocks are positively correlated over time, then the precautionary term will grow over time and cause the discount rate to decline (Gollier 2014). If there is parametric uncertainty regarding the process underlying consumption growth or the other values in the Ramsey formula, this can also lead to a declining discount rate.

The use of historical data to estimate a declining discount rate schedule is shown by Newell and Pizer (2003). They use historical data on U.S. interest rates and assumptions regarding their future path to characterize uncertainty and compute a certainty equivalent rate. In this case, uncertainty in the individual components of the Ramsey equation is not being modeled explicitly. This is attractive as a descriptive approach because it does not require specifying uncertainty over the consumption growth rate and parameters of the Ramsey formula, but its results are sensitive to the selection of a model to represent the stochastic interest rate process (Groom et al., 2007).

\textsuperscript{166} See, for example, Nordhaus (2017).

\textsuperscript{167} This holds regardless of whether or not the estimated investment effects are predominantly measured in terms of private capital or consumption.

\textsuperscript{168} Gollier and Zeckhauser (2005) reach a similar result using a model with decreasing absolute risk aversion.

\textsuperscript{169} While this explanation is motivated by uncertainty over long-term consumption growth, a similar result arises when there is persistent uncertainty about preferences or heterogeneity in preferences. See Heal and Millner, 2014.
Some modelers and government bodies have used fixed step functions for the discount rate term structure to approximate more rigorously derived declining discount rate schedules and to reflect non-constant economic growth,
intergeneration equity concerns, and/or heterogeneity in future preferences.\textsuperscript{170} This method acknowledges that a constant discount rate does not adequately reflect the reality of fluctuating and uncertain growth rates over long time horizons. However, no empirical evidence suggests the point(s) at which the discount rate declines, so any year selected for a change in the discount rate will be ad-hoc.

6.3.3.3 Consistency issues and declining discount rates

Another concern regarding declining discount rates is the potential for time inconsistency in policy recommendations over time (Arrow et al., 2014). Time inconsistency means that a net-beneficial policy today may not be net-beneficial if evaluated in the future, even when nothing has changed except for the date of the evaluation. The use of fixed step functions can exacerbate the problem. Therefore, whether an analysis shows the policy to be net-beneficial will be sensitive to the point in time the analysis is conducted. Text box 6.6 provides an illustration of this time consistency problem.

If the analyst obtains new information between the time the original and updated analysis are conducted, the results of the analysis may have changed. However, if a fixed declining discount rate schedule is adopted and not updated between analyses to reflect the arrival of new information, that could lead to a potential time inconsistency problem (Arrow et al., 2014).

6.3.3.4 Calibration and Challenges

A wide range of potential approaches for calibrating a discount rate or schedule of declining discount rates are available for discounting intergenerational costs and benefits. More complex analysis is justified when the proportion of costs and benefits occurring far out on the time horizon and the temporal separation of costs and benefits over the planning horizon are large. While strong theoretical and empirical evidence shows that a declining discount rate schedule is appropriate when considering effects that occur over long time horizons, calibration complications and concerns with time inconsistency remain notable challenges.

One possible response to such challenges is to select a constant but slightly lower discount rate when discounting costs and benefits that are expected to occur far out in the time horizon, reflecting a certainty equivalent discount rate. Independent of the approach or rate selected, the same discount rate should be applied to all benefits and costs that occur in the same year for both intra- or intergenerational consequences to ensure consistency in the analysis (Arrow et al., 2013).

6.4 The Role of Private Discounting in Economic Analysis

This chapter focuses on social discounting, which is discounting from the broad society-as-a-whole perspective embodied in BCA. By contrast, private discounting is discounting of expected future benefits or costs (e.g., revenues or expenditures) from the perspective of private individuals or firms. Private discount rates reflect the preferences of specific individuals for consumption over time, as well as the prices that individuals and firms pay to borrow and lend money. These rates vary among firms, industries, and individuals due to differences in preferences, tax treatments, and costs of borrowing. Section 6.2.1 describes why market interest rates differ from the consumption rates of interest.

\textsuperscript{170} For instance, in the United Kingdom the Treasury recommends the use of a 3.5 percent discount rate for the first 30 years followed by a declining rate over future time periods until it reaches 1 percent for 301 years and beyond. The guidance also requires a lower schedule of rates, starting with 3 percent for zero to 30 years, where the pure rate of time preference in the Ramsey framework (the parameter \(r\) in our formulation) is set to zero. For details see HM Treasury (2008) \textit{Intergenerational wealth transfers and social discounting: Supplementary Green Book Guidance}. Additionally, Weitzman (2001) presents a novel approach to calibrating a fixed step discount rate schedule based on uncertainty using survey data.
As previously stated, private discount rates should not be used to estimate the NPV of the social net benefits of policies and projects because the intertemporal preferences of society as a whole (as measured by the social rate of time preference) are not likely to be equal to private market lending rates or individual or firm preferences.

6.4.1 Predicting private behaviors and choices

Private discounting should be used to predict behaviors and choices of individuals and firms in response to policy, and how investment in the economy and consumption (broadly defined) are expected to change as a result. Individuals and firms can be expected to make decisions based on their own opportunity costs rather than those of society as a whole. For example, from the viewpoint of a private firm, the change in a stream of future profits due to the adoption of a pollution abatement project would be evaluated at the rate at which the firm can borrow. Similarly, the expected consumption behavior of individuals and households should be modeled consistently with how they make purchasing decisions. To predict the purchase of durable goods, for example, private evaluation and perception of the consumer’s benefits and costs from using these goods over time should be used. Failure to account for choices based on appropriate private discount rates will lead to inconsistencies between the behavior of individuals and firms in the analysis and their expected behavior in the real world. Therefore, private discount rates should be used to evaluate how firms and individuals will respond to policy.

6.4.2 Treatment of interest payments

Any changes in the amount of interest paid for borrowing (e.g., loans) as the result of a potential regulation should not be included in the calculation of its estimated social benefit or cost. Interest payments themselves do not reflect the use of real resources such as labor, capital, and materials in an economy. Rather, the interest payment is a transfer between the borrower and lender and would net out of a social benefit-cost analysis. Private interest rates, in part, reflect the opportunity cost to society of any changes in the timing of consumption as a result of a regulation, but this opportunity cost is already accounted for in social discounting as discussed above. However, interest payments should be

---

171 This guidance applies both the regulated sources and any individuals and firms meaningfully affected by the behavior of the regulated sources.

172 For this same reason, using a social discount rate to model how firms and individuals evaluate private benefits and costs can lead to misspecification of the baseline over time and/or a mistaken projection of their responses to a policy.

173 Administrative charges on a loan (e.g., origination fees) may include the cost of preparing and administering any loans. Changes in these costs, if
accounted for when evaluating the incidence and economic impacts of a regulation. For example, if a firm must take out a loan to comply with a regulation, the interest payment on that loan should be accounted for when estimating the effect of the regulation on the firm’s profits.\textsuperscript{174} See Chapter 9 for further discussion of how to determine the incidence of a regulation.

### 6.4.3 Selecting private discount rates

Selecting which discount rate to use to represent household or firm behavior presents challenges. An appropriate discount rate may be observed from market behavior, but different households and firms borrow at different interest rates, and even within a household or firm borrowing (and lending) occurs at different rates.\textsuperscript{175} For example, firms may borrow at different rates depending on whether they are financing investments through debt or equity. The choice of discount rate used to represent private behavior should be well explained and, if necessary, sensitivity analyses to different assumed rates may be conducted.

### 6.5 Recommendations and Guidance

The following recommendations are intended as practical and plausible default assumptions rather than comprehensive and precise estimates of social discount rates to apply in all situations. Some cases may present compelling reasons to gather necessary data and develop a more realistic model and bring to bear more accurate empirical estimates of the various factors that are most relevant to the specific policy scenario under consideration. In such cases, these estimates should be presented along with the rationale in the description of the methods, and any appropriate peer review. Results based on default assumptions should also be included for comparison purposes and for consistency with overarching OMB guidance. With this caveat in mind, recommendations for discounting are below.

- Display the full time paths of benefits and costs as they are projected to occur over the time horizon of the policy, i.e., without discounting.
- When determining the net-benefits of a regulation, the analysis should compare the entire the discounted value of the entire time path of benefits and costs. It is inappropriate to characterize the effect of a regulation with only the costs or benefits for a limited period of time (e.g., a single year), because benefits and costs may occur during other periods. Similarly, it is inappropriate to compare an annual value to an annualized value.
- To the extent that a regulation is expected to displace capital investment, the shadow price of capital approach is the analytically preferred method for discounting as it will discount any future consumption from displaced capital investments in a theoretically consistent manner with other consumption equivalent values present in the analysis. However, there is disagreement on the extent to which private capital is displaced by EPA regulatory requirements, and the shadow price of capital approach can be difficult to determine, should be accounted for in a benefit-cost analysis.

\textsuperscript{174} When evaluating the incidence of a regulation over time, it may also be important to recognize the annualization of any capital investment.

\textsuperscript{175} As discussed in the behavioral economics literature, individual behavior is not always consistent with the conventional discounting framework. For example, households may consume and save different sources of wealth differently, and therefore are applying different discount rates to those sources of wealth, even when the sources of wealth are fungible (Thaler, 1990). There is also evidence that discount rates for individuals decline over time, are lower the larger the magnitude of the future value, are higher for gains than for losses, and that individuals may prefer a stream of benefits that increase over time over one that is constant over time despite each having the same nominal values (Fredrick et al. 2002). Alternative behavioral frameworks have been proposed that are consistent with these observed patterns of discounting (e.g., Lowenstein and Prelec, 1992; Laibson, 1998). Conventional discounting should be used to represent individual, household or firm behavior in the economic analysis, although alternative discounting frameworks to represent the behavior of individuals or households may be provided in a sensitivity analysis, provided the alternative framework is well studied in the literature in settings comparable to that of the regulation. Care should be taken when applying alternative discounting models to predict behavior as observed behavior that can at first appear inconsistent with the conventional framework as they may actually be consistent with the perceived discrepancy being due to other modeling omissions. For example, an individual’s discount rate may appear to change over time due to issues such as perceived uncertainty about future outcomes being valued, even though their strict rate of time preference may not be changing (Fredrick et al. 2002).
implement in practice. If the shadow price of capital approach is not applied, either explicitly or implicitly using a general equilibrium framework, in the analysis then analysts should conduct a bounding exercise as follows:

- Calculate present or annualized value using the consumption rate of interest. This is appropriate for situations where all costs and benefits occur as changes in consumption flows rather than changes in capital stocks, i.e., capital displacement effects are negligible. Prior estimates by OMB are that the real consumption rate of discount is close to 3 percent.\(^{176}\)

- Also calculate present or annualized value using the rate of return to private capital. This is appropriate for situations where all costs and benefits occur as changes in capital stocks rather than consumption flows. OMB has previously estimated a real rate of 7 percent for the opportunity cost of private capital.\(^{177}\)

- If the policy has a long time horizon where net benefits vary substantially over time (e.g., most benefits accrue to one generation and most costs accrue to another), then the analysis should use the consumption rate of interest as well as additional approaches. These approaches include:
  - Discounting at a constant rate somewhat lower than those used in the conventional case if a time-declining approach cannot be implemented.

When discounting future benefits and costs, the following principles should be kept in mind:

- Regardless of the approach or rate selected, the same discount rate should be applied to all benefits and costs that occur in the same year, independent of whether the policy has intra- or intergenerational consequences, to ensure consistency in the analysis.

- Private discount rates should be used to predict behavior of individuals and firms and to evaluate economic impacts and incidence, but they should not be used in place of the social discount rate to assess social benefits and costs of a policy.

- The discount rate should reflect marginal rates of substitution between consumption in different time periods and should not be confounded with factors such as uncertainty in benefits and costs or the value of environmental goods or other commodities in the future (i.e., the “current price” in future years).

- The lag time between a change in regulation and the resulting welfare impacts should be accounted for in the economic analysis. This includes accounting for expected changes in human health, environmental conditions, ecosystem services, and other related factors.

### Chapter 6 References


\(^{176}\) OMB Circular A-4 (2003).

\(^{177}\) OMB Circular A-4 (2003).


Chapter 7

Analyzing Benefits

The aim of an economic benefits analysis of an environmental policy or action is to clearly describe the environmental changes resulting from that policy or action, and to estimate the social benefits resulting from the environmental changes. Willingness to pay (WTP) is the preferred measure of these changes and is the only measure consistent with the potential Pareto criterion that underlies benefit-cost analysis. WTP provides a full accounting of individual preferences across trade-offs between wealth and benefits and is measured in monetary terms needed to allow the calculation of net benefits — the sum of all monetized benefits minus the sum of all monetized costs. Net benefits are used to compare across policy options and the relevant baseline to assess whether a policy proposal represents a net improvement in societal welfare.

This chapter provides analysts with an overview of the benefits analysis process, focusing first on the quantification of benefits but primarily on their monetization. The methods and approaches for monetizing benefits are described in the context of an EPA policy, program, or regulation that leads to changes in emissions or discharges of contaminants. Although social benefits of an environmental policy or action are defined by economists as the favorable effects society gains from it, not every component need be positive. Even so, the same theory and toolkit of approaches for valuing the changes in environmental quality apply regardless of the direction of those changes.

The discussion below emphasizes the benefit transfer approach most often used by the Agency for monetizing benefits in economic analysis and highlights important considerations for performing benefits analysis. Recognizing that there are often benefits that cannot be monetized due to lack of available values or quantification, this chapter also includes a discussion of what analysts can do to incorporate these endpoints more fully into the analysis. Chapter 11 on the “Presentation of Analysis and Results” discusses how to carry forward information on non-monetized benefits to help inform the policy-making process.\(^1\)

\(^1\) Other methods, such as cost-effectiveness analysis (CEA), can also be used to evaluate policies. CEA does not require monetization of benefits but rather divides the costs of a policy by a particular effect (e.g., number of lives saved). CEA can be used to compare proposed policy changes, but unlike BCA, cannot be used to calculate a single, comprehensive measure of the net effects of a policy, nor can it compare proposed policy changes to the status quo. While cost effectiveness analysis is not covered extensively in these Guidelines, other methods for evaluating policies (e.g., distributional analyses) are covered in Chapter 10.
Figure 7.1 - A Conceptual Model for Benefits Analysis

Figure 7.1 presents a conceptual model for benefits analysis. After policy options have been identified, the first step is to identify the changes in environmental contaminants or stressors that are likely to result from policy options relative to the baseline. These may be measured as changes in emissions or in concentrations of contaminants, but they can also be considered more broadly. For example, “stressors” can be the number of hazardous waste sites, and the benefits analysis may be built upon changes in this metric.

Changes in contaminants or stressors often lead to changes in environmental quality such as a change in ambient air quality. Environmental quality should be interpreted broadly for this conceptual model, including exposure to contaminants. Often, a great deal of analysis is required to project how changes in contaminants or stressors affect environmental quality, including modeling the transport of the pollutants through the environment along a variety of pathways, including movement through the air, surface water, and groundwater; deposition in soils; and ingestion or uptake by plants and animals (including humans). In many cases, explicit modeling of human intake or exposure might be another intermediate step in the conceptual model that precedes quantifying changes in benefit endpoints.

The next step is to identify the benefit endpoints that may be affected by changes in environmental quality. Benefit endpoints are organized in the Guidelines into broad categories: human health improvements, ecological improvements, aesthetic improvements, and reduced materials damages (Section 7.2). Table 7.1 lists examples of benefit endpoints in each of these categories. Once changes in benefit endpoints are identified, valuation follows well-defined economic principles (Section 7.2) using well-established economic methods (7.3). Commonly used methods for each type of benefit are also described in Table 7.1.

Finally, the aggregate value for all benefits provides the basis for characterizing the benefits of each policy option.

Ideally, the benefits analyses would comprehensively assess all welfare-improving effects – all benefit endpoints attributable to a rule or policy decision, including potential interactions and feedbacks between effects. This may be possible to an extent with the use of integrated assessment models (IAMs) (see Text Box 7.1). However, the modeling and data required for such a comprehensive assessment make it impossible or impracticable to do so in most circumstances.
Table 7.1 - Types of Benefits Associated with Environmental Policies: Categories, Examples, and Commonly Used Valuation Methods

<table>
<thead>
<tr>
<th>Category</th>
<th>Examples</th>
<th>Commonly Used Valuation Methods</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Human Health Improvements</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mortality risk reductions</td>
<td>Reduced risk of:</td>
<td>Averting behaviors</td>
</tr>
<tr>
<td></td>
<td>Cancer fatality</td>
<td>Hedonics</td>
</tr>
<tr>
<td></td>
<td>Acute fatality</td>
<td>Stated preference</td>
</tr>
<tr>
<td>Morbidity risk reductions</td>
<td>Reduced risk of:</td>
<td>Averting behaviors</td>
</tr>
<tr>
<td></td>
<td>Cancer</td>
<td>Cost of illness</td>
</tr>
<tr>
<td></td>
<td>Asthma</td>
<td>Hedonics</td>
</tr>
<tr>
<td></td>
<td>Cognitive Impairment</td>
<td>Stated preference</td>
</tr>
<tr>
<td><strong>Ecological Improvements</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Market products</td>
<td>Food; Fuel; Timber; Fish</td>
<td>Production function</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Demand analysis for consumer benefits</td>
</tr>
<tr>
<td>Recreation activities and</td>
<td>Wildlife viewing</td>
<td>Production function</td>
</tr>
<tr>
<td>aesthetics</td>
<td>Fishing and hunting</td>
<td>Averting behaviors</td>
</tr>
<tr>
<td></td>
<td>Boating</td>
<td>Hedonics</td>
</tr>
<tr>
<td></td>
<td>Swimming</td>
<td>Recreation demand</td>
</tr>
<tr>
<td></td>
<td>Hiking</td>
<td>Stated preference</td>
</tr>
<tr>
<td></td>
<td>Scenic views</td>
<td></td>
</tr>
<tr>
<td>Valued ecosystem services</td>
<td>Climate moderation</td>
<td>Production function</td>
</tr>
<tr>
<td></td>
<td>Flood moderation</td>
<td>Averting behaviors</td>
</tr>
<tr>
<td></td>
<td>Groundwater recharge</td>
<td>Hedonics</td>
</tr>
<tr>
<td></td>
<td>Sediment trapping</td>
<td>Recreation demand</td>
</tr>
<tr>
<td></td>
<td>Soil retention</td>
<td>Stated preference</td>
</tr>
<tr>
<td></td>
<td>Nutrient cycling</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pollination by wild species</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Biodiversity, genetic library</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Water filtration</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Soil fertilization</td>
<td></td>
</tr>
<tr>
<td>Non-use values</td>
<td>Relevant species populations,</td>
<td>Stated preference</td>
</tr>
<tr>
<td></td>
<td>communities, or ecosystems</td>
<td></td>
</tr>
<tr>
<td><strong>Other Benefits</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aesthetic improvements</td>
<td>Visibility</td>
<td>Averting behaviors</td>
</tr>
<tr>
<td></td>
<td>Taste</td>
<td>Hedonics</td>
</tr>
<tr>
<td></td>
<td>Odor</td>
<td>Stated preference</td>
</tr>
<tr>
<td>Reduced materials damages</td>
<td>Reduced soiling</td>
<td>Averting behaviors</td>
</tr>
<tr>
<td></td>
<td>Reduced corrosion</td>
<td>Production / cost functions</td>
</tr>
</tbody>
</table>

Note: “Stated preference” refers to all valuation studies based on hypothetical choices. The other methods, in various ways, rely upon observations of actual choices and mostly fall under the general category of “revealed preference.”
Text Box 7.1 - The Use of Integrated Assessment Models for Benefits Valuation

At times, it may be possible to estimate the benefits of a policy with the use of an integrated assessment model (IAM). In the broadest sense, IAMs are “approaches that integrate knowledge from two or more domains into a single framework” (Nordhaus 2013), and the label has been applied to models used in many disciplines, such as earth sciences, biological sciences, environmental engineering, economics, and sociology, among others. In environmental economics, the focus is on IAMs that combine natural processes and economic systems into a single modeling framework. These models “connect economic activity with environmental consequences, and ultimately, with valuation” (Keiser and Muller 2017). A full IAM aims to capture four components – behavior that generates emissions/pollutant loadings, pollution fate and transport, environmental and human outcomes, and valuation – as well as feedbacks within and across these components. The goal of IAMs is to provide a transparent, reproducible way to capture the importance of interdisciplinary consequences of highly complex problems.

IAMs have been used in the empirical environmental economics literature for decades to study both stock pollutants, primarily greenhouse gases (GHGs) (e.g., Nordhaus 1993), and flow pollutants, such as air pollution (e.g., Mendelsohn 1980) and water pollution (e.g., conceptual work by Freeman 1979, 1982). Current IAMs vary significantly in structure, geographic resolution, and the degree to which they account for feedbacks and include valuation of changes in physical endpoints and regulatory compliance costs. In some areas, research is focused on improving representation of interactions and feedback loops. For example, IAMs have been used to study the interaction between GHG mitigation and urban and regional air pollution policies (e.g., Reilly et al. 2007), the dynamic economic and ecosystem general equilibrium effects of fisheries management policy (e.g., Finnoff and Tscherhart 2008), and linkages in the food-water-energy nexus that have a bearing on policy outcomes (Kling et al. 2017). The choice of what type of IAM is most appropriate to use depends on the research or policy question.

One area in which IAMs are used in benefit-cost analysis is the valuation of changes in GHG emissions. IAMs that combine representations of climate and economic systems are used to develop monetized estimates of the damages associated with incremental emissions of carbon dioxide (CO$_2$), denoted as the social cost of carbon dioxide (SC-CO$_2$), allowing the social benefits of actions that are expected to change these emissions to be incorporated into BCA. Specifically, the SC-CO$_2$ is the present value of the stream of future economic damages associated with an incremental increase (by convention, one metric ton) in CO$_2$ emissions in a particular year. It is intended to be a comprehensive measure and includes economic losses due to a wide range of anticipated climate impacts, such as net changes in agricultural productivity, human health risks, property damages from increased flood frequencies, the loss of ecosystem services, etc. Analogous metrics estimate the monetary value of climate impacts associated with other non-CO$_2$ GHGs, such as methane and nitrous oxide.

IAMs used to estimate the SC-CO$_2$ and other GHGs are necessarily highly simplified and limited by the current state of the rapidly expanding climate economics literature. In January 2017, The National Academies of Sciences, Engineering, and Medicine issued a report recommending specific criteria for future updates to the SC-CO$_2$ estimates, a modeling framework, and both near-term updates and longer-term research needs pertaining to various components of the estimation process. *

Notes:

* Since the framework used to estimate the social cost of methane and nitrous oxide is the same as that used for SC-CO$_2$, the Academies’ recommendations on how to update many of the underlying modeling assumptions also apply to the estimates of the social cost of non-CO$_2$ GHGs.

Benefits analysis need not proceed by enumerating all benefit endpoints separately or follow the specific sequence described in Figure 7.1, particularly if valuation estimates are linked to effects further upstream in the model. For example, rather than monetizing enumerated health benefit endpoints, the hedonic property method (Section 7.3.1.3) may estimate the total value to residents of changes in the presence of hazardous waste sites – a change in a stressor in figure 7.1 – by linking policy changes to changes in property values. This valuation estimate could then be used in benefits analysis. This method of assessing benefits can be viewed as a reduced form approach to the modeling. Even when viewed as a reduced form approach, however, it is important to think through the conceptual model to assess

---

* There are many other ways this sort of reduced form approach may be appropriately used including estimates of the benefits per unit (e.g., benefit per ton emitted) of environmental contaminants.
whether there are benefit endpoints not reflected in the reduced form valuation estimate that should be included through additional analysis.

**A general approach to benefits estimation**

Ultimately, benefits analysis should link policy changes to the value of benefits that can be meaningfully attributed to those changes. This is most often done using a pragmatic, general approach aligned with the conceptual model in Figure 7.1 by tracing policy-related changes through a set of models to predict changes in specific benefit endpoints and then by valuing each endpoint, or sometimes sets of endpoints, separately. An overall estimate of total benefits is the sum of these separate components.

In short, the goal is to monetize those benefit endpoints that can be monetized, to quantify those that can be quantified but not monetized, and to provide qualitative characterizations of what cannot be quantified.\(^{180}\)

This general approach can be divided into three steps:

**Step 1: Identify relevant benefit endpoints associated with the policy;**

**Step 2: Quantify significant changes in these benefit endpoints to the extent feasible;**

**Step 3: Monetize the changes using appropriate valuation methods or by drawing on values from existing studies.**

Each step in this approach is discussed in more detail in the sections that follow. Collaboration with appropriate experts will be necessary to execute these steps meaningfully.\(^{181}\)

**Step 1: Identify relevant benefit endpoints**

The first step is to conduct an initial assessment of the types of benefits associated with the policy options being considered. This requires evaluation of how conditions and ultimately benefit endpoints differ between each policy option and baseline conditions (Chapter 5). The goal for this step is to enumerate the full set of benefit endpoints and to identify those that should be further developed for quantification and valuation. In this assessment, analysts should, to the extent feasible:

- **Develop an understanding of the changes in environmental contaminants or stressors resulting from policy options.** Initially, the range of policy options being considered may be very broad. Collaboration among all analysts and policy makers involved in the policy analysis can help ensure that all potential effects are recognized. It is important to account for both contaminant or stressor changes directly targeted by the policy options and those that will occur even if not directly targeted.\(^{182}\)

- **Identify the benefit endpoints likely to be affected by policy options.** This step often requires considering the transport of contaminants through the environment along many pathways, including movement through the air, surface water, and groundwater. Along these pathways, the pollutants can have detrimental effects on natural resources, such as affecting oxygen availability in surface water or reducing crop yields. Pollutants can also have direct or indirect effects on human health, for example, affecting cancer incidence through direct inhalation or through ingestion of contaminated food. This step is inherently multi-disciplinary and will include consulting with risk assessors and other experts involved in the rule or policy, sometimes as part of a formal workgroup (US EPA, 2014).

---

\(^{180}\) See Chapter 11 for more detail on presenting qualitative, quantified, and monetized benefits.

\(^{181}\) A summary of a large-scale benefits exercise that followed these steps is described later in Text Box 7.5.

\(^{182}\) See Chapter 5, section 5.1.3, for additional discussion of considering benefits that arise from changes in pollutants other than those that would be directly regulated by the policy.
• **Evaluate the potential changes in benefit endpoints resulting from each policy option.** If policy options differ only in their level of stringency, then each option may have an impact on all identified endpoints. Where policy options are more complex, however, the options may have an impact on some endpoints but not on others.

• **Determine which benefit endpoints warrant further investigation** in the overall benefits analysis using at least the following four framing questions:

  - **Which benefit endpoints are likely to be large relative to total benefits or are otherwise important for informing policy decisions?** This determination should be based on an assessment of the importance of each benefit endpoint to the benefits analysis, including its potential magnitude, the extent to which it can be quantified, and the extent to which it can be monetized. Preliminary assessments should be made using the best, readily available quantitative information; however, as a practical matter, these decisions are often based on professional judgment.

  - **Which benefit endpoints should be included even if they may not be large relative to total benefits?** Some benefit endpoints may not be captured by the first criteria but are important and informative for other reasons. For instance, benefit endpoints necessary to evaluate how minority, disadvantaged, or susceptible groups are affected in distributional analyses (Chapter 10) may not be large at a national level, but may be very important at a smaller scale. Benefit endpoints may also be important because they reflect Agency priorities, are closely related to the underlying motivation for the rule, or are otherwise of particular interest to decision makers.

  - **Which benefit endpoints are likely to differ across policy options?** Analysts should assess how the effects of each policy option will differ. Benefits categories should be meaningfully attributed to policy with some degree of confidence, while recognizing that there will always be uncertainty and that this uncertainty can be characterized in the benefits analysis. Again, this may be done as part of an interdisciplinary team working on the rule or policy.

  - **What are the costs of undertaking analysis to characterize the benefit endpoint?** The costs of quantifying and monetizing benefit endpoints may be minimal if existing data and models can be applied. If existing data and models are insufficient, value of information considerations are important. A benefit endpoint may not be worth a great deal of further investigation if the costs to quantify and monetize it exceed its informational value. However, consideration should also be given to other current and future rulemaking efforts that would rely on this endpoint for benefits.

The outcome of this step can be summarized in a list or matrix that describes the changes expected from the policy options being considered, defines associated benefit endpoints, and identifies the endpoints that warrant further investigation.

The list of benefit endpoints should be as comprehensive as possible and may be lengthy at first, encompassing all of those that reasonably can be expected to occur regardless of whether they can be quantified and/or put in dollar terms. Analysts should preserve and refine this list as the analysis proceeds. Maintaining the full list of potential effects facilitates later revisions if new information warrants it. Equally important, benefits that can only be characterized qualitatively should be presented along with quantitative information in the benefits analysis (See Chapter 11.)

**Step 2: Quantify changes in significant benefit endpoints**

Next, the analysis should quantify changes in the benefit endpoints identified in Step 1 as warranting further investigation, focusing on changes attributable to each policy option relative to the baseline. Expertise from a wide array of disciplines in addition to economics is usually needed in this step, including human health and ecological risk assessment, engineering, and natural sciences. Quantifying endpoints generally requires a function relating changes in emissions, concentrations, and/or exposure to changes in specific ecological services, health effects, or risks. Data are
usually needed on the magnitude, duration, frequency, and severity of the endpoints. For example, changes in cancer risks typically come from human health risk assessments, and the benefits analysis will need information on baseline risks, risk changes associated with each option, the timing of the risk changes, fatality rates, and the size and age distribution of affected populations. If visibility is the attribute of concern, needed information includes the geographical areas affected, the baseline visibility, and the change in visibility resulting from each policy option.

Sometimes data or modeling constraints will prohibit quantifying significant benefit endpoints. In these cases, it is useful to quantify changes in environmental stressors or measures of environmental quality that would lead to benefits. These changes can be informative in the overall characterization of benefits even if they cannot be aggregated with benefit endpoints.

Analysts should consider the following recommendations when quantifying changes in benefit endpoints.

- **Ensure endpoints are appropriate for benefits evaluation.** A principal role of the economist at this stage is to ensure that the endpoints are characterized in ways that are consistent with principles of economic analysis and the specific models used for benefits analysis. They should also be characterized in a manner that avoids double counting. Focus on the needs of economic analysis is particularly important at the early stages of ecological or human health risk assessments, and it is generally useful for economists to be part of a cross-disciplinary team for planning and scoping these assessments. The ability to monetize or even quantify benefits analysis may be limited if effects are described too broadly, overlap with other benefits categories, cannot be linked to human well-being, or are otherwise incompatible with economic analysis. Text Box 7.2 provides a more detailed discussion on integrating risk assessment and economics.

- **Consider how behavior affects benefit endpoints.** One area where economists may lend unique insights at this stage is on assessing how endpoint quantification is affected by behaviors in the baseline and potential behavioral changes from the policy. These behaviors often drive, for example, how and how much individuals are exposed to environmental contaminants. Changes in behavior due to changes in environmental quality (e.g., staying indoors to avoid detrimental effects of air pollution) can be significant and economists need to ensure they are considered in benefits analysis.

**Step 3: Estimate the monetary value of the endpoints**

The next step is to estimate the monetary value to all affected individuals of the quantified benefit endpoints to obtain the total social benefits of each policy option. This starts with identifying valuation estimates for quantified benefit endpoints. Importantly, it may not be sufficient to multiply a change in endpoint by a single value for that endpoint, particularly in the presence of uncertainty or nonlinearities; valuation must be guided by economic theory (Section 7.2). For estimating total benefits, it is typical to use a representative agent approach, where values are calculated for an “average” or representative individual in the relevant population and then multiplied by the number of individuals in that exposed population.

---

183 See, for example, The EPA’s Framework for Human Health Risk Assessment to Inform Decision Making (US EPA 2014).
Text Box 7.2 - Integrating Economics and Risk Assessment

Health and ecological risk assessments are designed to support the setting of standards or to rank the severity of different hazards. However, measures from these assessments can be difficult or impossible to incorporate into benefits analyses for several reasons. First, the measures may not be probabilistic expressions of risk, but instead indicate how exposures compare to reference levels that are not associated with any quantitative level of risk. It may be that the modeled endpoints cannot be directly related to health outcomes or ecological services that can be valued using economic methods. Also, risk assessments sometimes focus on outcomes near the tails of the exposure and/or risk distribution for highly sensitive endpoints, leading to biased benefits estimates if extrapolated to the general population.

Because economists rely on risk assessment outcomes as key inputs into benefits analysis, it is important to coordinate risk assessment and economic valuation. As described in the EPA’s Ecological Benefits Assessment Strategic Plan (US EPA 2006) and Framework for Human Health Risk Assessment (US EPA 2014), this coordination should begin early in the planning process for any risk assessments, starting with the Planning and Scoping and Problem Formulation stages where a conceptual model is developed specifying key factors for the assessment including specific endpoints to be addressed. The EPA’s Generic Ecological Endpoints for Ecological Risk Assessment (US EPA 2016) contains specific guidance to assist ecological risk assessors and economists in identifying ecological services that are amenable to economic analysis (US EPA 2016).

Throughout the risk assessment process economists can contribute information and insights into how behavioral changes may affect realized risk changes. For example, if health outcomes in a particular risk assessment are such that early medical intervention could reduce the chances of illness, economists may be able to estimate the probability that individuals will seek preventive care. Even in cases where the economists’ contribution to the risk characterization is not direct, economists and risk assessors should communicate frequently to ensure that economic analyses are complete. Specifically, risk assessors and economists should strive to:

- Identify a set of human health and ecological endpoints that are economically meaningful, linked to human well-being, and are monetizable using economic valuation methods. Risk assessors may be required to model more or different outcomes than if they were attempting to capture only the most sensitive endpoint. This also may require risk assessors and economists to convert human health or ecological endpoints measured in laboratory or epidemiological studies to other effects that can be valued in the economic analysis.

- Estimate changes in outcome probabilities (human health or ecological) or changes in continuous outcomes (e.g., IQ) as exposure changes, rather than safety assessment measures (e.g., reference doses) whenever possible.

- Work to produce expected or central estimates of risk, rather than bounding estimates as in safety assessments. At a minimum, any expected bias in the risk estimates should be clearly described.

- Attempt to estimate the timing between changes in emissions or exposures and associated changes in health and ecological risks or outcomes. For health outcomes these time lags are referred to as cessation lag (the time between reduced exposure and reduced health risks) or latency (the time between increases in exposure and increased health risks.)

- Attempt to characterize the full uncertainty distribution associated with risk estimates. Not only does this contribute to a better understanding of potential regulatory outcomes, it also enables economists to incorporate risk assessment uncertainty into a broader analysis of uncertainty.

When estimating monetary value of effects, analysts should:

- **Determine which valuation methods are best suited for each endpoint.** When possible, the value estimate should be based on willingness to pay (WTP), but other measures (e.g., cost of illness) may be used when there are no available WTP estimates. Valuation methods are not unique to specific endpoints, and often a given endpoint can be valued through several methods. Table 7.1 shows general benefit categories, examples of specific benefit endpoints, and associated valuation methods commonly used. Sometimes time and resources may be available to conduct original research using these methods, but more often the analysis will need to draw upon existing value estimates in a process called benefit transfer. Details on valuation methods are in Section 7.3. Benefit transfer is described in Section 7.4.

- **Identify valuation estimates and how they are to be used.** Valuation estimates available for benefits analysis will not always match perfectly the policy context being considered. Benefit transfer is the exercise of both identifying valuation estimates that sufficiently relate to the policy context and then transferring the results to the policy analysis. It is important that this is done in ways that are consistent with economic reasoning and theory, and it is not always sufficient to simply apply a single, fixed value. Section 7.4 contains information on both general steps for benefit transfer and specific transfer methods to consider.

- **Describe the source of estimates and confidence in those sources.** Valuation estimates always contain a degree of uncertainty; using them in a context other than the one in which they were initially estimated can only increase that uncertainty. If many studies of the same effect have produced comparable values, analysts can have more confidence in using these estimates in their benefits calculations. In other cases, analysts may have only a single study, or even no directly comparable study, to draw from. In all cases, the benefits analysis should clearly describe the sources of the value estimates used and provide a qualitative discussion of the reliability of those sources.

- **Avoid double-counting to the extent possible.** Double counting may arise for at least two reasons. First, different valuation methods often incorporate different subsets of total benefits, so some types of benefits may be counted twice when aggregating across values. Second, endpoints may be defined in ways that overlap. For example, a human health endpoint of avoided “emergency room visits” is likely to overlap with an endpoint of avoided heart attacks so valuing these endpoints separately and aggregating them would introduce double-counting. It is important to avoid double-counting when possible and to clearly acknowledge any potential overlap when presenting the aggregated results.

- **Characterize uncertainty.** The analysis should include a quantitative uncertainty assessment when possible using sensitivity analysis or other methods. As with other aspects of the analysis the depth and scope of this assessment should be commensurate with the scale of the benefits analysis. In some cases, it may be sufficient to focus on a few key parameters. Important considerations for analysis of uncertainty are provided in Chapter 5, and principles for presenting information on uncertainty are in Chapter 11. The analysis should ultimately present both the aggregate monetized values as well as the value of each specific benefit endpoint. The monetized benefits estimate should be supplemented by displaying benefits that could be quantified but lack valuation estimates, and a characterization of benefits that can only be qualitatively described. When data or modeling limitations prevent quantitative characterization of benefits endpoints it can be useful to provide quantitative data related to benefits (e.g., changes in stressors or environmental quality). Chapter 11 discusses the presentation of information on benefits. When the policy or regulation under consideration is expected to result in important feedbacks and interactions between various physical and economic endpoints, analysts should consider whether available integrated approaches for analyzing the specific policy are more appropriate than quantifying each specific endpoint in a separate analysis.
7.2 Economic Value and Types of Benefits

Economic valuation is based on the traditional economic theory of human behavior and preferences, centered on the “utility” (or “welfare”) that people realize from consumption of goods and services, both in market and non-market settings. Core to this approach is the principle of consumer sovereignty, in which values used for benefit-cost analysis respect the preferences individuals have for these goods and services rather than being based, for example on the preferences of the analyst or policy maker. Different levels and combinations of goods and services provide different levels of utility for any one person. Also, because people have different preferences, utility derived from sets of goods and services will vary across people.

Utility is inherently subjective and cannot be measured directly; however, to give “value” an operational definition in benefits analysis, it must be expressed in a quantifiable metric. Dollars are convenient for direct comparison of benefits to costs and for summing benefits across different effects, but this choice for the unit of account has no special theoretical significance. Table 7.1 summarizes the types of benefits most often associated with environmental protection policies and provides examples of each of the benefits types, as well as valuation methods commonly used to monetize the benefits for each type.

Economic theory suggests that when goods and services are bought and sold in competitive markets, optimizing consumers maximize their level of utility subject to constraints on their budget by equating the ratio of the marginal utility (the utility afforded by the last unit purchased) of any two goods that a person consumes with the ratio of the prices of those goods. If it were otherwise, that person could reallocate her budget to buy a little more of one good and a little less of the other good to achieve a higher level of utility.

The benefits of an environmental improvement are illustrated graphically in Figure 7.2 which shows marginal abatement costs (MAC) and marginal damages (MD) of emissions. Reducing emissions from \( e_0 \) to \( e_1 \) produces benefits equal to the shaded area under the marginal damages curve. Many environmental goods and services, such as air quality and biological diversity, are not traded in markets. The challenge of valuing non-market goods that do not have prices is to relate them to one or more market goods that do. This can be done either by determining how the non-market good contributes to the production of one or more market goods (often in combination with other market good inputs), by observing the trade-offs people make between non-market goods and market goods, or by asking people directly about the tradeoffs they are willing to make. Section 7.3 provides a discussion of the various revealed and stated preference valuation methods. Of course, some methods will be more suitable than others in a given scenario for a variety of reasons, and some will be better able to capture certain types of benefits than others.

The economic valuation of an environmental improvement is the dollar value of the private goods and services that individuals would be willing to trade for the improvement at prevailing market prices. The willingness to trade compensation for goods or services can be measured either as willingness to pay (WTP) or willingness to accept (WTA). WTP is the maximum amount of money an individual would voluntarily pay to obtain an improvement. WTA is the least amount of money an individual would accept to forego the improvement.\(^{184}\) The key theoretical distinction between

---

\(^{184}\) For simplicity, the discussion in this section is restricted to the case of environmental improvements, but similar definitions hold for environmental damages. For a more detailed treatment of WTP and WTA and the closely related concepts of compensating variation, equivalent variation, and Hicksian and Marshallian consumer surplus, see Hanley and Spash (1993), Freeman et al. (2014), Just et al. (2005), and Appendix A of these Guidelines.
WTP and WTA is their respective reference utility levels. For environmental improvements, WTP uses the level of utility without the improvement as the reference point while WTA uses the level of utility with the improvement as the reference point. (Freeman et al. 2014).

Economists generally expect that the difference between WTP and WTA will be negligible, provided the values are small relative to household wealth and substitutes are available for the market or non-market goods in question (Willig 1976, Hanemann 1991). However, there may be instances in which income and substitution effects are important (such as for some environmental goods). Ultimately, economists use the valuation estimates to assess policy outcomes by applying the Kaldor-Hicks compensation test (see Appendix A). In short, the test asks whether hypothetically the gainers from a policy could fully compensate the losers and still be better off and conversely whether the losers could pay the winners to avoid the change altogether and still be as well off. Since WTP is the consistent measure for this test and to simplify the presentation, the term WTP is used throughout the remainder of this chapter to refer to the underlying economic principles behind both WTA and WTP.

WTP for environmental quality can also be non-linear. For example, Figure 7.2 illustrates a case in which marginal damages increase with emissions. When this occurs, it is important to account for baseline environmental quality when valuing the benefits of incremental improvements. Otherwise, inconsistent results can occur when estimating the benefits from a series of separate actions. In addition, sometimes environmental regulations yield relatively small average changes in health or the environment that may not be noticeable to the public until multiple regulations have achieved a large aggregate improvement. Just as it is important to account for small average costs imposed by regulations—which can be economically significant when aggregated over a sufficiently large population—it is conceptually correct to account for even very small improvements in environmental quality. Chapter 5 provides more discussion of analyzing multiple related rules. Text Box 7.6 in section 7.4 discusses the issue of estimating multiple improvements in environmental quality using benefit transfer.

The types of benefits that may arise from environmental policies can be classified in multiple ways (Freeman et al. 2014). As shown in Table 7.1, these Guidelines organize benefits into the following categories: human health improvements, ecological improvements, aesthetic improvements and reduced materials damages.

In addition, commonly used valuation methods are provided for reference. The list is not meant to be exhaustive, but rather to provide examples and commonly used methods for estimating values. The sections below provide a more detailed discussion of each of the benefit categories listed in Table 7.1.

---

185 For more information see Appendix A and Hanemann (1991). Younjun et al. 2015, Freeman et al. (2014) and Horowitz and McConnell (2003) discuss and evaluate various explanations for the disparity between WTP and WTA, and other studies have estimated the size of the disparity, e.g., Tuncel and Hammitt (2014). Kneisner, et al. (2014).
7.2.1 Human Health Improvements

Human health improvements from environmental policies include effects such as reduced mortality rates; decreased incidence of non-fatal cancers, chronic conditions, and other illnesses; and reduced adverse reproductive or developmental effects. These categories of outcomes are discussed separately below.

7.2.1.1 Mortality

Some EPA policies will lead to reductions in human mortality risks due to health conditions such as cancer or cardiovascular disease. In considering the impact of environmental policy on mortality risk, it is important to remember that environmental policies do not protect specific, identifiable individuals from death due to environmental causes. Rather, they generally lead to small reductions in the probability of death for many people.

The value of the mortality risk reductions reflects estimates of individuals' WTP for these small reductions in the risk of dying. When aggregated over the affected population, this value has typically been referred to as the "value of statistical life" (VSL) although other terms have been used. Regardless of terminology, it is important to recognize that it represents the tradeoff between wealth or income and small changes in mortality risks and is not the value of life itself.

For consistency and added transparency across analyses, EPA policy is to apply a single VSL estimate for the calculation of benefits of mortality risk reductions experienced by all affected populations associated with all EPA programs and policies. Appendix B describes this recommended value, its distribution and derivation, and details its application. To reduce public confusion and misunderstanding, analysts should not use the misleading term "value of life" in Agency analyses as that term does not accurately describe what the VSL represents.

As discussed in Appendix B, analysts should address the impact of risk and population characteristics on the VSL qualitatively. In addition, analysts should account for timing considerations, including

- the effects of latency -- delayed manifestation of health or other effects,
- cessation lags -- time frame between a reduction in exposure to an environmental contaminant and the reduced risk to health, and
- income growth over time, discounting appropriately where warranted.\(^{186}\)

Valuing mortality risk changes in children is particularly challenging. The EPA’s *Handbook for Valuing Children’s Health Risks* (2003b) provides some information on this topic, including key benefit transfer issues to consider when using adult-based studies. In addition, OMB's *Circular A-4* advises:

“For rules where health gains are expected among both children and adults . . . the monetary values for children should be at least as large as the values for adults (for the same probabilities and outcomes) unless there is specific and compelling evidence to suggest otherwise” (OMB 2003).

Reviews of the literature by Gerking and Dickie (2013) and Robinson et al. (2019) provide support for this position.

Methods for valuing mortality risk changes

Because individuals make risk-wealth trade-offs in different contexts, the value of mortality risk changes can be estimated using a variety of data sources and modeling approaches. The estimate recommended in Appendix B is derived from a combination of hedonic-wage and stated preference studies. In the hedonic wage or wage-risk method, value is inferred from the income-risk trade-offs made by workers for risks faced on the job. Stated preference studies, in which income-risk trade-offs are solicited directly through surveys, are also used to estimate WTP for reduced mortality risks. Key considerations in these studies include the extent to which individuals know and understand the

---

186 See Chapter 6 for more information on discounting.
risks involved, and the ability of the study to control for aspects of the actual or hypothetical transaction that are not risk-related.

There are additional methods that can be used to derive information on risk trade-offs. Averting behavior studies value risk changes by examining purchases of goods that can affect mortality risk (e.g., bicycle helmets). However, isolating the portion of the purchase price associated with mortality risk reductions from other benefits or joint products provided by the good is a challenging hurdle for this literature. Also, of potential importance is short term avoidance behavior -- altering one's activities, including the timing and frequency of activities, to reduce exposure.\textsuperscript{187} Another approach is to examine trade-offs between types of risks to estimate relative preferences for risk reduction. This approach may make the valuation task more manageable for the respondent but requires multiple steps to obtain a risk-dollar tradeoff.\textsuperscript{188}

**Important considerations**

The analyst should keep three important considerations in mind when estimating mortality benefits (each described in more detail below):

- Characterizing and measuring mortality effects;
- Heterogeneity in risk and population characteristics; and
- The timing of health risk changes.

**Characterizing and measuring mortality effects**

Although reduced mortality risks associated with an environmental policy are typically small, they are generally aggregated over the affected population and reported in terms of “statistical lives.” Suppose, for example, that a policy affects 100,000 people and reduces the risk of premature mortality in the coming year by 1 in 100,000 for each individual. Summing these individual risk reductions across the entire affected population shows that the policy leads to the equivalent of one premature fatality averted, or one statistical life “saved,” in the coming year.

An alternative metric seeks to capture the remaining life expectancy, or "quantity of life" saved associated with the risk reductions (Moore and Viscusi 1988) and is typically expressed as “statistical life years.” Looking again at the policy described above for reducing risk in the coming year, suppose the risks were spread over a population in which each individual had 20 years of remaining life expectancy. The policy would then "save" 20 statistical life years (1 statistical life x 20 life years). In practice, estimating statistical life years saved requires risk information for specific subpopulations (e.g., age groups or health status). Statistical life years may be used as an outcome measure in cost-effectiveness analysis (IOM 2006). However, consistent with past SAB advice, the use of a constant monetized value for a statistical life year is not supported by the literature and is not recommended for benefits analysis (US EPA 2007).

**Heterogeneity in risk and population characteristics**

The WTP to avoid mortality risks can vary both by risk characteristics and by the characteristics of the affected population. Key risk characteristics include voluntariness (i.e., whether risks are voluntarily assumed), timing (immediate or delayed), risk source (e.g., natural vs. man-made), and the causative event (e.g., cancer or trauma). Population characteristics include those generally expected to influence WTP for any good (e.g., income and education) as well as those more closely related to mortality risks such as baseline risk or remaining life expectancy, health status, risk aversion, and familiarity with the type of risk. The empirical and theoretical literature on the effect of many of these characteristics on willingness to pay is incomplete or ambiguous. For example, some studies suggest that older populations are willing to pay less for risk reductions (Viscusi and Aldy, 2007a); others find this effect to be small if it exists at all (Alberini et al. 2004). Still others suggest older populations have higher WTP (Kniesner, Viscusi, and Ziliak


\textsuperscript{188} See Nielsen, et al. (2019) for an overview and application of risk-risk tradeoff method.
Similarly, some studies have found that reductions in fatal cancer risks garner a higher WTP than other kinds of fatal risks (e.g., Viscusi et al. 2014) while others do not find evidence of a "cancer premium" (e.g., Hammitt and Haninger 2010). Appendix B contains a more complete discussion of risk and population characteristics and how they may affect WTP. Although mortality risk valuation estimates used in economic analyses could reflect these differences in WTP quantitatively with sufficient empirical evidence, Agency policy is to apply a single VSL estimate to all populations and mortality risks and to qualitatively describe population characteristics and risk attributes. One reason for this position is that the empirical evidence in the literature on the relationship between WTP and the various population and risk characteristics is inconclusive. In addition, population characteristics become less relevant for applications of VSL in benefits assessments of national regulations affecting broad spectrums of the population.

**Timing of health risk changes**

Environmental contamination can cause immediate or delayed health effects. If individuals typically prefer health improvements earlier in time rather than later, all else equal, then the WTP for reductions in exposure to environmental pollutants will depend on when the resulting health risk changes will occur.\(^{189}\)

The effects of timing on the present or annualized value of reduced mortality risk can be considered in the context of a lifecycle consumption model with uncertain lifetime (Cropper and Sussman 1990, Cropper and Portney 1990, and U.S. EPA 2007g). In this framework reductions in mortality risk are represented as a shift in the survival curve — the probability an individual survives to future ages as a function of current age — which leads to a corresponding change in life expectancy and future utility.

If the basis for benefit transfer is a marginal WTP for contemporaneous risk reductions, then calculating the benefits of a policy with delayed risk reductions requires three steps:

1. estimating the time path of future mortality risk reductions;
2. estimating the annual WTP for all future years; and
3. calculating the present value of these annual WTP amounts.

The first step should account for all the factors that ultimately relate changes in exposure to changes in mortality risk as defined by shifts in the survival curve.

**7.2.1.2 Morbidity**

Morbidity benefits consist of reductions in the risk of non-fatal health effects ranging from mild outcomes, such as headaches and nausea, to very serious illnesses such as cancer (see Table 7.1). Non-fatal health effects also include conditions such as birth defects, low birth weight, and reduced cognitive function. Morbidity outcomes need not be so severe to prevent affected individuals from participating in normal activities but are expected to affect quality of life and labor productivity or earnings for workers (Graff Zivin and Neidell 2013; 2018). Availability of existing valuation estimates for morbidity outcomes varies considerably, and the WTP to avoid many health outcomes are not available in the economics literature.

WTP to reduce the risk of experiencing an outcome is the preferred measure of value for morbidity effects. As described in Freeman et al. (2014), this measure consists of four additive components:

- “Averting costs” to reduce the risk of illness;
- “Mitigating costs” for treatments such as medical care and medication;

\(^{189}\) The description here focuses on mortality risk, but the same principles apply to non-fatal health risks.
• Indirect costs such as reduced earnings from paid work, or lost time maintaining a home and pursuing leisure activities; and
• Monetary equivalent of the disutility of illness (e.g., costs of discomfort, anxiety, pain, and suffering.)

Methods used to estimate WTP vary in the extent to which they capture these components. For example, cost-of-illness (COI) estimates generally only capture mitigating and indirect costs and omit averting expenditures and lost utility associated with pain and suffering. Consequently, COI estimates generally understate WTP to reduce a risk or avoid a given health effect. Some studies have estimated that total WTP can be two to four times as large as COI even for minor acute respiratory illnesses (Alberini and Krupnick 2000). Still, there is not any broadly applicable “scaling factor” that relates COI to WTP.

Methods for Valuing Morbidity

Researchers have developed a variety of methods to value changes in morbidity risks. Some methods measure the theoretically preferred value of individual WTP to avoid a health effect. Others (e.g., cost of illness) do not measure WTP but can provide useful data; however, those data must be interpreted carefully if they are to inform economically meaningful measures. Methods also differ in the perspective from which values are measured (e.g., before or after the incidence of morbidity), whether they control for the opportunity to mitigate the illness (e.g., before or after taking medication) and the degree to which they account for the four components of total WTP set out above. The three primary methods most often used to value morbidity in an environmental context are stated preference (Section 7.3.2), averting behavior (Section 7.3.1.4), and COI (Section 7.3.1.5). Hedonic methods (Section 7.3.1.3) are used less frequently to value morbidity from environmental causes.

Benefits analysis may also be informed by approaches that do not estimate WTP of reduced morbidity directly. As noted above, risk-risk trade-offs, for example, do not directly estimate dollar values for risk reductions, but rather, provide rankings of relative risks based on consumer preferences. Risk-risk trade-offs can be linked to WTP estimates for related risks.190

Other methods for valuing morbidity outcomes suffer from certain methodological limitations and are therefore generally less useful for policy analysis. For example, health-state indices, such as Quality Adjusted Life Years (QALYs) are composite metrics that combine information on quality and quantity of life lived under various scenarios often used for cost-effectiveness or cost-utility analyses (see Section 7.5.2.1). These measures are consistent with WTP measures only under very strict conditions that generally do not hold in practice (Bleichrodt and Quiggin, 1999). They should not be used for deriving monetary estimates for use in BCA (Hammitt 2003, and Institute of Medicine (IOM) 2006). Another commonly suggested alternative is jury awards; these also generally should not be used in benefits analysis, for reasons explained in Text Box 7.3.

---

190. EPA analyses have used risk-risk trade-offs for non-fatal cancers in conjunction with VSL estimates as one method to assess the benefits of reduced carcinogens in drinking water [U.S. EPA 2005a].
Text Box 7.3 - Non-Willingness to Pay Measures

Economic measures of value calculate willingness to pay for environmental changes. WTP is defined as that amount of money an individual would voluntarily pay for an environmental improvement. WTP is a valid measure of “economic value” because it can be used in potential compensation tests of Kaldor and Hicks. Measures of economic value that do not measure WTP and cannot be related to changes in utility are not valid. Others should be used only in a limited set of circumstances. Some examples are provided below.

**Replacement cost.** A common consequence of environmental deterioration is damage to assets. Some analysts suggest that the economic value of the damage is the cost of replacing the asset. In the context of benefit-cost analysis this is not generally true. It is only true if: (1) damage to the asset is the only cost incurred; and (2) the least expensive way to achieve the level of satisfaction realized before the deterioration would be to replace the asset (Freeman, Herriges, and Kling 2014). If the first condition is not met, consideration of replacement costs may be useful but should be combined with assessments of other costs. If the second condition is not met, replacement costs might overestimate the consequences. Suppose that water pollution kills fish in a pond. Replacing those fish with healthy, edible ones might prove extremely expensive: the pond might need to be dredged and restocked. However, people who are no longer able to catch fish in the pond might be compensated simply by giving them enough money to purchase substitutes in the market.

**Proxy costs.** A closely related concept to replacement cost is the cost of a substitute for the damaged asset. Ecologist H.T. Odum (1996) calculated the number of barrels of petroleum required to provide the energy to replace the services of wetland ecosystems. However, since there is no reason to suppose that people would be willing to pay for oil to replace services of damaged wetlands, this number is economically irrelevant. A similar argument can be made against the interpretation of “ecological footprints” as an estimate of economic consequences (Wackernagel and Rees 1996). Dasgupta (2002) interprets these approaches as single-factor theories of value, fallacies that were disproved in general by Samuelson’s (1951) “non-substitution theorem.”

**Cost-of-illness (COI).** The value of avoiding adverse health effects is often proxied by the “cost of illness” -- the total costs of treatment and lost productivity to an illness or adverse health outcome. COI captures components of WTP, but not all: (1) COI does not record other expenses incurred in efforts to avoid illness; (2) health insurance may drive a wedge between the costs incurred to treat illness and WTP to avoid it; and (3) COI ignores factors such as discomfort and dread that patients would also be willing to pay to avoid. Section 7.3.1.5 provides more details on the COI method and its use in benefits analysis.

**Jury awards.** Attempts are sometimes made to value environmental improvements using jury awards. Using jury awards in this way may prove problematic for at least two reasons. First, cases only go to trial if both sides prefer the expected value of an adjudicatory outcome to the certainty of a pre-trial settlement. Cases that go to juries are “atypical” by definition. Second, since adjudication does not always occur and can never be infallible, jury awards often do, and arguably should (Shavell 1979), embody “punitive” as well as “compensatory” elements. Juries make examples of guilty defendants to try to deter others from committing similar offenses. For this reason, jury awards may overstate typical damages. Finally, jury awards reflect a certain outcome and not the probability of experiencing an adverse event and therefore include the influence of characteristics typically not included in statistical analysis (e.g., pain, suffering, and likeability). These estimates are not appropriate for application to ex ante evaluation of the value associated with a statistical probability.

**Important considerations**

The analyst should keep two important considerations in mind when estimating morbidity benefits:

- Characterizing and measuring morbidity effects; and
- Third party costs

in addition to heterogeneity in risk and population characteristics and timing of health risk changes discussed above under mortality effects but equally applicable here.
Characterizing and measuring morbidity effects

Key characteristics that will influence the valuation of morbidity effects are their severity, frequency, duration, and symptoms. Severity defines the degree of impairment associated with the illness. Examples of how researchers have measured severity include daily limitations such as “restricted activity,” “bed disability,” and “lost work.”\textsuperscript{191} Severity can also be described in terms of health state indices that combine multiple health dimensions into a single measure.\textsuperscript{192} For duration, the primary distinction is between acute effects and chronic effects. Acute effects are discrete episodes usually lasting only a few days, while chronic effects last much longer and are generally associated with long-term illnesses. The frequency of effects also can vary widely across illnesses. Some effects, such as a gastrointestinal illness are one-time events that are unlikely to recur. Other effects, such as asthma, do recur or can be exacerbated regularly, causing disruptions in work, school, or recreational activities.

For chronic conditions or more serious outcomes, morbidity effects are usually measured in terms of the number of expected cases of illness. Given the risks faced by each individual and the number of people exposed to this risk, an estimate of “statistical cases” can be defined analogously to “statistical lives.” In contrast, morbidity effects that are considered acute or mild in nature can be estimated as the expected number of times a particular symptom associated with an illness occurs within an affected population over a specified period of time (e.g., annually), where individual members of the population may experience the effect more than once. These estimates of “symptom days” may be used in benefits analysis when appropriate estimates of economic value are available, although a richer characterization of combinations of symptoms, severity, duration, and episode frequency would be an improvement over much of the existing literature. Some studies have attempted to address these complexities in a more systematic manner (Cameron and DeShazo 2013).

Third party costs

The widespread availability of health insurance and paid sick leave shifts some of the costs of illness from individuals to others. While this cost-shifting can be addressed explicitly in COI studies, it may lead to problems in estimating total WTP especially in stated preference studies. If the researcher does not adequately address these concerns, individuals may mis-state their WTP, assuming some related costs will be borne by others. Some stated preference studies are designed to avoid capturing third party or insurance costs in which case the results would be additive to COI. Regardless, to the extent third party costs represent diversions from other uses in the economy, they represent real costs to society and should be accounted for in the analysis.

7.2.2 Ecological Benefits

Many EPA policies will produce ecological benefits by enhancing the delivery of ecosystem services, defined here as “direct or indirect contributions that ecosystems make to the well-being of human populations” (Thompson et al., 2009). Examples of EPA policies affecting ecosystem services include: reducing acid precipitation that may acidify forests and freshwater ecosystems. Controlling pesticides and other environmental contaminants that affect pollinators such as bees, as well as predators of pests. Reducing nutrient pollution from municipal wastewater treatment plants, septic systems, fertilizer and manure runoff, and atmospheric deposition may lead to changes in the composition and attributes of receiving water bodies.

\textsuperscript{191} As Cropper and Freeman (1991) note, these descriptions are essentially characterizations of a behavioral response to the illness. Lost workdays, for example, in some cases require a decision on an individual’s part not to go to work due to illness. Such a response may depend upon various socioeconomic factors as well as the physical effect of the illness.

\textsuperscript{192} The difference in the indices is intended to reflect the relative difference in disutility associated with symptoms or illnesses. There are serious questions about the theoretical and empirical consistency between these “health-related quality of life” index values and WTP measures for improved health outcomes (Hammitt 2002). Still the inclusion of some aspects of these indices may prove useful in valuation studies (Van Houtven et al. 2006). Examples of economic analyses that have employed some form of health state index include Desvousges et al. (1998) and Magat et al. (1996).
In each of these examples, environmental regulation may not directly affect goods or services in the household utility function. Instead they affect ecological inputs into the processes that generate such goods and services. The valuation of ecosystem services is not fundamentally different than the valuation of other productive assets on which our economy depends (Polasky 2008, Barbier 2012). The relevant endpoints are goods or services that enter the household utility function directly. Ecosystem services that contribute to the production of those assets, but are not directly valued by households, should be recognized as inputs in an ecological production function (EPF) and monetization should value their marginal product. Making the distinction between final ecosystem services and ecological inputs and identifying the relevant linkages is a challenging task facing analysts of environmental policy.

7.2.2.1 Ecological production functions

An ecological production function is a description of how ecosystems combine inputs to produce ecosystem services that consumers enjoy directly or are used in the production of goods or services that are enjoyed by consumers. The natural science literature provides guidance for some cases on the form of the ecological production function (MacArthur and Wilson 1967; Kingsland 1985) and numerous examples (Hamel et al. 2015; Reddy et al 2015; Kremen et al. 2007; Jaramillo et al. 2010).

Knowledge of the relevant ecological linkages may be essential to predicting the effects of environmental policies on ecosystem service provision and an economic analyst may benefit from collaborating with ecologists or other natural scientists to predict the effects of the proposed policy. The Agency’s Ecological Benefits Assessment Strategic Plan describes an interdisciplinary approach for conducting ecological benefits assessments (U.S. EPA 2006a). To familiarize themselves with “benefit relevant indicators” of ecological endpoints that may be affected by policy measures, analysts may also wish to consult National Ecosystem Services Partnership (2015) or National Ecosystem Services Classification System (US EPA 2015a).

There are several sources an analyst might consult for potentially useful ecological production functions. In addition to searching the scientific literature on the topic of interest, some large-scale research ventures maintain suites of models of pollination, storm protection, pollution treatment, groundwater recharge, and other phenomena. The Natural Capital Project, for example, maintains 19 models of Integrated Valuation of Ecosystem Services and Tradeoffs (InVEST). Another suite of such models is the Artificial Intelligence for Ecosystem Services platform (ARIES). Bagstad, et al. (2012) review and compare InVEST and a number of other models that might be used to model the generation of ecosystem services. In addition, the EPA’s Office of Research and Development has compiled lists of potentially useful models in its Ecoservice Models Library (ESML 2018).

Using an “off-the-shelf” ecological production function in benefits analysis may have the disadvantage of not being tailored to the specific application, but the advantages could include:

- Some EPFs have been extensively peer-reviewed and found to be of practical value;
- Using an existing EPF may save considerable time and effort;
- In some instances, it may be possible to calibrate the parameters of an ecological production function, even if the parameters are unknown, using readily available data or summary statistics for the policy case (e.g., Massey et al. 2017);
- Alternatively, it may be possible to develop useful bounding results that hold regardless of particular parameter values (e.g., Simpson, et al. 1996; Simpson 2016).

Even when using established and tested ecological production functions, there is no substitute for substantive familiarity with the subject matter. The analyst should understand the logic behind the form of the ecological production function and consult with an expert on the subject (often a biologist or other natural scientist) before deciding to adopt one for her work. Another consideration is that arguments of the production function should be relevant to the problem the
analyst is addressing. Many ecosystem service models relate habitat area to the provision of a service; this may not inform questions of how pollutants affect the provision of the service.

Ecological production functions sometimes exhibit what may seem to be counterintuitive effects. Consider as an example, nutrients (primarily reactive nitrogen and phosphorus) from municipal, agricultural, or other sources that may enter water bodies. The term “eutrophication” describes the consequences of excessive delivery of nutrients. Marine biologists have documented that, in some circumstances, increased nutrient availability may enhance some desirable endpoints, such as the support of larger populations of fish caught by commercial and recreational fishers (e.g., Breitburg, et al., 2009). In another example, bees that carry pollen between orange groves are a necessary component of orange productions, however, carrying pollen between different groves may at times hybridize fruits resulting in a lower value crop (Sagoff 2011). These examples are unusual, but they underscore a point: preserving systems in, or restoring them to, a more “natural” state may not always enhance the value of the services they provide.

7.2.2.2 Estimating Benefits Using Ecological Production Functions

Knowing only the ecological production function is generally insufficient to conduct economic valuation. While ecological production functions are analogous to production functions that are a staple of textbook microeconomics, they often differ in one important respect: the inputs and outputs of ecological production functions are often not traded in markets (EPA 2009b). Consequently, rather than being able to observe prices, we must infer them using the tools of nonmarket valuation. Massey et al. (2006); Newbold and Massey (2010); Smith and Crowder 2011; and Finnoff and Tschirhart (2011) exemplify how these linkages can be made for commercial fisheries and recreational anglers.

Monetizing the ecological benefits of environmental regulations using ecological production functions proceeds in three phases (Bateman, 2012). The first is to project changes in the ecological inputs caused by the regulation. This phase may require its own extensive modeling effort such as hydrological models that predict the effect of land use changes on nutrient and sediment loadings to lakes, rivers, and streams. The second phase employs the ecological production function to project how the changes in those inputs affect the provision of final ecosystem services. To use the nutrient and sediment pollution example again, this would require a model of aquatic ecosystems to project changes in environmental goods that people value such as fish to catch and water amenities like clarity and odor. Finally, changes in final ecosystem services are valued using nonmarket valuation methods. Revealed and stated preference approaches to nonmarket valuation are described in detail in sections 7.3.1 and 7.3.2. When resource constraints prevent an original nonmarket valuation study, benefit transfer can be used to apply values estimated in other contexts; see section 7.4 for a detailed discussion.

7.2.2.3 Benefits Estimation When the Ecological Production Function is not Known

When ecological production functions are not known, it may be easier and/or more defensible to infer ecosystem service values from other relationships. If only the changes in the ecological inputs are known, these can be used in revealed preference approaches to valuation by observing their impact on complementary market behaviors.

Fundamental results in economics establish that these production relationships may be equivalently expressed as profit functions and that profits may be capitalized into the value of assets such as advantageously located property. As such, hedonic valuation methods are frequently proposed for ecosystem service valuation (see, for example, Swinton et al. 2007, Bishop and Timmins 2018). Several researchers have conducted hedonic property value studies to estimate values of assets such as forest cover (Kim and Johnson 2002, Tyrvainen and Miettinen 2000, Mansfield et al. 2005, and Sander et al., 2010), wetlands (Tapsuwan et al. 2009, Mahan et al. 2000, Woodward and Wui 2001, Bin and Polasky 2005), or other varieties of “open space” (Sander and Polasky 2009, Cho et al. 2006, Irwin and Bockstael 2002, Irwin 2002, and Thorsnes 2002).

The estimation of recreational demand, or, more generally, locational choice models (e.g., Kuminoff et al., 2013) are based on similar underlying principles: choices of where to visit or live are made to maximize utility (or profits), and the
ecological attributes of an area affect such choices (McConnell 1990; Parsons 1991; Phaneuf, et al., 2008). Hedonic price, recreational demand, or locational choice models may be regarded as “reduced form” representations of ecological production from which the analyst can infer the values individuals ascribe to ecosystem services by observing the choices they make, provided that the analyst can adequately control for potentially confounding factors. These approaches are discussed further in sections 7.3.1.2 and 7.3.1.3.

7.2.3 Other Benefits

Other types of potential benefits from environmental policies include aesthetic improvements and reduced material damages.

Aesthetic improvements include effects such as the improved taste and odor of tap water resulting from water treatment requirements and enhanced visibility resulting from reduced air pollution. Increased visibility due to improved air quality can be divided into two types of benefits: residential visibility benefits and recreational visibility benefits. Improvements in residential visibility are generally assumed to only benefit residents living in the areas in which the improvements are occurring, while all households in the United States are usually assumed to derive some benefit from improvements in visibility in areas such as national parks. The benefits received, however, may decrease with the distance from the recreational area in which the improvements occur.

Reduced materials damages include welfare impacts that arise from changes in the provision of service flows from human-made capital assets such as buildings, roads, and bridges. Materials damages can include changes in both the quantity and quality of such assets. Benefits from reduced material damages typically involve cost savings from reduced maintenance or restoration of soiled or corroded buildings, machinery, or monuments.

Methods and previous studies

Changes in the stock and quality of human-made capital assets are assessed in a manner similar to their “natural capital” counterparts. Analytically, the valuation of reduced materials damages parallels the methods for valuing the tangible end products from managed ecosystems such as agriculture or forestry. For example, effects from changes in air quality on the provision of the service flows from physical resources are handled in a fashion similar to the effects from changes in air quality on crops or commercial timber stocks. The most common empirical applications involve air pollution damages and the soiling of structures and other property.

Linking changes in environmental quality with the provision of service flows from materials can be difficult because of the limited scientific understanding of the physical effects, the timing of the effects, and the behavioral responses of producers and consumers. An analysis of reduced materials damages often begins with an environmental fate and transport model to determine the direct effects of the policy on the stocks and flows of pollutants in the environment. Then stressor-response functions are used to relate local concentrations of pollutants to corrosion, soiling, or other physical damages that affect the production (inputs) or consumption (output) of the material service flows. The market response to these impacts serves as the basis for the final stage of the assessment, in which some type of structural or reduced-form economic model that relates averting or mitigating expenditures to pollution levels is used to value the physical impacts. The degree to which behavioral adjustments are considered when measuring the market response is important, and models that incorporate behavioral responses are preferred to those that do not. Adams and Crocker (1991) provide a detailed discussion of this and other features of materials damages benefits assessment. Also see the EPA’s benefits analysis of household soiling for an example that employs a reduced-form economic model relating defensive expenditures to ambient pollution (U.S. EPA 1997f).

---

7.3 Economic Valuation Methods for Benefits Analysis

For goods bought and sold in undistorted markets, the market price indicates the marginal social value of an extra unit of the good. Often, there are no markets for environmental goods. While some natural products are sold in private markets, such as timber and fish, the analyst’s concern will typically be with non-market inputs, which are, by definition, not traded in markets. To overcome this lack of market data, economists have developed a number of methods to value environmental quality changes. Most of these methods can be broadly categorized as either revealed preference or stated preference methods.

In cases where markets for environmental goods do not exist, WTP can often be inferred from choices people make in related markets. Specifically, because environmental quality is often a characteristic or component of a private good or service, it is sometimes possible to disentangle the value a consumer places on environmental quality from the overall value of a good. Methods that employ this general approach are referred to as revealed preference methods because values are estimated using data gathered from observed choices that—when combined with several important auxiliary assumptions (individuals have complete and stable preferences, are expected utility maximizers, have all relevant information, etc.) -- reveal the preferences of individuals. Revealed preference methods include production or cost functions, travel cost models, hedonic pricing models, and averting behavior models. This section also discusses COI methods, which are sometimes used to value human health effects when estimates of WTP are unavailable.

In situations where no markets for environmental or related goods exist to infer WTP, economists sometimes rely on survey techniques to gather choice data from hypothetical markets. The methods that use this type of data are referred to as stated preference methods because they rely on choice data that are stated in response to hypothetical situations, rather than on choice behavior observed in the real world. Stated preference methods include contingent valuation, conjoint analysis, and contingent ranking.

As a general matter, revealed preference methods have the advantage that they are based on actual tradeoffs and decisions made by individuals. Stated preference methods sometimes have the advantage that the choice question can be tailored to obtain values that more closely align with the needs of benefits analysis. Each of these revealed and stated preference methods is discussed in detail below, starting with an overview of the method, a description of its general application to environmental benefits analysis, and a discussion of issues involved in interpreting and understanding valuation studies. The discussion concludes with a separate overview of benefit-transfer methods.

It is important to keep in mind that research on all of these methods is ongoing. The limitations and qualifications described here are meant to characterize the state of the science at the time these Guidelines were written. Analysts should work with NCEE to determine the usefulness of additional resources as they become available. In practice, analyses will often need to draw upon values from multiple methods to value benefits. Text Box 7.4 briefly describes original valuation studies using multiple methods conducted by the Agency to estimate benefits of improved water quality in the Chesapeake Bay using many of the methods discussed here.

7.3.1 Revealed Preference Methods and the Cost of Illness

A variety of revealed preference methods for valuing environmental changes have been developed and are widely used by economists. While these methods all use observable data to estimate or infer value, they each have their own set of advantages and limitations. The following common types of revealed preference methods are discussed in this section:

There are examples in which environmental goods have been traded in markets. The Clean Air Act Amendments of 1990, for example, initiated a market in sulfur dioxide (SO\textsubscript{2}). Similarly, studies in the computational economics literature have performed ex-ante calibrated analyses by constructing hypothetical markets for environmental goods. The non-market good valuation is determined through shadow prices in imposed regulation-induced scarcity in the environmental good. However, prices in such markets are determined by ex-ante policy driven quantity constraints, and not through empirically based statistical methods by considerations of marginal utilities or marginal products.
Text Box 7.4 - Benefits Analysis of the Chesapeake Bay TMDL

In 2010, the EPA established the Chesapeake Bay total maximum daily load, a comprehensive “pollution diet” to restore clean water to the Bay and area streams, creeks, and rivers. The EPA’s National Center for Environmental Economics was tasked with assessing the TMDL’s benefits in a multi-faceted analysis of recreational and aesthetic amenities. NCEE began by conducting a scoping exercise to inform which detailed benefits analyses to pursue. Categories yielding small benefits in previous analyses were shelved (U.S. EPA 2002c, 2009b). EPA’s CB Program Office provided data on water clarity and pollutant loadings with and without the TMDL. NCEE engaged external experts on CB fisheries and water quality to obtain their best professional judgments of potential stock size changes relative to current water quality conditions, holding all other influences constant. NCEE used an extension of ORD’s SPARROW model to predict nutrient loadings and chlorophyll in lakes (Moore et al. 2011). Summaries appear below. Estimates are not additive across studies; overlaps may exist among homeowners, recreators, and respondents.

Hedonic property value analysis: Walsh et al. (2017) used spatially explicit water quality data paired with economic, geographic, and demographic variables to analyze the value of water clarity to home buyers using over 200,000 property sales in Maryland. Klemick et al. (2018) then used meta-analysis to synthesize the value of clarity improvements in Maryland and to transfer the results to properties in Delaware, Virginia, and the District of Columbia. Together, they found that predicted water clarity improvements from the TMDL would result in a 0.7 to 1.3 percent increase in property value for waterfront homes. Properties farther from the water had smaller effects. Total near-waterfront property values could increase by about $458 to $802 million from water clarity improvements, which is equivalent to an annualized value of $14 to $56 million at discount rates of 3 and 7 percent.

Market analysis: Like many fresh goods, fish and shellfish are highly perishable; producers cannot easily adjust supply in the short run to respond to changes in demand. Moore and Griffiths (2018) developed a two-stage inverse demand model to describe how prices respond to supply changes in other commodity groups. The model allowed NCEE to estimate consumer welfare impacts of an increase in CB fish and shellfish harvests while allowing other areas’ harvests to act as substitutes. The estimated annual value of expected harvest improvements is $14.2 million.

Fishing model: NCEE estimated benefits to recreational anglers using a linked participation and site-choice recreation demand model. The model relied on historic catch rate data from the Marine Recreational Fisheries Statistics Survey, an intercept survey that uses weights based on historic visitation frequencies at each intercept site. The data were used to estimate a random utility site-choice model and trip counts from respondent zip codes were used to estimate a participation model conditional on the inclusive value of all sites as estimated by the site-choice model. The resulting estimates of recreational fishing benefits range between $5.7 and $67.6 million per year.

Other recreation demand: NCEE used a recreation demand model to estimate the benefits from other outdoor recreation activities using data on total visitor counts to national and state parks in Maryland, Virginia, and Delaware, supplemented with survey data on the number of recreation trips taken to the CB area. The marginal effects of water quality on recreationists’ site choices were estimated in a second-stage regression, using estimates of site-specific constants from the first-stage site-choice model as the dependent variable and measures of average water quality conditions and other fixed site attributes as explanatory variables. The estimated annual outdoor recreation benefits (exclusive of recreational fishing) range from $120 to $321 million.

Stated preference survey: Moore, et al. (2018) conducted an SP survey linking forecasted water quality changes to ecological endpoints to estimate use and nonuse values for aesthetic and ecological improvements in the CB and watershed lakes. The survey estimated WTP for improvements in water clarity; populations of three CB species (striped bass, blue crab, and oysters); and the condition of freshwater lakes in the CB Watershed. They found that benefits to watershed lakes and nonuse values account for a large proportion of total WTP and would significantly affect the benefit-cost ratio of pollution reduction programs. Estimated benefits from the projected environmental improvements after the TMDL range from $4.47 billion to $7.79 billion per year.

- Production or cost functions;
- Travel cost models;
- Hedonic models; and
- Averting behavior models.
This section also discusses the Cost of Illness (COI) approach to valuation.

7.3.1.1 Production and Cost Functions

Discrete changes in environmental circumstances generally cause both consumer and producer effects, and it is common practice to separate the welfare effects brought about by changes in environmental circumstances into consumer surplus and producer surplus.\(^{195}\) Marginal changes can be evaluated by considering the production side of the market alone.

**Economic foundations of production and cost functions**

Inputs to production contribute to welfare indirectly. The marginal contribution of a productive input is calculated by multiplying the marginal product of the input by the marginal utility obtained from the consumption good, in whose production the input is employed. The marginal value of a consumption good is recorded in its price. While marginal products are rarely observed, the need to observe them is obviated when both inputs and outputs are sold in private markets because *prices* can be observed. Environmental goods and services are typically not traded in private markets, and therefore the values of environmental inputs must be estimated indirectly.

Production possibilities can be represented in three equivalent ways:

- As a production function relating output to inputs;
- As a cost function relating production expenses to output and to input prices; and
- As a profit function relating earnings to the prices of both output and inputs (Varian (1992) describes the relationships among these functions).

The value of a marginal change in some environmental condition can be represented as a marginal change in the value of production, as a marginal change in the cost of production, or as a marginal change in the profitability of production.\(^{196}\) It should be noted, however, that problems of data availability and reliability often arise. Such problems may motivate the choice among these conceptually equivalent approaches, or favor of another approach.

Note that derivation of values *on the margin* does not require any detailed understanding of consumer demand conditions. To evaluate marginal effects via the production function approach, the analyst needs to know the price of output and the marginal product of the environmental input. To derive the equivalent measure using a cost function approach, the analyst needs to know the derivative of the cost function with respect to the environmental input. In the profit function approach, the analyst needs to know the derivative of the profit function with respect to the environmental input.

In the statements above, note the emphasis that marginal effects are being estimated. Estimating the net benefits of larger, non-marginal, changes represent a greater challenge to the analyst. In general, this requires consideration of changes in both producer and consumer surplus.

**Links between production and hedonic and other models**

Note a fourth equivalent way to estimate environmental effects on production possibilities. Such effects are reflected in the profitability of enterprises engaged in production. That profitability also can be related to the return on fixed assets such as land. The value of a parcel of land is related to the stream of earnings that can be achieved by employing it in its “highest and best use.” Its rental value is equal to the profits that can be earned from it over the period of rental (the terms “rent” and “profit” are often used synonymously in economics). The purchase price of the land parcel is equal to the expected discounted present value of the stream of earnings that can be realized from its use over time. Therefore,

---

\(^{195}\) See Appendix A for more detail.

\(^{196}\) For a review of statistical procedures used for estimating production, cost, and profit functions see Berndt (1991).
the production, cost, and profit function approaches described above are also equivalent to inferences drawn from the effects of environmental conditions on asset values. This fourth approach is known as “hedonic pricing,” and will be discussed in detail in Section 7.3.1.3. It is introduced now to show that production, cost, or profit function approaches are generally equivalent to hedonic approaches.

“Production" as a term is broad in meaning and application, especially with regard to hedonic pricing. While businesses produce goods and services in their industrial facilities, one might also say that developers “produce” housing services when they build residences. Therefore, hedonic pricing approaches can measure the value of the environment in “production,” whether they are focusing on commercial or residential properties. Similarly, households may “produce” their health status by combining inputs such as air and water filtration systems and medical services along with whatever environmental circumstances they face. Or they “produce” recreational opportunities by combining “travel services” from private vehicles, their own time, recreational equipment purchases, and the attributes of their destination. Much of what is discussed elsewhere in this section is associated with this type of production analysis. This is not to say that estimation of production, cost, or profit functions is necessarily the best way to approach such problems, but rather, that all of these approaches are conceptually consistent.

**General application of production and cost functions**

Empirical applications of production and cost function approaches are diverse. Among other topics, the empirical literature has addressed the effects of air quality changes on agriculture and commercial timber industries. It also has assessed the effects of water quality changes on water supply treatment costs and on the production costs of industry processors, irrigation operations, and commercial fisheries.¹⁹⁷ Production, cost, or profit functions have found interesting applications to the estimation of some ecological benefits. Probabilistic models of new product discovery from among diverse collections of natural organisms can also be regarded as a type of “production.” Finally, work in ecology points to “productive” relationships among natural systems that may yield insights to economists as well.

**Considerations in evaluating and understanding production and cost functions**

The analyst should consider the following factors when estimating the values of environmental inputs into production:

**Data requirements and implications.** Estimating production, cost, or profit functions requires data on all inputs and/or their prices. Omitted variable bias is likely to arise absent such information and may motivate the choice of one form over another. Econometricians have typically preferred to estimate cost or profit functions. Data on prices are often more complete than are data on quantities, and prices are typically uncorrelated with unobserved conditions of production, whereas input quantities are not.

**The model for estimation.** Standard practice involves the estimation of more flexible functional forms (i.e., functions that can be regarded as second-order approximations to any production technology). The translog and generalized Leontief specifications are examples¹⁹⁸. Estimation often will be more efficient if a system of equations is estimated (e.g., simultaneous estimation of a cost function and its associated factor demand equations), although data limitations may impose constraints.

**Market imperfections.** Most analyses assume perfectly competitive behavior on the part of producers and input suppliers and assume an absence of other distortions. When these assumptions do not hold, the interpretation of welfare results becomes more challenging. While there is an extensive literature on the regulation of externalities under imperfect competition, originating with Buchanan (1969), analysts should exercise caution and restraint in attempting to correct for departures from competitive behavior. The issues can become quite complex and there is typically no direct evidence of the magnitude of the departures. In many circumstances it might reasonably be argued that departures from perfect competition are not of much

---

¹⁹⁷ See, for example, Price and Heberling (2018) and the studies reviewed therein on source water quality.

practical concern (Oates and Strassman 1984). Perhaps a more pressing concern in many instances will be the wedge between private and social welfare consequences that arise with taxation. An increase in the value of production occasioned by environmental improvement typically will be split between private producers and the general public through tax collection. The issues here also can become quite complex (see Parry et al. 1997), with interactions among taxes leading to sometimes surprising implications. While it is difficult to give general advice, analysts may wish to alert policy makers to the possibility that the benefits of environmental improvements in production may accrue to different constituencies.

### 7.3.1.2 Travel Costs

Recreational values associated with an environmental improvement constitute a potentially large class of use benefits (see Table 7.1 for examples). However, measuring these values is complicated by the fact that the full benefits of recreation activities are rarely reflected in the price to access them. Travel cost models address this problem by inferring the value of changes in environmental quality through observing the trade-offs recreators make between environmental quality and the cost of visiting sites. A heuristic example is choosing between visiting a nearby recreation location with low environmental quality versus a more distant location with higher environmental quality. The outcome of the decision of whether to incur the additional travel cost to visit the location with higher environmental quality reveals information about the recreator’s WTP for environmental quality. Travel cost models are often referred to as recreation demand models because they are most often used to value the availability or quality of recreational opportunities.

**Economic foundation of travel cost models**

Travel cost models of recreation demand focus on the choice of the number of trips to a given site or set of sites that a traveler makes for recreational purposes. Because there is generally no explicit market or price for recreation trips, travel cost models rely on the assumption that the “price” of a recreational trip is equal to the cost of visiting the site. These costs include both participants’ monetary cost and opportunity cost of time. Monetary costs include all travel expenses. For example, when modeling day trips taken primarily in private automobiles, travel expenses would include roundtrip travel distance in miles multiplied by an estimate of the average cost per mile of operating a vehicle, plus any tolls, parking, and admission fees.

A participant’s opportunity cost of time for a recreational day trip is the value of the participant’s time spent traveling to and from the recreation site plus the time spent recreating. Although estimates of the opportunity cost of time ranging from zero to more than 100 percent of wage rates can be found in the literature, time spent traveling for recreational purposes is generally valued at some fraction of an individual’s full wage rate. The fraction of the wage rate used is important because it directly affects estimates of willingness to pay. As the fraction of the wage rate assumed to represent the opportunity cost of time rises, it causes total travel cost estimates to rise, which in turn cause estimates of willingness to pay to also rise.

Most commonly, researchers have used one third of a person’s annual hourly wage as an estimate of participants’ hourly opportunity cost of time, although estimates of two thirds of the full wage rate can also be found (Parsons 2003b, English, Leggett, and McConnell 2015, Phaneuf and Requate 2017). Within that range, the U.S. Department of Transportation guidance recommends valuing recreational travel at 50 percent of the hourly median household income for local travel and 70 percent for intercity travel (U.S. DOT 2016). A number of researchers have also developed methods for estimating recreators’ opportunity cost of time endogenously, although no one method has yet been fully embraced in the literature. Unless compelling reasons for deviating from the standard wage rate assumptions are present, analysts should generally rely on the standard one third of the wage rate opportunity cost assumption when estimating recreation travel. Conducting analyses using one-half of the wage rate can also be justifiable in some cases.
when done in addition to the one third assumption as a way to check the sensitivity of estimates to opportunity cost assumptions.

Even among studies that use the same fraction of the wage rate to estimate the opportunity cost of time, care must still be taken in comparing estimates across studies. First, researchers in the literature vary in their use of personal or household income in calculating opportunity costs. Household income tends to be greater on average than personal income resulting in larger opportunity cost estimates. Second, when researchers do not have recreators’ self-reported incomes they have often used population median or average income levels. Average income is generally higher than median income because higher incomes in the tail of the distribution tend to pull the average up. Lastly, in cases where household income is used, opportunity cost estimates will depend on whether costs are assumed to accrue to adults and children or only to adults. The literature is not clear on the preferred choices specification of opportunity costs, so the analyst must use best professional judgement to decide what is best on a case by case basis.

Hourly opportunity costs are multiplied by round trip travel time and time on-site to calculate a person’s full opportunity cost of time. Total travel costs are the sum of monetary travel costs and full opportunity costs. Following the law of demand, as the cost of a trip increases the quantity of trips demanded generally falls, all else equal. This means that participants are more likely to visit a closer site than a site farther away.

While travel costs are the driving force of the model, they do not completely determine a participant’s choice of sites to visit. Site characteristics, such as parking, restrooms, or boat ramps; participant characteristics, such as age, income, experience, and work status; and environmental quality also can affect demand for sites. Changes in the measures of environmental quality are generally the focus of economic analyses done in support of the regulatory decision-making process. The identification and specification of the appropriate site and participant characteristics are generally determined by a combination of data availability, statistical tests, and the researcher’s best judgment. Ultimately, every recreation demand study strikes a compromise in defining sites and choice sets, balancing data needs and availability, costs, and time.  

**General application by type of travel cost model**

Travel cost models can logically be divided into two groups: single-site models and multiple-site models. Apart from the number of sites they address, the two types of models differ in several ways. The basic features of both model types are discussed below.

**Single-site models.** Single-site travel cost models examine recreators’ choices of how many trips to make to a specific site over a fixed period of time (generally a season or year). It is expected that the number of trips taken will increase as the cost of visiting the site decreases and/or as the benefits realized from visiting increase. The price of close substitute sites could also affect demand. Income and other participant characteristics act as demand curve shifters. For example, avid outdoor recreators (fishermen or birders for example) may be more likely to take more trips than non-avid recreators, all else equal. Most current single-site travel cost models are estimated using count data models because the dependent variable (number of trips taken to a site) is a non-negative integer. See Haab and McConnell (2003) and Parsons (2003a) for detailed discussions and examples of recreation demand count data models.

Single-site models are most commonly used to estimate the value of a change in access to a site, particularly site closures (e.g., the closure of a lake due to unhealthy water quality). The lost access value due to a site closure is given by the area under the demand curve between a participant’s current price and the price at which trip demand falls to zero. Estimating the value of a change in the cost of a site visit, for example the addition or increase of an admission fee, is possible but more complex.
fee, is another common application of the model. Although it is possible with alternative data and model structures, single site models are not generally used for valuing changes in site quality.

A weakness of the single-site model is its inability to deal with large numbers of substitute sites. If, as is often the case, a policy affects several recreation sites in a region, then traditional single-site models are required for each site. In cases with large numbers of sites, defining the appropriate substitute sites for each participant and estimating individual models for each site can impose overwhelming data collection and computational costs. Because of these difficulties, most researchers have opted to refrain from using single-site models when examining situations with large numbers of substitute sites.\textsuperscript{202}

**Multiple-site models.** Multiple-site models examine a recreator’s choice of which site to visit from a set of available sites (known as the choice set) on a given choice occasion and, in some cases, can also examine how many trips to make to each specific site over a fixed period of time. Compared to the single-site model, the strength of multiple-site models lies in their ability to account for the availability and characteristics of substitute sites. By examining how recreators trade the differing levels of each site characteristic and travel costs when choosing among sites, it is possible to place a per trip (or choice occasion) dollar value on site attributes or site availability for single sites or multiple sites simultaneously.

The two most common multiple-site models are the random utility maximization (RUM) travel cost model and Kuhn-Tucker (KT) system of demand models. Both models may be described by a similar utility theoretic foundation, but they differ in important ways. In particular, the RUM model is a choice occasion model while the KT model is a model of seasonal demand.

**Random utility maximization models.**\textsuperscript{203} In a RUM model each alternative in the recreator’s choice set is assumed to provide the recreator with a given level of utility, and on any given choice occasion the recreator is assumed to choose the alternative that provides the highest level of utility on that choice occasion.\textsuperscript{204} The attributes of each of the available alternatives, such as the amenities, environmental quality, and the travel costs, are assumed to affect the utility of choosing each alternative. Because people generally do not choose to recreate at every opportunity, a non-participation option is often included as a potential alternative.\textsuperscript{205} From the researcher’s perspective, the observable components of utility enter the recreator’s assumed utility function. The unobservable portions of utility are captured by an error term whose assumed distribution gives rise to different model structures. Assuming that error terms have a type 1 extreme value distribution leads to the closed form logit probability expression and allows for maximum likelihood estimation of utility function parameters. Using these estimated parameters, it is then possible to estimate WTP for a given change in sites’ quality or availability.

\textsuperscript{202} Researchers have developed methods to extend the single-site travel cost model to multiple sites. These variations usually involve estimating a system of demand equations. See Bockstael, McConnell, and Strand (1991) and Shonkwiler (1999) for more discussion and other examples of extensions of the single-site model.

\textsuperscript{203} English (2018) is a highly scrutinized RUM model conducted for the damage assessment following the Deepwater Horizon spill and demonstrates a number of sensitivity analyses. Additionally, the public archive for the case contains a wealth of information. See www.doi.gov/deepwaterhorizon/adminrecord under the heading “5.10 Lost Human Use;” see in particular section 5.10.4 for technical reports discussing issues surrounding RUM estimation.

\textsuperscript{204} While the standard logit recreation demand model treats each choice occasion as an independent event, the model can also be generalized to account for repeated choices by an individual.

\textsuperscript{205} In a standard nested logit RUM model, recreators are commonly assumed to first decide whether or not to take a trip, and then conditional on taking a trip, to next choose which site to visit. By not including a non-participation option, the researcher in effect assumes that the recreator has already decided to take a trip, or in other words, that the utility of taking a trip is higher than the utility of doing something else for that choice occasion. In other words, models lacking a participation decision only estimate the recreation values of the segment of the population that participates in recreation activities (i.e., recreators), while models that allow for non-participation incorporate the recreation values of the whole population (i.e., recreators and non-recreators combined). Because of this, recreation demand models without participation decisions tend to predict larger per person welfare changes than models allowing non-participation.
However, because the RUM model examines recreation decisions on a choice occasion level, it is less suited for predicting the number of trips over a time period and measuring seasonal welfare changes. A number of approaches have been used to link the RUM model’s estimates of values per choice occasion to estimates of seasonal participation rates. See Parsons, Jakus, and Tomasi (1999) for a detailed discussion.

The nested logit and mixed logit models are extensions of the basic logit model. The nested logit model groups similar alternatives into nests where alternatives within a nest are more similar to each other than they are to alternatives outside of the nest. In very general terms, recreators are first assumed to choose a nest and then, conditional on the choice of nest, they then choose an alternative within that nest. Nesting similar alternatives allows for more realistic substitution patterns among sites than is possible with a basic logit. The mixed logit is a random parameter logit model that allows for even more flexible substitution patterns by estimating the variation in preferences (or correlation in errors) across the sample. If preferences do not vary across the sample, then the mixed logit collapses to a basic logit.206

The Kuhn-Tucker (KT) model. The KT model is a seasonal demand model that estimates recreators’ choice of which sites to visit (like a multiple-site model) and how often to visit them over a season (like a single-site model). The model is built on the theory that people maximize their seasonal utility subject to their budget constraint by purchasing the quantities of recreation and other goods that give them the greatest overall utility. Similar to the RUM model, the researcher begins by specifying the recreator’s utility function. Taking the derivative of this utility function with respect to the number of trips taken, subject to a budget and non-negative trip constraint, yields the “Kuhn-Tucker” conditions. The KT conditions show that trips will be purchased up to the point that the marginal rate of substitution between trips and other spending is equal to the ratio of their prices. In cases where the price of a trip exceeds its marginal value none will be purchased. Given assumptions on the form of the utility function and the distribution of the error term, probability expressions can be derived, and parameter estimates may then be recovered. While recent applications have shown that the KT model is capable of accommodating a large number of substitute sites (e.g., von Haefen et al. 2004), the model is computationally intensive compared to RUM models.207

Considerations in evaluating and understanding recreation demand studies

Definition of a site and the choice set. The definition of what constitutes a unique site has been shown to have a significant effect on estimation results. Ideally, one could estimate a recreation demand model in which sites are defined as specific points such as exact fishing location, campsites, etc. The more exact the site definition, the more exact the measure of travel costs and site attributes, and therefore WTP, that can be calculated. However, in situations with a large number of potential alternatives, the large data requirements may be cost and time prohibitive, estimation may be problematic, and aggregation may be required. The method of aggregation has been shown to have a significant effect on estimated values. The direction of the effect will depend on the situation being evaluated and the method of aggregation chosen (Parsons and Needleman 1992; Feather 1994; Kaoru, Smith, and Liu 1995; and Parsons, Plantinga, and Boyle 2000).

In addition to the definition of what constitutes a site, the number of sites included in a recreator’s choice set can have a significant effect on estimated values. When defining choice sets, the most common practice in the literature has been to include all possible alternatives available to the recreator. In many cases availability has been defined by location within a given distance or travel time.208 This strategy has been criticized on the grounds that people may not know about all possible sites, or even if they do know they exist they may not seriously consider them as alternatives. In response to this, a number of researchers have suggested methods that either restrict choice sets to include only those

---


207 For a basic application of the KT model see Phaneuf and Siderelis (2003). For more advanced treatments of the models see Phaneuf, Kling, and Herriges (2000), and von Haefen and Phaneuf (2005).

208 Parsons and Hauber (1998) explore the implication of this strategy by expanding the choice set geographically and find that beyond some threshold the effect of additional sites is negligible.
sites that the recreators seriously consider visiting (e.g., Peters et al. 1995, and Haab and Hicks 1997) or that weight seriously-considered alternatives more heavily than less-seriously-considered alternatives (e.g., Parsons, Massey, and Tomasi 2000).

**Multiple-site or multipurpose trips.** Recreation demand models assume that the particular recreation activity being studied is the sole purpose for a given trip. If a trip has more than one purpose, it almost certainly violates the travel cost model's central assumption that the “price" of a visit is equal to the travel cost. The common strategy for dealing with multipurpose trips is simply to exclude them from the data used in estimation.\(^2\) See Mendelsohn et al. (1992) and Parsons (2003b) for further discussion.

**Day trips versus multi-day trips.** The recreation demand literature has focused almost exclusively on single-day trip recreation choices. Adding the option to stay longer than one day adds another choice variable in the estimation, thereby greatly increasing estimation difficulty. Also, as trip length increases multipurpose trips become increasingly more likely, again casting doubt on the assumption that a trip's travel costs represent the “price" of one single activity. A few researchers have estimated models that allow for varying trip length. The most common strategy has been to estimate a nested logit model in which each choice nest represents a different trip length option. See Kaoru (1995) and Shaw and Ozog (1999) for examples. The few multi-day trip models in the literature find that the per-day value of multi-day trips is generally less than the value of a single-day trip, which suggests that estimating the value of multi-day trips by multiplying a value estimated for single-day trips value by the number of days of will overestimate the multi-day trip value.

### 7.3.1.3 Hedonic Models

Hedonic pricing models use statistical methods to measure the contribution of a good’s characteristics to its price. These models are applicable to goods that can be thought of as "bundling" together many attributes that vary in quantity and quality. Houses differ in size, layout, location, and exposure to environmental contaminants. Labor hours can be thought of as "goods" differing in attributes like safety risks and supervisory nature that should be reflected in wages. Hedonic pricing models use variation in prices of such goods to estimate the value of these attributes.

**Economic foundations of hedonic models**

Hedonic pricing studies estimate economic value by weighing the advantages against the costs of different choices. A standard assumption underlying hedonic pricing models is that markets are in equilibrium, which means that no individual can improve her welfare by choosing a different home or job. For example, if an individual changed location she might move to a larger house, or one in a cleaner environment. However, to receive such amenities, the individual must pay for a more expensive house and incur transaction costs to move. The more the individual spends on her house, the less she has to spend on food, clothing, transportation, and all the other things she wants or needs. So, if the difference in prices paid to live in a cleaner neighborhood is observable, then that price difference can be interpreted as the WTP for a better environment.

One key requirement in conducting a hedonic pricing study is that the available options differ in measurable ways. To see why, suppose that all locations in a city’s housing market are equally polluted or all jobs in a labor market cause workers to be exposed to the same risks. The premium that homeowners place on environmental quality or that workers place on lower occupational risks could not be measured in this case. A hedonic pricing study requires a comparison to purchases of more expensive houses in less polluted neighborhoods, or to wages in lower-paying but safer jobs. However, there is also a practical limit on the heterogeneity of the sample. Workers in different countries earn very different wages and face very different job risks, but this does not mean it is possible to value the difference in job risks by reference to

---

\(^2\) Excluding any type or class of trip (like multiple-site or multipurpose) will produce an underestimate of the population's total use value of a site. The amount by which benefits will be underestimated will depend on the number and type of trips excluded.
international differences in wages. This is because: (1) there are many other factors that differ between widely separated markets; and (2) people are not very mobile between disparate sites. For these reasons it is important to exercise care in defining the market in which choices are made.\footnote{Michaels and Smith (1990) offer guidance for defining the extent of the market.}

A related issue is that only environmental attributes or health risks that market participants are aware of and understand can be valued using hedonic pricing methods (and revealed preference methods more generally). If homeowners are unable to recognize differences in health outcomes, visibility, and other consequences of differences in environmental quality at different locations, or if workers are unaware of differences in risks at different jobs, then a hedonic pricing study would not be suitable for estimating the values for those attributes. For example, groundwater contamination—such as from leaking underground storage tanks—can be difficult for homeowners to detect (Zabel and Guignet 2012). In contrast, stated preference surveys can directly ask respondents how they value changes in specific environmental commodities or health risks.

**General application by type of hedonic pricing study**

**Hedonic wage studies**, also known as wage-risk or compensating wage studies, are based on the premise that individuals make trade-offs between wages and occupational risks of death or injury. Hedonic wage studies can assess the value to workers of changes in workplace morbidity and mortality risks, which may then be applied to environmental risks using benefit-transfer techniques. Viscusi (2013) provides an overview of the method. Most current hedonic wage studies begin with estimation of the risk, calculated as workplace fatalities per worker. The Bureau of Labor Statistics (BLS) Census of Fatal Occupational Injuries (CFOI) is the most common source for workplace risk information, a complete record of US workplace fatalities since 1992. CFOI reports these fatalities by 3-digit occupation and 4-digit industry classifications, as well as the circumstances of the fatal events.\footnote{More information on the CFOI data is available at: http://www.bls.gov/iif/oshfat1.htm.} Typically, these data are used to construct the number of annual fatalities within categories such as a given industry, occupation, or industry-occupation cell. This is the numerator for the annual risk rate for the hedonic wage study. Other data sources, most commonly from the Current Population Survey also conducted by BLS, are used to estimate the number of workers in these categories, providing the denominator for the annual risk rate, as well as characteristics of workers, including wage rate. The estimating equation then uses the wage rate as the dependent variable, usually in a linear or semi-log specification, and the coefficient on the risk measure provides the basis for the implicit wage-risk tradeoff for mortality risk valuation.

There are questions about the applicability of hedonic wage study results to environmental benefits assessment. For example, hedonic wage estimates are derived from populations that are working age and able to work, and they reflect preferences of those who have chosen relatively risky professions. These characteristics may differ from populations affected by environmental contaminants. There is also a difference in risk context between fatal workplace accidents and environment-related mortality from, for example, cancer. Still, hedonic wage studies have been widely used to estimate the value of fatal risk reductions, because they provide revealed-preference information on how people trade off risks for money.\footnote{For example, the EPA’s SAB has recognized the limitations of these estimates for use in estimating the benefits of reduced cancer incidence from environmental exposure. Despite these limitations, however, the SAB concluded that these estimates were the best available at the time (U.S. EPA 2000d).} Historically, the EPA has used a VSL estimate primarily derived from hedonic wage studies. For more information on the Agency’s VSL estimate, see Appendix B.

**Hedonic property value studies** measure the contributions of various characteristics to property prices. These studies are typically conducted using residential housing data, but they have also been applied to commercial and industrial property, agricultural land, and vacant land. Bartik (1988) and Palmquist (1988, 1991) provide detailed discussions of benefits assessment using hedonic methods. Property value studies require large amounts of data. Market data on
individual housing units' prices and other attributes are strongly preferred to aggregate data such as census tract median home values and characteristics.

Hedonic property value studies have examined the effects of air quality (e.g., Smith and Huang 1995, Bishop and Timmins 2018), water quality (e.g., Leggett and Bockstael 2000, Walsh et al. 2017, Guignet et al. 2019), natural amenities (e.g., Landry and Hindsley 2011, Guignet et al. 2017), and land contamination (e.g., Messer et al. 2006, Guignet 2013, Walsh and Mui 2017) on property values. As discussed in section 7.1, the hedonic property approach can value changes in stressors, contaminant releases or media concentrations, or other intermediate endpoints linked to environmental benefits. The type of environmental amenity included in the analysis is often driven by data availability.

Other hedonic studies. Applicability of the hedonic pricing method is not limited to the property and labor markets. For example, hedonic pricing methods can be combined with travel cost methods to examine the implicit price of recreation site characteristics (Brown and Mendelsohn 1984, Phaneuf et al. 2008, Kovacs 2012). Hedonic analysis of light-duty vehicles has been used to examine the trade-offs between automobile price, fuel economy, and other features (Espey and Nair 2005; Fan and Rubin 2010).

Considerations in evaluating and understanding hedonic pricing studies

There are numerous statistical issues associated with applying hedonic pricing models to value changes in environmental quality and health risks. Below, we highlight the issues of identification and causality, spatial correlation, defining and measuring the environmental amenity, and interpretation of the estimates.

Identification and causality. A common challenge in hedonic pricing studies is establishing the direction of causality between the independent variable of interest (environmental quality or safety risks) and the dependent variable (e.g., home prices or wages). People choose among different houses not only because they can trade off differences in environmental amenities against price, but also because of other characteristics, like curb appeal, school quality, and crime. If these other characteristics are not included in the hedonic regression and are correlated with environmental quality, then the analysis may not identify the causal impact of environmental quality on prices. In this situation, endogeneity or omitted variable bias would lead to incorrect estimates of the value of environmental quality to home buyers (Taylor, Phaneuf, Liu 2016). Similarly, if the risk of accidental death is correlated with the risk of serious, nonfatal injuries, the premium estimated in a hedonic wage equation would overstate WTP for reductions in mortality if these other risks were omitted from the regression.

Approaches to identify causal effects in the hedonic property value literature include repeat sales models, which can identify the effect of changes in environmental quality over time using the sample of homes that sold multiple times during the study period, and quasi-experimental approaches, which rely on "natural experiments" in which environmental quality varies for reasons that are exogenous to home prices. Quasi-experimental approaches include instrumental variables, regression discontinuity, and difference-n-difference models (e.g., Greenstone and Gayer 2009; Greenstone and Gallagher 2008; Gamper-Rabindran and Timmins 2013). Spatial fixed effects denoting discrete geographic units such as Census tracts or counties can also help control for difficult-to-measure local characteristics, but environmental quality must vary within this spatial unit for these models to yield useful valuation estimates (Abbott and Klaiber 2010). While fixed effects alone may not mitigate omitted variable bias if unobserved characteristics correlated with environmental quality also vary within the Census tract or other spatial unit, research has found that a combination of spatial fixed effects, quasi-experimental identification, and temporal controls can greatly reduce bias (Kuminoff et al. 2010).

Spatial correlation. Some property characteristics depend on neighborhood or geographic features that vary over space. Excluding these characteristics from the econometric model may cause dependence across the error terms of the model. There may also be spatial correlation in the dependent variable of the model if home prices are directly affected by the prices of nearby homes (for example, due to the home appraisal process). Spatial econometrics techniques allow
Defining and measuring the environmental amenity. Another important issue is the way that the environmental amenity or health risk included in a hedonic model is defined and measured. The ideal measure is an indicator that market participants value and that can be linked to a change in environmental policy, but such measures are not always available. For instance, available water quality indicators may not fully reflect water quality or ecosystem health (Griffiths et al., 2012). Water clarity has been shown to positively affect property prices (Michael et al., 2000, Gibbs et al., 2002, Walsh et al., 2017), but it is not always a good indicator of ecosystem health (Shaw, Mechenich, and Klessig 2004). Furthermore, data on water clarity may contain errors because clarity cannot be accurately measured under cloud cover (Olmanson, Bauer and Brezonik 2008). If water clarity is measured with error or is not a good proxy for home buyers’ perceptions of water quality, then measurement error could produce valuation estimates that are biased toward zero due to attenuation (Greene 2000), though empirical research has found that objective measures of water clarity have higher predictive power than individuals’ subjective measure of water clarity (Poor et al. 2001).

Interpretation of the estimates. Understanding how to interpret hedonic model estimates is important. As with many results in economics, hedonic pricing models are best suited to the valuation of small, or marginal, changes in attributes. Under such circumstances, the slope of the hedonic price function can be interpreted as WTP for a small change. Hedonic price functions typically reflect equilibria between consumer demands and producer supplies for fixed levels of the attributes being evaluated. Public policy, however, is sometimes geared to larger, discrete changes in attributes. When this is the case, using hedonic model estimates to calculate benefits is more complicated. Palmquist (1991) describes conditions under which exact welfare measures can be calculated for discrete changes. See Freeman et al. (2014) and Ekeland, Heckman, and Nesheim (2004) for treatments.

Studies that compare prices before and after a change in environmental quality using repeat sales and quasi-experimental approaches raise particular challenges for interpretation (Klaiber and Smith 2013). These approaches are sometimes called "capitalization" rather than hedonic studies because they estimate the extent to which changes in amenities are capitalized into prices over time. The capitalization effect only equals WTP if WTP remains stable over the study time horizon (Kuminoff and Pope 2010). If marginal WTP for environmental quality is increasing (decreasing) over time in the study area, then the capitalization estimate will tend to overestimate (underestimate) the benefits of cleanup.

In property value studies, if gentrification or re-sorting occurs such that people with a higher WTP move to neighborhoods with improving environmental quality, pre- and post-cleanup housing prices reflect the preferences of two distinct groups of people. In addition, the capitalization estimate from repeat sales and quasi-experimental models represents the average rather than marginal change in property values that occurs in response to a change in an amenity (Parmeter and Pope 2013). If residents do not re-sort, their preferences and incomes are not changing over time, and WTP is linear in environmental quality, then a capitalization estimate can be interpreted as a measure of WTP. These conditions are less likely to hold in a study that examines a large change in environmental quality over a relatively long timespan. For example, Parmeter and Pope (2013) argue that Chay and Greenstone’s (2005) quasi-experimental study of the housing price effects of improvements in air quality in nonattainment counties after passage of the 1970 Clean Air Act Amendments provides a capitalization rather than a WTP estimate because of the 10-year timespan of the study and the non-marginal reduction in air pollution. However, the assumption that a capitalization estimate provides a

---

213 Residential sorting models provide another alternative to hedonic and capitalization studies in the property value literature. These models derive estimates of WTP explicitly accounting for residential sorting behavior and resulting changes in a variety of neighborhood amenities (e.g., Klaiber and Phaneuf 2010; Kuminoff and Jarah 2010).
good approximation of WTP might be reasonable for studies covering relatively short periods of time and examining small changes in environmental quality.

### 7.3.1.4 Averting Behaviors

The averting behavior method infers values for environmental quality from observations of actions people take to avoid or mitigate the increased health risks or other undesirable consequences of reductions in environmental quality. Examples of such defensive actions can include the purchase and use of air filters, the activity of boiling water prior to drinking it, and the purchase of preventive medical care or treatment. By analyzing the expenditures associated with these averting behaviors, economists can attempt to estimate the value individuals place on small changes in risk or environmental quality. Dickie (2017) provides a detailed overview of the approach.

#### Economic foundations of averting behavior methods

Averting behavior methods can be best understood from the perspective of a household production framework. Households can be thought of as producing health outcomes by combining an exogenous level of environmental quality with inputs such as purchases of goods that involve protection against health and safety risks (Freeman et al. 2014; Dickie 2017). To the extent that averting behaviors are available, the model assumes that a person will continue to take protective action as long as the expected benefit exceeds the cost of doing so. If there is a continuous relationship between defensive actions and reductions in health risks, then the individual will continue to avert until the marginal cost just equals her marginal WTP for these reductions. Thus, the value of a small change in health risks can be estimated from two primary pieces of information: (1) The cost of the averting behavior or good, and (2) its effectiveness, as perceived by the individual, in offsetting the loss in environmental quality.

One approach to estimation is to use observable expenditures on averting and mitigating activities to generate values that may be interpreted as a lower bound on WTP. As noted earlier in section 7.2.1.2, WTP for small changes in environmental quality can be expressed as the sum of the values of four components: changes in averting expenditures, changes in mitigating expenditures, lost time, and the loss of utility from pain and suffering. The first three terms of this expression are observable, in principle, and can be approximated by calculating changes in these costs after a change in environmental quality. The resulting estimate can be interpreted as a lower bound on WTP that may be used in benefits analysis (Shogren and Crocker 1991; Quiggin 1992).

#### General application of averting behavior method

Although the first applications of the averting behavior method estimated the benefits of reduced soiling of materials from environmental quality changes (Harford 1984), recent research has primarily focused on health risk changes. Conceptually, the averting behavior method can provide WTP estimates for a variety of other environmental benefits such as damages to ecological systems and materials.

Some averting behavior studies focus on behaviors that prevent or mitigate the impact of specific symptoms (e.g., shortness of breath or headaches), while others have examined averting expenditures in response to specific episodes of contamination (e.g., groundwater contamination). Because many contaminants can produce similar symptoms, studies that estimate values for symptoms may be more amenable to benefit transfer than those are episode-specific and don’t value specific symptoms or illnesses. The latter could potentially be more useful, however, for assessing the benefits of a regulation expected to reduce the probability of similar contamination episodes.

#### Considerations in evaluating and understanding averting behavior studies

**Perceived versus actual risks.** As in other revealed preference methods, analysts should remember that consumers base their actions on perceived benefits from their behaviors. Many averting behavior studies explicitly acknowledge that their estimates rest on consistency between the consumer’s perception of risk reduction and actual risk reduction. While there is some evidence that consumers are rational regarding risk — for example, consumer risk-reduction
expenditures increase as risk increases — there is also evidence that there are predictable differences between consumers’ perceptions and actual risks. For example, individuals tend to overestimate risks that are very small or that are novel or unfamiliar (Renner, et al. 2015). Thus, averting behavior studies can produce biased WTP estimates for a given change in objective risk. Surveys may be necessary to determine the benefits individuals perceive they are receiving when engaging in defensive activities. These perceived benefits can then be used as the object of the valuation estimates. For example, if surveys reveal that perceived risks are lower than expert risk estimates, then WTP for risk reduction can be estimated with the lower, perceived risk (Blomquist 2004).

**Data requirements and implications.** Data needed for averting behavior studies include information detailing the effect(s) being averted (e.g., specific illnesses, exposure to environmental contaminants); actions taken to avert or mitigate damages; the costs of those behaviors and activities; and other variables that affect health outcomes, like age, health status, or chronic conditions.

**Separability of joint benefits.** Analysts should exercise caution in interpreting the results of studies that focus on goods in which there may be significant joint benefits (or costs). Many defensive behaviors not only avert or mitigate environmental damages, but also provide other benefits. For example, air conditioners obviously provide cooling in addition to air filtering, and bottled water may not only reduce health risks, but also taste better. Conversely, it also is possible that the averting behavior may have negative effects on utility. For example, wearing helmets when riding bicycles or motorcycles may be uncomfortable. Failure to account for these “joint” benefits and costs associated with averting behaviors will result in biased estimates of WTP.

**Modeling assumptions.** Restrictive assumptions are sometimes needed to make averting behavior models tractable. Analysts drawing upon averting behavior studies will need to review and assess the implications of these assumptions for the valuation estimates.

### 7.3.1.5 Cost of Illness

A frequent alternative to WTP estimates is the avoided cost of illness (COI), which estimates the resource costs associated with an adverse health effect (an “illness”). Though WTP is the preferred valuation measure for BCA, the COI method is straightforward to implement and explain to policy makers, and has been widely applied, particularly in health economics or studies of the burden of disease (e.g., Trasande, et al. 2016). COI estimates for many illnesses are readily available from existing studies. COI estimates are usually less expensive to develop than WTP estimates using stated or revealed preference approaches, so it may be feasible to develop new COI estimates for a given benefits analysis even with typical time and resource constraints. Jo (2014) and Tarricone (2006) provide overviews of the method.

**Economic foundations of COI studies**

**Relating cost of illness to WTP.** The COI method does not estimate WTP but is based on estimating the market value of goods and services used to treat illness and the lost productivity due to that illness. It does not incorporate any disutility from illness, the costs of averting behaviors taken to avoid the illness, or risk preferences that would be inherent in estimates of WTP to reduce risks of illness. Because of these limitations COI is best viewed as a proxy for WTP when WTP estimates are not available, and is generally considered to be a lower bound on WTP, especially for more serious illnesses.\(^\text{214}\) Available comparisons of COI and total WTP estimates suggest that the difference can be large but varies greatly across health effects; COI estimates cannot be simply "scaled up" to approximate WTP.

In some cases, COI may be additive to a WTP estimate that did not account for certain costs. COI estimates capture medical expenses passed on to third parties such as health insurance companies and hospitals, whereas WTP estimates

\(^{214}\) However, any particular COI estimate is not necessarily going to be lower than WTP for a given health condition. Depending on the design of the studies, WTP could reflect avoidance costs that are lower than the cost of illness once the illness has been contracted.
generally do not. COI estimates can capture the value of lost productivity, something that may be overlooked in WTP estimates — especially when derived from consumers or employees covered by sick leave.

**Types of costs considered in COI studies.** COI studies generally distinguish between direct costs (costs related to medical treatment) and indirect costs (costs related to lost productivity). Many COI studies estimate both direct and indirect costs, but some may focus solely on costs of treatment while other studies, broadly categorized as COI, only reflect lost productivity.

- **Direct Costs** are those related to treatment and care for the illness. These costs include the value of goods and services spent for items such as physicians' services, testing, hospitalization, medications, and medical devices. But it also includes the value of household expenditures, transportation, accommodation, and other resources spent on care for the illness. COI studies may not capture all of these costs. For example, studies relying solely on databases of medical expenditures might not capture the costs of household expenditures.

- **Indirect costs** refer to productivity losses associated with the illness, most often measured by the human capital approach where earnings reflect the value of productive time. That is, assuming the wage equals the value of marginal product. Losses to productivity, therefore, are a social cost and can be measured by the wage rate.\(^\text{215}\) Lost productivity may be focused on the short-term, e.g., for illnesses where the losses are associated with a loss of work days or, for more serious illnesses a permanent loss of income.

In principle, indirect costs should also consider the costs of lost home productivity and the value of leisure, but this is not always done in COI studies. Lost productivity for home health care, e.g., the time spent by members of the household in caring for family members or accompanying patients to medical appointments, should also be included in indirect costs.

Note that the human capital approach applies not just to lost work time at a given wage in what would be considered a traditional COI study but for any impact on productivity associated with adverse health effects. For example, lowered IQ – an effect associated with exposures to many pollutants – has been related to labor participation and lower lifetime earnings, a loss of human capital (Salkever 1995; Lin et al. 2018). This relationship can be useful in economic analyses to value the benefits of avoiding IQ losses. Additionally, exposure to ozone has been linked to loss of productivity among agricultural workers (Crocker and Horst 1981; Graff Zivin and Neidell 2012).\(^\text{216}\)

**General application by type of COI study**

**Prevalence-based estimates.** Prevalence-based COI estimates are derived from the costs faced by all individuals who have a sickness in a specified time period. For example, an estimate of the total number of individuals who currently have asthma, as diagnosed by a physician, reflects the current prevalence of physician-diagnosed asthma. Prevalence-based COI estimates for asthma include all direct and indirect costs associated with asthma within a given time period, such as a year. Prevalence-based COI estimates are a measure of the financial burden of a disease but will still generally underestimate total WTP for avoiding the disease altogether. They are most applicable for valuation of policies that reduce or eliminate morbidity associated with existing cases of illness.

**Incidence-based estimates.** By contrast, incidence-based COI estimates reflect expected costs for new cases of an illness in a given time period. For example, the number of individuals who receive a new diagnosis of asthma from a physician in a year reflects the annual incidence of physician-diagnosed asthma. Incidence-based COI estimates reflect the expected

\(^{215}\) The EPA has a similar approach for cost analysis that is also based on the opportunity cost of time: U.S. EPA “Valuing Time Use Changes Induced by Regulatory Costs and Other EPA Actions” [currently undergoing peer-review].

\(^{216}\) For examples of how productivity estimates have been used in economic analyses, see the primary benefits analysis for the 2011 Transport Rule as well as the supplemental benefits analysis for the 2015 Ozone NAAQS.
value of direct medical expenditures and lost income and productivity associated with a disease from the time of
diagnosis until recovery or death. Because these expenses can occur over an extended time period, incidence-based
estimates should be discounted to the year the illness is diagnosed and expressed in present value terms. Incidence-based
COI estimates are most applicable for valuation of policies that reduce the expected number of new cases of disease,
which is often the case for environmental regulations.

**Bottom-up, top-down, and econometric approaches**. There are three primary methods for estimating COI for a given
health condition. The "bottom up" approach constructs a typical profile of treatment for the condition and then uses
unit costs to estimate total treatment costs over time, usually based on databases of medical expenditures. The "top-
down" approach, on the other hand, typically starts with aggregate expenditures across a number of illnesses and then
attributes these expenditures across that set of illnesses. Finally, the econometric approach to COI typically uses data on
total costs for a given sample over a given time period and then econometrically estimates the difference in costs
between those with and without a given health condition. The difference provides an estimate of the cost of treatment
for the illness. Bottom-up or econometric approaches are generally best-suited for benefits analysis.

**Considerations in evaluating and understanding COI studies**

**Technological change.** Medical treatment technologies and methods are constantly changing, and this could push the
true cost estimate for a given illness either higher or lower. When using previous COI studies, the analyst should be sure
to research whether and how the generally accepted treatment has changed from the time of the study.

**Measuring health care costs.** The COI literature uses a variety of methods to measure health care costs.¹²¹ One
important distinction is between medical expenditures and medical charges. Expenditures are the better indicator of
social costs because they better represent actual resources used by healthcare providers rather than the "list price,"
which is often discounted. Studies that rely on medical charges may use them as-is or try to approximate expenditures
using hospital-specific cost-to-charge ratios. For benefits analysis adjusted charges are better than unadjusted charges,
but studies that use expenditure data are even more preferred.

**Measuring the value of lost productivity.** The value of lost productivity in many studies may only reflect persons in the
work force, omitting the productivity costs of those persons not involved in paid jobs. Homemakers’ household upkeep
and childcare services, retired persons’ volunteering efforts, and students’ time in school all directly or indirectly
contribute to the productivity of society. In cases where an affected individual requires a caregivers’ assistance, e.g.,
when children, elderly or impaired individuals are affected, the caregiver may also incur time away from work and lost
productivity. The value of lost leisure time to an individual and their family is not included in most COI studies. A second
set of considerations is the choice of wage rate in the study, which will reflect the study population and may not match
the wage rate of the population in the policy case.

**7.3.2 Stated Preference**

The distinguishing feature of stated preference methods compared to revealed preference methods is that stated
preference methods rely on people’s responses to hypothetical questions while revealed preference methods rely on
observations of actual choices. Stated preference methods use surveys that ask respondents to consider one or a series
of hypothetical scenarios that describe a potential change in a non-market good. The advantages of stated preference
methods include their ability to estimate non-use values and to incorporate hypothetical scenarios that closely
correspond to a policy case. The main disadvantage of stated preference methods is that they may be subject to
systematic biases that are difficult to test for and correct.

*The Report of the NOAA Panel on Contingent Valuation* is often cited as a primary source of recommendations for best
practices for stated preference studies. Often referred to as the “NOAA Blue Ribbon Panel,” this panel, comprised of five

---

distinguished economists including two Nobel Laureates, deliberated on the usefulness of stated preference studies for policy analysis for the National Oceanic and Atmospheric Administration (NOAA) (Arrow et al. 1993). The panel focused on a rather narrow application of stated preference - the use of contingent valuation to estimate non-use values for litigation in the US. In the years since, stated preference research has advanced significantly and its applications have expanded to more diverse contexts. In 2012 The Journal of Economic Perspectives published a symposium on contingent valuation including an assessment of the state of the science. More recently, Johnston et al. (2017) published an updated set of guidelines that reflects contemporary stated preference research, changes in survey methods and technology, and the transfer of primary estimates to different policy scenarios.

7.3.2.1 Economic Foundation of Stated Preference Methods

The role of non-use value in benefit cost analysis has been well established since the 1990s (see Kopp 1992, Bishop and Welsh 1992, Freeman 1993 for early discussions of non-use value and welfare theory). Further, ignoring non-use value in environmental regulatory analysis can lead to large omissions in benefits estimation and a misallocation of resources. A regulatory analysis should carefully consider when non-use values might be substantial and, given stated preference is the only valuation approach that captures them, what studies are available to draw from and how to evaluate the validity of their results.

The responses elicited from stated preference surveys, if truthful, unbiased and well-informed, are either direct expressions of WTP or can be used to estimate WTP for the good in question. However, the caveats listed above are paramount. While many environmental economists believe that respondents can provide truthful answers to hypothetical questions and therefore view stated preference methods as useful and reliable if conducted properly, a non-trivial fraction of economists are more skeptical of the results elicited from stated preference surveys. Due to this skepticism, it is important to employ validity and reliability tests of stated preference results before applying them to policy decisions.

If the analyst decides to conduct a stated preference survey or use stated preference results in a benefit transfer exercise, then a number of survey design issues should be considered. Stated preference researchers have attempted to develop methods to make individuals’ choices in stated preference studies as consistent as possible with market transactions or consequential referenda. Reasonable consistency with the framework of market transactions is a guiding criterion for ensuring the validity of stated preference value estimates. Three components of market transactions need to be constructed in stated preference surveys: the commodity, the payment, and the scenario (Fischoff and Furby 1988).

Stated preference studies need to carefully define the commodity to be valued, including characteristics of the commodity such as the timing of provision, certainty of provision, and availability of substitutes and complements. The definition of the commodity generally involves identifying and characterizing attributes of the commodity that are relevant to respondents. Commodity definition also includes defining or explaining baseline or current conditions, property rights in the baseline, the policy scenarios, as well as the source of the change in the environmental commodity. Respondents also must be informed about the transaction context, including the method, timing, and duration of payment. The transaction must not be coerced, and the individual should be aware of her budget constraint. The payment vehicle should be described as a credible and binding commitment should the respondent decide to purchase the good (Carson et al. 1997). The timing and duration of a payment involves individuals implicitly discounting payments and calculating expected utility for future events. The transaction context and the commodity definition should describe and account for these temporal issues.

218 Depending on the scenario, the description of the commodity may produce strong reactions in respondents and could introduce bias. In these cases, the detail with which the commodity of the change is specified needs to be balanced against the ultimate goals of the survey. Regardless, the commodity needs to be specified with enough detail to make the scenario credible.
The hypothetical scenario(s) should be described to minimize potential strategic behavior such as “free riding” or “yea-saying.” In the case of free riding, respondents will underbid their true WTP for a good if they believe it will be provided regardless of their response. In the case of yea-saying, respondents pledge amounts greater than their true WTP with the expectation that they will not be made to pay for the good but believing that their response could influence whether the good will be provided. Incentive-compatible choice scenarios and attribute-based response formats have been shown to mitigate strategic responses. Both are discussed below.

It is recognized in both the experimental economics and the survey methodology literatures that different survey formats can elicit different responses. Changing the wording or order of questions can also influence responses. Therefore, the researcher should provide a justification for her choice of survey format and include a discussion of the ramifications of that choice.

### 7.3.2.2 General Application by Type of Stated Preference Study

Two main types of stated preference survey format are currently used: direct WTP questions and stated choice questions. Stated choice questions can be either dichotomous choice questions or multi-attribute choice experiments. Because survey formats are still evolving, and many different approaches have been used in the literature, no definitive recommendations are offered here regarding selection of the survey format. Rather, the following sections describe some of the most commonly used formats and discuss some of their known and suspected strengths and weaknesses. Researchers should select a format that suits their topic, and should strive to use focus groups, pretests, and statistical validity tests to address known and suspected weaknesses in the selected approach.

#### Direct/open-ended WTP questions

Direct/open-ended WTP questions ask respondents to indicate their maximum WTP for the specific quantity or quality changes of a good or service that has been described to them. An important advantage of open-ended stated preference questions is that the answers provide direct, individual-specific estimates of WTP which requires a smaller sample size and simpler estimation approach. While these advantages could lower the cost of the study, early stated preference studies found that some respondents had difficulty answering open-ended WTP questions and non-response rates to such questions were high. Such problems are more common when the respondent is not familiar with the good or with the idea of exchanging a direct dollar payment for the good. An example of a stated preference study using open-ended questions is Brown et al. (1996).

Various modifications of the direct/open-ended WTP question format have been developed to help respondents arrive at their maximum WTP estimate. In iterative bidding respondents are asked if they would pay some initial amount, and then the amount is changed up or down depending on whether the respondent says “yes” or “no” to the first amount. This continues until a maximum WTP is determined for that respondent. Iterative bidding has been shown to suffer from “starting point bias,” wherein respondents’ maximum WTP estimates are systematically related to the dollar starting point in the iterative bidding process (Rowe and Chestnut 1983, Boyle et al. 1988, and Whitehead 2002). A payment card is a list of dollar amounts from which respondents can choose, allowing respondents an opportunity to look over a range of dollar amounts while they consider their maximum WTP. Mitchell and Carson (1989) and Rowe et al. (1996) discuss concerns that the range and intervals of the dollar amounts used in payment card methods may influence respondents’ WTP answers.

#### Stated choice questions

While direct/open-ended WTP questions are efficient in principle, researchers have generally turned to other stated preference techniques in recent years. This is largely due to difficulties respondents face in answering direct WTP questions and the lack of easily implemented procedures to mitigate these difficulties. Researchers also have noted that direct WTP questions with various forms of follow-up bidding may not be “incentive compatible.” That is, the respondents’ best strategy in answering these questions is not necessarily to be truthful (Freeman et al. 2014).
In contrast to direct/open-ended WTP questions, stated choice questions ask respondents to choose a single preferred option or to rank options from three or more choices. When analyzing the data, the dependent variable will be continuous for open-ended WTP formats and discrete for stated choice formats.\(^{219}\) In principle, stated choice questions can be distinguished along three dimensions:

- **The number of alternatives each respondent can choose from in each choice scenario** — surveys may offer only two alternatives (e.g., yes/no, or “live in area A or area B); two alternatives with an additional option to choose “don’t know” or “don’t care;” or multiple alternatives (e.g., “choose option A, B, or C”).

- **The number of attributes varied across alternatives in each choice question (other than price)** — alternatives may be distinguished by variation in a single attribute (e.g., mortality risk) or multiple attributes (e.g., mortality risk, length and severity of illness, source of risk, etc.).

- **The number of choice scenarios an individual is asked to evaluate through the survey.**

Any particular stated choice survey design could combine these dimensions in any given way. For example, a survey may offer two options to choose from in each choice scenario, vary several attributes across the two options, and present each respondent with multiple choice scenarios. The statistical strategy for estimating WTP is largely determined by the survey format adopted, as described below.

The earliest stated choice questions were simple yes/no questions. These were often called *referendum* questions because they were often posed as, “Would you vote for . . ., if the cost to you were $X?” However, these questions are not always posed as a vote decision and are now commonly called *dichotomous choice* questions.

Stated preference researchers have also adapted a choice question approach used in the marketing literature called *conjoint analysis*. These are more complex choice questions in which the respondent is asked repeatedly to pick her preferred option from a list of two or more options. Each option represents a package of product attributes. Including cost as an attribute and varying it across options allows researchers to estimate marginal WTP for each attribute of the good. Holmes and Adamowicz (2003) refer to this as attribute-based stated choice.

**Dichotomous choice WTP questions.** Dichotomous choice questions present respondents with a specified environmental change costing a specific dollar amount and then ask whether they would be willing to pay that amount for the change. The primary advantage of dichotomous choice WTP questions is that they are easier to answer and less prone to manipulation than direct WTP questions, because the respondent is not required to determine her exact WTP, only whether it is above or below the stated amount. Sample mean and median WTP values can be derived from analysis of the frequencies of the yes/no responses to each dollar amount. Bishop and Heberlein (1979), Hanemann (1984), and Cameron and James (1987) describe the necessary statistical procedures for analyzing dichotomous choice responses using logit or Probit models. Dichotomous choice responses will reveal an interval containing WTP, and in the case of a ‘yes’ response this interval will be unbounded from above. As a result, significantly larger sample sizes are needed for dichotomous choice questions to obtain the same degree of statistical efficiency in the sample means as direct/open-ended responses that reveal point-values for WTP (Cameron and James 1987).

To increase the estimation efficiency of dichotomous choice questions, some studies have used what is called a double-bounded approach. In double-bounded questions the respondent is asked whether she would be willing to pay a second amount, higher if she said yes to the first amount, and lower if she said no to the first amount.\(^{220}\) Sometimes multiple

---

219 Some researchers use the term “contingent valuation” to refer to direct WTP and dichotomous choice/referendum formats and “stated preference” to refer to other stated choice formats. In these Guidelines the term “stated preference” is used to refer to all valuation studies based on hypothetical choices (including open-ended WTP and stated choice formats), as distinguished from “revealed preference.”

220 Alberini (1995) illustrated an analysis approach for deriving WTP estimates from such responses and demonstrated the increased efficiency of double-bounded questions. The same study showed that the most efficient range of dollar amounts in a dichotomous choice study design was one that covered the mid-range of the distribution and did not extend very far into the tails at either end.
follow-up questions are used to try to narrow the interval around WTP even further. These begin to resemble iterative bidding style questions if many follow-up questions are asked. Similar to starting point bias in iterative bidding questions, the analyses of double-bounded dichotomous choice question results suggest that the second responses may not be independent of the first responses (Cameron and Quiggin 1994, 1998; and Kanninen 1995).

Multi-attribute choice questions. In multi-attribute choice questions (also known as conjoint analysis), respondents are presented with alternative choices that are characterized by different combinations of attributes and prices. Multi-attribute choice questions ask respondents to choose the most preferred alternative (a partial ranking) from multiple alternative goods (i.e., a choice set), in which the alternatives within a choice set are differentiated by their attributes including price (Johnson et al. 1995 and Roe et al. 1996). The analysis takes advantage of the differences in the attribute levels across the choice options to determine how respondents value marginal changes in each of the attributes. To measure WTP, a cost (e.g. a tax or measure of travel costs) is included in multi-attribute choice questions as one of the attributes of each alternative. This price and the mechanism by which the price would be paid need to be explained clearly and plausibly, as with any payment mechanism in a stated preference study. Boyle and Özdemir (2009) examine the impact of question design choices, such as the ordering of attributes and the number of alternatives in a single question, on the mean WTP estimate.

There are many desirable aspects of multi-attribute choice questions, including the nature of the choice being made. To choose the most preferred alternative from some set of alternatives is a common decision experience in posted-price markets, especially when one of the attributes of the alternatives is a price. One can argue that such a decision encourages respondents to concentrate on the trade-offs between attributes rather than taking a position for or against an initiative or policy. This type of repeated decision process may also diffuse the strong emotions often associated with environmental goods, thereby reducing the likelihood of yea-saying or of rejecting the premise of having to pay for an environmental improvement. Presenting repeated choices also gives the respondent some practice with the question format, which may improve the overall accuracy of her responses, and gives her repeated opportunities to express support for a program without always selecting the highest price option. One challenge of attribute based methods is representing the environmental changes in small number of separable attributes. Extensive focus group research is required to choose the most salient attributes and find the best way to convey those changes to respondents. To estimate marginal values, the attributes must be able to change independent of one another without respondent rejecting the scenario.

7.3.2.3 Considerations in Evaluating Stated Preference Results

Survey mode. The mode used to administer a survey is an important component of survey research design because it is the mechanism by which information is conveyed to respondents, and likewise determines the way in which individuals can provide responses for analysis. Common survey modes include telephone, in-person, mail, electronic surveys administered by computer or smart phone. Telephone surveys are primarily conducted with a trained interviewer using random digit dialing (RDD) to contact households. In-person surveys are conducted in a variety of ways, including door-to-door, intercepts at public locations, and via telephone recruiting to a central facility. Mail surveys are conducted by providing written survey materials for respondents to self-administer. Electronic surveys are computerized and can be self-administered at a central facility, at home via the internet, or on a smart phone wherever internet access is available. As technology and society have changed, so has the preference for one mode over the other. With the influx of market research, telemarketing, telephone scams and the abandonment of landlines, the telephone has become a less convenient way to administer surveys. The same can be said of mail surveys. People are quick to ignore unsolicited mail. With increased prevalence of smartphone technology, internet access and email accessibility, computerized surveys have emerged as an expedient means of survey administration. Researchers may also choose to combine modes

221 Some applications of multi-attribute survey formats include Layton and Brown (2000), Boyle et al. (2001), Morey et al. (2002), and Moore et al. (2018). Studies that investigate the effects of multi-attribute choice question design parameters include Johnson et al. (2000) and Adamowicz et al. (1997).
using one for recruiting and the other for survey administration. With every survey mode mentioned, there are inherent biases. These biases are generally classified as social desirability bias, sample frame bias, avidity bias, and non-response bias. See Maguire (2009), Loomis and King (1994), Mannisto and Loomis (1991), Lindberg et al. (1997), and Ethier et al. (2000) for a discussion of different biases in survey mode.

Framing issues. An important issue regarding survey formats is whether information provided in the questions influences the respondents’ answers in one way or another. For example, Cameron and Huppert (1991) and Cooper and Loomis (1992) find that mean WTP estimates based on dichotomous choice questions may be sensitive to the ranges and intervals of dollar amounts included in the WTP questions. Kanninen and Kriström (1993) show that the sensitivity of mean WTP to bid values can be caused by model misspecification, failure to include bid values that cover the middle of the distribution, or inclusion of bids from the extreme tails of the distribution.

Selection of payment vehicle. The payment vehicle in a stated preference study refers to the method by which individuals or households would pay for the good described in a particular survey instrument. Examples include increases in electricity prices, changes in cost of living, a one-time tax, or a donation to a special fund. It is imperative that the payment vehicle is incentive compatible and does not introduce strategic responses or bias. Incentive compatibility means that the individual is motivated to respond truthfully and does not use their responses to try to influence a particular outcome (e.g., state a WTP value that is higher than their true WTP to try to make sure a particular outcome succeeds).

Strategic behavior. Adamowicz et al. (1998) also suggests that respondents may be less likely to behave strategically when responding to multi-attribute choice experiments. Repeatedly choosing from several options gives the respondent some practice with the question format that may improve the overall accuracy of her responses and gives her repeated opportunities to express support for a program without always selecting the highest price option.

Yea-saying. Respondents may overstate their true WTP in order to show support for the situation described in survey questions. For example, Kanninen (1995) finds some evidence of yea-saying in dichotomous choice responses through testing in follow-up questions. The extent of this potential problem is not well established, but it may provide an explanation for the fact that mean WTP values based on dichotomous choice responses tend to be equal to or higher than values from direct WTP questions for the same good (Cummings et al. 1986, Boyle et al. 1993, Brown et al. 1996, Ready et al. 1996, and Balistreri et al. 2001).

Treatment of “don’t know” or neutral responses. Based on recommendations from the NOAA Blue Ribbon panel (Arrow et al. 1993), many surveys have included “don’t know” or “no preference” options. The Contemporary Guidance for Stated Preference (Johnston et al. 2017) recommends including a no-answer option for sensitive topics but not necessarily for all applications. There have been questions about how such responses should enter the empirical analysis. Examining referendum-style dichotomous choice questions, Carson et al. (1998) found that when those who chose not to vote were coded as “no” responses, the mean WTP values were the same as when the “would not vote” option was not offered. Offering the “would not vote” option did not change the percentage of respondents saying “yes”. Thus, they recommend that if a “would not vote” option is included, it should be coded as a “no” vote, a practice that has become widespread. Stated preference studies should always be explicit about how they treat “don’t know,” “would not vote,” or other neutral responses.

Reliability, in general terms, means consistency or repeatability. If a method is used numerous times to measure the same commodity, then the method is considered more reliable if the variability of the results is lower than an alternative.

• Test-retest approach. Possibly the most widely applied approach for assessing reliability in stated preference studies has been the test-retest approach. Test-retest assesses the variability of a measure between different

- **Meta-analysis of stated preference survey results** for the same good also may provide evidence of reliability. Meta-analysis evaluates multiple studies as though each was constructed to measure the same phenomenon. Meta-analysis attempts to sort out the effects of differences in the valuation approach used in different surveys, along with other factors influencing the elicited value. For example, Boyle et al. (1994) use meta-analysis to evaluate eight studies conducted to measure values for groundwater protection. (Also see Section 7.4.)

**Validity tests** seek to assess whether WTP estimates from stated preference methods behave as a theoretically correct WTP should. Three types of validity discussed below are: content validity, criterion validity, and convergent validity.

- **Content validity.** Content validity refers to the extent to which the estimate captures the concept being evaluated. Content validity is largely a subjective evaluation of whether a study has been designed and executed in a way that incorporates the essential characteristics of the WTP concept.

  To evaluate a survey instrument, analysts look for features that researchers should have incorporated into the survey scenario. First, survey should clearly define the environmental change being valued. The description should include a careful exposition of the conditions in the baseline case and how these would be expected to change over time if no action were taken. Next, the action or policy change should be described, including an illustration of how and when it would affect aspects of the environment that people might care about. Boyd and Banzahf (2007), and Boyd and Krupnick (2013) put a finer point on this concept and advocate developing the valuation scenario based on “ecological endpoints” rather than intermediate goods that are less clearly associated with outcomes of interest. For example, if respondents ultimately care about the survival of a certain species, it is more sensible to structure questions to ask about WTP for the species’ survival than to ask about degradation of habitat, as respondents are unlikely to know the relationship between habitat attributes and species survival. Respondent attitudes about the provider and the implied property rights of the survey scenario can be used to evaluate the appropriateness of features related to the payment mechanism (Fischhoff and Furby 1988). Survey questions that probe for respondent comprehension and acceptance of the commodity scenario can offer important indications about the validity of the results (Bishop et al. 1997).

- **Criterion validity.** Criterion validity assesses whether stated preference results relate to other measures that are considered to be closer to the concept being assessed (WTP). Ideally, one would compare results from a stated preference study of use values (the measure) with those from actual market data (the criterion). Another approach would be to estimate a sample of individuals’ WTP for a commodity using a stated preference survey and then later give the same sample of individuals or a different random sample of individuals drawn from the same population a real opportunity to buy the good. (See Mitchell and Carson 1989, Carson et al. 1987a, Kealy et al. 1990, Brown et al. 1996, and Champ et al. 1997 for examples.)

  When unable to conduct such comparisons, sensitivity to scope and income has been used to assess criterion validity. “Scope tests” are concerned with how WTP responds to changes in the amount of the referenced good provided in the valuation scenario (Smith and Osborne 1996, Rollins and Lyke 1998, and Heberlein et al. 2005). If the referenced good is indeed a “normal good,” utility theory implies that WTP should increase with the provision of the good. For the same reason one would expect WTP to exhibit positive income elasticity (McFadden 1994, and Schlapfer 2006). Neither test is necessary or sufficient to establish criterion validity (Heberlein et al. 2005), but either can serve as a useful proxy when an alternate measure of WTP for the same good is unavailable. Diamond (1996) suggests that stronger scope tests can be conducted by comparing departures from strict “adding up” of WTP for partial changes and relating them to the income elasticity of
WTP. Other researchers, however, argue that the Diamond test may not be practicable and imposes a specific structure on the preference function which may not be appropriate (Carson et al. 2001).

- **Convergent validity.** Convergent validity examines the relationship between different measures of a concept.\(^{222}\) This differs from criterion validity in that one of the measures is not taken as a criterion upon which to judge the other measure. The measure of interest and the other measure are judged together to assess consistency with one another. If they differ in a systematic way (e.g., one is usually larger than another for the same good), it is not clear which one is more correct. If stated preference estimates are being compared with revealed preference estimates, care should be taken that the same values are being captured by both approaches. Stated preference estimates often include non-use values whereas revealed preference estimates do not capture that portion of total economic value.

**Hypothetical bias** occurs when the responses to hypothetical stated preference questions are systematically different than what individuals would pay if the transactions were to actually occur. Widely cited as one of the most common problems with the stated preference method (List and Gallet 2001, and Murphy and Allen 2005), researchers have made advances in techniques to minimize such bias. These techniques include the use of “cheap talk” methods to directly tell respondents about the potential for hypothetical bias (Cummings and Taylor 1999, and List 2001); calibrating hypothetical values (List and Shogren 1998, and Blomquist et al. 2009); and allowing respondents to express uncertainty in their responses and restricting the set of positive responses to those about which the respondent was most certain (Vossler et al. 2003). Several studies have shown that attribute-based choice experiments reduce hypothetical bias in the bid amounts and the marginal value of attributes relative to other elicitation methods (Carlsson and Martinsson 2001, Murphy and Allen 2005, and List et al. 2006).

Tests for hypothetical bias often involve a comparison of actual payments and responses to hypothetical scenarios that use the same solicitation approach. The actual payments typically occur in one of three scenarios. Market transactions are the most common (Cummings et al. 1995, and List and Shogren 1998) but generally involve payments for private goods while most stated preference applications are concerned with public or quasi-public goods. Simulated markets can be used to solicit actual donations for public good provision (Champ et al. 1997). However, donation solicitations are subject to free riding, so while it may be possible to test for hypothetical bias using this approach, both the actual and hypothetical payment scenarios lack incentive compatibility and may not represent total WTP. In rare instances comparisons have been made between actual referenda for public good provision and hypothetical responses to the same scenario, but the conditions for a valid comparison of this sort are exceedingly difficult to satisfy (Johnston 2006).

**Non-response bias** is introduced when non-respondents would have answered questions systematically differently than those who did answer. Non-response bias can take two forms: item non-response and survey non-response.

- **Item non-response bias** occurs when respondents who agreed to take the survey do not answer all of the choice questions in the survey. Information available about respondents from other questions they answered can support an assessment of potential item non-response bias for the WTP questions that were unanswered. The key issue is whether there were systematic differences in potential WTP-related characteristics of those who answered the WTP questions and those who did not. Characteristics of interest include income, gender, age, expressed attitudes and opinions about the good or service, and information reported on current use or familiarity with the good or service. Statistically significant differences may indicate the potential for item non-response bias, while finding no such differences suggests that the chance of significant non-response bias is lower. However, the results of this comparison are only suggestive because respondents and non-respondents may only differ in their preference for the good in question (McClelland et al. 1991).

\(^{222}\) Mitchell and Carson (1989) define convergent validity and theoretical validity as two types of construct validity. Construct validity examines the degree to which the measure is related to other measures as predicted by theory.
• **Survey non-response bias** is created when those who refuse to take the survey have preferences that are systematically different from the preferences of those who do respond. Although it is generally thought that surveys with high response rates are less likely to suffer from survey non-response bias, it is not a guarantee.\(^{223}\) For survey non-respondents, there may be no available data to determine how they might systematically differ from those who responded to the survey. The most common approach is to examine the relevant measurable characteristics of the respondent group, such as income, resource use, gender, age, etc., and to compare them to the characteristics of the study population. Similarity in mean characteristics across the two groups suggests that the respondents are representative of the study population and that non-response bias is expected to be minimal.

A second way to evaluate potential survey non-response bias is to conduct a short follow-up survey with non-respondents. This can sometimes be accomplished through interviews conducted during the recruiting phase. Such follow-ups typically ask a few questions about attitudes and opinions on the topic of the study as well as collecting basic socioeconomic information. Questions need to match those in the full survey closely enough to compare non-respondents to respondents. The follow-up must be very brief, or response rates will be low (OMB 2006).

### 7.3.3 Combining Revealed and Stated Preference Data

Instead of looking at revealed preference and stated preference data as two separate methods for estimating environmental benefits, some researchers have used them in combination. The practice has been in use much longer in the marketing and transportation literature and many of the lessons learned by those researchers are now being employed in environmental economics. In theory, the strengths of each data type should help overcome some of the weaknesses of the other. As described by Whitehead et al. (2008) in an assessment of the state of the science, the advantages of combining revealed preference and stated preference data include:

- Helping to ground the hypothetical stated preference data with real world behavior, potentially decreasing any hypothetical bias;
- Providing the ability to test the validity of both data sources;\(^{224}\)
- Increasing the range of historical stated preference data to include conditions not observed in the past and thereby reducing the need to make predictions outside of the sample;
- Increasing the sample size;
- Extending the size of the market or population to include larger segments than captured by either method alone; and
- Exploiting the flexibility of stated preference experimental design to overcome revealed preference data’s potential multicollinearity and endogeneity problems (von Haefen and Phaneuf 2008).

The different strategies for combining revealed preference and stated preference data can be grouped into three categories. The first two methods rely on joint estimation. If the revealed preference and stated preference data have similar dependent and independent variables and the same assumed error structures, then they can simply be pooled together and treated as additional observations (Adamowicz et al. 1994; Boxall, Englin, and Adamowicz 2003; and

---

\(^{223}\) Note that OMB’s Guidance on Agency Survey and Statistical Collections (OMB 2006) has fairly strict requirements for response rates and their calculation for Agency-sponsored surveys, recommending that “ICRs for surveys with expected response rates of 80 percent or higher need complete descriptions of the basis of the estimated response rate...ICRs for surveys with expected response rates lower than 80 percent need complete descriptions of how the expected response rate was determined, a detailed description of steps that will be taken to maximize the response rate...and a description of plans to evaluate non-response bias” (pp. 60-70).

\(^{224}\) Herriges, Kling, and Phaneuf (2004) point out that revealed preference may not always be valid for estimating WTP for quality changes when weak complementarity cannot be assured.
Morgan, Massey, and Huth 2009). If the revealed preference and stated preference data sources cannot be pooled, it is sometimes possible to use them in a jointly estimated mixed model that relies on a utility theoretic specification of the underlying WTP function (Huang, Haab, and Whitehead 1997; Kling 1997; and Eom and Larson 2006). If the data cannot be combined in estimation, it can still be useful to estimate results separately and then use them to test for convergent validity between the two data sources (Carson et al. 1996, and Schlapfer et al. 2004).

### 7.4 Benefit Transfer

As noted at the outset of this chapter, benefit transfer is the approach most often used by the Agency for monetizing benefits in economic analysis. Benefit transfer refers to the use of estimated values of environmental quality changes from primary studies to the evaluation of similar changes that are of interest to the analyst (Freeman et al. 2014). The case under consideration for a new policy is referred to as the “policy case.” Cases from which estimates are obtained are referred to as “study cases.” A benefit transfer study identifies stated preference or revealed preference study cases that sufficiently relate to the policy context and transfers their results to the policy case.

Benefit transfer is necessary when it is infeasible to conduct an original study focused directly on the policy case and is the most common approach for completing a BCA at the EPA. Given the time and analytical resource constraints under which most regulatory analysis activities are conducted, conducting new revealed or stated preference studies that are tailor-made to examine all of the (sometimes numerous) endpoints changed by the policy or regulation in question is near impossible (Newbold et al. 2018a). Because original studies are time consuming and expensive, benefit transfer can reduce both the time and financial resources required to develop estimates of a proposed policy’s benefits. Benefit transfer might also be useful as a scoping exercise to predict the approximate magnitude of benefits that might then be more precisely estimated with an original study.

In general, the advantages of benefit transfer in terms of time and cost savings should be weighed against the disadvantages in terms of potential reduced reliability of the final benefit estimates. Ideally benefit transfer would only be used as a last resort, and whenever it is used, a clear justification for using a benefit transfer approach over conducting original valuation studies should be provided (OMB 2003).

While there is no universally accepted single approach for conducting benefit transfer, there are some generalized steps involved in the process. These steps are described below.

### Steps for Conducting Benefit Transfer

1. **Describe the Policy Case**
   
   The first step in a benefit-transfer study is to clearly describe the policy case with respect to the baseline so that its characteristics and consequences are well understood. Are human health risks reduced by the policy intervention? Are ecological benefits expected (e.g., increases in populations of species of concern)? It is also important to identify to the extent possible the population affected by the proposed policy and to describe their demographic and socioeconomic characteristics (e.g., users of a particular set of recreation sites, children living in urban areas, or older adults across the United States). Information on the affected population is generally required to translate per person (or per household) values to an aggregate benefits estimate.

2. **Select Study Cases**
3. **Transfer Values**
4. **Report Results**
**Step 2: Select study cases.**

A benefit-transfer study is only as good as the study cases from which it is derived, and it is therefore crucial that studies be carefully selected. First, the analyst should identify potentially relevant studies conducting a comprehensive literature search. Online searchable databases summarizing valuation research may be especially helpful at this stage. Because peer-reviewed academic journals may be more likely to publish work based on methodological contributions, analysts should be aware of the potential for publication bias. Some studies of interest may be found in government reports, working papers, dissertations, unpublished research, and other “gray literature” (Rosenberger and Stanley 2006; Johnston and Rosenberger 2010; Johnston et al 2015). While including studies from the gray literature may help mitigate publication bias, use of valuation estimates that are not published in peer-reviewed journals may necessitate subsequent peer review in some form (i.e., formal peer review or a less formal peer input) of the benefit transfer. See U.S. EPA (2015a) for more guidance (in particular, sections 3.5.6 and 3.5.7).

Next, the analyst should develop an explicit set of selection criteria to evaluate each of the potentially relevant studies for quality and applicability to the policy case. The quality of the value estimates in the study cases will in large part determine the quality of the benefit transfer. As a first step, the analyst should review studies according to the criteria listed for each methodology in the previous sections in this chapter. Results from study cases must be valid as well as relevant. Concerns about the quality of the studies, as opposed to their relevance, will generally hinge on the methods used. Valuation approaches commonly used in the past may now be regarded as unacceptable for use in benefits analysis. Studies based on inappropriate methods or reporting obsolete results should be removed from consideration.

It is unlikely that any single study will match perfectly with the policy case; however, study cases potentially suitable for use in benefit transfer should be similar to the policy case in their: (1) definition of the environmental commodity being valued (include scale and presence of substitutes); (2) baseline and extent of environmental changes; and (3) characteristics of affected populations. Analysts should avoid using benefit transfer in cases where the policy or study case is focused on a “good” with unique attributes or where the magnitude of the change or improvement across the two cases differs substantially (OMB 2003). It is crucial to remember that economic value is determined on the margin and depends upon how scarce something is relative to the demand for it at the time and place it is provided (Simpson 2017).

The analyst should determine whether adjustments should and can be made for important differences between each study and policy case. For example, some case studies will be based on stated preference methods while others may be based on revealed preference methods. The ability of the analyst to make these adjustments will depend, in part, on both the number of value estimates for suitably similar study sites and the method used to combine these estimates. These methods are now discussed in turn.

**Step 3: Transfer values.**

There are several approaches for transferring values from study cases to the policy case. These include unit value transfers and function transfers, and they may use techniques from meta-analysis if multiple studies are available. Transfers may also be structural or non-structural (referring to a utility-theoretic structure). Each of these approaches is typically used to develop per person or per household value estimates that are then aggregated over the affected

---

225 For example, the Environmental Valuation Reference Inventory (EVRI) is maintained by Environment Canada and managed by a cross-county working group. EVRI contains summaries of over 4,000 studies that can be referenced according to keyword, study type, region, and environmental asset. EVRI also provides a bibliography on benefit transfer. See www.evri.ca for more information.

226 Newer, unpublished research may also be on the cutting edge of methods.

227 In some cases, the transfer method itself may inform the choice of study cases to include. For example, meta-analysis approaches (discussed below) can facilitate some forms of statistical validity testing. For example, Moeltner (2015, 2019) uses Bayesian methods in a meta-analysis to identify optimal pooling of studies.
population to estimate total benefits. In general, when reporting transfer results, researchers should provide information on
the background of the problem, the strategy for selecting studies, analytic methods used, results, discussion, and conclusions.

- **Unit value transfers** are the simplest of the benefit-transfer approaches. They take a point estimate of WTP
for a unit change in the environmental resource from a study case or cases and apply it directly to the policy
case. The point estimate may be a single estimated value from a single case study, but it can also be the
(otherwise unadjusted) average of a small number of estimates from a few case studies. For example, a study
may have found a WTP of $20 per household for a one-unit increase on a water quality scale. A unit value
transfer would estimate total benefits for the policy case by multiplying $20 by the number of units by which
the policy is expected to increase water quality and by the number of households who will benefit from the
change. This approach can be useful for developing preliminary, order-of-magnitude estimates of benefits,
but it should be possible to base final benefit estimates on more information than a single point estimate
from a single study. If multiple studies are available, the mean or median WTP value may provide a useful
point estimate, though analysts should consider weighting estimates by inverse variance or sample size to
give more weight to more precise estimates when calculating the mean (Nelson and Kennedy 2009; Nelson et
al. 2013). Point estimates reported in study cases are typically functions of several variables, and simply
transferring a summary estimate without controlling for differences among these variables can yield
inaccurate results. It is important to recognize that unit value transfer assumes that the original good, as well
as the characteristics and tastes of the population of beneficiaries, are the same as the policy good.

- **Function transfers** use information on other factors that influence WTP to adjust the unit value for quantifiable
differences between the study case and the policy case. This is accomplished by transferring the estimated
function upon which the value estimate in the study case is based to the policy case. This approach implicitly
assumes that the population of beneficiaries to which the values are being transferred has potentially different
characteristics but similar tastes as the original one and allows the analyst to adjust for these different
characteristics. Generally, benefit function transfers may be preferable to unit value transfers as they
incorporate information relevant to the policy scenario (OMB 2003; Johnston and Rosenberger 2010).

- To implement a function transfer, suppose that in the hypothetical example above the $20 unit value was the
result of averaging the results of an estimated WTP function over all individuals in the study case sample,
where the WTP function included income, the baseline water quality level, and the change in the water quality
level for each household. A function transfer would estimate total benefits for the policy case by:

1. Applying the WTP function to a random sample of households affected in the policy case using each
household’s observed levels of income, baseline water quality, and water quality change;
2. Averaging the resulting WTP estimates; and
3. Multiplying this average WTP by the total number of households affected in the policy case.

- If the WTP function is nonlinear and statistics on average income, baseline water quality, and water quality
changes are used in the transfer instead of household level values, then bias will result. Feather and
Hellerstein (1997) provide an example of a function transfer that attempts to correct for such bias. Although
function transfers can adjust and compensate for small differences between the case and policy study
populations, they are subject to the same basic usage rules governing unit value transfers. Function transfers
should only be used if the case and policy studies are evaluating sufficiently similar environmental goods,
change in environmental levels, and affected populations.

- **Meta-analysis** uses results from multiple valuation estimates in a new unit or function transfer. Meta-analysis
is an umbrella term for a suite of techniques that synthesize the results of empirical research. This could include

---

228 A typical meta-analysis combines estimates from many studies, but meta-analyses that combine multiple estimates from one study or more than one
a simple ranking of results, a meta-analytic average or other central tendency estimate, or a multivariate regression. The advantage of these meta-analytic methods is that they incorporate and can potentially control for more information than transfers based on a single estimate. This approach has been widely used in environmental economics (see Rolfe et al. 2015 and Johnston et al. 2018).

There are several guidelines for meta-analyses that outline protocols that should be followed in conducting or evaluating a study. See Bergstrom and Taylor (2006); Nelson and Kennedy (2009); Nelson et al. (2013); Nelson (2015), and Boyle et al. (2013, 2015) for more information. Some may choose to follow a systematic review protocol as described in Text Box 7.5. The EPA's Peer Review Handbook (U.S. EPA 2015b) recommends that meta-analyses used in regulatory analysis should generally be peer-reviewed. Boyle and Wooldridge (2018) emphasize that the purpose of a meta-analysis for benefit transfer is prediction and the purpose of a traditional meta-analysis is to summarize a literature. This latter paper provides a number of technical suggestions to "provide the best econometric prediction of value for a benefit-transfer application."

- **Structural benefit transfer** involves deriving a benefit transfer function from an assumed form of the direct or indirect utility function and calibrating or estimating the form of the transfer function using insights from economic theory.229 The advantages of structural transfer functions are that they can accommodate different types of economic value measures (e.g., WTP, WTA, or consumer surplus) and can be constructed to satisfy certain theoretical consistency conditions (e.g., WTP bounded by income). Using a structural benefit transfer or preference calibration approach is one way to ensure that that the adding-up condition holds (See Text Box 7.6). However, Johnston et al. (2018) discuss the tradeoff between theory and accuracy of the transfer in structural benefit transfer. They conclude that core concepts such as diminishing marginal utility are necessary but there can be a tradeoff between empirical accuracy of transfers and imposing a specific functional form to satisfy stronger theoretical restrictions; and there is not a consensus in the empirical literature on the appropriate balance.

- **Which benefit transfer method to choose** is not always obvious. Boyle et al. (2013) note, there is no consensus on which method works best. There have been numerous studies comparing the (convergent) validity and reliability of transfers (see Rosenberger 2015 and Kaul et al. 2013 for summaries). Some general lessons are "that function transfers tend to be more accurate than value transfers; transfers of values for environmental quantity changes tend to be more accurate than those for quality changes; geographic similarity between sites improves the accuracy of transfers, especially for value transfers; combining information from multiple studies improves the accuracy of transfers; and that transfers based on stated preference valuation formats with more options per question, such as choice experiments, have larger transfer errors than methods with fewer choices per question, such as contingent valuation surveys" (Newbold et al. 2018a). Few studies test the validity and reliability of meta-analytic transfers, however. Johnston et al. (2018) describe how increasingly complex methods may not always be worthwhile, noting

---

229 See Smith and Pattanayak 2002; and Smith, Pattanayak, and van Houtven 2006 for descriptions on the method; see Newbold et al. 2018b, discussed in Text Box 7.6, for an example of a functional form of a meta-analysis being based on theory.
Text Box 7.5 - Systematic Review Approaches and Benefit Transfer

EPA risk assessments have increasingly adopted systematic review approaches as recommended by the National Research Council (NRC 2014), and these approaches may provide a useful model for identifying and evaluating literature for a benefit transfer or meta-analysis. The Institute of Medicine has defined systematic review as “a scientific investigation that focuses on a specific question and uses explicit, prespecified scientific methods to identify, select, assess, and summarize the findings of similar but separate studies” (IOM 2011). A key element in conducting a systematic review is preparation of a protocol which details in advance the methods that will be used in conducting the review. Major advantages of systematic review include improved documentation and transparency, as well as minimization of potential bias in how the review is conducted (NRC 2014).

The steps in conducting a systematic review, as outlined by the NRC (2014), are:

- Problem Formulation: define the study question (roughly equivalent to “Describe the policy case”);
- Develop a protocol for conducting the systematic review: the protocol defines the methods to be used (e.g., search strategy, inclusion/exclusion criteria, study evaluation criteria);
- Evidence Identification: conduct the literature search and screen the literature search results (apply the search strategy and the inclusion/exclusion criteria from the protocol to identify relevant studies);
- Evidence Evaluation: evaluate quality of studies by applying the criteria specified in the protocol to each included study;
- Evidence Integration: develop conclusions from the included studies to answer the study question.

that, while more flexible transfer functions tend to outperform unit value transfers, value transfers may outperform other types of transfers when the sites are very similar. The further highlight evidence that simple function transfers that adjust estimates for a few key variables (e.g., income elasticity) may have lower transfer error than complex function transfers that control for numerous characteristics. The benefit transfer literature is large and diverse, but the EPA will continue to monitor it and update these recommendations as necessary.

Step 4: Report the results.

In addition to reporting the final benefit estimates from the transfer exercise, the analyst should clearly describe all key judgments and assumptions, including the criteria used to select study cases and the choice of the transfer approach. The uncertainty in the final benefit estimate should be quantified and reported when possible. Any limitations should also be discussed.230 (See Chapter 11 on Presentation of Analysis and Results.)

7.5 Accommodating Non-Monetized Benefits

It often will not be possible to quantify every significant physical impact for all policy options. For example, animal studies may suggest that a contaminant causes severe illnesses in humans, but the available data may not be adequate to determine the number of expected cases associated with different human exposure levels. Likewise, it often is not possible to quantify the various ecosystem changes that may result from an environmental policy. While Chapter 11 discusses how to present these benefits so as to provide a fuller accounting of all effects, this section discusses what analysts can do to incorporate these endpoints more fully into the analysis.

---

Text Box 7.6 - The Adding-Up Condition in Benefit Transfer

When a benefit transfer function is estimated in a meta-analysis or in some other fashion, it is important to carefully consider the form of the estimating equation used to relate study characteristics to willingness-to-pay. For consistency in regulatory accounting, a willingness-to-pay function must satisfy a basic adding-up condition (Kling and Phaneuf 2018). In other words, WTP for good X, plus WTP for good Y given good X, must equal WTP for X and Y together.

A benefit transfer function that violates the adding-up condition can lead to inconsistent policy evaluations. For example, an omnibus policy that appears to have lower net benefits than a set of component policies that, when combined, yield the same water quality outcomes and have the same total cost as the omnibus policy does not satisfy the adding up condition. In this case, the policy change appears to have larger benefits if it is broken up into several smaller policy changes. A benefit transfer function that violates the adding-up condition also could lead to inconsistent policy rankings, since independently evaluating the provision of goods X and Y could pass a benefit-cost test while evaluating the provision of both goods X and Y together could fail a benefit-cost test.

Analysts who use meta-analysis to estimate a benefit transfer function or apply a benefit transfer function developed in a previous study should ensure that the resulting willingness-to-pay function satisfies the adding-up condition. If the function fails to satisfy the adding-up condition, the analyst should consider re-estimating the benefit transfer function using a different functional form that does satisfy the adding-up condition. One way to ensure that a benefit transfer function complies with the adding-up condition is to use a "structural benefit transfer" or "preference calibration" approach, as described in the main text.

Newbold et al. (2018b) examine existing valuation studies and document violations of the adding-up condition and impacts on benefit-cost results because of these violations. They further describe a structural meta-analytic model that meets the adding-up condition and compare it to a non-structural model that does not. They find that the nonstructural model produces much larger benefits estimates than the structural model and that the violations of the adding-up condition are severe in the non-structural model.

7.5.1 Qualitative Discussions

When there are potentially important effects that cannot be quantified, the analyst should include a qualitative discussion of the benefit endpoints. The discussion should explain why a quantitative analysis was not possible and the reasons for believing that these non-quantified effects may be important for decision making. Chapter 11 discusses how to describe benefit categories that are quantified in physical terms but not monetized.

7.5.2 Alternative Analyses

Alternative analyses exist that can support benefits valuation when robust value estimates and/or risk estimates are lacking. These analyses, including cost-effectiveness, break-even, and bounding analysis, can provide decision makers with some useful information. However, analysts should remember that because these alternatives do not estimate the net benefits of a policy or regulation, they fall short of BCA in their ability to identify an economically efficient policy. This shortcoming and any others should be discussed when presenting results from these analyses to decision makers.

7.5.2.1 Cost-Effectiveness Analysis

Cost-effectiveness analysis (CEA) is most useful when outcome measures from a policy are not in dollar terms, for example, a number of tons of emissions reduced. Cost-effectiveness is calculated by dividing the annualized cost of the option by the annual non-monetary outcome measure, resulting in a ratio of cost per unit (e.g., dollar per ton reduced). Because the outcome is a ratio, it is more sensitive to how benefits and costs are characterized. Whereas a net benefits outcome (benefits minus costs) is robust to whether negative benefits are counted as costs, the same is not true for a ratio. This is one reason net benefits are generally preferred to benefit-cost ratios.

The preferred option from CEA is clear when all options achieve the same result (e.g., the same number of tons reduced): the option with the lowest cost per unit is the most cost-effective option. More typically, however, options vary not only in costs but also in the outcomes they produce, and there is no generally accepted criterion defining the
preferred option in this case. Each cost-effectiveness ratio represents a different tradeoff between the outcome measure and costs, but there is no information on which, if any, of the options is efficient. Still, cost-effectiveness on outcomes can inform decision-making in the absence of monetized benefits.

Because cost-effectiveness is defined by cost-per-unit, CEA requires a single outcome measure. It is not possible to perform a CEA where there are two separate outcomes, which is often the case for environmental regulations. For example, if a program reduces both hydrocarbon and nitrogen oxide emissions, it is probably not possible to develop a cost per ton of hydrocarbons reduced and a separate cost per ton of nitrogen oxides, because the same costs produce both outcomes. For health and safety regulations, however, there are a number of measures that integrate disparate health outcomes into a single metric for cost-effectiveness calculations. These metrics were largely developed for comparing public health or medical interventions. The most common metric is a quality-adjusted life year (QALY), which combines health-related quality of life with longevity. Cost-effectiveness using QALYs is sometimes referred to as “cost-utility analysis” (CUA) because the health-related quality of life component is based on stated preferences about the impact of different health conditions.

The wide application of QALYs to regulatory analysis is relatively recent but has been recommended in OMB Circular A-4 and has been evaluated in detail by the Institute of Medicine (IOM, 2006.) It is important that cost-effectiveness analysis using QALYs be distinct from benefit-cost analysis. QALYs should not be converted to a monetary value using a “cost per QALY” because they are not fully consistent with utility theory underlying benefit-cost analysis (IOM 2006; Hammitt, 2002). When there is a benefit-cost analysis, cost-effectiveness analysis should be considered a complement that provides a different perspective on the tradeoffs of a regulatory action.

7.5.2.2 Break-Even Analysis

Break-even analysis is one alternative that can be used when either risk data or valuation data (but not both) are lacking.231 Analysts who have per unit estimates of economic value, but lack risk estimates cannot quantify net benefits. They can, however, estimate the number of cases (each valued at the per unit value estimate) at which overall net benefits become positive, or where the policy action will break even.232 Consider a proposed policy that is expected to reduce the number of cases of endpoint X with an associated cost estimate of $1 million. Further, suppose that the analyst estimates that WTP to avoid a case of endpoint X is $200, but that because of data or modeling limitations it is not possible to estimate the number of cases of this endpoint reduced by the policy. In this case, the proposed policy would need to reduce the number of cases by 5,000 in order to “break even.”233 This estimate then can be assessed for plausibility either quantitatively or qualitatively. Policy makers will need to determine if the break-even value is acceptable or reasonable.

The same sort of analysis can be performed when analysts lack valuation estimates, producing a break-even value that should again be assessed for credibility and plausibility. Continuing with the example above, suppose the analyst estimates that the proposed policy would reduce the number of cases of endpoint X by 5,000 but does not have an estimate of WTP to avoid a case of this endpoint. In this case, the policy can be considered to break even if WTP is at least $200.

In some cases, it may be possible to assess the credibility of economic break-even values by comparing them to valuation estimates for effects that are considered to be more or less severe than the endpoint being evaluated. For the break-even value to be plausible, it should fall between the estimates for these more and less severe effects. For the example above, if

231 Boardman et al. (2011) describes determining break-even points under the general subject of sensitivity analysis and includes empirical examples.

232 Circular A-4 (OMB 2003) refers to these values as “switch points” in its discussion of sensitivity analysis. Section 5.4.4 on uncertainty analysis also contains related discussions on switch points.

233 Note that this assumes that endpoint X is the only effect of the policy, or that the $1 million cost is the net costs after other benefits have been accounted for.
the estimate of WTP to avoid a case of a more serious effect was only $100, the above break-even point may not be considered plausible.

Break-even analysis is most effective when there is only one missing value in the analysis, or at least when it is being applied to what is likely the largest benefit endpoint. For example, if an analyst is missing risk estimates for two different endpoints (but has valuation estimates for both), then they will need to consider a “break-even frontier” that allows the number of both effects to vary. It is possible to construct such a frontier, but it is difficult to determine which points on the frontier are relevant for policy analysis.

7.5.2.2 Bounding Analysis

Bounding analysis can help when analysts lack value estimates for a particular endpoint. As suggested above, reducing the risk of health effects that are more severe and of longer duration should be valued more highly than those that are less severe and of shorter duration, all else equal. If robust valuation estimates are available for effects that are unambiguously “worse” and others that are unambiguously “not as bad,” then one can use these estimates as the upper and lower bounds on the value of the effect of concern. Presenting alternative benefit estimates based on each of these bounds can provide valuable information to policy makers. If the sign of the net benefit estimate is positive across this range, then analysts can have some confidence that the program is welfare-enhancing. Analysts should carefully describe judgments or assumptions made in selecting appropriate bounding values.

Chapter 7 References


7-60


Chapter 8

Analyzing Costs

This chapter discusses estimating costs for use in benefit-cost analyses (BCA). In a BCA, estimates of the costs of regulatory (or deregulatory) actions are compared to estimated benefits to illuminate key differences across the options under consideration. While the estimation of costs is often portrayed as relatively straightforward — particularly compared to estimating benefits — it must be guided by economic theory. For example, errors can easily occur if the analyst confuses transfers with costs or ignores pre-existing regulation or taxes in the affected market.

Estimating the costs of regulation involves a series of decisions. The analyst must determine the scope of the analysis to appropriately capture the range of anticipated effects from the regulation or policy. Analysts must determine which types of costs are most relevant within a specific regulatory context and choose the best way to measure them based on the best available data and methods. Both the scope of the analysis and how costs are measured will affect the choice of economic model. Models vary in their ability to capture certain types of costs; whether they are static or dynamic; their level of geographic and sectoral detail; and their scope; among others. After selecting one or more economic models, analysts face a series of implementation decisions, such as how to best parameterize the model and how to account for uncertainty.

8.1 The Economics of Social Cost

The appropriate measure of cost to use in a BCA is social cost. Social cost represents the total burden that a regulation will impose on the economy, defined as the sum of all opportunity costs incurred as a result of the regulation. These opportunity costs consist of the value to society of the goods and services no longer produced and consumed as resources are reallocated away from other activities towards pollution abatement. To be complete, an estimate of social cost should include the opportunity costs of current and future consumption and leisure that will be foregone as a result of the regulation (e.g., effects in the future could occur because of effects on capital investment). The social cost of a

---

234 Several executive and legislative mandates require that different aspects of costs be considered in a regulatory analysis. For instance, EO 12866 specifies that an assessment of the costs of a regulation should include “any adverse effects on the efficient functioning of the economy and private sector (including productivity, employment, and competitiveness).” The UMRA of 1995 requires that cost estimates account for indirect and implicit costs on state and local governments. Many of these types of “costs” are categorized by the EPA as economic impacts and therefore discussed in Chapter 9.

235 Comparing the social cost of different regulatory options may provide some sense of the relative burden they would impose on the economy, but without evaluating the social benefits this exercise alone would not indicate which, if any, of the options may be worthwhile from an efficiency standpoint.

236 A model is a “simplification of reality that is constructed to gain insights into select attributes of a particular physical, biologic, economic, or social system” (NRC, 2009) and the simplifications necessary to tractably model complex systems will introduce uncertainty.

237 For a more detailed treatment of the material in this section, see Pizer and Kopp (2005).
regulation is generally not the same as its effects on gross domestic product (GDP) or other broad measures of economic activity.\footnote{GDP is defined as the sum of the value (price times quantity) of all market goods and services produced in the economy and is equal to either Consumption (C) + Investment (I) + Government (G) + (Exports (X) – Imports (M)), or Labor (L) + Capital (K) + Taxes (T).} See Section 8.2.2 for more discussion.

Social cost is distinct from but generally includes the cost of compliance borne by the regulated entity. The \textbf{compliance cost} is the private cost that a regulated entity incurs to reduce or prevent pollution to comply with an environmental regulation - for instance, through the installation and operation of pollution abatement equipment. To estimate social cost, analysts may use one or some combination of compliance cost, partial equilibrium and/or general equilibrium approaches. A compliance cost approach assesses the costs of abatement and other actions taken to comply (e.g., monitoring, reporting, and recordkeeping requirements) for the directly regulated sector.\footnote{The term direct cost is sometimes used to refer to the costs incurred by regulated entities to comply with the regulation. The term indirect cost is sometimes used to refer to costs incurred in related markets or experienced by consumers or government not subject to the regulation, often transmitted through changes in the prices of goods or services produced by the regulated sector.} A partial equilibrium approach models the supply and demand responses of the regulated sector to these compliance costs and may be extended to consider a small number of related sectors (e.g., markets that supply intermediate goods to the regulated sector, markets for substitute or complementary products, or markets that supply abatement equipment or other services to comply with requirements). When broader economy-wide impacts are expected due to a regulation, a compliance cost or partial equilibrium approach will miss these impacts. In this case, a general equilibrium approach is needed to more fully estimate social cost. Models that utilize one of these three approaches to estimate social cost are discussed in Section 8.3.

### 8.1.1 Compliance Cost Approach

A compliance cost approach estimates the direct compliance expenditures incurred by regulated entities (e.g., individual emitting units or facilities) when installing and operating abatement technologies or processes to comply with a regulation, conditional on a given level of output. It does not attempt to estimate welfare impacts associated with a change in the amount of production or use of inputs but generally assumes that regulated sources are cost-minimizing in their compliance behavior. Its primary advantage is the ability to generate highly detailed and, when data are available, relatively specific information on compliance options and their associated costs that reflect the heterogeneity of regulated entities. This detailed information can be very useful, as many stakeholders are keenly interested in understanding the anticipated cost of meeting regulatory requirements. Furthermore, reporting detailed assessments of how regulated entities are expected to respond to a regulation can generate useful public comments that further EPA’s understanding of available compliance options and their costs.

A compliance cost approach typically does not account for other producer or consumer behavioral changes that may be incentivized by a new regulation. However, it can still provide a reasonable estimate of the social cost of a regulation when changes in a regulated sector’s outputs and input mix (aside from direct compliance activities) are expected to be minimal. On the other hand, when significant changes on the producer or consumer side are expected to occur, a compliance cost approach may substantially misestimate the social cost of regulation.\footnote{The degree to which a demand response influences social cost depends on a variety of factors such as the magnitude of the price change, the price elasticity of demand for output of the regulated sector, and the degree of competition in the market. An elasticity is a measure of how responsive a firm or consumer is to a change in price. In the case of demand, it is the percentage change in the quantity of the product that is demanded by consumers divided by the percentage change in the product’s price. See Appendix A for more discussion of elasticities.} Likewise, it does not capture supply side responses, such as changes in the composition of goods produced by the industry or changes in product quality. A key question for analysts is whether it is worth expending additional resources to expand beyond a
compliance cost approach to capture other potentially substantial costs within the sector itself, related sectors, or to the economy as a whole.

8.1.2 Partial Equilibrium Approach

In contrast to a compliance cost approach, a partial equilibrium approach to cost estimation accounts for market changes in the regulated sector. Market responses to the regulation may include reduced industry output or higher prices as firms pass on some costs directly to consumers. The goal of a partial equilibrium approach is to measure the net change in consumer and producer surplus relative to the pre-regulatory equilibrium.\textsuperscript{241}

In theory, in the absence of market distortions (e.g., pre-existing taxes, market power), the social cost of a regulation can be assessed with a partial equilibrium approach of the regulated market (Just et al., 2004; Harberger 1964).\textsuperscript{242} This is because while a policy may have effects in many other markets, market clearing conditions effectively cancel out these effects with regard to aggregate welfare (Farrow and Rose 2018).\textsuperscript{243} Thus, a partial equilibrium approach is sufficient for estimating social cost when the analyst expects that a regulation will result in appreciable changes in market activities, but the effects will be confined primarily to a single market or a small number of markets. The use of a partial equilibrium approach assumes that the effects of the regulation in all other markets, outside of those being modeled, will be minimal and therefore can be ignored.

Figures 8.1 and 8.2 illustrate how social cost can be defined in partial equilibrium. Figure 8.1 shows a competitive market before the imposition of an environmental regulation. The intersection of the supply ($S_0$) and demand ($D$) curves determines the equilibrium price ($P_0$) and quantity ($Q_0$). The shaded area below the demand curve and above the equilibrium price line is consumer surplus. The area above the supply curve and below the price line is producer surplus.\textsuperscript{244} The sum of these two areas defines the total welfare generated in this market (i.e., the net benefits to society from producing and consuming the good or service represented in this market). Note that total welfare as depicted ignores the negative pollution externality arising in this market, which the regulation is designed to correct.\textsuperscript{245}

In this market, the imposition of a new environmental regulation raises firms' production costs. Each unit of output is now more costly to produce because of expenditures incurred to comply with the regulation. As a result, firms will respond by reducing their level of output. For the industry, this will appear as an upward shift in the supply curve. This is shown in Figure 8.2 as a movement from $S_0$ to $S_1$. The effect on the market of the shift in the supply curve is to increase the equilibrium price to $P_1$, and to decrease the equilibrium output to $Q_1$, holding all else constant. As seen by comparing Figures 8.1 and 8.2, the overall effect on welfare is a decline in both producer and consumer surplus.\textsuperscript{246}

\textsuperscript{241} Consumer surplus is the sum of consumers' net benefits — i.e., what they are willing to spend on a good or service over and above market price. Thus, it is the area under the market demand (marginal benefit) curve but above market price. Producer surplus is the sum of producers' revenues over and above the cost of production. Thus, it is the area above the market supply (marginal cost) curve but below market price. See Appendix A.

\textsuperscript{242} As previously defined in Chapter 5, market distortions are factors such as pre-existing taxes, externalities, regulations, or imperfectly competitive markets that move consumers or firms away from what would occur under perfect competition.

\textsuperscript{243} In theory, impacts in undistorted related markets are "pecuniary" and do not need to be included if the social costs have been correctly measured in the primary market, but pecuniary effects are important to consider in inefficient related markets (Boardman et al. 2006). In addition, it is likely that most regulations will result in winners and losers. Economic impact analysis evaluates how different groups are impacted by a regulation. See chapter 9.

\textsuperscript{244} Producer surplus may also be interpreted as the profits of regulated sources. Profits equal total revenues (price multiplied by quantity of total output) minus total private costs (the area under the supply curve). Profits are also the return on fixed investments by firms. Over time, the share of costs attributable to investments that are fixed generally declines and the supply curve becomes more elastic. See Section 8.2.3.2.

\textsuperscript{245} Appendix A presents a graphical representation of how to account for this externality. Reduction of the negative externality would be quantified in the benefits portion of an analysis. The supply curve in Figure 8.1 corresponds to the marginal private cost (MPC) curve described in Figure A.5.

\textsuperscript{246} The figure depicts an equal distribution of welfare between consumers and producers in both the old and new equilibria. Depending on the elasticities of supply and demand, this may not be the case. The elasticities determine the magnitude of the price and quantity changes induced by the cost

8-3
In the long run (i.e., when all costs are variable), compliance costs in this market are equal to the area between the old and new supply curves, bounded by the new equilibrium output, \( Q_1 \).\(^{247}\) Several useful insights about the total costs of the regulation can be derived from Figures 8.1 and 8.2. First, when consumers are price sensitive — as reflected in the downward sloping demand curve — a higher price causes them to reduce consumption of the good. If costs are estimated ex ante and this price sensitive behavior is not taken into account (i.e., the estimate is based on the original level of output, \( Q_0 \)), compliance costs will be overstated. Extending the vertical dotted line in Figure 8.2 from the original equilibrium to the new supply curve \( S_1 \) illustrates this point. A second insight derived from Figures 8.1 and 8.2 is that realized compliance costs are usually only part of the total costs of a regulation. The “deadweight loss” (DWL) shown in Figure 8.2 is an additional, real cost arising from the regulation. It reflects the foregone net benefit (or opportunity cost) due to the reduction in output. Moreover, DWL will be a component of social cost in future periods.

Under the assumption that impacts outside this market are not significant, the social cost of the regulation is equal to the sum of the compliance costs and the deadweight loss (shown in Figure 8.2). This is exactly equal to the reduction in producer and consumer surplus from the pre-regulation equilibrium (shown in Figure 8.1). This estimate of social cost would be the appropriate measure to use in a BCA of the regulation.

The preceding discussion describes the use of a partial equilibrium approach when the regulated market is perfectly competitive. In many cases, however, some form of imperfect competition (e.g., monopolistic competition, oligopoly, or monopoly) may better characterize the regulated market. Firms in imperfectly competitive markets will adjust differently to the imposition of a new regulation, which can alter the estimate of social cost.\(^{248,249}\) If the regulated market is imperfectly competitive, this may significantly influence compliance behavior and costs, in which case the market structure should be reflected in the analysis. See section 8.2.3.6 for more discussion.

When the effects of a regulation are expected to impact a limited number of markets beyond the regulated sector, it still may be sufficient to use a partial equilibrium approach to estimate social cost. A multi-market approach extends a single-market, partial equilibrium representation of the directly regulated sector to include closely related markets. These may include the upstream suppliers of major inputs to the regulated sector (including pollution abatement equipment or services), downstream producers who use the regulated sector’s output as an input, and producers of

---

\(^{247}\) In the long run, when all costs are variable, they are fully represented in the movement of the supply curve. In the short or medium run, fixed costs do not affect the supply curve but do contribute to compliance costs. See Tietenberg (2002).

\(^{248}\) For further discussion of the welfare effects of environmental regulation in the context of imperfectly competitive markets, see Chapter 6 of Baumol and Oates (1988), Requate (2006) and Chapter 6 of Phaneuf and Requate (2017).

\(^{249}\) See Ryan (2012), Ferris, et al. (2014), Wolverton, et al. (2019) for examples where accounting for the way market structure affected firm decision-making would have potentially led to a different estimate of costs.
8.1.3 General Equilibrium Approach

A general equilibrium approach to cost estimation concurrently considers the effect of a regulation across all sectors in the economy. It is structured around the assumption that, for some discrete period of time, an economy can be characterized by a set of equilibrium conditions in which supply equals demand in all markets. When the imposition of a regulation alters conditions in one market, a general equilibrium approach will determine a new set of prices for all markets that will return the economy to equilibrium. These prices in turn determine the outputs and consumption of goods and services in the new equilibrium. In addition, a new set of prices and demands for the factors of production (labor, capital, and land), the returns to which compose the income of businesses and households, will be determined in general equilibrium. The social cost of the regulation can then be estimated by comparing the value of variables in the pre-regulation “baseline” equilibrium with those in the post-regulation, simulated equilibrium.251

When the imposition of an environmental regulation is expected to have appreciable effects in markets beyond those that are directly subject to the regulation, a partial equilibrium approach may be insufficient to adequately estimate social cost. A general equilibrium approach, which captures linkages between markets across the entire economy, is most likely to add value when both cross-price effects and pre-existing distortions (e.g., taxes, regulations, market power) are expected to be significant (U.S. EPA 2017).252 253

Consider as an example a regulation that imposes emission limits on the electric utility sector. In the long run, we expect at least some, if not all, compliance costs are passed through to consumers as increases in the electricity price. Because electricity is used as an input in the production of many goods, the prices of these products may also increase to reflect the increase in their marginal cost of production. Households are affected through two channels: as consumers of these goods and as direct consumers of electricity. Increases in prices may cause households to alter their choices. For example, their consumption of energy-intensive goods and services may decrease relative to other goods; and the number of hours they are willing to work may change (e.g., when goods become more expensive, households can afford less with the same income; thus, their real wage has declined. On the margin, they respond by changing the number of hours worked). When an environmental regulation affects the real wage such that individuals opt to work fewer hours, it can exacerbate pre-existing tax distortions in the labor market (Goulder, et al., 1997). The impacts of a regulation also may interact with pre-existing distortions in other markets, which may cause additional impacts on welfare.254 In cases such as these, a general equilibrium approach is capable of identifying the nature and magnitude of the costs of complying with a regulation as they flow through the

250 Just, et al. (2005) detail methods for evaluating PE welfare changes across multiple related markets (see also Bullock 1993). Estimating welfare is only possible when the relevant relationships among the sectors (e.g., cross-price elasticities) are correctly specified. Plizer and Kopp (2005) and Kokoski and Smith (1987) provide additional discussion of when these methods are suitable for estimating social cost.

251 Computable general equilibrium (CGE) models are discussed in section 8.3.3. Analyses that use CGE models to estimate social cost of environmental regulation include Hazilla and Kopp (1990), Jorgenson and Wilcoxen (1990) U.S. EPA (1997a), U.S. EPA (2011), and Marten, et al. (2019).

252 Cross-price effects are measured by elasticities. For example, the cross-price elasticity of demand is defined as the percentage change in the quantity of product X demanded by consumers in response to a change in the price of product Y. If two markets are unrelated, then the cross-price effect is expected to be near or equal to zero. If they are substitutes, then the cross-price elasticity is positive. If they are complements, then it is negative.

253 The previous section shows how the social cost of a regulation can be estimated in a single market using PE analysis. The example demonstrates how a regulation may cause DWL in that market, reflecting a decline in economic welfare as measured by consumer and producer surplus. In reality, DWL already exists in many, if not most, markets as a result of taxes, regulations, and other distortions. When the imposition of a regulation causes a new distortion in one market, it may interact with pre-existing distortions in other markets, which may cause additional impacts on welfare.

254 See Text box 8.4 for a discussion of interactions that could also affect benefits estimation.
economy, including changes in substitution among factors of production, trade patterns, endogenous demands, and even inter-temporal consumption. These effects are partially or wholly missed by compliance cost and partial equilibrium approaches.

**Figure 8.3 - Labor Market with Pre-Existing Distortions**

![Diagram of labor market with pre-existing distortions](image)

Figure 8.3 illustrates how a regulation can interact with pre-existing tax distortions in the labor market. A pre-existing tax on wages causes the net, after-tax wage \( W' \) to be lower than the gross, pre-tax wage \( W \) by the amount of the tax. With this tax distortion, the quantity of labor supplied is \( L_0 \) and there is a DWL. When a new regulation is imposed in another market, raising production costs, one of its effects may be an increase in the price level. This increase in the price level will reduce the real wage and, given an upward sloping labor supply curve, the amount of labor supplied.\(^{255}\) This is shown in Figure 8.3 as a decrease in the net wage to \( W'' \) and a decrease in the amount of labor supplied to \( L_1 \). The interaction between new and pre-existing distortions is especially pronounced in the labor market. As shown in Figure 8.3, even a small reduction in the amount of labor supplied will result in a large increase in DWL.\(^{256}\) Similar interactions are likely to occur in other markets with significant pre-existing distortions (e.g., capital markets). In cases where they are likely to have a significant impact, analysts should incorporate these distortions into models used to estimate social cost.\(^{257}\)

### 8.2 Estimating Social Cost

When estimating social cost, the objective is to measure the incremental cost for each regulatory option under consideration. Incremental cost is defined as the additional cost associated with a new requirement relative to a baseline.\(^{258}\) Often when specifying a baseline from which to measure costs, the analyst needs to first identify what abatement activities are already in place. The costs associated with previously installed abatement controls are not counted toward the cost of the rule, as these occurred prior to the regulation under consideration and are therefore in the baseline.\(^{259}\) Understanding what abatement controls are already in place also aids the analyst in identifying what additional abatement control options may be available to further reduce emissions and the compliance costs associated with each regulatory option. See chapter 5 for a detailed discussion of baseline.

---

\(^{255}\) In GE analysis, all prices and wages are real, i.e., measured relative to a numéraire, a specific single price or weighted average of prices such as the GDP deflator. Here, the consumer price level rises relative to the numéraire. The result is a fall in the real wage — the nominal wage divided by the consumer price level.

\(^{256}\) The labor tax distortion affects individual labor supply decisions at the margin. While full-time workers may not change (or be able to change) hours worked in response to a fall in the real wage, part-time workers, workers in households with more than one full-time worker, or potential retirees may be more likely to adjust the number of hours they work. Parry (2003) discusses the theoretical and empirical basis for this depiction of the labor market.

\(^{257}\) Economists have long recognized these interaction effects (Ballard and Fullerton 1992), and a rich body of work has focused on them in the context of environmental regulation. In this literature, these interactions are known as the “tax-interaction effect.” If an environmental regulation raises revenue through a tax on pollution or other revenue raising provision, and the revenue is used to reduce pre-existing distortions such as taxes on wages, the tax-interaction effect may be offset. This is known as the “revenue recycling effect.” The offset may be partial, complete, and in some cases, the overall efficiency of the tax system may actually improve. The net result is an empirical matter, depending on the nature of the full set of interactions across the economy and how the revenue is raised. Goulder (2000) provides an accessible summary of the early literature. See also Parry and Bento (2000); Murray, et al. (2005); and Bento and Jacobsen (2007).

\(^{258}\) While this chapter focuses on the anticipated social costs of regulation, the same approach also applies in a retrospective setting (see Text box 8.1).

\(^{259}\) Other issues relevant to defining a defensible baseline for cost estimation include ensuring consistency in key assumptions across costs and benefits; treatment of anticipatory actions to meet regulatory requirements; and compliance with pre-existing regulations. See Chapter 5 for more discussion.
Text Box 8.1 - Planning for Retrospective Analysis

The same principles for prospective cost analysis also apply to estimating costs retrospectively for regulations that are already in place. Retrospective analyses may lead to improved methods for prospective analysis and ultimately improvements in regulatory design. However, while the importance of retrospective analysis in policy evaluation and reform is well-recognized, ex post studies of the costs and/or benefits of EPA regulations are relatively rare (U.S.EPA, 2014; Aldy, 2014).

Retrospective studies of the cost of EPA regulations are often relatively narrow in scope. Typically, retrospective studies are only able to evaluate a subset of the questions of interest: What mix of compliance strategies were utilized? Were ex ante cost estimates fairly accurate for the compliance strategies chosen? Were there unanticipated costs that were unaccounted for in the prospective analysis? Did costs of compliance change over time? Did technological innovation play a role in the compliance strategies chosen or the cost of compliance?

One challenge in conducting high quality retrospective cost analysis is a lack of data. The U.S. Census Bureau previously conducted the Pollution Abatement Costs and Expenditures (PACE) survey regularly. PACE collected establishment-level information from about 21,000 manufacturing, mining and electric utility plants on operating costs and capital expenditures disaggregated by media (i.e., air, water, solid waste) and pollution activity (e.g., end-of-pipe controls, production process enhancements to prevent pollution, disposal, and recycling). However, it has not been conducted since 2005 (U.S. Census, 2019).

Detailed source-specific information on compliance strategies and/or costs is useful in estimating the ex post cost of specific EPA regulations. If the analyst can identify the specific sources that were regulated, it may be possible to approximate the incremental cost of a regulation by comparing pollution abatement costs before and after a regulation becomes effective. Given sufficient data, analysts can use statistical techniques to control for other exogenous factors that affected the abatement strategies chosen as well as the cost of compliance. In addition, if a set of similar facilities remains unregulated over the relevant time period, then it may also be possible to compare the regulated firms’ behavior to a counterfactual. If cost data for several years before and after the regulation became effective is available, it may also be possible to understand how pollution abatement costs change over time. This would also potentially allow one to estimate how regulations induce technological change and affect employment.

Absent a plan to reinstate PACE or some other mechanism to collect disaggregated data from a wide array of regulated plants, retrospective analysis of the cost of regulations will be conducted opportunistically based on data availability. In addition, retrospective cost assessments have struggled with issues such as “how to evaluate a highly heterogeneous industry with a limited set of information, how to form a reasonable counterfactual, and how to disentangle the costs of compliance from other factors” (U.S. EPA, 2014).

The EPA could improve its ability to conduct retrospective cost evaluation by identifying analytic requirements when a regulation is promulgated. Data needs could be identified and avenues for ex post data collection integrated into the regulation (while also accounting for the cost and time needed for firms to collect such information). In this way, the EPA could learn from past experience and improve both policy designs and analytic approaches to prospective cost estimation.

It is important that analysts derive the most defensible central estimates of the compliance costs associated with identified abatement strategies, as they are the building block for developing social cost estimates. Social cost estimates should continue to rely on central assumptions and inputs that are well supported by standard engineering practice and the published scientific literature. In addition, analysts should ensure that: (1) the information supporting cost estimation is relevant for its intended use; (2) the scientific and technical procedures, measures, methods and/or models employed to generate the information are reasonable for, and consistent with, the intended application; and (3)

---

260 Note that the expected abatement strategies that underpin estimates of social costs also affect the expected change in the level and exposure to the environmental contaminant and therefore the benefits of the regulation.
the data, assumptions, methods, quality assurance, sponsoring organizations, and analyses employed to generate the information are well documented.²⁶¹,²⁶²

8.2.1 Compliance Cost Estimation

Recall that compliance costs are the additional costs that regulated entities incur to reduce or prevent pollution to comply with the regulation. There are a variety of different types of compliance costs including but not limited to:

- Treatment/Capture: The cost of any method, technique, or process designed to remove pollutants, after their generation in the production process, from air emissions, water discharges, or solid waste.
- Recycling: The cost of on-site or off-site processing of waste for an alternative use.
- Disposal: The cost of the final placement, destruction, or disposition of waste after pollution treatment/capture and/or recycling has occurred.
- Prevention: The cost of preventing pollution from being generated or contamination from occurring during the production process.

Entities that directly incur compliance costs to meet regulatory requirements may include firms, households, and government agencies. For example, firms normally incur costs to purchase and operate pollution control equipment; households might incur the costs of periodic inspections of pollution control equipment on vehicles; and government agencies may implement, administer, monitor, and enforce a regulation. In the case of product standards, compliance costs include the incremental cost of designing and manufacturing the compliant product relative to the already existing noncompliant product.

It is relatively straightforward to infer a value for compliance costs where an explicit monetary payment (e.g., purchasing pollution abatement equipment) is made. Compliance costs for which monetary values are not readily available are often more difficult to quantify. For example, the value and length of time households spend on vehicle inspections may be uncertain. Guidance on how to value time spent on such activities is discussed in Section 8.2.4. Compliance costs are also more difficult to quantify when, instead of installing abatement equipment, firms modify production processes to prevent emissions. Regardless of the ease with which various compliance costs can be estimated or what terminology is used to characterize them, if the compliance activities require resources that are redirected from other activities relative to the baseline, the value of those resources should be accounted for in compliance costs.

A compliance cost estimate reasonably approximates the social cost of a regulation when the value of the resources used for compliance generally reflect their social opportunity cost and prices or other producer and consumer behavior are not expected to change significantly as a result of a regulation. Determining whether compliance activities change prices or behavior requires: (1) estimates of supply and demand conditions, (2) an assessment of how compliance costs affect production costs, and (3) evidence of whether producers in the sector significantly change their level of production relative to one another. When compliance costs are used to estimate social costs, the analysis should provide evidence that justifies their choice.

It is common to refer to different categories of compliance costs, such as fixed or variable costs, as a way to systematically identify the costs that may result from a regulation. In practice, these categories of compliance costs may not be entirely distinct.

²⁶¹ At times, EPA uses externally derived (e.g., contractor, industry association, or advocacy group) cost estimates for its regulatory analyses. Any cost estimate produced by an external source and used by EPA in its analyses should meet these criteria.

²⁶² Some statutes require EPA to choose a regulatory option that is demonstrably affordable. In this case, analysts should continue to rely on the most defensible central estimate of costs. Estimating an upper bound (instead of a central estimate) of the compliance cost associated with the chosen option to demonstrate affordability will bias the net benefits of the regulation downward and/or could result in artificially low levels of stringency.
8.2.1.1 Fixed Costs

Fixed costs do not change with the level of production or abatement over a specific time period, often referred to as the short run. They are often one-time costs, or costs that are only occur once over the time horizon of the analysis, such as the installation of pollution control equipment. However, fixed costs may also refer to recurring costs that are independent of the level of production or abatement over a given time period. (Note that in the long run, virtually all fixed costs are variable.) Two common categories of fixed costs are described below.

**Capital costs** are costs related to the installation or retrofit of structures or equipment. These expenditures include materials and labor used for equipment installation and startup. Once equipment is installed, capital costs generally do not change with the level of abatement or production. Capital costs may also include changes to the production process.

**Research and development (R&D) costs** are incurred to develop new products, processes, or techniques. These costs are in addition to capital costs and should be accounted for when estimating compliance costs. Similarly, if a supplier to the regulated entity is expected to incur R&D expenditures in response to the regulation those costs should also be included in the estimate of the social costs.

In the case where a supplier incurs the R&D expense, care should be taken to avoid double counting; in the long run, the supplier will reflect these R&D costs in the price charged to the regulated entity. If social costs are estimated using the prices regulated entities pay, accounting for these additional R&D investments by the supplier, or any other resources reflected in prices charged to the regulated entity for that matter, do not need to be added to the estimate of social costs again.

R&D costs that have been incurred by a regulated entity or supplier prior to the announcement of the regulation should not be counted as a cost of the regulation as these costs are sunk. If past R&D costs are reflected in market prices of inputs sold by the supplier because the supplier now retains some market power based on its past R&D then, if possible, the cost associated with the R&D should be excluded from the social cost estimate. This is because the higher price, to the extent that it represents market power, reflects a transfer (defined below) from the regulated source to the supplier.

8.2.1.2 Variable Costs

Variable costs change with the level of production or abatement. They are the sum of the marginal cost for each unit that is produced. Common categories of variables costs are described below.

**Operating costs** are recurring expenditures associated with the operation and maintenance of equipment, including salaries and wages, energy inputs, materials and supplies, purchased services, and maintenance of equipment associated with pollution abatement. In general, operating costs increase with the level of abatement or the amount of production or use.

**Monitoring, reporting, and recordkeeping** costs are incurred to demonstrate or assure compliance with a regulation. They may be incurred by regulated entities or regulators and generally reflect the use of resources that should be

---

263 “Sunk costs” are costs that have already been incurred and cannot be reversed (e.g., existing pollution control equipment not currently in use or fully utilized). Care should be taken when identifying whether certain costs are sunk because they may actually be, in part, reversible, in which case there is an opportunity cost of using those resources under a tightening regulation (or value of avoiding use of those resources under a loosened regulation). For example, there may be scrappage value of pollution abatement equipment or avoidable variable costs such as operating pollution control equipment.

264 Use of some resources, especially energy, also can cause negative environmental or other externalities. Techniques for non-market valuation can be applied even when impacts are counted on the cost side of the ledger in a BCA (see Chapter 7).
Transactions costs are the costs incurred when buying or selling a good or service. They may include the costs of searching out a buyer or seller, bargaining, and enforcing contracts. Transaction costs reflect the use of real resources (e.g., time, equipment) and should be included in an estimate of social costs.

8.2.1.3 Transfers

Environmental regulations may also affect transfers. Transfers are shifts in money or resources from one part of the economy (e.g., a group of individuals, firms or institutions) to another in a way that does not affect the total resources that are available to society. In other words, the loss to one part of the economy is exactly offset by the gain to another. Since social cost represents the total burden that a regulation imposes on the economy, it nets out transfers. Examples of transfers include payments for most taxes, subsidies received, as well as higher prices that reflect the exercise of market power.

While transfers should be excluded from an estimate of social costs, the conditions leading to the transfer may create additional costs that should be accounted for in a partial or general equilibrium framework. For example, taxes are generally thought of as transfers between households or firms and government. However, when environmental regulation interacts with them in ways that distort behavior relative to what would occur absent government intervention in the marketplace, the welfare loss from these distortions should be included in an estimate of cost.

8.2.2 Measuring Social Cost

In instances when compliance costs do not fully represent all opportunity costs of a regulation, other metrics can be used based on partial or general equilibrium analysis. For instance, it is possible to estimate social cost by adding up the net change in consumer and producer surplus in all affected markets. Consumer’s equivalent variation (EV) and compensating variation (CV) are other measures that have been utilized. Both EV and CV are monetary measures of the change in household utility brought about by changes in prices and incomes resulting from the imposition of a regulation. As households are the ultimate beneficiaries of government and investment expenditures, the EV and CV measures focus on changes in consumer welfare, rather than on changes in demand.

The ability for EV or CV to capture the true social costs of a regulation depends on the inclusion of all relevant markets in the consumer’s utility function. In practice, analysts are often faced with incomplete data or limited knowledge on the interactions in assumed consumer preferences. In these cases, the EV or CV (and any other welfare-based) metric will be incomplete and represent only the opportunity costs from changes in explicitly represented prices. Efforts should be

---

265 Transfers are important for understanding how a regulation affects the private cost of a regulation for different groups. Thus, they are included in an economic impact analysis. See Chapter 9.

266 An exception is when one group has economic standing in the analysis, and the other does not. See Chapter 5.

267 Regulations may also create scarce compliance assets such as allowances in cap-and-trade systems. Generally, the gratis receipt of or any payments for allowances are also a transfer (see, for example, Burtraw and Evans 2009).

268 Appendix A describes the relationship between consumer surplus, equivalent variation and compensating variation. EV and CV are particularly well suited for partial and general equilibrium analysis because both modeling frameworks require the explicit characterization of consumer preferences. Calculating EV and CV requires only pre- and post-policy price and utility levels.

269 The difference between them is based on whether one assumes that the change will occur (EV) or is not yet in place (CV). For instance, EV measures what a consumer would be willing to pay to avoid an increase in prices (and thus, a decline in real income) resulting from a regulation going into effect. In contrast, CV measures how much a consumer would need to be compensated to accept an increase in price (and a decline in real income) such that the consumer achieves the same level of utility experienced absent the regulation being analyzed.
made to be as inclusive as possible in defining consumer preferences and to acknowledge missing relevant components.\textsuperscript{270}

The social cost of a regulation is generally not the same as a change in GDP or aggregate consumption (U.S. EPA 2017). As measures of social cost, changes in GDP and aggregate consumption both miss potentially important regulatory effects, such as impacts on leisure demand of the demand for nonmarket goods.\textsuperscript{271} GDP is comprised of more than just changes in consumption as it is a measure of total economic output.\textsuperscript{272} Thus, changes in GDP do not necessarily correlate well with impacts on individual welfare (Paltsev and Capros, 2013). For instance, a regulation that requires firms to install new capital in a given year will see an increase in investment. However, capital also affects the availability of goods and services that can be consumed over a much longer time period. As a result, GDP effectively double counts the new capital installed since investment and consumption are both components of GDP.

8.2.3 Evaluating Costs Over Time

After the imposition of a new environmental regulation, the economy can be expected to adjust over time to a new long-run equilibrium. Compliance costs are treated as permanent additions to the cost of production for a firm, while effects in other sectors outside of those directly regulated by the EPA are often incurred once the economy adjusts to a new equilibrium.

However, in some contexts it is possible that firms and/or consumers may incur additional short-term costs during the period when the economy is adjusting to the new equilibrium. These are known as transition costs. Examples include costs to train workers to use new equipment, search costs as some workers seek employment in other sectors, and additional costs associated with initially limited availability of new monitoring or abatement equipment. It is also possible that at least some factors of production are fixed initially, limiting the ability of firms to respond quickly to new regulatory requirements. For instance, contractual or technological constraints may prevent firms from fully adjusting their input mix or output decisions until those contracts expire or technology is ready to be replaced. If these types of adjustment costs are substantial, a sole focus on long run costs may underestimate the total social cost of regulation.

Thus, it is important to consider both short- and long-run effects when measuring costs over time. In addition, analysts must make choices about the time horizon of the analysis, the use of a static versus a dynamic framework, discounting, and technical change, employment effects, and effects on market structure.

8.2.3.1 Time Horizon

The time horizon for calculating producer and consumer adjustments to a new regulation should be considered carefully. The analyst should strive to estimate the present value of all future costs of a regulation (see Chapter 6). If the analyst is only able to estimate a regulation’s costs for one or a few representative future years, she must take care to ensure that the year(s) selected are truly representative, that no important transitional costs are effectively dismissed by assumption, and that no one-time costs are assumed to be on-going.

In the short run, at least some factors of production and consumer demand are fixed. If costs are evaluated over a short period of time, then contractual or technological constraints can prevent firms from responding quickly to increased compliance costs by adjusting their input mix or output decisions. In the long run, by contrast, all factors of production

\textsuperscript{270} EV and CV can also provide a complete welfare metric (incorporating both benefits and costs) if non-market goods are explicitly accounted for in consumer utility functions. However, these metrics are often only used to assess the social cost of a regulation because traditional economic models do not yet incorporate non-separable benefits, or explicit linkages between environmental quality and economic costs (see Text box 8.2).

\textsuperscript{271} See Paltsev et al. (2009), U.S. EPA (2011) and Paltsev and Capros (2013) for examples of how these measures differ in specific policy contexts.

\textsuperscript{272} It is also the case that transfer payments, which are excluded from BCA, are subsumed within G. In addition, while changes to trade patterns due to a regulation may be reflected in both GDP and welfare, they are not necessarily equivalent measures (Paltsev et al., 2009; Paltsev and Capros, 2013).
are variable. Firms can adjust any of their factors of production in response to a new regulation and can even change their production processes. Similarly, consumers, including producers in other sectors, may not be able to adjust demand for the output of the regulated sector in the short run but have more flexibility in the long run. The time horizon for the analysis should be long enough to capture any flexibility the regulation provides firms in their compliance approach. However, if transition costs seem likely, analysts should also consider presenting evidence that sheds light on the length of the transition period and the magnitude of these costs. In some cases, regulatory requirements are phased in gradually over time, either explicitly through graduated compliance dates or requirements or implicitly through characteristics like vintage differentiation (i.e., varying regulatory requirements based on the age of the plant). For example, consider a regulation that enacts more stringent requirements on new sources of a pollutant. Selecting a time period of analysis that is early in the program when only a few new sources of production are affected may not accurately capture a future year in which most sources of production are new for the purposes of the regulation. A regulation also may influence the rate at which old sources are replaced by new ones. Chapter 5 contains additional discussion regarding how to determine the most appropriate time horizon for analysis.

8.2.3.2 Dynamics

One key decision for the analyst is whether to assume that economic conditions are invariant over time (i.e., static) or attempt to account for expected future changes in prices and economic activity (i.e., dynamic). Costs that are estimated at a given point in time or for a selection of distinct points in time and compared to the baseline are static. They provide snapshots of costs faced by firms, government, and households but do not allow behavioral changes from one time period to affect responses in another time period. A dynamic framework, one that explicitly captures trade-offs across time periods, allows for this possibility.

In most cases, a regulation will continue to have economic impacts after its initial implementation. If these intertemporal impacts are likely to be significant, they should be included in the estimation of social cost. Pizer and Kopp (2005) note that static productivity losses from environmental regulations are amplified over time due to their effect on capital accumulation (a lower capital stock over time reduces economic output and therefore welfare). A static model would miss this effect. In some cases, the potential effect of a regulation on long-term growth may be significantly larger than its effect on the regulated sector alone. In addition to these capital-induced growth effects, the evaluation of costs in a dynamic framework may be important when a proposed regulation is expected to affect product quality, productivity, innovation, and/or changes in markets indirectly affected by the environmental policy. Dynamic effects also impact net levels of measured consumer and producer surplus over time. See section 8.4 for how a regulation’s potential dynamic impacts affect model choice.

Conceptually, a dynamic framework allows the analyst to specify the process by which the economy moves between equilibria in response to a regulation across time. In practice, however, economists have more experience characterizing long-run equilibria than the pathways between them. While shorter run equilibria can be approximated by treating some factors of production as fixed (e.g., labor or capital), very near-term transitional costs are typically ignored in the modeling approaches discussed in Section 8.3.

8.2.3.3 Discounting

---

273 Note that a comparative static framework compares snapshots of key economic outcomes before and after a change in an exogenous factor such as a regulatory requirement.

274 Pizer and Kopp (2005) estimated that the “additional cost of this accumulation effect on welfare can be as much as 40 percent above the static cost that ignores changes in capital stock.” Hazilla and Kopp (1990) and Jorgenson and Wilcoxen (1990) also showed that this effect is potentially significant. It is important to note, however, that this conclusion is based on studies of large-scale changes in environmental regulation (i.e., the welfare effects of the 1972 Clean Water and 1977 Clean Air Acts).
Costs that occur over time must be properly and consistently discounted to allow for legitimate comparisons with benefits. Procedures for social discounting in economic analyses are reviewed in considerable detail in Chapter 6.

There are two applications of discounting that are closely related to the modeling of social costs. First, when modeling firms’ behavior, the analyst should use a discount rate that reflects the industry’s cost of capital, just as a firm would. The social cost of the regulation, on the other hand, is calculated using the social discount rate, the same discount rate used for estimating the benefits of the regulation. Section 6.4 provides additional details on the choice of discount rate when modeling behavior, such as firms' compliance decisions. Second, when a dynamic general equilibrium model is used to estimate social costs, any displacement of investment due to the regulation has already been accounted for and the social cost estimates should only be compared to present value estimates of benefits discounted at the consumption discount rate.

8.2.3.4 Technical Change and Learning

Estimating the social cost of an environmental regulation over a relatively long time horizon requires assumptions about future technological change. Jaffe et al. (2002) lay out a conceptual framework for understanding how technological change in response to environmental regulation may affect the relationship between inputs and output, ultimately reducing the unit costs of production. Consider an economy where growth in aggregate production over time is a function of the rates of growth in inputs such as labor, capital, and environmental quality. If an environmental regulation changes overall productivity over time in an unbiased way, then inputs will be used in the same proportion as before to produce output. An environmental regulation that instead affects the growth rate of one or more inputs over time, for instance only capital, changes the relative productivity of inputs to production.

Compliance with an environmental regulation may result in the adoption of existing technology, improvement or application of existing technology to a new use, and/or development of entirely new technologies or processes (Sue Wing, 2006). Whether it is more appropriate to capture these compliance responses as affecting overall productivity or the relative productivity of one or more inputs is an empirical matter.

Despite its importance as a determinant of economic welfare, the process of technical change is not well understood. Different approaches to environmental regulation present widely differing incentives for how compliance is achieved and the relative role of technological innovation (e.g., Fischer et al., 2003). As a result, the same environmental end may be achieved at significantly different costs, depending on the pace and direction of technical change.

The empirical economics literature also has noted that variable costs of production or environmental abatement tend to decline over time with cumulative experience. While the explanations for why this occurs vary (e.g., workers learn from mistakes and determine shortcuts; ad hoc processes become standardized), the evidence for such “learning” is compelling enough that analysts should consider the potential for learning effects when analyzing the cost of regulation. The EPA’s Advisory Council on Clean Air Compliance Analysis recommends that default learning effects be

---

275 It is equally important to properly discount cost estimates of different regulatory approaches to facilitate valid comparisons.

276 For instance, the realized costs of Title IV of the 1990 Clean Air Act Amendment’s SO₂ Allowance Trading program are considerably lower than initial predictions, in part due to incentives to innovate in response to the policy (e.g., Bellas and Lange 2011, Frey 2013). See chapter 4 for a discussion of how different regulatory approaches may affect innovation.

277 OMB’s Circular A-4 asserts that a cost analysis should incorporate credible changes in technology over time, stating that "...retrospective studies may provide evidence that 'learning' will likely reduce the cost of regulation in future years" (OMB 2003).
applied even when sector- or process-specific empirical data are not available (U.S. EPA 2007b).\textsuperscript{278,279} See Section 5.4.1.4 for additional discussion of technical change and learning.

### 8.2.3.5 Social Cost and Employment Effects

Recall that compliance costs include the value of labor for activities such as the installation and maintenance of abatement technologies as well as monitoring, recordkeeping, and reporting. Conceptually, the social cost of a regulatory action also incorporates the value of induced changes in consumption and deadweight loss from changes in the use of time (i.e., due to pre-existing tax distortions in the labor market). Employment effects more generally, such as those driven by labor-leisure choice, are not part of social costs under two commonly-held assumptions: if the economy is at full-employment (i.e. every worker who wants a job at the prevailing wage has one) and with \textit{de minimus} transition costs (Ferris and McGartland 2014). Typically, a regulation reallocates workers among economic activities - increasing employment in some industries and decreasing it in others - rather than affect the general employment level (Arrow, et. al 1996).

However, there are circumstances when the assumptions of full-employment and no transition costs are less compelling due to the potential for unmeasured welfare effects of regulation on both short run and long run employment. For instance, in the short run workers employed in labor markets that experience changes in demand due to a regulation may incur transition costs as they change jobs or workforce status.\textsuperscript{280} Transition costs may be higher during a recessionary period, when overall labor demand is already reduced due to nationwide declines in production, which can lengthen the time needed to locate new employment.

Economists do not have a unified theory that incorporates employment impacts into BCA (Ferris and McGartland 2014). There also are many conceptual and practical challenges to evaluating the social costs associated with employment effects. For instance, transition costs for workers experiencing involuntary job loss and unemployment may overstate social costs if they include economic rents (e.g., if displaced workers are highly paid relative to their productivity, then these rents are in fact transfers) (Davis, et al. 2011). More generally, the costs to displaced workers to relocate across labor markets and the impacts of job loss on health or well-being are not well-measured (Smith 2015).\textsuperscript{281} For these reasons, the social costs of changes in employment due to regulation are difficult to quantify. Therefore, they typically are not included in the BCA. However, they should be characterized in the economic impact analysis, as explained in Chapter 9.

### 8.2.3.6 Effects on Market Structure and Entry and Exit

Market power or existing market regulations are examples of distortions that may influence the social cost of a new regulation. It is common to assume that regulated markets as well as other markets affected by the regulation are perfectly competitive, and both producers and consumers are price takers. However, if either producers or consumers in the market hold some market power or existing regulations restrict the behavior of market participants (e.g., public utilities), that structure can influence the response of market participants to regulatory requirements in ways that affect

\textsuperscript{278} A useful description of the calculations used to identify a learning curve are found in van der Zwaan and Rabl (2004). Learning rates for 26 energy technologies are described in McDonald and Schrattenholzer (2001). The U.S EPA (2016) reviews learning rates in the published literature for manufacturing and electric utilities with a specific focus on the production of transportation-related goods (e.g., cars, ships, trucks). Note that the empirical estimates in the literature represent a biased sample, since they only represent technology that has been successfully deployed (Sagar and van der Zwaan 2006).

\textsuperscript{279} Note that cost decreases associated with technological change and learning may have additional costs associated with them such as training costs. See Section 8.2.3 for a discussion of transition costs.

\textsuperscript{280} These costs may be higher for certain categories of workers such as those whose skills are specially adapted for the sector experiencing reduced labor demand. The distribution of costs across different types of workers is discussed in Chapter 9.

\textsuperscript{281} For more discussion, see Davis, et al. (2011), Ferris and McGartland (2014), and Smith (2015). Efforts to quantify the social cost of employment effects due to environmental regulation include Smith (2015), Rogerson (2015), and Hafstead and Williams (2019).
social cost. This can occur when the baseline market structure affects the response of affected sources to regulation or when regulation leads to a change in market structure, changing the deadweight loss associated with this type of pre-existing market distortion.

Increases in production costs due to compliance requirements may lead to a reduction in output for the affected industry, which - if met by a net reduction in the number of firms - could reduce competition. The effect of competition on social costs is distinct from the effect of increased costs due to compliance requirements shown in Figure 8.2. While a regulation may affect the number of producers in an industry, and therefore the market concentration in that industry, it is not necessarily the case that this will lead to an increase in market power.

Environmental regulations can potentially affect the number of firms and market structure of the regulated sector by raising production costs, modifying economies of scale, or affecting barriers to entry. For example, spatial heterogeneity in the stringency of environmental regulations or compliance costs, and in turn the impact on production costs, can lead to market consolidation at existing firms (e.g., Gray and Shadbegian, 2007). Market structure can also be affected by the impact of compliance activities and abatement technologies on the minimum efficient scale for firms in the industry. Positive economics of scale for abatement technologies can lead to reduced entry and greater exit, as reviewed by Millimet et al. (2009). Similarly, larger firms in the industry may have a competitive advantage in the presence of economies of scale (Dean, et al. 2000). Differences in product offerings by producers may also affect market structure. If some firms subject to new product standards already have compliant products, they will have a distinct advantage over others. Regulations can also create barriers to entry either due to vintage differentiated standards, whereby new entrants have stricter standards, or through the control of patents on abatement technologies held by incumbents who innovated as a result of the regulation.

Existing economic regulation may also affect market structure and the response of affected sources to an environmental regulation and, in turn, social costs. A well-known example is in the context of electric utility regulation where investment and retail prices are subject to regulation in some states to assure that producers do not exercise market power. Electricity producers subject to state-level investment and price regulations have a history of complying with environmental regulations in ways that differ from producers in states without these types of regulations (e.g. Parry, 2005; Burtraw and Palmer 2008; Fowlie 2010).

Analysts should assess the existing market structure in the baseline and appropriately account for this when estimating the compliance behavior of market participants. Similarly, analysts should assess any expected changes to market structure as a result of the regulation and how they may impact social cost. See Chapter 9 for discussion of how changes in market structure may also affect the composition and distribution of costs within a sector.

### 8.2.4 Valuing Time

Compliance with environmental regulations changes the use of productive resources, including people’s time. Often, these changes occur at the workplace where labor is required to undertake pollution control activities. Less often, time outside of the workplace is also affected; for example, product bans might cause consumers to switch to substitutes that occupy more time. Changes in time use can affect the social costs of a regulation. The EPA has produced separate guidance on how to value work time and nonwork time in regulatory analyses. We summarize the recommendations below, but analysts should consult U.S. EPA (2020) for a detailed discussion.

The opportunity cost of worktime is determined by the value of the marginal product that would have occurred absent the regulation. As a proxy for this opportunity cost, analysts should use the employer’s cost of employing a worker,

---

282 It is theoretically ambiguous as to whether a reduction in output will be accompanied by a net reduction in the number of firms, through increased exit and reduced entry, for common regulatory designs such as performance standards (e.g., Requate, 1997; Lahiri and Ono, 2007).
consisting of the wage, fringe benefits, and any overhead costs.\footnote{Overhead costs are employer costs associated with labor, but not paid directly to workers, such as the value of personnel services and training activities. For more information on how wages, fringe benefits, and overhead costs are defined, see Section 2.1 of U.S. EPA (2020).} The value of worktime will vary based on the industries and occupations affected by the regulation. If overhead data are not available, U.S. EPA (2020) recommends that analysts use a default multiplier applied to an unloaded wage rate for the value of worktime.\footnote{The default multiplier in U.S. EPA (2020) reflects the national average 46\% ratio of fringe benefits to wages reported by BLS’s ECEC. It is also reflective of multiplier values used in prior analyses that are based on industry and occupation-specific benefit and overhead rates affected by EPA regulations.} EPA (2020) provides links to data sources for wages, benefits and overhead rates that normally are included when valuing worktime.

Nonwork time includes time spent on leisure, household production, or other unpaid activities. Its opportunity cost may vary by the types of activities foregone; the utility derived from the activity that occupies time; whether workers have a continuous choice over their hours of paid work; the socioeconomic characteristics of affected individuals; and more. As a proxy for the opportunity cost of nonwork time, analysts should add the value of voluntary fringe benefits to the the wage net of any taxes paid by workers to federal, state, and local governments on earned income.\footnote{Voluntary fringe benefits are the categories of employer-paid benefits that are not legally required and include paid leave, supplemental pay (e.g., for overtime), insurance, and retirement and savings plans.} Table 8.1 summarizes the recommended approach and data sources for estimating work time and non-work time.

In unusual circumstances, analysts may have access to information that allows an alternative approach to estimating the value of work time or nonwork time. If utilized, analysts should explain why the alternative is preferred to the approach recommended here and in U.S. EPA (2020).

### 8.2.5 Compliance Assumptions

In most cases, analysts should develop baseline and policy scenarios that assume full compliance with existing and newly enacted (but not yet implemented) regulations. Assuming full compliance focuses the analysis on the incremental effects of the new regulatory action without double counting benefits and costs already accounted for in previous regulatory analyses. That said, it is important to determine whether specific policy options are more likely to result in compliance issues or may be more difficult to enforce. In such cases, it is important to evaluate these effects (e.g., options that require monitoring and reporting may have higher costs, but compliance is easier to verify) and explore whether alternative options would result in improved compliance and/or easier enforcement.

Assumptions about compliance behavior in the baseline and policy scenarios should be clearly explained in the analysis. When compliance rates are uncertain or expected to vary across policy options, analysts should explore the sensitivity of the results to these assumptions. See Section 5.4.2 in Chapter 5 for a more in-depth discussion.

### 8.3 Models Used in Estimating the Costs of Environmental Regulation

Several types of models have been used to estimate the social costs of environmental regulation. They range from models that estimate costs in a single industry (or part of an industry) to models that estimate costs for the entire U.S. economy. In this section, we focus on three main model types: compliance cost models, partial equilibrium (PE) models, and computable general equilibrium (CGE) models. Input-output and input-output econometric models should not be used to estimate social cost; however, these approaches and their limitations are also described. Analysts are encouraged to consult with the National Center for Environmental Economics (NCEE) when choosing a model for cost estimation.
Table 8.1 Estimating Work and Non-Work Time

<table>
<thead>
<tr>
<th>Type of Time Affected</th>
<th>Displaced activity</th>
<th>Estimation approach</th>
<th>Data sources</th>
</tr>
</thead>
<tbody>
<tr>
<td>Worktime: Tasks</td>
<td>Other market work</td>
<td>Employer costs of</td>
<td>• BLS OES(^1) or ECEC data(^1) on wages and fringe benefits</td>
</tr>
<tr>
<td>completed while</td>
<td>in the same industry</td>
<td>labor = Wages + Fringe benefits + Overhead costs</td>
<td></td>
</tr>
<tr>
<td>working for pay</td>
<td>and occupation as</td>
<td></td>
<td>• For overhead costs, use industry specific data as available</td>
</tr>
<tr>
<td></td>
<td>workers asked to</td>
<td></td>
<td>• If overhead are not available, use the recommended multiplier to obtain a fully-loaded wage</td>
</tr>
<tr>
<td></td>
<td>complete the</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>required tasks</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nonwork Time: Tasks</td>
<td>Other nonmarket</td>
<td>Individual valuation of time = (Wages – Taxes on earned income) + Voluntary fringe benefits</td>
<td></td>
</tr>
<tr>
<td>completed outside</td>
<td>activities such as</td>
<td></td>
<td>• BLS OES or ECEC data on wages and voluntary benefits</td>
</tr>
<tr>
<td>of paid worktime</td>
<td>leisure and</td>
<td></td>
<td>• Adjust wage estimates using Census CPS(^1) data on median household income before and after taxes to estimate average income tax rate(^1)</td>
</tr>
<tr>
<td></td>
<td>nonmarket work</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>


In practice, some models are simple enough to be implemented in a spreadsheet. Others may consist of systems of hundreds or even thousands of equations that require highly specialized software. Many models are data intensive.\(^{286,\,287}\) Given model complexity, a simple model that captures key economic features may be useful to identify which aspects of the regulatory options under consideration likely matter from a cost perspective and therefore warrant further investigation in a more complicated model. Likewise, the use of a simple analytic general equilibrium approach is a less resource-intensive way to build intuition within an internally consistent framework before utilizing a CGE model (U.S. EPA 2017). Analysts should rely on a model that is “no more complicated than necessary to inform the regulatory decision” (U.S. EPA 2009).

In some cases, use of more than one type of model may be warranted. Specifically, a more aggregate CGE analysis may complement the cost estimates of a detailed compliance cost or PE model (U.S. EPA 2017). For example, direct cost estimates from a compliance cost model can be used as an input into a PE or CGE model. In some cases, models also can be linked to combine the sectoral detail of a PE approach with the economy-wide features of a CGE model. Text Box 8.2 discusses linking models. Table 8.2 summarizes key attributes by model type.

When selecting a type of model, it is important to evaluate whether it is the most appropriate for the question at hand (i.e., fit for purpose) and does a reasonable job of approximating the market(s) and behavioral responses of interest. Most model types involve tradeoffs between different strengths and weaknesses. Below are several factors that may be

\(^{286}\) Data requirements for these models vary, though advances in computing power, data availability, and more user-friendly software packages continually reduce the barriers to sophisticated model-based analysis. Refer to Chapter 9 for a discussion of the public and private data sources that can be used for cost estimation.

\(^{287}\) Analysts should take great care in ensuring the quality of a model’s data and specifications. See section 8.4 for a discussion of approaches to parameter selection, and ways to address parameter and model uncertainty.
Text Box 8.2: Linking Models

CGE models are aggregate representations of the economy that allow an analyst to capture the interactions of producers and consumers as changes in prices and quantities in the regulated sector percolate through the rest of the economy. These economy-wide interactions are captured through exogenously specified elasticities of substitution that approximate detailed demand and supply responses from the policy. However, CGE models often do not explicitly represent details that may be important for identifying how regulated entities may respond to a regulation, such as the types of compliance methods that are available.

Partial equilibrium and compliance cost approaches typically do not suffer from a lack of detail. They often have technology-rich representations that reflect the range of salient characteristics for regulated sources as well as installation and operation costs for each individual compliance technology. However, often demand and supply are specified in a very simplified way and interactions with other potentially affected markets are not considered.

Much could be gained if these two modeling approaches were linked in a coherent and sensible way to take advantage of the technological detail of compliance cost or partial equilibrium models and the theoretically consistent economic structure of CGE models (Böhringer and Rutherford, 2008). There have been a number of studies, many in the energy context, that have leveraged such linkages (e.g., Cai and Arora 2015; Rausch and Karplus 2014; Kiuila and Rutherford 2013; Lanz and Rausch 2011; Sue Wing 2006; Schafer and Jacoby 2005; McFarland, et al. 2004). The SAB (U.S. EPA 2017) recommended that EPA make linking more aggregate CGE models to more detailed models of households, industries, or sectors a research priority. It signaled a clear preference for two-way linkages between models: the CGE model simulates prices and investment for use as inputs to the compliance cost or PE model, while the compliance cost or PE model computes technology capacities and output supplies that are used as inputs by the CGE model. The two models are run in an alternating fashion, iterating until both solutions converge.

It is important to note that, in practice, any linking exercise is dependent on the information available from the sector model and the representation of relevant sectors and markets in the CGE model. As the information and available models differ significantly across separate regulatory analyses, any application of linking also may present unique challenges and considerations.

To link a compliance cost model with a CGE model, the accounting of outputs and inputs between the two models needs to be sufficiently aligned. To do this, it is important to disaggregate compliance costs into the factors (e.g., labor, capital, energy, materials) that correspond to the inputs to the sector’s production function as specified in the CGE model. However, this is often not a straightforward exercise. For instance, the fixed cost of a compliance method may include both the capital used for a compliance technology and the labor to install it. Likewise, variable costs may include materials as well as labor for maintenance. However, in both cases the shares of the compliance cost from the specific inputs are rarely available. It is also a challenge to aggregate compliance cost information up to the sector level for the purpose of linking to the CGE model. A compliance cost model often provides information on the expected compliance behavior and cost for each affected entity. The CGE model usually represents a sector with a single representative firm.

Many of the challenges of linking compliance cost models to CGE models also apply to linking CGE models with PE models. A further challenge is that PE and CGE models may have different baseline forecasts and elasticities of demand and supply for various goods and factors that need to be reconciled. Similarly, PE and CGE models may represent demand and supply for these markets in different ways (e.g., different functional forms and underlying behavioral assumptions). Another challenge is ensuring that assumptions with regard technological innovation are consistent between the models.

helpful in choosing a model. Criteria for ensuring any specific model is appropriate and of sufficient quality for analyzing the effects of a regulation are also discussed in Text Box 5.1.

-----

This list of factors is informed by Industrial Economics, Inc. (2005).
• **Types of impacts being investigated.** Models differ in their abilities to estimate different types of costs.

• **Geographic scope of expected impacts.** Some models are well-suited for examining regional or local impacts but may not capture the full range of costs at the national level and vice versa.

• **Sectoral scope of expected impacts.** Some models are highly aggregate and lack the detail necessary to capture important aspects of compliance behavior within a single sector. Likewise, highly detailed sector models often do not capture effects on other sectors and may not adequately capture demand response.

• **Expected magnitude of impacts.** A model that is well suited for estimating the cost of a regulation with large effects may have difficulty estimating the cost of a regulation with relatively smaller expected effects, and vice versa.

• **Expected importance of interactions and feedbacks with other sectors.** When regulations are expected to have substantial effects on the broader economy, it is important to choose a model that can capture those effects.

Usually, some combination of the above factors will determine the most appropriate type of model for a specific application. Analysts should present a reasoned discussion of the factors that inform their model choice. Analysts should describe the main upstream and downstream sectors affected, whether close substitutes to the regulated good are available, the extent to which the goods affected are substitutes or complements to leisure, and the existence of pre-existing distortions in affected sectors (e.g., subsidies, imperfect competition, other regulations or externalities). Evidence from the literature such as supply and demand elasticities that indicate market responsiveness (e.g., of consumers, input markets, substitutes and complements) will aid the analyst in justifying model choice. Ultimately, models need to be supported by the data: for example, a single-market PE analysis requires demand and supply elasticities, while a multi-market or CGE analysis requires cross-price elasticities.

### 8.3.1 Compliance Cost Models

Compliance cost models are used to estimate the direct costs of compliance with a regulation. Estimates by engineers and other experts are used to produce algorithms that characterize the changes in costs resulting from the adoption of various compliance options and are usually determined for individual facilities or for categories of model facilities with varying baseline characteristics. To estimate the control costs of a regulation for an entire sector, disaggregated data that adequately reflect the industry’s heterogeneity are used as an input into the model. The disaggregated cost estimates are then aggregated to the industry sector level. These models are most informative when the data are available to capture heterogeneity across facilities, both in terms of individual characteristics (e.g., facility age and production technology, input costs) and compliance options.

The structure of compliance cost models can vary depending on the scope of an analysis. For instance, compliance cost models may include many of the categories of costs previously described in section 8.2.1 (e.g., capital costs, operating and maintenance expenditures, monitoring, measurement, and reporting costs). Moreover, some compliance cost models are designed to allow the integrated estimation of control costs for multiple pollutants and multiple regulations. Some models account for cost changes over time, including technical change and learning. While most compliance cost models are for facilities within a specific industry, they may also be models of households.

While precise estimates of compliance costs are an important component of any analysis, recall that, in cases where the regulation is not expected to significantly affect market supply and demand in the regulated market, compliance costs can be considered a reasonable approximation of social cost. Compliance cost models usually focus on the supply side because regulations are typically imposed on producers. In circumstances where producer and consumer behavior are appreciably affected, these models are not able to provide estimates of changes in industry prices and output resulting from the imposition of a regulation.
Advantages:

- Compliance cost models often contain significant industry detail and can provide relatively precise estimates of the costs incurred by regulated sources (or categories of regulated sources) when complying with a regulation.
- Once constructed, compliance cost models require a minimum of resources to implement and often are relatively straightforward to use and easy to interpret.

Limitations:

- As they usually focus on the supply side and do not capture changes in production among affected sources, compliance cost models can only provide estimates of social cost in certain limited cases.
- Compliance cost models are often limited to estimating the costs of complying with regulatory requirements for a single industry.

Linear Programming Models

Often linear programming models are used in the analysis of EPA regulations to estimate compliance costs. Linear programming models minimize (or maximize) a linear objective function by choosing a set of decision variables, subject to a set of linear constraints. In EPA’s regulatory context, the objective function is usually to minimize compliance costs incurred by the regulated sources. The decision variables represent the production and compliance choices available to the regulated entities. The constraints may include available technologies, productive capacities, fuel supplies, and regulations on emissions.

Although linear programming models can be constructed to examine multiple sectors or even economy-wide effects, they are commonly focused on a single sector. For the regulated sector, a linear programming model can incorporate a large number of technologies and compliance options, such as end-of-pipe controls, fuel switching, and changes in plant
operations. Similarly, the model’s constraints can include multiple regulations that require simultaneous compliance. The objective function usually includes the fixed and variable costs of each compliance option.

In addition to compliance costs, the outputs from the model may include other related variables, such as projected input use, emissions, and demand for new capacity in the regulated industry. In some cases, linear programming models may also include supply and demand representations (e.g., elasticities) of multiple markets and therefore more closely resemble the partial equilibrium models described in Section 8.3.2.

While the estimated change in expenditures incurred by the regulated sector may be of policy interest, it is not equal to social cost when input or output prices change. If the linear programming model captures changes in market prices in response to the policy, then it is possible to use the model outputs to estimate a partial equilibrium estimate of social cost (e.g., changes in producer and consumer surplus).

### 8.3.2 Partial Equilibrium Models

In cases where the effects of a regulation are confined to a single or a few markets, partial equilibrium models that incorporate anticipated demand and supply responses can be used to estimate social cost.

Inputs into a partial equilibrium model may include regulatory costs estimated using a compliance cost model and the supply and demand elasticities for the affected market. The model then can be used to estimate the change in market price and output. Changes in producer and consumer surplus reflect the social cost of the regulation.

In a partial equilibrium model, the magnitude of the impacts of a regulation on the price and quantity in the affected market depends on the shapes of the supply and demand curves in the region at which expected changes are to occur. The shapes of these curves reflect the underlying elasticities of supply and demand. These elasticities either can be estimated from industry and consumer data or taken from previous studies.

If the elasticities used in an analysis are drawn from previous studies, they should reflect:

- a similar market structure and level of aggregation;
- the appropriate spatial resolution (i.e., local, regional, or national);\(^{289}\)
- current economic conditions; and
- the appropriate time horizon (i.e., short or long run).

In some cases, if the effects of a regulation are expected to spill over into adjoining markets (e.g., suppliers of major inputs or consumers of major outputs), partial equilibrium analysis can be extended to these additional markets as well.

**Advantages:**

- Because they usually simulate only a single market, partial equilibrium models generally have fewer data requirements relative to a CGE approach and are more straightforward to construct.
- Partial equilibrium models are comparatively easy to use and interpret.

**Limitations:**

- Partial equilibrium models are limited to cost estimation in a single or small number of markets and do not capture broader effects in the overall economy.

---

\(^{289}\) For instance, Bernstein and Griffin (2006) estimated short-run price elasticities of demand for electricity in the United States that varied from -0.04 to -0.31 by region, and long-run price elasticities of demand for electricity that varied from about -0.05 to almost -0.6 by region.
8.3.3 Computable General Equilibrium Models

The most appropriate type of model to estimate the social cost of a regulation in a general equilibrium framework is a computable general equilibrium model. This type of model is comprehensive and internally consistent, accounting for budgetary and resource constraints operating throughout the economy. A key advantage over the other types of models discussed in this section is its ability to capture interactions between economic actors (often delineated with multiple sectors and regions) and with pre-existing distortions (e.g., taxes, other regulations or externalities) across the entire economy. Relative to PE and compliance cost approaches, however, CGE models are highly aggregate depictions of the economy and offer less detail on each sector. They may also be more difficult to parameterize.

CGE models assume that an economy can be characterized by a set of conditions in which supply equals demand in all markets. When the imposition of a regulation alters conditions in one market, a general equilibrium model determines a new set of relative prices for all markets that return the economy to a long-run equilibrium. These prices in turn determine changes in sector outputs and household consumption of goods, services, and leisure in the new equilibrium. In addition, the model determines a new set of relative prices and demand for factors of production (e.g., labor, capital, and land), the returns to which compose business and household income. The social cost of the regulation is estimated in CGE models as the change in economic welfare in the post-regulation, simulated equilibrium compared to the pre-regulation, “baseline” equilibrium.290

CGE models are built using structural micro-theoretic foundations to capture behavioral responses.291 In canonical CGE models,292 firms are generally assumed to be profit maximizers with constant returns to scale in production; households maximize utility from the consumption of goods and services using a specific functional form; and markets are perfectly competitive. Multiple household types can be included in the model (for instance, differentiated based on geography or income) to calculate distributional impacts of policy changes. A common feature in many models is an underlying model of international trade following Armington (1969) where preferences for goods are differentiated by country of origin to allow for two-way trade for otherwise identical goods. Labor and capital are typically fully mobile between sectors with labor fully employed and no involuntary unemployment.

CGE models are generally more appropriate for analyzing medium- or long-term effects of regulation, when most inputs are free to adjust and consumers can modify purchasing and labor-leisure decisions in response to new prices. A longer time horizon also affords greater opportunities for firms to change their production processes (i.e., to innovate). The time required to move from one equilibrium to another after a policy shock is not defined in a meaningful way (and is

---

290 Regulatory compliance creates a need for additional inputs to produce goods in the regulated sector along with pollution abatement. While the total cost of these additional inputs can be derived from detailed compliance cost estimates, it is not always clear how to allocate the total cost among the inputs specified in the CGE model because CGE models are by their nature an aggregated, parsimonious representation of the economy.

291 Structural models explicitly specify underlying preferences, production, and resource allocation in ways that are consistent with economic theory. The calibration of structural or behavioral model parameters with actual data ensures that the model represents important economic features while remaining in agreement with the underlying theory (Woodford 2009).

292 Here, the term "canonical" is indicative of off-the-shelf models, or models with features that are most common in the literature. In reality, a CGE model may contain several hundred sectors or only a few and may include a single "representative" consumer or multiple household types. It may focus on a single economy with a simple representation of foreign trade, or contain multiple countries and regions linked through an elaborate specification of global trade and investment. The behavioral equations that govern the model allow producers to substitute among inputs and consumers to substitute among final goods as the prices of commodities and factors shift. The behavioral parameters can be econometrically estimated, calibrated, or drawn from the literature. In some models, agents may be able to make intertemporal trade-offs in their consumption and investment choices.
usually assumed to be an instantaneous adjustment). As such, CGE models are generally not well-suited for analyzing transition costs as the economy moves to the new equilibrium unless a transition path can be appropriately specified.\(^{293}\)

The case for using CGE models to evaluate a regulation’s effects is strongest when the regulated sector has large linkages to the rest of the economy and the regulation is expected to affect most of the firms in a broadly defined sector. Narrowly targeted regulations are more difficult to capture without explicitly linking a CGE model to a detailed PE sector model (U.S. EPA 2017). Linking models is discussed in Text box 8.2. The extent to which CGE models will add value to the analysis also depends on data availability (see Text box 8.3 on input-output data efforts).\(^{294}\) When developing their plan for analysis, analysts should consult with NCEE if they anticipate using a CGE model to evaluate the effects of a regulation.

Note that absent a credible way to represent environmental externalities in a CGE model - or the benefits that accrue to society from mitigating them – a CGE model’s economic welfare measure is incomplete.\(^{295}\) However, the inability to account for interactions between costs and benefits in a CGE model does not invalidate their use to estimate costs or make it impossible to design consistent approaches to cost and benefit estimation (U.S. EPA 2017). The possibility of incorporating benefits into a CGE framework is discussed in Text box 8.4.

**Advantages:**

- CGE models are best suited for estimating the cost of policies that will have a broad set of economy-wide impacts, especially when indirect and feedback effects are expected to be significant.
- CGE models are most appropriate for analyzing medium- or long-term effects of policies or regulations.

### 8.3.4 Other Input-Output Based Models

Several other economy-wide approaches are referenced in the literature, including input-output (I-O) models and I-O econometric models. However, these methods should not be used to estimate the social cost of environmental policy (U.S. EPA 2017).\(^{296}\)

#### 8.3.4.1 Input-Output Models

I-O models are highly disaggregated empirical descriptions of the interrelated flows of good and factors of production.\(^{297}\) I-O models are generally static and assume a fixed, strictly proportional relationship between inputs and outputs via multipliers.\(^{298}\) Although their specifications can sometimes be partially relaxed, input-output models embody the assumptions of fixed prices and technology, which do not allow for the substitution that normally occurs when goods become more or less scarce. Similarly, most input-output models are demand driven and not constrained by limits on

---

\(^{293}\) For instance, Williams and Hafstead (2018) embed short run transitional unemployment costs in the context of a general equilibrium model.

\(^{294}\) Data limitations are a significant obstacle for all of the modeling approaches discussed in section 8.3, both in terms of achieving the granularity needed to adequately represent a regulation and to estimate its effects.

\(^{295}\) A growing use of these models has been to quantify previously unrecognized welfare costs that can occur when environmental policies interact with pre-existing distortions in the economy. An expanding body of work also has begun to include non-market goods into CGE models (Smith et al. 2004, and Carbone and Smith 2008).

\(^{296}\) Dynamic stochastic general equilibrium (DSGE) models, while conceptually similar to CGE models, lack sectoral detail and are not recommended at this time, though they may be useful in the future to examine modeling uncertainty and how regulations interact with business cycles (U.S. EPA SAB 2017).

\(^{297}\) Miller and Blair (2009) is a standard reference on input-output analysis.

\(^{298}\) The assumption that output changes translate directly to proportional changes in inputs is not empirically founded and therefore should not be used, even in the short run, because it ignores the potential for factor substitution. Such shifts may change the labor-, capital-, or energy-, or materials-intensity of production.
**Text Box 8.3 - Input-Output Data and Open Source Initiatives**

Input-output data are a basic input into any CGE model. An I-O table assembles data in a tabular format that describes the interrelated flows of market goods and factors of production over the course of a year. It may consist of hundreds of sectors or just a few sectors. In the United States, the Bureau of Economic Analysis provides a time series of national level I-O accounts with multiple levels of sectoral aggregation (between 15 and 405 sectors) based on North American Industry Classification System codes. For more information on constructing I-O tables, see Miller and Blair (2009), Horowitz and Planting (2009), and https://www.bea.gov/industry/input-output-accounts-data.

Below is an example of an aggregated I-O table for the United States for the year 2017 based on BEA data. The columns for the individual sectors denote how much of each commodity is used to produce that sector’s output (cost of annual production). A given sector’s cost schedule (also denoted as upstream sectoral linkages) is composed of intermediate inputs, factors of production (labor and capital) and tax payments. Payments to factors (wages and profits) and tax payments comprise sectoral value added. For instance, the agricultural sector’s intermediate input costs consisted of $97 billion of agricultural inputs, $112 billion of manufactured inputs, $145 billion of service inputs (among others), and $181 billion of value added, for a total of $608 billion in inputs. The row for each sector shows how that sector’s output is consumed (also known as downstream sectoral linkages). In the case of the agricultural sector, $448 billion is consumed as intermediate inputs for sectoral production, while $158 billion is consumed as final demand (i.e., C+G+I+X-M). In this framework, the total output receipts must equal total input costs.

### I-O Table for the United States (2017)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Agriculture</td>
<td>97</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>304</td>
<td>0</td>
<td>43</td>
<td>148</td>
<td>-6</td>
<td>13</td>
<td>3</td>
<td>608</td>
<td></td>
</tr>
<tr>
<td>Mining</td>
<td>2</td>
<td>44</td>
<td>47</td>
<td>26</td>
<td>337</td>
<td>1</td>
<td>58</td>
<td>0</td>
<td>110</td>
<td>-82</td>
<td>-12</td>
<td>530</td>
<td></td>
</tr>
<tr>
<td>Utilities</td>
<td>4</td>
<td>6</td>
<td>27</td>
<td>6</td>
<td>54</td>
<td>14</td>
<td>253</td>
<td>267</td>
<td>1</td>
<td>-31</td>
<td>-1</td>
<td>602</td>
<td></td>
</tr>
<tr>
<td>Construction</td>
<td>2</td>
<td>4</td>
<td>10</td>
<td>1</td>
<td>17</td>
<td>6</td>
<td>259</td>
<td>297</td>
<td>1029</td>
<td>0</td>
<td>-2</td>
<td>1623</td>
<td></td>
</tr>
<tr>
<td>Manufacturing</td>
<td>112</td>
<td>64</td>
<td>47</td>
<td>571</td>
<td>2251</td>
<td>133</td>
<td>1633</td>
<td>3659</td>
<td>134</td>
<td>1223</td>
<td>-821</td>
<td>-414</td>
<td>8591</td>
</tr>
<tr>
<td>Transportation</td>
<td>64</td>
<td>81</td>
<td>11</td>
<td>2</td>
<td>309</td>
<td>130</td>
<td>323</td>
<td>224</td>
<td>43</td>
<td>-22</td>
<td>-2</td>
<td>1165</td>
<td></td>
</tr>
<tr>
<td>Services</td>
<td>145</td>
<td>97</td>
<td>89</td>
<td>181</td>
<td>3233</td>
<td>273</td>
<td>7030</td>
<td>9023</td>
<td>2944</td>
<td>1013</td>
<td>267</td>
<td>-205</td>
<td>24090</td>
</tr>
<tr>
<td>Labor</td>
<td>58</td>
<td>71</td>
<td>157</td>
<td>543</td>
<td>1026</td>
<td>368</td>
<td>8198</td>
<td>1623</td>
<td>8591</td>
<td>1165</td>
<td>24090</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Capital</td>
<td>113</td>
<td>141</td>
<td>181</td>
<td>278</td>
<td>1003</td>
<td>222</td>
<td>5840</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prodn. Taxes</td>
<td>10</td>
<td>23</td>
<td>33</td>
<td>10</td>
<td>56</td>
<td>18</td>
<td>452</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Total Inputs</strong></td>
<td>608</td>
<td>530</td>
<td>602</td>
<td>1623</td>
<td>8591</td>
<td>1165</td>
<td>24090</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Source: Numbers are based on use and supply tables from the BEA. All values are in billions of 2017 dollars. Note that zeros capture small numbers that round to zero while missing entries reflect actual zeros in the data. C is household consumption (excluding leisure), G is government expenditures, I is investment, and X-M is exports minus imports. The sum across these demand accounts equals U.S. GDP in 2017.

In a CGE framework, the columns for the individual sectors determine the input shares used to calibrate production functions. The columns for final demand determine the expenditure shares used to calibrate household expenditure functions. These data, along with “transactions and transfers between institutions related to the distribution of income in the economy,” form the basis of the social accounting matrix (Miller and Blair 2009). Constructing a social accounting matrix requires reorganizing the data in Table 8.3 to link sources of household income to expenditures.

I-O accounts with sub-national or international detail are not provided by the BEA but have been established by others. For instance, the Global Trade and Analysis Project (GTAP) compiles and reconciles data from numerous sources to have consistent sectoral and agent aggregations across included countries (https://www.gtap.agecon.purdue.edu/). Another initiative, the Wisconsin National Data Consortium (WiNDC), develops consistent subnational I-O tables based on publicly available data (http://windc.wisc.edu/).
Text Box 8.4 - Separability between Benefits and Costs

When estimating the benefits and costs of environmental regulation it is almost always assumed that the two are separable, such that the beneficial impacts of the policies do not meaningfully affect the factors that determine the cost of the policy, and vice versa. This is due, in part, to a lack of empirical evidence regarding the sign or importance of the relationship between environmental quality, which is typically not priced in the marketplace, and market goods (Carbone and Smith 2008).

Benefits and costs are non-separable when either the compliance costs borne by firms or households interact with and alter the valuation of environmental benefits (other than through changes in environmental contaminants), and/or the beneficial impacts of the regulation alter the costs. Non-separability may occur for several reasons. The costs of an environmental policy may alter the budget constraint for households - for instance, when compliance costs are passed on to consumers as higher prices for goods such as electricity, and this, in turn, affects their willingness to pay for the beneficial impacts of the policy. It may also be the case that changes in environmental quality and health status lead to changes in household behavior, which for large policies could affect relative prices in equilibrium and the cost of complying with the policies. For example, ongoing work suggests that reductions in mortality risk may affect how households smooth consumption over time (i.e., through savings), which may interact with pre-existing capital taxes or affect the price of investment in pollution abatement capital (Marten and Newbold, 2017). The fact that changes in environmental quality and health status can affect behavior in markets underpins the revealed preference approaches for estimating willingness to pay discussed in Chapter 7.

As noted by the SAB (U.S. EPA 2017), when either the costs or benefits of a regulation are estimated while holding the other constant, any potential non-separability between costs and benefits is missed, which complicates comparing them and calculating social net benefits. The specific magnitude and ultimate impact of non-separability on the net benefits of environmental regulations is an empirical question that requires additional study and is the subject of an emerging literature (Sue Wing 2011). The SAB noted that potential non-separability for large policies does not invalidate estimates of costs and benefits using existing methods. However, caution should be applied when obvious interactions exist.

supply, which would normally be transmitted through increases in prices. While some of the rigidities in the models may be reasonable assumptions in the very short run or for regional analysis with limited ties to the broader national economy, they limit the applicability of I-O models for evaluating medium- to long-run effects or national issues. For instance, the lack of resource constraints and substitution effects that occur over the longer run means that I-O models tend to overestimate the effects of a policy.²⁹⁹ Importantly, the I-O approach does not necessarily account for shifts in economic activity towards the pollution abatement sector (e.g., when the directly regulated sector purchases pollution abatement equipment or services to comply with the regulation). Because input-output models do not include flexible supply-demand relationships or the ability to estimate changes in producer and consumer surpluses, they are not appropriate for estimating social cost.³⁰⁰

8.3.4.2 Input-Output Econometric Models

I-O based econometric models integrate the high level of detail from an input-output model with the forecasting properties of a macro-economic forecasting model. Unlike standard I-O models, this approach accounts for supply-demand conditions in the economy, including resource constraints, through a series of accounting (e.g., savings equal investment) and econometrically estimated relationships (Hahn and Hird, 1991). Feedbacks between supply and demand occur via econometric equations. (CGE models accomplish this via a price mechanism and market clearing assumptions (West, 1995).) The predictions generated by this type of model “are integrated and simultaneously determined...price increases in one sector are translated into cost and price increases in other sectors” (Portney, 1981). This is a key advantage over standard I-O models that assume away these effects. In addition, I-O econometric models

²⁹⁹ Studies that rely on I-O models often calculate some combination of direct, indirect, and induced effects. Direct effects are the changes in output that result from an increase in the cost of inputs (e.g., fuel) in the directly regulated sectors, using the fixed, proportional relationship mentioned above. Indirect effects of a regulation are calculated by using the I-O relationship between outputs in the directly affected sectors and required inputs in related sectors (e.g., suppliers). Induced effects are general re-spending effects that result from changes in household income.

can estimate changes in demand for and production of intermediate goods due to their coupling with a detailed input-output model.

While CGE models assume full market clearing, I-O econometric models assume imperfect knowledge of product and factor markets, with an emphasis on tracking short run disequilibrium, such as business cycles or cyclical unemployment, over time (West 1995). This makes them particularly attractive for analyzing transition costs. In the long run, I-O econometric models are consistent with neoclassical growth theory in that supply side effects dominate model outcomes (Arora 2013). A major drawback of I-O econometric models is that, because they are reduced-form models predicated on historical relationships, they cannot take into account the possibility that a firm or consumer may modify behavior in response to changes in policy (referred to as the Lucas critique). The inability to account for this dependence invalidates these models for purposes of policy evaluation outside of short-term forecasting (Schmidt and Wieland 2013; Arora 2013; Fischer and Heutel 2013; U.S. EPA 2017).

8.4 Modeling Decisions and Challenges

Even when the analyst has determined what types of models are most appropriate for the estimation of social cost, several important modeling decisions remain, including deciding on the level of sectoral and regional aggregation, whether to use a static or dynamic framework, and how to parameterize the model. In addition, analysts should evaluate key uncertainties and take care not to double count, particularly when using outputs from one type of model as an input into another.

8.4.1 Aggregation

The level of sectoral and regional aggregation assumed in a model will determine what aspects of the sector or economy can and cannot be captured explicitly in a regulatory analysis. Matching the level of aggregation in a model to the level needed to evaluate a policy's main effects is important to ensure that the analysis does not miss important contributors to the cost. For example, consider the effects of a new regulation on refrigerant gases in the frozen bakery products sector. In a CGE model, the frozen bakery products sector is not typically separated out as its own sector. Instead, it is captured in a more aggregate category, food products, along with many other related industries such as soft drinks, cereal, and chewing gum. As a result, the frozen bakery products that are affected by the policy “may be too small a part of the model’s food products sector to give meaningful results due to ‘aggregation bias.’”\(^ {301}\) Put another way, there are too many products in the model’s sector to accurately isolate the frozen bakery products industry” (Rivera 2003).

The level of aggregation can affect sectoral and economy-wide results. For instance, sectoral disaggregation allows for a differentiated representation of production technologies, behavioral parameters (e.g., elasticities), and emission intensities that may matter for estimating costs and other impacts (Alexeeva-Talebi, et al. 2012).\(^ {302}\) However, models that are highly specialized for capturing impacts in a specific sector will usually miss impacts on a broader set of sectors. It is also important to consider how costs are allocated spatially (and temporally) to avoid a mismatch between affected facilities' locations and the scale of the model. While proficient at capturing major impacts and interactions between sectors, CGE models generally are not well suited for focusing on a single or small number of specialized sectors because of their level of aggregation.

---

\(^{301}\) Caron (2012) defines “aggregation bias” for a specific variable as the difference in its value from an aggregated model relative to a disaggregated model that has been re-aggregated after the fact to a comparable level.

\(^{302}\) Alexeeva-Talebi, et al. (2012) and Caron (2012) found that the range and standard deviation of sectoral impacts increased with disaggregation. In some cases, even the direction of the estimated impacts was reversed relative to more aggregate results. However, while a highly aggregated model may not be a reliable predictor of sub-sectoral impacts, for many applications, they found that these models produce satisfactory estimates of the overall impacts on the economy.
8.4.2 Choosing Between a Static and Dynamic Framework

It is possible to construct static or dynamic versions of all three types of economic models discussed in this chapter (i.e., compliance cost, PE, and CGE). In a compliance cost framework, the analyst may assume that economic conditions are static or dynamic. If future economic conditions are expected to change meaningfully, a dynamic framework should be used because compliance decisions may be influenced by future economic conditions even when the regulation is not expected to meaningfully influence production or prices. For example, if an affected source anticipates operating for a long time it may choose a more capital-intensive compliance option over a less-capital intensive option because there is a longer period over which it can recover the cost of that investment. Similarly, if the number of affected sources is anticipated to change over time, then the cost of complying may change over time. A dynamic compliance cost framework may also be preferred if, for example, regulated sources may make anticipatory investments prior to a regulation’s compliance dates or to account for the potential for technological change (see section 8.2.3.4).

When the analyst expects intertemporal effects of a regulation to be confined to the regulated sector or a few related sectors, it may be appropriate to simply apply partial equilibrium analysis to multiple periods. As with compliance cost models, relevant conditions, like expected changes in market demand and supply over time, should be taken into account in the analysis. The costs in individual years can then be discounted back to the initial year for consistency.

If the intertemporal effects of a regulation on non-regulated sectors are expected to be significant, analysts also can estimate social cost using a dynamic CGE model. Dynamic CGE models can capture the effects of a regulation on affected sectors throughout the economy. They can also address the long-term impacts of changes in labor supply, savings, factor accumulation, and factor productivity on the process of economic growth. In a dynamic CGE model, social cost is estimated by comparing values in the simulated baseline (i.e., in the simulated trajectory of the economy without the regulation) with values from a simulation with the regulation in place.

Analysts should keep in mind that the evolution of variables in a dynamic model sometimes depends on exogenously imposed assumptions that are not always easy to validate. For instance, modelers sometimes need to constrain the pace at which some variables in the model change (e.g., how quickly technology changes) based on an external assessment of what is technically feasible. Key exogenous assumptions should be clearly documented and explained. In some cases, it also may be useful to explore the robustness of cost estimates to alternative assumptions.

8.4.2.1 Expectations

Dynamic models must specify the ways in which firms and consumers formulate and update expectations about future prices, returns, growth, or other key economic variables. There are a variety of ways to formulate expectations about the future, but they generally fall into two general categories: backward-looking and forward-looking. With advances in computer power, forward-looking expectations are the more common assumption in CGE models.

The two main backward-looking formulations are myopic and adaptive expectations. Myopic economic agents do not anticipate future changes to the economy or regulatory setting, and do not make investments or change consumption and savings behavior until the period when the change takes effect (Paltsev and Capros, 2013). Firms and consumers with adaptive expectations base their expectations about the future primarily on past experiences and are, therefore, relatively slow to modify behavior in response to new information.

In forward-looking models, economic agents have either perfect foresight or rational expectations. A consumer or firm with perfect foresight knows what the future values of key economic variables will be with certainty and incorporates this information immediately into current decisions (Paltsev and Capros, 2013). Rational expectations allow for uncertainty; in this case, economic agents incorporate all relevant information, both past and future, into decision-making and are assumed to get future values correct on average. In other words, they do not systematically make forecasting errors.
There are several analytic implications tied to the degree of model foresight assumed in a dynamic model. For instance, a backward-looking model may lead to higher estimates of compliance costs and welfare impacts compared with a forward-looking model since it restricts the response flexibility of consumers and firms. However, in cases where agents have less than perfect foresight, a forward-looking model may lead to underestimates of the compliance costs and welfare impacts of regulation. Note that many EPA regulations phase in standards or allow for intertemporal smoothing of compliance (e.g., banking of emissions allowances) that could at least partially alleviate this concern.

Another consideration is the large number of variables and constraints that must be simultaneously determined in a forward-looking model. This, in turn, restricts the level of detail that can be included in the model, which may be critical to adequately assessing the social cost of a regulation. As such, a more aggregate forward-looking CGE model should be viewed as a complement to analysis supported by detailed compliance cost or PE sector model.

### 8.4.2.2 Time Steps

Static models provide cost estimates for one period, typically a year. They either assume that conditions are invariant over time or that the cost estimate is indicative of a typical or representative period. Static models exist for all three frameworks discussed in this chapter (i.e., compliance cost, PE, and CGE). As discussed above, if economic conditions are expected to change over time, or if changes in behavior to come into compliance and/or to new market equilibria take time, static models may provide incomplete estimates of costs (and benefits).

Most dynamic models operate using discrete time steps. Time steps between periods are chosen to provide enough detail regarding the adjustment to policy over time, while using a manageable number of time periods for computational reasons. For instance, because dynamic CGE models are often solved over periods of 50 years or more, it is not always practical to solve the model for each individual year. However, when using a dynamic CGE model, the year in which a regulation comes into effect may not be explicitly modeled. Due to the expense and time required to adjust the model and baseline, adding a new solution year may not be an option. In this instance, analysts may use the model year closest to the year in which the regulation will come into effect as a proxy. Regulations that are introduced gradually or vary timing of compliance by region or state pose additional challenges for model representation.

In addition, if the end-year chosen for a dynamic model stops short of capturing important regulatory effects, the social cost estimate may be biased downward. When compliance costs cannot be estimated for all future years, a forward-looking model may smooth them over time, which can also lead to biased social cost estimates, though the direction of the bias will depend on what is assumed about future compliance costs.

### 8.4.3 Model Parameterization

Regardless of the chosen modeling framework, there is a distinction between values determined within the model (those that are endogenous) and values determined outside of the model (those that are exogenous). Model parameterization is concerned with the latter. Which values are imposed exogenously depends on the type of model used; which values are most appropriate depends on assumptions about functional forms and the context of a given model.

In general, model parameterization takes place in two steps. First, the analyst attempts to accurately represent the current structure of the sector(s) or markets of interest. For compliance cost models, this step typically relates to specifying relevant compliance options and constraints (e.g. production capacities). In the case of CGE models, this step consists of characterizing a baseline and calibrating model functions. Second, parameter values are chosen that best characterize economic relationships (i.e., the curvature of different functions) in the model. In the case of compliance cost models, the analyst may need to specify constraints on economic behavior (e.g., production levels, other regulatory

---

303 Specifically, parameters are “terms in the model that are fixed during a model run or simulation but can be changed in different runs, either to conduct sensitivity analysis or to perform an uncertainty analysis when probabilistic distributions are selected” (U.S. EPA 2009).
requirements) and cost functions (e.g. slopes, nonlinearities). For partial and general equilibrium models, this second step is more difficult. Analysts must first make assumptions about functional forms in the model and then select appropriate parameters for them (e.g., elasticities).

Parameter values can be estimated or based on existing values from the literature. While basing parameter values on the existing literature is the more common approach, inconsistencies between the underlying structure of the model and the empirical analyses from which values are drawn can lead to inaccuracies.\textsuperscript{304} When borrowing estimates from the literature to parametrize a model, analysts should discuss the reasons for choosing one value over another and discuss any limitations. In cases where there is no clear consensus in the literature on the most defensible estimates to use, sensitivity analysis to understand the robustness of cost estimates to the parameters chosen is recommended.

Often, a regulation covers many highly heterogeneous facilities where both the compliance options available and abatement costs vary widely. When highly disaggregated source-level data are unavailable, analysts may pursue a model plant approach to estimate compliance costs, where a subset of individual facilities sharing certain characteristics (e.g., plant age, type of production process, industrial sector) are represented by a single model plant. Analysts also may use a model plant approach to reduce the computational requirements of a compliance cost model.

The model plant is intended to represent the typical conditions of a group of facilities. While this provides a way to overcome data limitations and simplify the model, parameterization can still prove challenging. This is particularly true when conditions vary significantly across seemingly similar facilities. For example, if an abatement technology exhibits positive economies of scale, the compliance cost of an average-sized facility will not equal the average cost of all the facilities represented by that model plant. This is because, with economies of scale, the higher cost to smaller facilities will outweigh the lower cost to larger facilities relative to the average-sized facility. In this case, the compliance costs of the facilities represented by the model plant will be under-estimated. It is therefore important for the analyst to carefully consider the number of model plants needed to capture the heterogeneity among constituent facilities that could affect compliance cost estimates.

Parameter assumptions are also important drivers in applied CGE analysis. Estimates of elasticities that help define potential production processes and agent preferences are of particular interest because model results are often sensitive to these parameters. CGE-derived social cost estimates are particularly sensitive to parameters that affect behavior in labor markets due to their pre-existing distortions, such as the assumed elasticities governing the labor-leisure choice of consumers and production elasticities between factors of production (Marten et al. 2019).\textsuperscript{305} Model results tend to be more sensitive to behavioral assumptions, for instance, the values chosen for elasticities relative to other data inputs such as the benchmark input-output data (see Text box 8.2) (Elliott et al. 2012).

### 8.4.4 Uncertainty

Clear communication of uncertainties is critical for transparency of the analysis. Uncertainty in the social cost estimates can arise from uncertainty regarding the baseline, the affected universe of facilities, policy responses, the number of affected markets, and the cost of compliance activities. The degree to which these and other factors affect the confidence placed in the social cost estimates should be carefully reported and quantified when appropriate and possible.

\textsuperscript{304} For CGE analysis illustrating this point, see Shoven and Whalley (1984) and Canova (1995). To alleviate some of these concerns, researchers have econometrically estimated the model parameters in a framework that is consistent with the underlying CGE model (e.g., Jorgenson et al., 2013).

\textsuperscript{305} Previous research has also illustrated these sensitivities in other contexts. For instance, Shoven and Whalley (1984) observe that results from CGE analyses of the U.S. tax system are sensitive to labor supply, saving, and commodity-demand elasticity assumptions. Fox and Fullerton (1991) find that estimates of welfare changes associated with tax reform are more sensitive to assumptions about the elasticity of substitution between labor and capital than the actual level of detail about the U.S. tax system in the model.
While some key uncertainties have implications for both benefits and costs (e.g., for the baseline or affected facilities), several are unique to social cost estimation. For instance, estimates of compliance costs are often "study" estimates, used by engineers to judge the economic feasibility of projects prior to engaging in a costly planning process, and are associated with an error (e.g., +/- 30%). In some cases, more precise cost estimates, described by engineers as "scoping" or "detailed" estimates, may be available. When compliance costs are used to approximate or generate social cost estimates, qualitative and quantitative information available on the degree of precision in the underlying estimates should be prominently discussed to provide appropriate context.

Uncertainty regarding the costs of compliance will propagate through to the estimate of social costs when used in a partial or general equilibrium model. Estimates of social costs may also be subject to model and parameter uncertainty. Model uncertainty refers to uncertainty in a model's ability to accurately represent underlying processes relevant to understanding how an intervention affects the system of interest (for example, due to simplifications necessary to tractably model complex systems) (NRC 2009). As noted in Section 8.4.3, challenges in parameterizing models, including the choice of functional form, also may be a prominent source of uncertainty. Conducting sensitivity analysis or more sophisticated probabilistic analysis across a tractable range of identified uncertainties can provide information on the robustness of the central social cost estimates.

As noted in Section 8.2.3.4, technical change and learning can have an important effect on future compliance costs. Estimates about the effect of innovation will be inherently uncertain and, in some cases, may not be available. Even so, the expectation is that technological change and learning will reduce social cost over time; uncertainty in this case is asymmetric, as innovation is unlikely to increase future costs.

Uncertainty may also affect social cost estimates when projecting the costs of regulations that are implemented by local or state jurisdictions in the future. For example, in illustrative attainment analyses conducted for some NAAQS, once all identified control technologies have been applied, some areas of the country may still be modeled as out of compliance with the air quality standard. In these cases, it is uncertain how attainment will be achieved and at what cost. Similarly, in the case of deregulatory actions how state and local jurisdictions respond, for example by potentially enacting protections in place of the forgone federal standards, can affect the ultimate cost (and benefits) of relaxing the federal standard. In these cases, sensitivity analysis is useful for understanding the robustness of social cost estimates to alternate assumptions.

### 8.4.5 Potential for Double Counting

Because a regulation may have multiple effects through the economy, the analyst should take particular care to avoid double-counting costs. For example, counting both the increased costs of production to firms resulting from a regulation and the attendant increases in prices paid by consumers for affected goods would mean counting the same costs twice, leading to an overestimate of social cost. Also, when reporting private costs for certain groups, the portion of those costs that reflect social costs versus transfers to other groups should be clearly identified in the analysis.

Even in a general equilibrium analysis, analysts must take care in selecting an appropriate measure of social cost. Calculating social cost by adding together estimates of the costs in individual sectors can lead to double counting. Instead, focusing on measures of changes in final demand, so that intermediate goods are not counted, can avoid the double-counting problem (see Section 8.2.2).

When analysts rely on multiple models that take fundamentally different approaches to cost estimation, care should be taken to separately report and characterize each model's output to avoid double-counting. For example, if a technology-rich PE model is linked to a CGE model, the estimate of social costs comes from the CGE model. The social cost is not the

---

306 For example, EPA's Air Pollution Control Cost Manual (U.S. EPA 2018) notes that "costs and estimating methodology in this Manual are directed toward the "study" estimate with a probable error of 30% percent."

307 See Chapter 5 for further discussion of uncertainty and sensitivity analyses.
sum of the costs from the CGE and PE models. Furthermore, the cost estimate from a compliance cost model, for example the increased expenditures on compliance activities in the sector, should not be reported as the social cost of the regulation without further elaborating what this cost estimate represents, why it provides a reasonable estimate of the social cost, and that it is not equivalent to the actual social cost of the rule.

Chapter 8 References


U.S. Census. 2019. Pollution Abatement Costs & Expenditures Survey. Available at:
https://www.census.gov/econ/overview/mu1100.html


Chapter 9

Economic Impacts

Analysis of who will experience gains and who will be burdened by a regulation, and of the nature and magnitude of regulatory impacts, provides important information for decision makers, stakeholders, and the broader public. An Economic impact analysis (EIA) identifies and quantifies a wide range of regulatory impacts including market-based impacts such as changes in employment, prices, profitability and plant closures; as well as impacts outside the marketplace (e.g., impacts on state and local governments). An EIA identifies specific groups that may benefit or be burdened by a policy and assesses the impacts they experience. Affected groups may include consumers, industries, small businesses, workers, communities, tribes, and governments. Using this definition of an EIA, this chapter discusses issues relevant to estimating the economic impacts of EPA policies. An EIA can be tailored to improve understanding of specific regulatory impacts. However, in some instances, an EIA is required, as explained in the first section of this chapter. Subsequent sections begin with a review of frameworks that provide a general understanding of economic impacts, followed by guidance for assessing each impact category.

This chapter primarily focuses on market impacts due to compliance costs. However, Section 9.5.6 is a discussion of the impacts of benefits (changes in environmental quality and public health) and several other sections, such as 9.5.2.5, briefly discuss specific beneficial impacts. Impacts on governments and nonprofits is discussed in Section 9.5.4. A consideration of economy-wide impacts from both costs and benefits is possible and is discussed in Section 9.5.5. Chapter 10 “Environmental Justice and Life Stage Considerations” complements the current chapter by discussing how regulation might change the distribution of environmental quality and health risks across minority and low-income populations, by life stage, and across generations.

9.1 Background

Analyzing economic impacts sheds light on the distribution across groups of costs, transfers, benefits, and other economic outcomes induced by regulation. An EIA may include a broad range of measures including monetized metrics such as profit or price changes, as well as non-monetized metrics such as changes in employment or the likelihood of plant closures. The crux of an EIA is understanding these changes experienced by specific groups. In contrast, a BCA focuses on measuring aggregate social net benefits and is concerned with economic efficiency which requires that benefits outweigh costs by as much as possible, irrespective of to whom net benefits accrue. Thus, the two types of

---

308 At the EPA, an EIA differs from a Regulatory Impact Analysis (RIA). The latter is frequently used interchangeably with “economic analysis” and may contain analyses of benefits, costs, and economic impacts; in other words, an EIA is often contained within an RIA. For more information, see Chapter 1.

309 Chapter 8 “Analyzing Costs” discusses the estimation of social costs.

310 Transfers are shifts of money or resources from one part of the economy to another such as tax payments. See Section 8.2.1.3 for a discussion of compliance costs and transfers.
analyses use different measures. Unlike aggregate benefit and cost measures calculated for a BCA, the impact measures included in an EIA need not be mutually exclusive. For example, an impact that appears simultaneously in two related markets, such as costs in the regulated sector and revenues in the pollution control sector, can be included and appear as two impacts in an EIA. Transfers must be excluded from a BCA due to its focus on aggregate welfare. However, because transfers affect who experiences gains or burdens from a policy, they may be included or even be key within an EIA (U.S. OMB 2003).

Despite these important differences, analyses of economic impacts in an EIA and of social benefits and social costs in a BCA are complimentary as both shed light on the consequences of regulation. When conducted for the same policy, both types of analyses should use a consistent baseline and set of assumptions. Generally, both analyses have similar scopes; that is, if it is appropriate for the analysis of social costs to extend to markets beyond the regulated industry then it would likely be appropriate for the EIA as well. Both analyses should explain underlying assumptions, explore the sensitivity of results to assumptions and inputs, strive for transparency, and include documentation and references (U.S. OMB 2003).

Whether regulatory consequences are measured in terms of economic impacts, changes in social welfare, or both, ultimately the focus is on how people are affected. An EIA that analyzes profitability, for example, is studying potential impacts on the income of firm owners or shareholders. Analysis of employment impacts sheds insight on impacts on workers. And an EIA that estimates changes in prices is concerned about impacts on consumers. To complicate matters, many impacts estimated in an EIA give insight into changes that might affect multiple groups. For example, an increased likelihood of plant closure affects both owners of firms and workers.

9.2 Statutes and Policies

Multiple statutes and policies contain requirements for an EIA that are applicable across media. OMB’s Circular A-4 states, “Where distributive effects are thought to be important, the effects of various regulatory alternatives should be described quantitatively to the extent possible, including the magnitude, likelihood, and severity of impacts on particular groups.” (U.S. OMB 2003) The following statutes and EOs, described more fully in Chapter 2, directly address economic impacts.

- Regulatory Flexibility Act of 1980 (RFA), as amended by the Small Business Regulatory Enforcement Fairness Act of 1996 (SBREFA);
- Unfunded Mandates Reform Act of 1995 (UMRA);
- EO 13132, “Federalism;”
- EO 13175, “Consultation and Coordination with Indian Tribal Governments;”
- EO 13211, “Actions Concerning Regulations That Significantly Affect Energy Supply, Distribution, or Use.”

Together with Circular A-4, these directives highlight features of affected entities that may be relevant for EIAs. Table 9.1 lists the features identified by these directives and offers examples of potentially affected groups.

---

311 The EPA’s Action Development Process (ADP) Library (http://intranet.epa.gov/adplibrary) is a resource for those who wish to access relevant statutes, EOs, or Agency policy and guidance documents. Besides the broadly applicable statutes and directives discussed in this section, there are also environmental statutes with specific applicability that require consideration of impacts on certain populations (e.g., see Section 9.4.2.4 Impacts on Labor), or that may require analysis of impacts for facilities potentially eligible for regulatory variances.
Table 9.1 - Features of Potential Relevance to Economic Impact Analyses as Identified by Statutes, Executive Orders, and Other Directives

<table>
<thead>
<tr>
<th>Feature</th>
<th>Statute, Order, or Directive</th>
<th>Examples of Potentially Affected Economic Groups</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sector</td>
<td>UMRA; EO 13132; EO 13175; OMB Circular A-4</td>
<td>Producers; industries; state, county, local, or tribal governments.</td>
</tr>
<tr>
<td>Entity size</td>
<td>RFA/SBREFA; UMRA; OMB Circular A-4</td>
<td>Businesses, governmental jurisdictions, not-for-profit organizations. Analyze small entities separately.</td>
</tr>
<tr>
<td>Time; Dynamics</td>
<td>OMB Circular A-4</td>
<td>Groups (e.g. consumers, workers, producers, firms, industries) experiencing transitional or long-run impacts.</td>
</tr>
<tr>
<td>Geography</td>
<td>UMRA; OMB Circular A-4</td>
<td>Regions, states, counties, non-attainment areas, local or regional markets.</td>
</tr>
<tr>
<td>Energy</td>
<td>EO 13211</td>
<td>Energy sector (i.e. developers, distributors, generators, or users of energy resources).</td>
</tr>
</tbody>
</table>

9.3 Connections between Economic Impacts and Frameworks of Distributional Effects

Virtually any economic measure of the consequences of a regulation may be included in an EIA.\(^{312}\) To accommodate this degree of flexibility, an EIA is not constrained or governed by an operating framework. However, there are conceptual frameworks that provide insight into the meaning and interpretation of impact categories. These frameworks, often presented in terms of welfare effects, are useful for understanding parts of an EIA because they illustrate the different pathways through which regulatory costs are distributed across population groups.\(^{313}\)

Compliance expenditures incurred by regulated entities may be passed on partially or fully to other groups.\(^{314}\) For example, costs may be experienced by firm owners or shareholders through lower profits or passed on to consumers through higher prices. Or, costs may be passed on to workers through changes in labor compensation, and/or on to the owners of other factors of production through reduced rates of return to land and capital.\(^{315}\) The portion of the cost experienced by these different groups depends on a variety of factors including the time-frame under consideration, the characteristics of the regulated market such as the elasticity of demand relative to the elasticity of supply, and whether there are barriers that prevent new firms or imports from entering the market. Some costs may trickle through to related markets. While in practice economists cannot always measure the extent of cost pass through, existing frameworks help shed light on the variety of ways that costs percolate through the economy.\(^{316}\)

\(^{312}\) For textbook discussions of the meaning and usefulness of impact analysis, see Field and Field (2005) and Tietenberg (2006).

\(^{313}\) If the regulated entity is not a profit-maximizing firm, then the principles discussed in this section are likely not relevant. We address impacts on governments and non-profits in Section 9.5.4.

\(^{314}\) For a more detailed discussion, see Tietenberg (2002) and (2006), which is the basis for the discussion in this paragraph. Useful textbook discussions are also provided by Kolstad (2000) and Field and Field (2005). For a review of the empirical literature, see Bento (2013).

\(^{315}\) Throughout this chapter, all factors of production are represented by either land (natural resources), labor (human resources), or capital (man-made resources).

A framework developed by Harberger (1962) to better understand the distributional effects (incidence) of taxation provides insights into who bears the costs of environmental regulation. Effects are separated into two broad categories: those falling on the sources of income including owners of firms, labor, capital, or land; and those falling on the uses of income, or consumption, due to changing prices. Harberger’s simple two sector, two good model representing a perfectly competitive closed economy with perfectly mobile factors of production suggests that a tax on one input could lead to either, or both, source-side and use-side effects. Adapting the model to represent an environmental tax shows a use-side burden on purchasers of the commodity in the taxed sector; and a source-side impact on factors affected by the tax (Fullerton & Muehlenkamp 2019). Many other existing frameworks also categorize distributional effects according to the route through which the effect is transmitted (product prices, profits, shifts in factor compensation) which is then traced to the group on which the effect falls (consumers; owners of firms, land, or capital; workers).

Figure 9.1 illustrates how Robinson et al. (2016) conceptualize one set of pathways through which total regulatory compliance costs may eventually be distributed across population groups. These pathways help contextualize metrics that often appear in an EIA. The groups experiencing economic impacts as described in Section 9.5 (consumers, producers, workers, other factors of production, communities, and the overall economy) are related to one or more of the three routes through which regulatory compliance costs flow. The groups themselves, however, do not always align perfectly with the three groups identified in the figure (consumers, employees, and owners). For example, the figure does not directly represent “producers,” yet impacts on producers are commonly analyzed at the EPA, and consideration of some producer impacts is even required by statute or executive order. Impacts on producers will ultimately be felt by all the people who together make up affected firms (owners/shareholders, workers, and other owners of productive factors). Other impact categories discussed in Section 9.5, such as impacts on labor or employees, are more directly represented by the figure. The right-hand box conceptualizes how costs might be experienced across different population groups; for example, among regions or among households with different demographic characteristics. This is a common endpoint for an EIA as explained in the sections below on specific impact categories; for example, Section 9.5.1 explains how price increases might be experienced differently by high versus low income consumer groups.

Fullerton (2016) offers a more nuanced framework for disaggregating regulatory consequences. He identifies the following potential cost-related effects on the regulated market: (1) an increased cost of production results in an increase in the price of the regulated good affecting people who purchase the good; (2) decreased production reduces revenues and changes relative returns to workers and owners of firms and factors of production; (3) restrictions on pollution create scarcity rents for owners of firms and/or of capital, land; (4) transitional impacts occur as the

---

317 The two remaining groups discussed in Section 9.5 are government and non-profit organizations which are structured differently than private firms and are not well represented by Figure 9.1.

318 For example, RFA/SBREFA and EO 13211 require consideration of impacts on small firms, and on energy producers, respectively.

319 For example, regulatory costs could lead suppliers of inputs (e.g. coal) to lower their prices, thinking that if they do not, the regulated facilities (e.g., power plants) could shut down.

320 Some “changes” in Figure 9.1 may be measured as economic impacts, welfare changes, or possibly both. For more context on Figure 9.1’s “changes in employee income and employment” see Text Box 8.2.3.5 on social costs and employment in Chapter 8.

321 Some of these effects may be negligible or may not occur at all. Fullerton (2016) also identifies channels through which distributional effects can occur on the benefits side. For example, asset prices can be affected by environmental quality improvements; e.g., improvements could be capitalized into land and housing prices (and some households could be dislocated due to higher rents). See Sections, 9.5.2.5, and 9.5.3; and Chapter 10 for more discussion.

322 Scarcity rents represent a measure of welfare: “This producer’s surplus which persists in long-run competitive equilibrium is called scarcity rent.” (Tietenberg 2006). For a discussion of scarcity rents created by environmental regulations through pollution restrictions and captured by firms in the form of higher profits, see Fullerton and Metcalf 2001. See Buchanan and Tullock (1975) for a discussion of the potential for scarcity rents under a quota or a cap-and-trade where permits are distributed for free. See Chapter 8 for more information on welfare measures like scarcity rents.
Figure 9.1 Example framework to map distribution of compliance costs (Robinson et al. 2016)

A few key insights for EIA at the EPA can be gleaned from these frameworks:

- **Differentiating between impacts that occur in the short- and long-run is important.** The short-run refers to the period in which only some factors of production are variable (e.g., labor) while others are fixed (e.g., capital equipment), and consumers are constrained by existing household assets, commitments, and information. In policy contexts, the short run is sometimes referred to as a transition period. The long-run refers to the period in which all factors of production are variable, the afore-mentioned consumer constraints are relaxed, and the economy returns to equilibrium (i.e., all prices and quantities have fully adjusted to the new regulation). There are likely to be different implications for the economic impacts of a policy in the short run compared to the long run. For example, in the long run, consumers are better equipped to switch to substitute goods, and firms to switch to producing different output and to make entry and exit decisions. These time frames also have different implications for workers (see Section 9.5.2.4).

- **The distribution of impacts among market participants depends on the nature of the affected market(s).** Market characteristics including the extent of competition and the elasticity of demand relative to the elasticity of supply determine the allocation of impacts among consumers, labor, and owners of firms, capital, and other resources. All things equal, competitive markets pass regulatory costs through to consumers to a greater extent than markets in which firms have monopoly power. Firms in very competitive markets do not earn excess profit and have no choice but to pass on costs if they want to stay in business. Of course, the reduced quantity demanded at higher prices may force them to close. Firms with market power have incentive to absorb a portion of regulatory costs since raising the price they charge reduces the

---

For an example, see Fullerton (2011) where this framework is applied to an example environmental policy (a carbon permit system) by linking measurable outcomes to welfare changes.
quantity consumers demand of their product and reduces their profits. Relative elasticities are also
important. In an imperfectly competitive market, the portion of the cost borne by producers increases with
a greater elasticity of demand relative to elasticity of supply (and the portion borne by consumers increases
with a greater elasticity of supply relative to elasticity of demand).  

- **Impacts may differ within market participant categories.** Substantial heterogeneity of a regulation’s
impacts is often experienced within groups. In practice, firms and their circumstances are not identical, so
compliance may be more burdensome for some firms than for others. For example, small firms may have
fewer units of production over which to spread compliance costs, or some firms may have technologies that
are more expensive to bring into compliance. Similarly, not all consumers purchase the same bundle of
goods and therefore will not be uniformly affected by price changes induced by regulation. Industries,
factors of production and other market participant categories can be affected differently as well. In Section
9.5 we discuss the conditions associated with divergent impacts for each impact category.

This section has discussed frameworks that shed light on the potential distribution of compliance costs. Several papers
also consider the distribution of health benefits or environmental quality (e.g., Fullerton 2016; Robinson, et al 2016;
Pearce 2006). For example, Robinson et al. trace the effects of hazard reduction on changes in human risks and the
valuation of those changes. See section 10.2.1 in Ch. 10 for a discussion of this literature.

### 9.4 Analytic Components of an Economic Impact Analysis

An EIA should develop a profile of baseline conditions among groups expected to experience important effects of the
rule. These are the conditions occurring in the absence of the rule or policy over the period of analysis. For example, the
profile could include the number of regulated firms, their average size, and their average profitability. These metrics
would be estimated for the year the rule takes effect and for the remaining timeframe of analysis. An EIA may also
include two additional components: an analysis to screen for the magnitude of incremental impacts and an in-depth
examination of expected important impacts. For each component of an EIA, analysts should highlight key analytic
limitations and uncertainties. This section discusses the baseline profile, the screening analysis, and the in-depth
examination, and identifies potentially useful data sources.

#### 9.4.1 Baseline Profile

An EIA should develop a baseline profile that describes the industries, consumers, workers, or other groups that are
expected to experience important incremental effects of a regulation. The profiles will overlap with baseline profiles
developed for other components of a regulatory analyses, such as the cost analysis.

The effects of some regulations may extend beyond participants in directly regulated markets, affecting, for instance,
upstream or downstream markets, or complementary or substitute product markets. Often the markets involved in
pollution control activities are affected. We will refer to the latter as the environmental protection sector and note that
it may overlap with upstream markets.

The following information can contribute to an industry profile:

---

324 For more details with a discussion of economic impacts for a representative firm and for the market, in a supply and demand model with perfect
competition and under monopoly, see Tietenberg (2006), pp 510-516.

325 See Fullerton and Metcalf (2002).

326 Heterogeneity in impacts may also be the result of regulatory design (e.g., differentiation of standards by facility vintage). This possibility is discussed
under Section 9.5.2.

327 For more about how to define and describe baselines, see Chapter 5. For more about developing a baseline for government or nonprofit
organizations, see Section 9.5.
• the affected North American Industrial Classification System (NAICS) industry codes. (NAICS is the standard used by Federal statistical agencies in classifying business establishments.)

• industry summary statistics, including total employment, revenue, costs, number of establishments, number of firms, and size of firms;

• baseline industry structure, including competitive structure, market concentration, degree of vertical integration within the industry;

• characteristics of supply and demand (e.g., relative elasticities);

• industry trends including growth rates, expected changes in technology, and financial conditions;

• openness to and reliance on international trade;

• pre-existing environmental and other regulations and associated compliance behavior;

• barriers to entry; and

• diversity of production technologies among firms.

The baseline socioeconomic characteristics of groups expected to experience consequential economic gains or burdens due to a regulation are also important and may include consumers, workers, business owners, shareholders, renters, community members, and others. Attributes to consider include:

• income and poverty levels,

• age distribution,

• employment status,

• community characteristics such as unemployment rate,

• geographic location and mobility.

The potential relevance of these market conditions and socioeconomic characteristics within the context of a specific impact category is discussed in Section 9.5.

9.4.2 Screening Analysis

During the early stages of regulatory analysis, screening for important impacts can be useful and may be as simple as systematically thinking through the expected impacts of a regulation and qualitatively describing them. When data are sparse, it may still be possible to roughly estimate some regulatory impacts. For example, to screen for significant impacts on small businesses, analysts can compare a rule’s estimated annualized costs per regulated facility to estimated annual revenues of affected small facilities to determine whether the ratio of regulatory costs to facility revenues violates established thresholds.

While the EPA has established thresholds that suggest when impacts on small entities are significant, in most cases the criteria for when an impact warrants additional analyses are not well defined and may depend on the condition of the economy. For example, during an economic recession, impacts on workers may be a concern. Or, the timing of regulatory impacts may be relevant, including the period of anticipation of an upcoming compliance date, and whether effects are expected to grow or diminish or affect different groups over time. The context or location within which economic impacts are experienced is important. For example, reduced demand for labor in a small town with declining

---

328 For more information on classifying industries by NAICS codes, see https://www.census.gov/eos/www/naics/.

job opportunities might have bigger labor market impacts than in a larger city with abundant work opportunities. Or, when a trade exposed industry is the subject of regulation, there may be concerns regarding potential loss of domestic market share. Finally, if analysts suspect important impacts beyond directly regulated industries, the scope of analysis can be broadened, even if data and tools permit only qualitative assessments.

9.4.3 In-Depth Examination

Analysts may conduct an in-depth examination of the impact categories identified as likely to be important by the screening analysis. Substantial adverse impacts deserve special attention. If possible, a partial equilibrium analysis of affected markets will yield greater insights into impacts relative to an engineering cost analysis alone.\(^3\) For example, with information on demand and supply elasticities in affected markets, analysts can move to a more refined analysis that examines the pathways through which costs would travel (e.g., consumer prices versus producer profits and input prices including wages). With regional and firm specific demand and supply information, analysts might also be able to shed light on how impacts vary across regions and firms. It may also be possible to link together several sector-specific partial equilibrium models with a multi-market model to examine linked impacts on regulated and related markets. If appropriate, a general equilibrium model can offer insights into impacts on a broad spectrum of markets and groups across the economy (see Section 9.5.5).

9.4.4 Data

Analysts may have access to proprietary data or detailed plant-level data (which may be confidential business information) collected through the rulemaking process that can be leveraged in an economic impact analysis. However, often data must be sought elsewhere. Table 9.2 describes publicly available data sources that might be useful to combine with compliance cost information to analyze economic impacts. Note that quantitative estimates of some economic impacts may not be possible because of inadequate household-, firm- or community-specific data (including elasticity estimates). Data that are available are often aggregated to the sector, or jurisdiction, level.

<table>
<thead>
<tr>
<th>Source</th>
<th>Examples of types of data</th>
<th>Examples of relevant groups</th>
</tr>
</thead>
<tbody>
<tr>
<td>U.S. Census Bureau – Longitudinal Employer-Household Dynamics: <a href="https://lehd.ces.census.gov/">https://lehd.ces.census.gov/</a></td>
<td>Statistics on employment, earnings, and job flows at detailed levels of geography and industry for different demographic groups.</td>
<td>Sectors or Industries, Labor, Producers, Government entities.</td>
</tr>
</tbody>
</table>

\(^3\) For a discussion of partial equilibrium and other market and engineering models, see Chapter 8 on Analyzing Costs.
### Source

**Published research specific to an industry or sector.**

University of Wisconsin – Wisconsin National Data Consortium:
http://windc.wisc.edu/

**U.S. Census States & Local Areas:**
https://www.census.gov/govs/local/

**U.S. Census State and County Quickfacts:**
https://www.census.gov/quickfacts/fact/table/US/PST045216

**U.S. Census Bureau – American Housing Survey:**
https://www.census.gov/programs-surveys/ahs.html

**U.S. Department of Housing and Urban Development Aggregated USPS Administrative Data on Address Vacancies:**
https://www.huduser.gov/portal/datasets/usps.html

**U.S. Census Bureau – American Community Survey:**
https://www.census.gov/programs-surveys/acs

**Trade Publications and Associations:**

**U.S. Census Statistics of U.S. Businesses:**
https://www.census.gov/programs-surveys/susb.html

**U.S. Bureau of Economic Analysis:**
https://www.bea.gov/data

**U.S. Census Bureau – Annual Survey of Manufacturers:**
https://www.census.gov/programs-surveys/asm.html

### Examples of types of data

**Demand and supply elasticities, regional supply and demand information, and other specific estimates of interest.**

**Open-source datasets for economic analysis, for U.S. states and counties, with state, sector and region economic activity.**

**Demographic and socioeconomic information.**

**Data on the housing and construction industry, homeownership, and characteristics of homes.**

**Occupancy rates.**

**Detailed population and housing information, by community**

**Market and technological trends, sales, location, ownership changes**

**National and subnational economic activity by enterprise size and industry**

**Economic statistics on U.S. production (e.g. GDP), consumption, investment, exports and imports, and income and saving. National, Regional, Industry, and International economic accounts.**

**Statistics for manufacturing establishments**

### Examples of relevant groups

Sectors or Industries, Consumers, Producers.

Sectors or Industries, Consumers, Producers, Government entities.

Consumers, Government entities.

Housing and Construction Industry, Consumers, Government entities, Communities.

Households, Communities.

Sectors or Industries, Labor, Producers, Consumers, Government entities, Communities.

Sectors or Industries.

Producers, Non-profits, Communities.

Sectors or Industries, Producers, Labor, Consumers, Government entities, Communities, International flows.

Manufacturing sector, Producers.
### Examples of types of data

<table>
<thead>
<tr>
<th>Source</th>
<th>Examples of relevant groups</th>
</tr>
</thead>
<tbody>
<tr>
<td>U.S. Department of Commerce: Economic Census: <a href="http://www.census.gov">www.census.gov</a></td>
<td>Sector-level sales, value of shipments, number of employees &amp; establishments, value added, cost of materials, capital expenditures, household and community characteristics</td>
</tr>
<tr>
<td>Risk Management Association, Annual Statement Studies: <a href="https://www.rmahq.org/annual-statement-studies/">https://www.rmahq.org/annual-statement-studies/</a></td>
<td>Income statement and balance sheet summaries, profitability, debt burden and other financial ratios by quartile for recent years (based on loan applicants only)</td>
</tr>
<tr>
<td>Dun &amp; Bradstreet Information Services: <a href="http://www.dnb.com">www.dnb.com</a></td>
<td>NAICS code, address, facility and parent firm revenues and employment</td>
</tr>
<tr>
<td>Standard &amp; Poors: <a href="http://www.standardandpoors.com">www.standardandpoors.com</a></td>
<td>Quarterly financial information for publicly-held firms, line-of-business and geographic segment information, S&amp;P ratings</td>
</tr>
<tr>
<td>Securities and Exchange Commission Filings and Forms: <a href="https://www.sec.gov/edgar.shtml">https://www.sec.gov/edgar.shtml</a></td>
<td>Income statement and balance sheet, working capital, cost of capital, employment, regulatory history, foreign competition, lines of business, ownership and subsidiaries, mergers and acquisitions</td>
</tr>
<tr>
<td>United States Utility Rate Database: <a href="https://openei.org/wiki/Utility_Rate_Database">https://openei.org/wiki/Utility_Rate_Database</a></td>
<td>Rate structure information for electric utilities in the United States. The URDB includes rates for utilities based on the authoritative list of U.S. utility companies maintained by the U.S. Department of Energy’s <a href="http://www.eia.gov">Energy Information Administration</a>.</td>
</tr>
<tr>
<td>U.S. Department of Commerce Pollution Abatement Costs and Expenditures Survey: <a href="https://www.census.gov/econ/overview/mu1100.html">https://www.census.gov/econ/overview/mu1100.html</a></td>
<td>Pollution abatement costs for manufacturing facilities by industry, state, and region. Data is limited to annually from 1973 to 1994, with the exclusion of 1987; and 1999 and 2005.</td>
</tr>
<tr>
<td>Standard and Poor's, Moody's, and Fitch state and city bond ratings.</td>
<td>Financial strength indicator.</td>
</tr>
</tbody>
</table>

---

331 Rates are posted annually by the National Renewable Energy Laboratory (NREL), under funding from the U.S. Department of Energy’s Solar Energy Technologies Program, in partnership with Illinois State University’s Institute for Regulatory Policy Studies.
9.5 Impact Categories

This section provides guidance for assessing specific impact categories. Categories discussed are not mutually exclusive; rather, they have a high likelihood of overlap. For example, impacts on producers (employees and owners) likely overlap with impacts on the communities where they are located. Impact categories discussed in this section are:

- Consumers,
- Producers and Factors of Production,
- Communities,
- Governments and Non-Profits,
- Economy-wide,
- Benefits of improved environmental quality or health.

The discussion that follows usually considers the impacts of new compliance activities. However, it is also relevant to reductions in compliance activities which generally would produce impacts going in the opposite direction.

9.5.1 Impacts on Consumers

To study impacts on consumers, analysts typically examine the expected impacts of a regulation on the prices of final goods. Also relevant are the characteristics of consumers purchasing the goods. Firms may also be consumers of regulated products and as such are covered in Section 9.5.2.6 “Impacts on Related Markets.”

New environmental requirements typically raise the cost of production in directly regulated industries, causing an upward shift in the market supply curve (that is, an increase in the price producers require for each quantity supplied). In response, consumers will do without or with less of the product, and/or pay a higher price, thus bearing some of the burden of regulatory costs.

A good starting point to analyze potential impacts on consumers purchasing output from the regulated sector is to gather information on the determinants of the elasticity of demand relative to the elasticity of supply for the affected goods. To gauge elasticity of demand, a useful consideration is whether the product is a necessity, has many substitutes, or its purchase makes up a substantial portion of the consumer budget. Consumer impacts may be smaller if there are good substitutes that are comparably priced causing a high demand elasticity and smaller price change. There also may be small changes in output prices if compliance expenditures are low relative to total production costs.

---

For more information about the determinants of elasticities, see Appendix A: Economic Theory, Section A.4.1 Elasticities

9-11
To gauge elasticity of supply, analysts should assess how easily firms can increase or decrease production quantities. Information on the flexibility of capital equipment and buildings for shifting into different types of production would be useful; for example, understanding whether excess capacity can be used to produce comparably valued output.

A qualitative discussion of the factors that can affect impacts on consumers may be useful, however analysts may be able to locate empirical estimates of demand and supply elasticities. If possible, analysts should select elasticity estimates that reflect the focus of analysis. For example, to understand potential differences in the pass-through of regulatory costs into prices over time, analysts should examine estimates of elasticity in the short-run compared to the long-run; or regional demand elasticity measures could illuminate differences in cost-pass-through across communities.

The characteristics of the regulated industry may also influence the share of costs passed on to consumers. A market consisting of producers that have different cost structures, perhaps because they use different technologies or are of different sizes or ages, may suggest different degrees of cost-pass-through. Or a market with barriers to entry such as a necessity for access to a scarce natural resource will have market power and the supply side will be capable of bearing more of the compliance costs than the amount that more competitive markets with lower profit margins (e.g., markets facing international competition) could bear.

Combining an estimated price increase with information on the share of the consumer’s budget spent on the product will improve understanding of the impacts on households. There is a possibility that budget shares may vary substantially across consumers. Even if price increases are small, specific groups of consumers may still be affected if the product is a necessity for which low income households spend a substantial portion of their budget. For example, the share of income spent on energy or water by low income households is larger than for others, so energy or water price increases may affect them more. This effect may be strengthened by the flexibility among higher income households to purchase substitutes with substantial upfront costs such as efficient appliances. However, it is also important to consider whether existing government programs may act to mitigate the impact of price increases on consumers.

If consumer impacts are expected to be nonnegligible, information on affected consumers such as their age distribution, income level, or residential location should be gathered to contribute to a baseline profile. Nationwide averages of these variables may be appropriate if consumers are broadly distributed across the country.

In some cases, assessing the impact of a regulation on consumers can be complex. Analyzing policies that affect goods consumed broadly across the country differs from analyzing goods with more limited use patterns such as pesticides or paint removers. Data for the latter might be less accessible. Other complicating factors are associated with goods for which price- or rate-setting is complex. For example, to explore the extent to which proposed air pollution control costs will be experienced by different electricity consumers, the analysis would need to include information on how the policy affects consumers served by cost-of-service utilities, compared to deregulated electricity providers. Any assistance available for low-income or other consumers to offset rate increases is also relevant; as is variability in consumption patterns among categories of customers. If regulatory costs are large, economy-wide models may lend additional insight into how impacts affect consumers across the economy (see Section 9.5.5). Such models may also examine the interaction with existing government transfer programs.

---

333 For cases where government or non-profit organizations are the producers, see Section 9.5.4.

334 The share of income spent on energy falls as income increases. Some studies have found that policies that increase energy prices are regressive, placing a greater burden on lower income households (e.g., Burtraw et al., 2009; Hassett et al., 2009; Williams et al., 2015). Other studies account for the indexing of transfer payments to inflation and find that the burden of a carbon tax is roughly proportional to permanent income, and so is neither regressive, nor progressive (Cronin et al., 2019). See Deryugina et al. (2019) for a discussion of some of these energy policy studies.

335 Cory and Taylor (2017) conduct a detailed analysis of spending by low-income households and explore the potential impacts on health spending caused by price changes induced by safe drinking water standards.

336 Some government transfer payments like Social Security are indexed to inflation and may provide some protection of purchasing power for lower income households.
9.5.2 Impacts on Producers and Factors of Production

Compliance activities typically increase production costs to regulated industries. This may affect many different impact categories which are listed below and discussed in this section:

• Production,
• Profitability and Plant Closures,
• Small Businesses,
• Labor,
• Land and Capital,
• Related Markets,
• Competitiveness.

Effects may vary by industry or firm characteristics, production technologies, pollution intensities, policy design, and more. There may be different effects in the long-run versus the short-run, and according to whether one-time, ongoing, or transitional costs are being considered. Ongoing costs are necessary to maintain the newly achieved state of environmental quality. Transitional costs stem from adjusting from one state of environmental quality to another (Baumol and Oates 1988). Consideration of the effect on small businesses is mandated by statute.337

If regulatory costs are small and/or distributed widely, there may be negligible impacts on producers. However, even if the average impact across firms is small, some producers, such as those facing the highest abatement costs, may be substantially affected. The following subsections discuss how to assess impacts on producers and factors of production.

9.5.2.1 Impacts on Production

In response to substantial regulatory costs, the supply curve in the directly regulated market may shift upward near the market price which typically leads to higher prices and lower output.338 Reductions in industry output are usually driven by a mix of increased and lowered operating rates at existing plants, closure of some plants, and/or reduced future growth in production relative to the baseline. This section discusses circumstances that influence changes in output at the firm or facility (for firms that own more than one plant) level. Such changes can be combined with industry characteristics such as the number and size or regional distribution of firms to assess total changes in production.

At least two conditions can cause environmental regulation to have different impacts across firms, and lead to changes in both the number and size of the average firm (Tietenberg 2006). The first is significant heterogeneity in firm or facility cost structures. Such variability can cause variation in the magnitude of regulatory costs and, while not always the case, can lead to differences in the magnitude and direction of changes in output across producers. For example, total industry output may decline or shift from the highest cost plants to more efficient competitors.

To better understand the extent of heterogeneity in how firms might adjust production in response to regulatory requirements, a profile of baseline conditions is useful. If available, detailed industry, firm, or plant-level information may provide insights into how production processes and baseline costs might vary across facilities and how this variation might lead to different incremental costs of a regulation. For example, the ease with which facilities can accommodate pollution control equipment may vary; or there could be variability in the ability to substitute less hazardous chemicals for more toxic ones. Ultimately, what analysts will need are the differences across firms in post-regulatory costs. Firms may be able to maintain or even increase production levels if after absorbing compliance costs, their production costs

337 See Chapter 2 and Section 9.1 which refers to the RFA as amended by the SBREFA.
338 If in the post-policy equilibrium, the production costs of the marginal firm are not notably affected by the regulation then it is possible that the production and price effects can be de minimus even if inframarginal firms face notable compliance costs.
fall below the highest cost firms. Or they may decrease production if after absorbing compliance costs, their production costs are among the highest in the market.

A second cause of variable impacts across firms are regulatory requirements that differ depending on firm characteristics. Vintage-based regulations that vary with the age of facilities may differentiate between existing and future pollution sources, with future sources regulated more stringently. In other cases, firms in regions with high baseline pollution may face stricter emission controls. In general, regulatory requirements that vary by firm characteristics will shift economies of scale and can affect the distribution of output among firms as well as firms’ average level of output.

For example, firms may respond to policies that differ across plant locations by relocating production to a less-regulated area within the U.S. The greater the degree to which firms take advantage of this ability to shift production across space to reduce compliance costs, the more likely it is that overall domestic production does not change substantially. The outcome could be plant closure(s) and accompanying plant opening(s) due to relocations, with distributional effects on affected areas. Shifts in production from domestic to foreign sources can also occur and are discussed in more detail in Section 9.5.2.8.

### 9.5.2.2 Impacts on Profitability and Plant Closures

Regulatory costs can reduce profits and increase the possibility of plant closures. The industry profile (see Section 9.4) describes baseline industry growth and financial conditions at regulated firms. To assess changes in profits due to a regulation, analysts should compare the expected change in market price to the change in production costs after accounting for compliance activities. This increment should be multiplied by expected changes in output to estimate how profits change.

Industries and firms that are relatively profitable in the baseline will be better able to absorb any new compliance costs that are not passed on to consumers. In cases where facilities have different baseline pollution controls or different production technologies, those with lower costs after meeting a new environmental standard will be better able to maintain profitability relative to other firms and may increase their market share. These firms may even be able to increase profitability if their costs of compliance increase by less than the increase in the market price.

Discussing the likelihood of baseline closures improves understanding about the likelihood of closures attributable to the regulation. Note that vertically or horizontally integrated facilities might not be viable as stand-alone operations but may continue to operate based on their contribution to the business line.

If pollution restrictions limit production of industry output, profitability may be affected. There may be different profitability impacts for new versus existing firms. This may be the case, for example, with vintage-differentiated regulation that “grandfathers in” existing firms, imposing less rigorous pollution controls on them. If market demand is increasing, new firms can enter but face higher costs which negatively impact profitability. Existing firms can benefit through newly created scarcity rents; with positive impacts on profitability. Over the long run, the likelihood of plant closures may change if older plants with higher emissions are kept in operation for longer than expected in the baseline.

---

339 The firms experiencing less stringent regulation might be more likely to see expanding market shares relative to their counterparts, though some empirical evidence suggests this is not the case (Tietenberg 2006 citing Pashigian 1984 and Pittman 1981, Greenstone 2002).

340 Shadbegian and Wolverine (2010) survey the plant location literature which suggests that firms reallocate production (Gray and Shadbegian 2010), plant entry (List et al 2003), or plant exit (Kahn 1997) in response to environmental regulations.

341 For example, the EPA’s Power Sector Modeling Platform v6 using IPM includes detailed information on power plants that have made public announcements of future closures, and this information can inform a baseline analysis. (U.S. EPA 2018b)

342 See Tietenberg (2006) Chapter 21 for more discussion and for references to literature finding evidence of a new-source bias in environmental regulations.
Analysis of impacts on regulated firms’ financial conditions involves the use of available financial data. Impacts can be assessed by examining direct compliance costs as a percent of a firm’s average revenues, profits, or sales. An upper-bound assumption is that compliance costs are borne entirely by the regulated industry (i.e., none are passed through to consumers).

Analyzing changes to firms’ and facilities’ financial conditions requires caution. First, economic models are simplified representations of complex economic systems built to assess relationships between economic factors. They can be useful for estimating effects on groups but often are not reliable predictors of firm or facility-level decisions. Second, common simplifying assumptions about firm decision-making include perfect foresight, where agents know precise values for all economic variables in all future years, and perfect information, where precise values drive decision-making so that a one-cent difference between costs and revenues can be the difference between continued operation versus closure. Such assumptions may perform well when describing aggregate behavior, but they often run counter to the everyday complex and uncertain decision-making by managers, which is remarkably difficult to model. There is typically little information regarding the economic decision maker’s expectations about the future (e.g., the firm’s profitability, costs, revenues, and market conditions) and how those expectations respond to new information, such as a new regulation. Indeed, many decisions are multi-faceted. For example, management decisions about plant closures often result from the cumulative effect of multiple factors, such as financial distress, unfavorable market conditions, and aging equipment, rather than any single factor such as a new environmental regulation. Finally, facility-specific, rather than firm-specific, financial information is preferred for assessing profitability and particularly for assessing the likelihood of plant closures. However, it is often difficult to find. For instance, while financial data for publicly held companies is available, it is often too aggregated to shed light on specific business practices or management decisions. For these reasons it is important for analysts to describe the main limitations of the analysis when evaluating the incremental impact of a regulation on firm profitability or the likelihood of plant closures.

### 9.5.2.3 Impacts on Small Businesses

The RFA requires agencies to define small business according to the Small Business Administration’s (SBA) small business size standard regulations. As another option, the RFA authorizes any agency to adopt an alternative definition of small business “where appropriate to the activities of the Agency” after consulting with the Chief Counsel for Advocacy of the SBA and after opportunity for public comment. If adopted, the agency must publish the alternative definition in the Federal Register. The analytical tasks associated with complying with the RFA include a screening analysis for “significant economic impacts on a substantial number of small entities” (SISNOSE). The small businesses to be included in the analysis are those that are directly regulated; that is, those that are subject to the rule’s requirements. If a small business does not have an obligation imposed directly by the regulation, then EPA guidance is that it should be excluded from the analysis. In order to determine SISNOSE, the EPA conducts a screening analysis for both proposed and final rules based on a percentage of sales as an economic impact for small businesses (a “sales test”) (U.S. EPA 2006). “Small Entities” are defined by the RFA but “substantial number” is not specified. The EPA has broad guidelines including example thresholds for determining SISNOSE certification, but generally recommends three factors in determining “significant impact” and “substantial number”:

---

343 Some models use “model plants” to represent specific plant or unit types and solve a linear programming problem by choosing compliance strategies to minimize costs across the model plants (e.g., see Section 8.4.3).

344 This is referred to in the literature as the “penny-switching effect.” See Krey and Riahi (2009).

345 The financial literature points to managers’ individual characteristics and biases that can affect corporate decision-making, e.g., risk aversion, confident or pessimistic approaches, misestimation of financial market data, or loss aversion. For a brief survey of the literature on behavioral corporate finance, see Malmendier and Tate (2015).


347 See also Chapter 2. For a discussion of the screening analysis for small governments and small non-profits, see Section 9.5.4.1.
1. Magnitude of economic impact that may be experienced by regulated small entities;
2. Total number of regulated small entities that may experience the economic impact; and
3. Percentage of regulated small entities that may experience the economic impact.

If the screening analysis reveals that a rule cannot be certified as having no SISNOSE, then the RFA requires a regulatory flexibility analysis be conducted for the rule, which includes a description of the economic impacts on small entities. Further analysis examining other types of impacts, as discussed elsewhere in this chapter, in relation to small businesses, may provide additional information for decision makers.  

9.5.2.4 Impacts on Labor

Evaluation of employment impacts is required by many of the major environmental statutes. Impacts can vary according to baseline labor market conditions; employer and worker characteristics such as industry, occupation, skill-level, and region; and the type of workforce adjustment or job transition. Employment impacts may occur in the regulated and environmental protection sectors, in upstream or downstream sectors, or in sectors producing substitutes or complements. As economic activity shifts in response to a regulation, typically there will be a mix of declines and gains in employment in different parts of the economy over time. This section focuses on labor demand and on employment impacts measured as changes in employment levels. To present a complete picture, an employment impact analysis will describe both positive and negative changes in employment. For most situations, employment impacts are assessed as part of an EIA, and should not be included in the formal BCA. See Section 8.2.3.5 for a discussion of labor impacts and BCA.

When the economy is at full employment as in long-run equilibrium, a regulation may reallocate employment among economic activities rather than affect the general employment level (Arrow et. al. 1996). In the short-run, regulations can lead to transitional employment effects including involuntary job losses, which may involve periods of unemployment for some workers (Smith 2015; Schmalensee and Stavins 2011; Congressional Budget Office 2011; and U.S. OMB 2015). Involuntary job loss refers to job displacement that results from employer decisions and that is unrelated to worker performance; e.g. plant closings, mass layoff events, and other firm-level employment reductions (Farber 2017, Sullivan and von Wachter 2009b).

Economic theory of labor demand indicates that employers affected by environmental regulation may increase their demand for some types of labor, decrease demand for other types, or for still other types, not change it at all. Morgenstern et al. (2002) decompose the labor consequences in a regulated industry facing increased abatement costs. They identify three separate components. First there is a demand effect caused by higher production costs raising market prices. Higher prices reduce consumption (and production) reducing demand for labor within the regulated industry. Second there is a cost effect: as production costs increase, plants use more of all inputs including labor to produce the same level of output. For example, pollution abatement activities require additional labor services to produce the same level of output. Third, there is a factor-shift effect: post-regulation production technologies may be more or less labor intensive (i.e., more/less labor is required per dollar of output). Deschênes (2014) describes environmental regulations as requiring additional capital equipment for pollution abatement that does not increase productivity. This can be included in a labor demand model as an increase in the rental rate of productive capital. These

---


349 Relevant statutes include the Clean Air Act (section 321(a), the Clean Water Act section 507(e), the Toxic Substances Control Act section 24, the Solid Waste Disposal Act section 7001(e), and the Comprehensive Environmental Response, Compensation, and Liability Act section 110(e).

350 See Section 9.5.6 and Chapter 7 Section 7.2.1.2 and Text Box 7.9 for examples of how environmental regulation may also affect labor supply through changes in worker health and productivity (e.g., Graff Zivin and Neidell 2012, 2013, 2018).

351 Except to the extent that labor costs are part of total costs in a BCA.
higher production costs induce regulated firms to lower output and decrease labor demand (an output effect) as well as shift away from the use of more expensive capital towards increased labor demand (a substitution effect). Berman and Bui (2001) discuss how affected firms’ overall labor demand could increase, decrease, or remain unaffected, depending, in part, on the labor-intensity of environmental protection activities needed for regulatory compliance compared to the labor-intensity of producing output. To study labor demand impacts empirically, a growing literature has compared employment levels at facilities subject to an environmental regulation to employment levels at similar facilities not subject to that environmental regulation; some studies find no employment effects, and others find significant differences. For example, see Berman and Bui (2001), Greenstone (2002), Ferris, Shadbegian, Wolverton (2014), Walker (2013), and Curtis (2018).

Workers affected by changes in labor demand due to regulation may experience a variety of impacts including job gains or involuntary job loss and unemployment. Transitional, or adjustment, costs may occur as workers shift out of current employment and into other, potentially less desirable, jobs (for example, jobs that are lower paying or in a less desirable location); or into unemployment; or exit the labor force sooner than otherwise. Workers displaced from declining industries or occupations, with long job tenure, or living in areas where labor mobility is low or unemployment is high, may be especially likely to face challenges in finding comparable re-employment (Baumol and Oates 1988). Involuntary job loss can lead to significant earnings losses for workers, as well as other impacts, such as negative health effects (Sullivan and von Wachter 2009a, 2009b). See Text Box 9.1 for a discussion of involuntary job loss, unemployment impacts, health and wealth effects. While involuntary job loss is a special concern to policy makers, it is a challenge to evaluate. Shifts in and out of unemployment may also affect transfers. US Department of Labor (2020) has recently compiled data on the magnitude of change in transfer payments with shifts in the numbers of individuals who are unemployed. In practice, an EIA evaluates potential changes and shifts in employment.

Workforce adjustments can be costly to firms as well as workers, so employers may choose to adjust their workforce over time through natural attrition (retirements, voluntary separations) or reduced hiring, rather than incur costs associated with job separations (layoffs or other firm-level employment reductions). Curtis (2018) estimates changes in industry employment levels over time due to an environmental regulation and finds that changes occurred slowly through reduced hiring rates, and not through increased job separations. Hafstead and Williams (2018) find a similar result for the regulated sector of employment levels decreasing through slow hiring and natural attrition rather than increased separations, when modeling a carbon tax.

As a result of shifts in the demand for labor, environmental regulation might also induce wage effects. However, firms generally avoid adjusting existing employees’ wages downward (Walker, 2013; Curtis, 2018). Nominal wage rigidity has been attributed to many causes, not least is the potential impact of lowering wages on employee morale (Howitt 2002). Another factor suggesting very limited wage impacts in the specific context of environmental regulation, is that regulated firms are often a fraction of employers in affected labor markets and thus are not influential enough to affect industry wage rates (Berman and Bui, 2001).

The remainder of this section describes practical approaches to employment impacts analysis.

---

352 For an overview of the neoclassical theory of production and factor demand, see Chapter 9 of Layard and Walters (1978). For a discussion specific to labor demand, see chapter 4 of Borjas (1996). When using this theoretic framework, authors have conceptualized regulation as an increase in the price of pollution (Greenstone 2002, Holland 2012), an increase in the price of capital (Deschénes 2014), an increase in energy prices (Deschénes 2012), an increase in pollution abatement costs (Morgenstern et. al 2002), or with pollution abatement requirements modeled as quasi-fixed factors of production (Berman and Bui 2001).

353 In practice, sticky wages are expected in virtually all labor markets. However, general equilibrium models often incorporate real wage adjustments through changes in the price level. For example, a CGE model might incorporate output price changes, perhaps caused by environmental regulation, that are significant enough to change the general price level in the economy including aggregate real wages (Hafstead and Williams, 2018). Most, if not all, environmental regulations have not had measurable impacts on general price levels.

354 Walker (2013) finds earnings effects only for workers who separate from their jobs, and zero wage effects for workers who remain employed. Curtis (2018) finds wage effects for new hires only.
Text Box 9.1: Unemployment Impacts, Health, and Wealth

Unemployment is associated with increased mortality risk for those in their early and middle careers, but “whether unemployment is causally related to mortality remains an open question... and recent research has begun to focus on possible confounding, mediating, and moderating factors.” (Roelfs et al., 2011, pg. 2). The figure below shows the complex relationships between workforce status, regulation, health, and wealth, focused on those most related to environmental regulation. As line (3) indicates, a bi-directional relationship exists between unemployment and health. Causality is difficult to identify for the unemployed population: increased mortality risk may be caused by unemployment itself, independent of pre-existing health status, or increased mortality risk may be caused by a decline in health that also resulted in a workforce status change (e.g. job loss, unemployment). The first causal pathway is potentially informative for regulatory analysis, but many studies lack detailed data to isolate this pathway.

A nascent economic literature uses detailed worker data to explore the effect of plant closures or mass layoff events on health outcomes. Sullivan and von Wachter (2009b) find increased mortality rates among displaced male workers with long job tenure in Pennsylvania and, in a study of displaced Austrian male workers, Kuhn et al. (2009) find that involuntary job loss negatively affected mental health. There may also be positive health impacts of moving out of unemployment into a job; for example, a decreased risk of depression (van der Noordt, et. al. 2014).

The usefulness of this literature for regulatory analysis depends on whether involuntary job loss and unemployment are expected impacts. This possibility is illustrated by line (2). If detailed financial data are available and such an impact is expected, the RIA may describe the likelihood of plant closures and employment impacts for affected workers. However, analysts should exercise caution if transferring empirical estimates from the literature on adverse health impacts into such a description. Studies often use selective samples that may not correspond well to affected workers in the policy scenario and some lack detailed information on important worker characteristics, such as the “involuntariness” of job separation.

While RIAs may estimate employment impacts of regulations, it is challenging to identify job displacement at the firm- or plant-level due to an environmental regulation. Both Curtis (2018) and Hafstead and Williams (2018) find workforce adjustments happen through reduced hiring rates rather than increased job separations. Reduced hiring rates could still imply that workers who might otherwise have been hired instead spend more time unemployed, though this may have a smaller impact than increased job separations. In a survey of firms experiencing mass layoffs, government regulation is rarely a stated reason (U.S. BLS 2011). More research is needed.

Finally, the economics literature has found connections between wealth and health (line (5)). Sullivan and von Wachter (2009a) find that higher variability of earnings is associated with increased mortality. Dobkin et. al. (2018) find that adverse health events measured by hospital admissions can lead to reduced earnings and increased risk of bankruptcy for those without health insurance.
**Estimating Labor Impacts:** An employment impact analysis provides a baseline profile of potentially affected employers and workers, labor market conditions, and possibly communities. The analysis discusses or estimates potential changes or shifts in employment. Both positive and negative employment changes should be examined, including for example, possible employment impacts in the regulated sector as well as the environmental protection sector. When feasible, analysts can describe direct changes expected in the use of labor by the regulated sector for compliance requirements. In cases where impacts are anticipated, and if data and modeling allow, analysts can describe employment impacts due to changes in production, revenues, or expenditures by the regulated sector and potentially also by related sectors.

A baseline employment profile may include the size of the affected labor force, the degree to which affected labor markets are concentrated among few employers, the amount of labor mobility, job turnover, job search rates, and the affected workers’ regional or occupational unemployment rates. Recent employment trends may be relevant. Characteristics of affected workers, such as sector, industry, occupation, earnings, experience, and job skills may be described. If employment impacts are expected to be concentrated in certain communities, those communities could be characterized. Table 9.2 lists examples of possible data sources that may be helpful in developing a baseline employment profile.

To examine the incremental impacts of a regulation on employment, analysts should keep in mind that labor demand may be affected differently in the short-run compared to the long-run. For example, the Clean Power Plan RIA includes employment impact estimates for the power sector both for short-run effects (e.g. construction, installation, or other “one-time” employment needs) as well as long-run, or ongoing effects, such as shifts in the use of fuels in electricity generation (U.S. EPA 2015).

For many regulations, assessing employment impacts will be limited to a qualitative discussion. It will include the baseline profile described above, and the likely direction of change of employment levels in affected sectors and occupations. A discussion of any concentrated employment impacts, regionally or otherwise, would be useful. Information on the ability or limitations of workers to respond to shifts in labor demand should be considered.

A quantitative analysis may project changes in employment in affected sectors by occupation or among other groups of workers (e.g., by region). The quantitative estimates can use information from the compliance cost analysis if the labor requirements for expected compliance activities including installation, operation and maintenance of pollution control equipment, monitoring, inspecting, reporting, recordkeeping are provided. For example, the RIA for the EPA’s recent Risk Management Plan rule included estimates of changes in the number of labor hours required for compliance activities among different occupations and for different sized facilities. Its analysis of labor impacts examined how many total labor hours on average per year would be required, and whether new workers would likely be hired. These data allow employment impact estimates to be developed for the regulated sector (e.g., chemical facilities) and related sectors (e.g., contract inspectors), both in the short-run (e.g., for rule familiarization) and long-run (e.g., for auditing activities).

In quantitative analyses, aggregated labor hours should be converted to estimates of annual average job-years or FTEs. When these estimates are small relative to average employment at a representative facility or firm, a reasonable

---

355 These labor costs (in dollars) are already included in the cost analysis of an RIA as they are costs to regulated firms. (See Chapter 8 for more information.) They can also be described within an employment impact analysis, and may be converted from dollar value labor costs to numbers of employees, or annual FTE, etc.

356 See, for example, Smith (2015) on local labor market conditions and unemployment, and Baumol and Oates (1988) Chapter 15, on reemployment prospects and consideration of workers in communities characterized by one or two large employers.


358 A job-year is not an individual job and is not necessarily a permanent or full-time job. Instead it is the work performed by one full time equivalent (FTE) employee in one year. For example, 20 job-years may represent 20 full-time jobs or 40 half-time jobs in a given year, or any combination of
assumption may be that existing employees or contractors would take on the tasks for regulatory compliance rather than a facility or firm adjusting the size of its workforce.

While transparent, the quantitative approach just outlined only addresses a subset of employment impacts as it does not address shifts in labor demand associated with potential changes in output in the regulated, or related, sectors. When a regulatory cost analysis projects shifts in output due to compliance costs or shifts in the composition of production within the regulated sector (e.g. shifts in the electricity generation fuel-mix) a more detailed analysis may be possible. In these specific cases, analysts can estimate employment impacts by multiplying the change in output by the average amount of labor per unit of output (or per value of shipments) in the sector. This gives an approximation of the output effect, a potentially important type of employment impact. The U.S. Census and U.S. Bureau of Labor Statistics (BLS) provide estimates of the units of labor associated with expenditures (or value of output/sales) at the industry-level. A limitation of this type of analysis is that in practice producer-level employment impacts will likely differ from aggregate, industry-level employment impacts. For example, relatively more efficient firms may expand output (and employment) to pick up the slack as less efficient producers contract (Jaffe et. al. 1995; Tietenberg 2002; and Christiansen and Tietenberg 1985).

Detailed industry information is useful to develop disaggregated employment estimates for related sectors. For example, as part of estimating labor impacts in regulatory analyses of air pollution regulations affecting the electric power sector, the EPA examined coal mining by region. The EPA combined estimates of changes in coal demand with detailed estimates of coal supply and regional coal mining productivity data available from the U.S. Energy Information Administration (US EIA). Labor productivity differed significantly across geographic regions, e.g. in 2018 labor productivity in Appalachia was 2.32 short tons of coal per labor hour, in the interior region it was 4.73, and in the West, it was 17.09 (U.S. EPA 2018). This level of detail informed the analysis of employment impacts.

Approaches for estimating the employment impacts of environmental regulation are rapidly evolving. Analysts are encouraged to engage the EPA’s National Center for Environmental Economics early in the process when developing a strategy for evaluating the employment impacts of a regulation. Analysts should describe the methods used in a quantitative employment impacts analysis – whether it analyzes changes in pollution abatement activities alone, or combined with changes in production - and explain analytical limitations, which might include:

- Use of an estimation approach that produces partial employment impacts and does not fully measure all potential changes in regulated and related sectors.
- Application of average labor-to-cost or labor-to-output ratios instead of the change in labor expected in response to incremental increases or decreases in costs or production.
- Estimation of labor-to-cost, or labor-to-output, ratios at the industry-level that reflect the labor component of pre-regulation costs or production rather than post-regulation costs or production. This is a limitation because such ratios can be influenced by the regulation.
- Use of available labor ratio data that may be for industrial sectors not well-aligned with the affected sectors.

---

full- and part-time workers such that the total is equivalent to 20 FTE employees. In practice, for example, if the cost analysis for a regulation estimates a need for 1 million labor hours per year in the regulated sector to conduct compliance activities, this could be converted to approximately 480 job-years by dividing 1 million by the annual work hours for a full-time employee which equals 2.080.

Data on labor per unit of output would be a proxy for the overall effect on labor demand in the regulated sector. These data are based on past production processes and therefore are not directly useful for measuring a substitution effect between labor and other productive inputs when compliance activities are required in the regulated industry.

359 The U.S. Census and U.S. Bureau of Labor Statistics (BLS) provide estimates of the units of labor associated with expenditures (or value of output/sales) at the industry-level.

• Heterogeneity of firm- or facility-level responses to regulation, especially those of marginal facilities operating at the tail end of productive efficiency, may be glossed over by labor ratio data typically available at the sector level only.

Cautionary Notes: Analysts should proceed with caution regarding the following approaches sometimes used to estimate quantitative employment impacts of regulation.

Transferring Certain Empirical Estimates: Morgenstern et al. (2002) estimated the effect of pollution abatement expenditures on the quantity of labor in four highly polluting and regulated industries. However, a later attempt to replicate and extend this research failed. Analysts should not rely on the empirical estimates from Morgenstern et al. (2002). Likewise, analysts should not rely on the estimates from Belova et al. (2013, 2015) as the authors “recommend that the EPA refrain from using these results until the underlying cause(s) for the implausibly large estimates in the employment effects found in Belova et al. (2013a) are uncovered and resolved.”


Input-Output Analysis: As described in Section 8.4.4, input-output analysis can provide employment impact estimates. This type of analysis is most suitable for analyzing detailed sectoral impacts of regional, state, or local policies in the short term. In general input-output models should not be used for estimating impacts of national regulations because they do not allow prices, production processes, or technologies to adjust over time. As a result, they represent a very short-term response to regulation and are better equipped to represent the response of a single region to a small regulatory change which is not expected to affect prices. They are of limited use for analyzing large regulatory changes or regulations that are national in scope.

Plant Closures and Employment: Section 9.5.2.2 discusses difficulties in assessing the likelihood of plant closures given a dearth of data and a limited ability to model key factors, such as expectations of future profitability. Even in cases when estimates of the likelihood of plant closures are available, estimating employment impacts from them can be difficult. Employment impacts associated with plant closures may differ from the projected decline in plant output. Firms face labor adjustment costs, and, for example, multi-plant firms may choose to transfer workers, potentially those more skilled and experienced, to other locations (Ferris, Shadbegian, Wolverton 2014). Or, as noted above, production and employment may shift between firms, away from higher cost plants towards more efficient competitors. Such heterogeneity implies that employment impacts at the firm or plant-level can differ in direction from industry-level employment impacts. Analysts should consider these possibilities.

9.5.2.5 Impacts on Other Productive Factors: Land and Capital

In addition to labor impacts, environmental regulation can lead to changes in the demand for, and value of, other factors of production employed by regulated firms. Economists label these other factors of production as land (any natural resource), and capital (any man-made resource). In general, environmental regulation is expected to have varying effects across factors, and tracing impacts back to specific factors is difficult (Fullerton 2009). Regulated firms often own these assets and disaggregating information on changes in firm-level profitability to allow estimation of impacts on specific types of capital or natural resources is often challenging. Estimating changes in the quantities demanded of broad categories of land and capital is more practical. There are two separate and valid ways to represent the value of

361 Quote is from Belova et al. (2015). Note that Belova et al. (2013a) in the quote is identical with Belova et al. (2013) cited above.
362 Even for regional analyses, input-output models tend to overestimate impacts. “They typically include exogenous multipliers that magnify direct effects on output and employment based on the assumption that all new economic activity will recirculate within the regional economy. Input-output models tend to ignore displacement of workers or resources that might occur outside the region under analysis” (U.S. EPA 2011).
363 The underlying data can be useful for showing related sectors, e.g. upstream and downstream.
364 Land and capital may also be rented or supplied under contract. When not owned by the regulated firm, the impacts are considered upstream, as
factors of production: earnings per period (also called rates of return) or asset values. The latter is the discounted present value of the future stream of earnings generated by the productive factor.

The relationship between changes in regulated firms’ price and quantity of output, and changes in their factor demands or factor returns, can be complicated. In response to stricter environmental regulation, factors used intensively by the regulated industry might experience reduced demand and/or returns. For instance, if a unit of capital is not perfectly mobile, or a type of natural resource cannot be taken out of production in the short-run, it may lose value and impose a burden on the owners if its use falls (Fullerton and Muehlegger 2019). Factors that are complements to pollution abatement might actually see an increase in demand or returns; while those that are highly mobile with similarly valued alternative uses should hold their value. There are two general expectations for the long-run response to environmental regulation. One is for land and capital to shift away from high-emission activities towards lower emitting ones, including the environmental protection sector; another is for land and capital to shift towards less regulated uses. Regionally differentiated impacts on capital and land are possible when the stringency of pollution control varies by region.

To estimate how the costs of compliance are passed through to and distributed across productive factors, analysts need the cross-price elasticities between these factors. When this type of information is not available, analysts can examine current production practices and the input biases of anticipated abatement activities to inform a qualitative discussion of likely impacts on productive factors.

In general, income earned from ownership of land and capital (or of firms) tends to make up a greater proportion of earnings for higher-income households. Thus, an increase in regulatory costs passed through to households via lower returns to capital tend to be progressive, placing a greater share of the burden on wealthier households. The magnitude of the impact on owners and investors depends on the proportion of their portfolio affected by the change.

A different impact on factors of production stems from improved environmental quality which can be capitalized into the price of nearby land, and buildings (including housing). The increase in property values accrues to the owners at the time of the improvement as a positive economic impact: an increase in asset value. The degree to which the land and buildings are owner occupied versus rented and the degree to which the increased value is passed on in the form of higher rents will influence who experiences positive versus negative impacts of the environmental improvement. If landlords increase rents to the point of forcing out renters, then the renters may experience transitional impacts from relocation activities. Identifying how owners and renters respond to improved environmental quality is a complicated exercise and quantitative analysis is challenging. A qualitative discussion can be useful. Related literature and modeling challenges are discussed in the final paragraphs of Section 10.2.1.

### 9.5.2.6 Impacts on Related Markets

An environmental regulation may affect markets other than those that are directly regulated. Related markets may be positively affected, such as those in the environmental protection industry or those producing substitutes; or negatively affected, such as those producing complements, or those who are up- or downstream from the regulated industry. (Note that the environmental protection sector may overlap with upstream markets.) If the regulation causes a firm to use different inputs or new technologies, then the producers of the new inputs will gain, while the producers of the old ones will be burdened. Consumers in the related markets may experience impacts as well (see section 9.5.1).

---

365 For example, if a regulation induces firms to switch from high-carbon coal to lower emitting natural gas, then the value of coal will decline, and the stock value of coal-intensive businesses could decline as well.

366 For more details, see Rausch et al. (2011) or Fullerton et al. (2011).

367 If land improvements are concentrated and substantial, there could be community-wide effects. See Section 9.5.3.
Downstream impacts may accrue to firms who purchase the regulated firms’ outputs. In general, when analyzing related markets, analysts should consider the same potential impacts as for directly regulated markets.

If substantial impacts on related industries are expected, it will be useful to include firm sizes, profit margins, growth rates, and more, in a baseline profile. For instance, when the regulated sector sells an intermediate good or service (e.g., electricity), questions that might be relevant include: What proportion of the purchasing firms are small or face narrow profit margins? Are substitute inputs readily available? What proportion of the purchasing firms’ spending goes to the regulated firms?

Partial equilibrium models that represent significantly affected, related markets may be useful, although their use may be limited by scarce data and resources. For regulations that are expected to substantially affect many related markets, a general equilibrium model as described in Section 9.5.5 might be considered, though the additional conditions described there should also be satisfied.

### 9.5.2.7 Impacts on Energy Supply, Distribution, or Use

EO 13211 requires agencies to prepare a Statement of Energy Effects for “significant energy actions,” which are defined as significant regulatory actions (under EO 12866) that also are “likely to have a significant adverse effect on the supply, distribution, or use of energy.” OMB guidance suggests that adverse effects could include any of the following:

- Reductions in crude oil supply in excess of 10,000 barrels per day;
- Reductions in fuel production in excess of 4,000 barrels per year;
- Reductions in coal production in excess of 5 million tons per year;
- Reductions in natural gas production in excess of 25 million mcf per year;
- Reductions in electricity production in excess of 1 billion KWH per year or in excess of 500 MW of installed capacity;
- Increases in energy use required by the regulatory action that exceed any of the thresholds above;
- Increases in the cost of energy production in excess of 1 percent;
- Increases in the cost of energy distribution in excess of 1 percent; or
- Other similarly adverse outcomes.

A regulatory action also may have adverse effects if it is likely to:

- Adversely affect, in a material way, productivity, competition or prices in the energy sector;
- Adversely affect, in a material way, energy productivity, competition or prices within a region;
- Create a serious inconsistency or otherwise interfere with an action taken or planned by another agency regarding energy; or
- Raise novel legal or policy issues adversely affecting the supply, distribution or use of energy arising out of legal mandates, the President’s priorities or the principles set forth in Executive Orders 12866 and 13211.

For actions that may be significant under EO 12866, particularly for those that impose requirements on the energy sector, analysts must be prepared to examine the energy effects listed above.

---

9.5.2.8 Impacts on Domestic and International Competitiveness

Competitiveness impacts are regulatory impacts that change the distribution of market power among firms or sectors, either domestically or internationally. Unfair advantage may accrue to producers that are free from regulatory constraints, or that face less expensive regulation. Or, high fixed costs that are incurred to comply with environmental regulation may cause production to become concentrated among fewer firms, enhancing their monopoly profits over the long run. Regulatory constraints may differ among specific subsets of sectors or firms: existing versus new; or small versus large.369 If some firms find it less costly to comply with a regulation, they may benefit competitively at the expense of other regulated firms. Analysts may wish to consider the extent to which production is shifted towards plants with higher than average productivity (Jaffe et al. 1995).

As with other impact categories, the extent to which a regulation leads to effects on competitiveness depends on the interaction between the regulated firms’ absorption of compliance costs and their market structure.370 In general, greater compliance flexibility is expected to reduce competitiveness effects.371

A first step to gauge the potential for competitiveness effects is the baseline profile of affected industries. The profile should identify which domestic and international firms compete with regulated entities, and their basic market structures. Do competitors face expensive environmental regulation? Is the output produced by regulated firms differentiated from that of competitors, potentially reducing impacts on competition? The literature suggests an increased likelihood of competitiveness effects for industries in which compliance costs are high relative to total production costs.372

Consideration of the impact of new environmental regulation in three key areas is particularly germane to competitiveness effects. First, lack of access to debt or equity markets to finance market entry including regulatory costs, can represent significant barriers to entry. Over the long run, this can change market structures and reduce competitiveness. Second, a regulation may have an impact on market concentration. A potentially useful measure of concentration is the Herfindahl-Hirschman index (HHI), which is the sum of the squares of the market shares of each firm in a given market. The Department of Justice uses the HHI to estimate changes in market concentration due to mergers and acquisitions. Post-merger HHI values that are below 1,000 are considered “unconcentrated,” between 1,000 and 1,800 are regarded as moderately concentrated, and above 1,800 are considered highly concentrated.373

Finally, the impact of regulation on the market position of domestic firms relative to their foreign counterparts is important. Domestic environmental regulations may have global economic implications because the costs of domestic producers is increasing relative to foreign producers.374 Analyses of impacts on international competitiveness have been concentrated on the most pollution- or energy-intensive and most trade-exposed industries because they are most likely to face regulatory requirements and least able to pass compliance costs to consumers.375 For example, in the context of unilateral climate policy, proposed legislation has focused on potential competitiveness impacts on trade-exposed

---

369 Section 9.5.2.2 discusses the impacts of differentiated regulation that occurs when existing firms are “grandfathered in.”

370 The importance of this interaction is discussed by Iraldo et al. (2011).

371 Evidence for this is presented by Iraldo et al. (2011) and Jaffe et al. (1995).


373 For more information, see https://www.justice.gov/atr/15-concentration-and-market-shares.

374 A related literature examines how differences in environmental regulation across countries, states, or sectors may result in increased emissions in less regulated countries, also called emissions leakage. For instance, see Bohringer, et al (2012) and Fischer and Fox (2012).

375 Carbone and Rivers (2017) discuss the impacts of environmental regulation on international competitiveness. In general, the literature has found relatively small effects (Jaffe, et al. 1995; Aldy and Pizer 2015; Carbone and Rivers 2017). Jaffe et al. (1995) point out that concerns about industry competitiveness may also ultimately affect consumers as net exports decline and in the long-run imported goods become more expensive as the economy returns to balanced trade.
domestic energy firms. Quantifying these effects can be complex and may require a multi-country computable general equilibrium model. There are three classes of indicators of impacts on international competitiveness: the degree to which net exports change, the degree to which production shifts overseas (i.e., pollution haven effect), and the relative change in investment from domestic (regulated) producers to producers in other countries (Jaffe et al. 1995).

### 9.5.3 Impacts on Communities

Environmental regulation may have significant impacts on communities or neighborhoods. Facility closures or production curtailments provide an example of locally concentrated economic impacts which could be acute in areas with limited economic opportunities. Regulation on coal-fired power plants could, for example, have negative impacts on coal-dependent communities. Plant closures and employment cuts can affect others in the community as the economic base and local tax revenues decline. In the longer term, there may be impacts on the provision and quality of local public and private goods.

Localized improvements in environmental quality, such as hazardous site cleanup, can reduce health risks and improve local property values thereby increasing the local tax base, and potentially in the long run, improving local public and private goods. If low-income residents are largely renters, then they could be burdened by increases in land values and subsequent increases in rent due to improved environmental quality. Low income property owners can be burdened through property tax increases, as realizing the benefits of increased property values requires selling and moving which typically imposes high transaction costs.

When localized impacts of environmental policies are expected, a baseline profile of affected communities will be informative. Data on the unemployment rate, average income level, the poverty rate, whether the community is rural or urban, and its growth rate can help inform policy makers as to the relative disadvantages faced by affected communities.

### 9.5.4 Impacts on Governments and Non-Profits

State and local governments and their residents, and non-profit organizations may incur costs or bear the burden of costs from EPA regulations. The frameworks and impacts discussed above apply to private markets. Governments and non-profits are distinctive because they are public, and they are not motivated by profits. Analysts should consider potential impacts to governments and non-profits, including short- and long-run impacts. Useful measures for evaluating impacts on these types of entities include assessments of the difficulty of paying regulatory costs and of continuing to provide services.

Examples of important impacts on government include water treatment costs paid by municipally-owned water authorities to comply with water quality standards. Air pollution controls required of power plants may affect municipally-owned electric companies. Implementation and enforcement costs associated with a variety of environmental regulations may impose costs on state or local government. If regulation affects the local tax base, then

---

376 For a discussion, see “Economy-Wide Modeling: Use of CGE Models to Evaluate the Competitiveness Impacts of Air Regulations,” Jun 17, 2016: https://yosemite.epa.gov/sab/SABPRODUCT.NSF/0/07E67CF77B54734285257BB0004F87ED/$File/competitiveness+memo_061716.pdf

377 See, for example, Morris, Kaufman, and Doshi (2019).

378 For a discussion of gentrification in housing markets, see Section 10.2.1 of this guidance document, Chapter 8 of EPA’s Handbook on the Benefits, Costs, and Impacts of Land Cleanup and Reuse (2011), and/or Banzhaf and McCormick (2012).

379 For details regarding examining environmental justice communities, see Chapter 10.

380 For a discussion on contributors to higher susceptibility, see the EPA Technical Guidance for Assessing Environmental Justice in Regulatory Analysis (U.S. EPA 2016), Section 4.3, which addresses susceptibility or vulnerability within groups such as communities.

381 In some cases, consideration of impacts on government and nonprofits is required. For example, UMRA requires assessment of impacts to state, local, and tribal governments and the RFA as amended by SBREFA requires assessment of impacts to small entities including governments and nonprofits (see Section 9.1 and Chapter 2).
there may be impacts on government revenues or expenditures that may affect the provision of local public or private goods and services.

To understand economic impacts on state, local, and tribal governments, analysts should develop a baseline profile potentially including the following relevant factors:

- Size of the population in the community;
- Property values;
- Household income levels (e.g., median or income range);
- Age distribution;
- Unemployment rate;
- Foreclosure rate;
- Revenue amounts by source.

If property taxes are the major revenue source, then the assessed value of property in the community and the percentage of this assessed value represented by residential versus commercial and industrial property may be important. If a government entity serves multiple communities, such as a regional water or sewer authority, then information for all the communities in the service area may be relevant.

To gain insight into the difficulty governments might experience when facing new regulatory costs, U.S. EPA’s Guidance for Financial Capability Assessment and Schedule Development for Combined Sewer Overflows (1997) suggests examining baseline financial capability by exploring indicators of debt, socioeconomic conditions, and success regarding financial management. Analysts can obtain the community’s bond or credit rating, which is itself determined by an assessment of financial health. For governments that rely on property taxes for income, analysts might consider the amount of debt that must be repaid through property taxes (known as net debt) per capita; or the net debt relative to the value of taxable properties. Property tax revenues relative to full market value of properties may be a useful indication of the property tax burden (U.S. EPA 1997). Table 9.3 provides thresholds used by the Office of Enforcement and Compliance Assurance (OECA) and the Office of Water (OW) to indicate weak, mid-range, or strong financial wellbeing of government entities.

To screen for significant impacts on governments, analysts may wish to consider new regulatory costs per capita, and new costs as a ratio to median household income. Depending on these values, further analysis might be desirable. Further analysis should consider a government entity’s options for funding new costs or how new process requirements could change operating procedures. For example, what is the availability of new loans or grants and user fees? Are there other viable routes for increasing funds available to finance new regulatory costs? Do new processes alter the quality or quantity of goods and services provided to residents? Other factors that are potentially relevant are the historic trend in government revenues; the capability of the revenue sources to shoulder additional financial burdens; and the magnitude of the benefits from the rule enjoyed by citizens.

Finally, indirect impacts on state, local, and tribal government may be important if a policy changes local property values or employment rates or has other community-wide impacts. For example, brownfield grants to assess or clean up land

---


383 For another source that explores approaches for assessing the health of a local government see McDonald (2018).

384 For instance, when assessing regulatory costs, OW and OECA consider financial impact as low if costs per household are less than 1 percent of median household income; mid-range if it is 1 to 2 percent of median household income; and high if it is greater than 2 percent (U.S. EPA 1997).
may cause small increases in local property values which could raise property tax revenues (Sullivan 2017). On the other hand, a policy that exacerbates unemployment, for example, could cause more spending on assistance programs. EPA regulations may also affect non-profit organizations. For example, non-profit hospitals face costs from hazardous waste disposal requirements. A baseline profile for non-profits should consider:

- Entity size and size of community served;
- Goods or services provided;
- Operating costs; and
- Amount and sources of revenue.

If the entity is raising its revenues through user fees or charging a price for its goods or services (such as university tuition), then the income levels of its clientele are relevant. If the entity relies on contributions, then it would be helpful to know the financial and demographic characteristics of its contributors and beneficiaries. If it relies on government funding (such as Medicaid) then possible future changes in these programs should be identified.

To screen for impacts on non-profits, analysts can compare regulatory costs to baseline revenues or operating expenses. Regulatory costs can also be compared to baseline asset values or, after accounting for debts, net asset values. If these ratios are large, insights would be gained from information on the relative importance, size, and growth rate of the nonprofit, the nature of the population being served, and the vulnerability of revenues and donors.

**Impacts on Small Governments and Non-Profits**

Consideration of impacts on small governments and small non-profits is required by the RFA as amended by SBREFA.\(^{385}\) The RFA defines a small governmental jurisdiction as the government of a city, county, town, school district, or special district with a population of less than 50,000. As with the definition of small business, the RFA authorizes agencies to establish alternative definitions of small government after opportunity for public comment and publication in the Federal Register. Any alternative definition must be “appropriate to the activities of the agency” and “based on such factors as location in rural or sparsely populated areas or limited revenues due to the population of such jurisdiction” (U.S. EPA 2006). Under the RFA, economic impacts on small governments are included in the screening analysis for

\(^{385}\)See Chapter 2 and Section 9.2 for more information.
significant economic impacts on a substantial number of small entities (SISNOSE), and any required regulatory flexibility analysis. In order to determine SISNOSE for small governments, the EPA conducts a screening analysis for both proposed and final rules based on annualized compliance costs as a percentage of revenue (U.S. EPA 2006).

The Unfunded Mandates Reform Act (UMRA) uses the same definition of small government as the RFA, with the addition of tribal governments. Section 203 of UMRA requires the Agency to develop a “Small Government Agency Plan” for any regulatory requirement that might “significantly” or “uniquely” affect small governments. In general, “impacts that may significantly affect small governments include — but are not limited to — those that may result in the expenditure by them of $100 million [adjusted annually for inflation] or more in any one year.” Other indicators that small governments are uniquely affected may include whether they would incur higher per-capita costs due to economies of scale, a need to hire professional staff or consultants for implementation, or requirements to purchase and operate expensive or sophisticated equipment.386

The RFA requires separate consideration of regulatory impacts on small non-profits and defines one as a non-profit “enterprise which is independently owned and operated and is not dominant in its field.” Agencies are authorized to establish alternative definitions “appropriate to the activities of the agency” after providing an opportunity for public comment and publication in the Federal Register. Under the RFA, direct economic impacts on small non-profit organizations are included in the SISNOSE screening analysis, and if required, the regulatory flexibility analysis for a rule. In order to determine SISNOSE for small non-profits, the EPA conducts a screening analysis for both proposed and final rules based on annualized compliance costs as a percentage of operating expenditures.387

9.5.5 Economy-Wide Impacts

The more interconnected a regulated sector is with the rest of the economy, the greater the likelihood that a regulation will affect related markets. If a regulation is expected to affect markets with (i) significant cross-price effects between markets, and (ii) significant pre-existing distortions, it may be appropriate to examine economy-wide impacts in a supplemental analysis (U.S. EPA 2017). Pre-existing market distortions that could be exacerbated by environmental regulations include taxes or subsidies on labor, energy, or capital; monopoly or monopsony power; price controls; or other government regulations that change the way markets operate.

Computable general equilibrium (CGE) models are particularly effective at assessing long-run economy-wide impacts.388 These include the allocation of employment or other factors of production across sectors, the distribution of output by sector, and the distribution of income among households. For example, regulations in the power sector may cause electricity prices to increase. The price increase will affect all industries that use electricity as an input to production, as well as households. A CGE model can assess the distribution of consequent changes in production and consumption. By design, the basic capacity to describe and evaluate these sorts of impacts exists to some extent within every CGE model. More detailed impacts (e.g., effects on a certain type of facility or on an environmental endpoint such as drinking water) are difficult to capture in a CGE model due to model convergence and/or data constraints.

The simplest CGE models typically include a single representative consumer, a set of relevant production sectors, and a government sector within a single-country, static framework. Additional complexities can be specified. A CGE model can be solved dynamically over a lengthy time horizon, incorporating intertemporal decision-making on the part of

---


387 See Table 1, “Recommended Quantitative Metrics for Economic Impact Screening Analyses” of U.S. EPA 2006.

388 CGE models assume that for some discrete period of time an economy can be characterized by a set of conditions in which supply equals demand in all markets. When the imposition of a regulation alters conditions in one market, the model determines a new set of relative prices that return the economy to its long-run equilibrium. While highly aggregate in nature, CGE models capture substitution possibilities between production, consumption and trade; interactions between economic sectors; and interactions with pre-existing distortions. Thus, they provide information on changes outside the directly-regulated sector. See Chapter 8 for more discussion.
consumers or producers. These decisions have implications for the treatment of savings, investment, and the long-term profile of consumption and capital accumulation. Consumers can be divided into income quintiles or deciles, and producers disaggregated into a variety of regions and sectors, each producing a set of unique commodities. The government, in addition to implementing a variety of taxes and other policy instruments, may provide a public good or run a deficit. CGE models can be international in scope, consisting of many countries or regions linked by international flows of goods and capital. The behavioral equations that characterize economic decisions may take on simple or intricate functional forms.

While CGE modeling is complex, the effort may be worthwhile when appropriate data (e.g., demand, supply, and cross-price elasticities) are available, and impacts are likely to be substantial and widespread. Text Box 5.2 and Chapter 8 discuss detailed criteria for judging model quality. Feedback from the Science Advisory Board (SAB) identified several guiding principles as to when economy-wide modeling is appropriate for assessing economic impacts of regulation (U.S. EPA 2017). Aspects of a CGE model that could affect suitability include: degree of temporal, sectoral, and geographic disaggregation; time horizon; the way in which firm and household expectations about the future are modeled; the types of impacts that can be forecast; and the approach for representing the policy instrument. CGE models may be useful as a supplement to other analytic approaches to evaluate sectoral effects (including shifts in labor or capital between sectors), impacts on energy supply and energy prices, and effects on consumers. In some instances, linking a CGE model to sector models may be a useful way to leverage the relative advantages of both approaches in a single comprehensive framework (U.S. EPA 2017).

CGE models have limitations. Many are not designed to illuminate certain types of impacts, such as short-run or transitional impacts. For example, a standard forward-looking CGE model that assumes full employment and instantaneous market adjustments is ill-suited to evaluate overall employment impacts or the potential for short-run disequilibria in labor and capital markets. Analysts interested in evaluating the short-run impacts of a policy should select a different framework for analysis. Finally, relatively few CGE models incorporate feedbacks from changes in pollution; instead they mainly focus on private markets.

A partial equilibrium model of multiple markets that considers the interactions between a regulated market and other closely related markets may be a practical alternative to a CGE model. As with a CGE model, such models require estimates of demand and supply elasticities and cross-price elasticities for included markets. Partial equilibrium models may be appropriate for regionally-based or sector-specific regulations that are too narrowly defined to be adequately captured in more aggregate CGE models.

The SAB recommends that analysts apply the simplest model that is adequate to address the policy question and consider a suite of models when possible (U.S. EPA 2017). A balance should be struck between capturing detail and complexity in the model versus transparency and tractability of the analysis.

As with all economic models, economy-wide and partial equilibrium models are simplified representations of complex economic systems built to assess relationships between economic factors. They are useful for estimating effects on groups but are not reliable predictors of firm or facility-level decisions. See Section 9.5.2.2 for further explanation of the common simplifying assumptions about firm decision-making.

### 9.5.6 Impacts of Benefits

As with costs, the benefits from improved environmental quality or health can accrue to, and may differ among, a wide variety of individuals. Environmental benefits are generally nonmarket effects and as such pose special analytic challenges. A key determinant of differential impacts is whether environmental improvements differ among affected groups (e.g., due to different exposure pathways, for example), or are uniform but have variable impacts due to

---

389 See Textbox 8.2 for more discussion of model linking.
differences in pre-existing factors such as baseline exposures or health (for more discussion, see U.S. EPA 2016, especially Sections 4.2 and 4.3).

The literature provides several potential frameworks for explicitly considering variability in the impacts of benefits across groups. Typically, these frameworks start with defining environmental damages as a function of exposure and individual susceptibility to environmental stressors, then they identify sources of susceptibility, and finally they assess the impacts from environmental regulation (see, for instance, Hsiang et al. 2019; Gee and Payne-Sturges 2004; and Morello-Frosch and Jesdale 2006).

Useful information to improve understanding of the distribution of regulatory benefits includes:

- the types of health effects or other benefits;
- population groups to whom the benefits are expected to accrue;
- how exposure varies across the affected groups; and
- how beneficial outcomes vary across population groups.

In addition to accruing to those who directly experience a reduced health risk, health and environmental quality benefits may also accrue to people who own homes near improved environmental quality, or to employers whose workers enjoy improved health and increased labor productivity, as well as to others.

Chapter 10 discusses how to analyze health effects and benefits for specific populations of concern (i.e., by income, race/ethnicity, and age). The data and methods discussed there may be relevant for analyzing the distribution of benefits on other categories of people, on communities, or on the general population. Sometimes analysts may wish to account not only for the ways in which changes in the regulated sector affect the distribution of benefits, but also how price and quantity responses across the economy affect the distribution of benefits, or how changes in environmental quality affect prices and quantities. Absent a partial-equilibrium or economy-wide model that explicitly incorporates benefits, relatively rare in the literature, these indirect impacts are difficult to evaluate.

Chapter 9 References


9-32


Morello-Frosch, R., and B.M. Jesdale. 2006. Separate and Unequal: Residential Segregation and Estimated Cancer Risks Associated with Ambient Air Toxics in U.S. Metropolitan Areas. *Environmental Health Perspectives*, 114(3), 386-393.


Evaluating a regulation’s economic and distributional impacts is an important complement to benefit-cost analysis. Instead of focusing on quantifying and monetizing total benefits and costs, these types of analyses examine how a regulation allocates benefits, costs, transfers and other outcomes across populations or other groups of interest. This chapter, along with Chapter 9, gives analysts guidance on how to describe the potential economic and distributional impacts of environmental policy. Chapter 9 primarily addresses economic impacts that stem from compliance costs. This chapter provides a broad overview of options for considering impacts that stem from changes in environmental quality and human health risks from an environmental justice (EJ) (i.e., on minority, low-income, or indigenous populations; see section 10.2 for a formal definition), life stage (i.e., on children, older adults), and intergenerational perspective. Note that consideration of costs may also be relevant when evaluating distributional impacts on these specific populations.

The chapter begins with a brief overview of Executive Orders (EOs) and policies related to distributional analyses. It then discusses the analysis of distributional impacts in the context of EJ and children’s health. The chapter concludes with a brief discussion of other distributional considerations, including impacts on older adults and across generations.

10.1 Executive Orders, Directives, and Policies

Consideration of distributional effects across population groups and life stages arises from several executive orders, directives, and other documents. The Agency also has developed separate guidance to provide direction to analysts on conducting environmental justice analyses. Together these orders, directives, and policies provide a solid foundation for considering distributional effects on population groups from an EJ and life stage perspective in the rulemaking process.

OMB’s Circular A-4 (OMB 2003) states that regulatory analyses “should provide a separate description of distributional effects (i.e., how both benefits and costs are distributed among populations of particular concern) so that decision makers can properly consider them along with the effects of economic efficiency.” It specifically calls for a description of “the magnitude, likelihood, and severity of impacts on particular groups” if the distributional effects are expected to be

---

390 This chapter recommends examining the distribution of benefits prior to monetization for reasons discussed in Section 10.1.

391 EPA’s Regulatory Management Division’s Action Development Process Library (http://intranet.epa.gov/adplibrary/adp) is a resource for accessing relevant statutes, executive orders, and EPA policy and guidance documents in their entirety (accessed on April 1, 2020).

392 Some environmental statutes also identify population groups that may merit additional consideration. See Plan EJ 2014 Legal Tools (U.S. EPA 2011a) for a review of legal authorities under the environmental and administrative statutes administered by the EPA.
important (OMB 2003). In addition, the following EOs, described more fully in Chapter 2, directly address distributional considerations for population groups of concern:

- EO 12898, “Federal Actions to Address Environmental Justice in Minority Populations and Low-Income Populations” (1994), calls on each Federal agency to make achieving EJ part of its mission “by identifying and addressing, as appropriate, disproportionately high and adverse human health or environmental effects of its programs, policies, and activities on minority populations and low-income populations.”
- EO 13045, “Protection of Children from Environmental Health Risks and Safety Risks” (1997), states that each Federal agency shall ensure that its policies, programs, activities, and standards address disproportionate risks to children that result from environmental health or safety risks.
- EO 13175, “Consultation and Coordination with Indian Tribal Governments” (2000), calls on Federal agencies to have “an accountable process to ensure meaningful and timely input by tribal officials in the development of regulatory policies that have tribal implications.”
- EO 12866, “Regulatory Planning and Review” (1993), explicitly allows for consideration of “distributive impacts” and “equity” when choosing among alternative regulatory approaches, unless prohibited by statute.

Table 10.1 summarizes the relevant dimensions identified by these orders and directives for distributional considerations and offers examples of potentially affected population groups of concern.

The EPA also has developed guidance for conducting environmental justice analysis for rulemakings, starting with the Guidance on Considering Environmental Justice During the Development of Regulatory Actions (U.S. EPA 2015). This guide is designed to help EPA staff incorporate EJ into the rulemaking process, from inception through promulgation and implementation. The guide also provides information on how to screen for EJ effects and directs rule-writers to respond to three basic questions throughout the rulemaking process:

1. How did the public participation process provide transparency and meaningful participation for minority populations, low-income populations, tribes, and indigenous peoples?
2. How did the rule-writers identify and address existing and/or new disproportionate environmental and public health impacts on minority populations, low-income populations, and/or indigenous peoples?
3. How did actions taken under #1 and #2 impact the outcome or final decision?

In addition, the Technical Guidance for Assessing Environmental Justice in Regulatory Actions (EJTG) (U.S. EPA 2016) provides detailed guidance on how to incorporate EJ into all aspects of the regulatory analytical process, including risk assessment and economic analysis. While the material discussed in this chapter is generally consistent with the EJTG, the

---

393 A Presidential memorandum to heads of departments and agencies that accompanied EO 12898 specifically raised the importance of procedures under the National Environmental Policy Act (NEPA) for identifying and addressing environmental justice concerns (White House 1994). The Council on Environmental Quality (CEQ) issued EJ guidance for NEPA in 1997 (CEQ 1997). The EPA issued guidance in 1998 for incorporating EJ goals into the EPA’s preparation of environmental impact statements and environmental assessments under NEPA (U.S. EPA 1998a).

394 This chapter addresses analytical components of EO 12898 and does not cover other components such as ensuring proper outreach and meaningful involvement.

395 The Presidential memorandum also states that existing civil rights statutes provide opportunities to address environmental hazards in minority communities and low-income communities (White House 1994).

396 A “covered regulatory action” under EO 13045 is any substantive action in a rulemaking that may be economically significant (i.e., have an annual effect on the economy of $100 million or more or would adversely affect in a material way the economy, a sector of the economy, or the environment) and concern an environmental health risk that an agency has reason to believe may disproportionately affect children.

397 EO 13563, issued in January 2011, supplements and reaffirms the provisions of EO 12866.
Table 10.1 Orders and Directives that Identify Dimensions Relevant for Distributional Consideration

<table>
<thead>
<tr>
<th>Dimension</th>
<th>Executive Order or Directive</th>
<th>Examples of Population Groups of Concern</th>
</tr>
</thead>
<tbody>
<tr>
<td>Income</td>
<td>EO 12898; OMB Circular A-4</td>
<td>Low-income groups; poverty status</td>
</tr>
<tr>
<td>Race/ethnicity</td>
<td>EO 12898; OMB Circular A-4</td>
<td>Minority groups</td>
</tr>
<tr>
<td>Age</td>
<td>EO 13045</td>
<td>Children, older adults</td>
</tr>
<tr>
<td>Sex</td>
<td>OMB Circular A-4</td>
<td>Male, female</td>
</tr>
<tr>
<td>Tribes</td>
<td>EO 13175</td>
<td>Indian Tribal governments</td>
</tr>
</tbody>
</table>

EJTG includes a detailed discussion of ways to incorporate EJ into risk assessment, while the main focus of this chapter is consideration of EJ in economic analysis. The EJTG suggests analysts attempt to answer three questions:

1. Are there potential EJ concerns associated with environmental stressors affected by the regulatory action for population groups of concern in the baseline?
2. Are there potential EJ concerns associated with environmental stressors affected by the regulatory action for population groups of concern for the regulatory option(s) under consideration?
3. For the regulatory option(s) under consideration, are potential EJ concerns created or mitigated compared to the baseline?

When data are available to assess both the baseline and policy options under consideration these questions provide for a robust assessment of distributional considerations. The extent to which an analysis can address all three questions will vary due to data limitations, time and resource constraints, and other technical challenges. Analysts are encouraged to document the key reasons why a specific question cannot be addressed. This will help identify future priorities for filling key data and research gaps. In addition, due to the inherent limitations and uncertainties associated with analyses of potential EJ concerns, it is important to conduct sensitivity analysis around key assumptions.

10.2 Environmental Justice

The EPA defines environmental justice as “the fair treatment and meaningful involvement of all people regardless of race, color, national origin, or income with respect to the development, implementation, and enforcement of environmental laws, regulations, and policies” (U.S. EPA 2015). EO 12898 specifically states that Federal agencies "should “…identify and address…disproportionately high and adverse human health or other environmental effects…on minority populations and low-income populations…” (U.S. EPA 2015).

The case is often made that there are no relevant EJ concerns for a rule that is strengthening an environmental standard. After all, environmental quality is improving. However, it is incorrect to conclude that tighter standards necessarily improve environmental quality for everyone. The nuances of a rule could change the distribution of

---

398 The economic analysis often incorporates risk assessments and analyses of risks changes associated with regulatory options. As with many aspects of the analysis, economists and risk assessors need to coordinate when conducting analyses of EJ issues.

399 Note that the term environmental stressor encompasses the range of chemical, physical, or biological agents, contaminants, or pollutants that may be subject to a regulatory action.
emissions across communities. For example, older, more polluting facilities that are grandfathered from the regulation may continue to operate and avoid upgrading equipment to avoid being subject to new emission requirements, which means emissions would not fall by as much and could even increase in those communities. In addition, while there may be few adverse environmental effects, other economic impacts (e.g., the incidence of costs) could affect population groups of concern disproportionately and may warrant examination. Finally, there may be regulatory options that address EJ concerns more effectively than other options by mitigating existing disparities or implementing the standard differently, even when all options improve environmental quality.

Distributional analysis also improves transparency of rulemaking and provides decision makers and the public with more complete information about a given policy's potential effects. Such documentation helps the EPA and the public track and measure progress in addressing EJ concerns. Analysts play a role in ensuring meaningful involvement by explaining distributional analysis in plain language, including key assumptions, methods, and results, and by asking for information from the public (e.g., asking for comment in the proposed rulemaking) on exposure pathways, end points of concern, and data sources that may improve the distributional analysis. Further guidance on ensuring meaningful engagement of environmental justice stakeholders in the rulemaking process can be found in U.S. EPA (2015).

10.2.1 Background Literature

This section provides a brief overview of EJ analysis from the economics and health literatures. Studies of EJ can vary by pollutant, the proxy used for risk or exposure, geographic area, and time period. In addition, studies vary in the extent to which they mainly characterize baseline conditions or attempt to examine the distributional implications of changes in environmental exposure or risk, making it difficult to directly apply general findings to a specific rulemaking. The studies described in this section identify possible methods for evaluating EJ impacts, indicate contexts where EJ concerns may be present, and illuminate some challenges in conducting EJ analyses, such as data availability and sensitivity of the results to assumptions and comparison groups. For a more comprehensive discussion see Ringquist (2005), Banzhaf (2012a), and Banzhaf (2012b).

It is common for EJ studies to ask whether greater amounts of pollution result in increased exposure or poorer health outcomes for certain population groups. However, there is also the possibility that some populations (e.g., children) are more susceptible to pollution for a given level of exposure and that socioeconomic factors (e.g., poorer households may less access to health care) may play a role. While studies that examine differences in susceptibility across population groups are not discussed here, Section 10.2.8.5 discusses various risk considerations potentially relevant to an EJ analysis, including susceptibility (also see U.S. EPA, 2016). In addition, both the EJ literature and this chapter tend to

---

400 U.S. EPA (2015) provides additional information on how an EJ concern may arise in the context of a rule.

401 Meaningful involvement occurs when “1) potentially affected populations have an appropriate opportunity to participate in decisions about a proposed activity (i.e., rulemaking) that may affect their environment and/or health; 2) the population’s contribution can influence the EPA’s rulemaking decisions; 3) the concerns of all participants involved will be considered in the decision-making process; and 4) the EPA will seek out and facilitate the involvement of populations potentially affected by the EPA’s rulemaking process” (U.S. EPA 2015).

402 EO 13166, “Improving Access to Services for Persons with Limited English Proficiency” (2000), may also be relevant when addressing meaningful engagement. EPA’s Order 1000.32 “Compliance with Executive Order 13166: Improving Access to Services for Persons with Limited English Proficiency” requires that the EPA ensure its programs and activities are meaningfully accessible to LEP persons.

403 For a discussion of the possible distributional effects of environmental policies by income (but not race or ethnicity), see Fullerton (2009).

404 The impacts of environmental regulations on economic growth, productivity, firm profitability, plant closures, and workers has been of keen interest to policymakers since the inception of the U.S. EPA (Ferris, et al, 2017). The rise in concern over EJ is often traced to demonstrations in Warren County, North Carolina in 1982 over the siting of a polychlorinated biphenyl landfill in a poor and minority community. Public attention on these issues gradually led to an increased focus in the economics literature on distributional issues in the context of race, poverty, and income.

405 An individual who is susceptible is one who is more responsive to exposure (Schwartz et al., 2011) or one who has an increased likelihood of sustaining an adverse effect (U.S. EPA, 2003).
focus on how a regulatory action affects the distribution of exposure or proximity to harm, and not risk per se such as an increase in the frequency or severity of health effects.  

Evidence exists of potential disproportionate impacts from environmental stressors on various population groups using a wide variety of proxies for exposure, many of which are proximity-based (e.g., distance to a polluting facility as a surrogate for exposure). These studies often find evidence that locally-unwanted land-uses such as landfills or facilities that treat, store, or dispose of hazardous waste are more likely to be concentrated in predominantly minority or low-income neighborhoods (for example, Bullard 1983; GAO 1983; UCC 1987; Boer et al. 1997; and Mohai et al. 2009).

Other studies attempt to better approximate exposure by examining whether existing emission patterns are related to socio-economic characteristics. These studies often focus on a specific pollutant and geographic area. They also often differ in how they define the relevant neighborhood and comparison group. As such, findings with regard to whether race and income are associated with potential exposure vary across studies. For example, after controlling for other factors, Aurora and Cason (1998) find that both the percentage of minority and poor households in a community are positively related to reported Toxic Release Inventory (TRI) emissions, although the significance of these relationships varies by region. Gray and Shadbegian (2004) find that communities with a higher percentage of poor households are exposed to more air and water pollution from pulp and paper mills on average, while communities with a greater percentage of minority households are exposed to less pollution on average. Some studies attempt to examine how changes in exposure or risk associate with race and income. For instance, Hamilton (1993, 1995) finds that expansion decisions for waste sites are unrelated to race and finds mixed evidence for income, while De Silva, et al. (2016) find that new plants that report to the TRI are more likely to locate in census tracts with a higher share of nonwhite and less educated individuals (and new plants that are non-TRI reporters are less likely to locate in these neighborhoods).

Finally, other studies attempt to more explicitly account for exposure and/or health risk. For example, Rosenbaum et al. (2011) combine information on ambient concentrations of diesel particulate matter in marine harbor areas throughout the United States with exposure and carcinogenic risk factors broken out by race, ethnicity, and income. They find that the most important factor in predicting higher exposure is population density and that low-income and minority individuals are over-represented in marine harbor areas that exceed risk thresholds. Likewise, Morello-Frosch and Jesdale (2006) combine estimates of outdoor air toxic concentrations with lifetime cancer risks by socioeconomic status and race. They find that even though lifetime cancer risks are high for all individuals (exceeding the 1990 Clean Air Act Amendment goal by several orders of magnitude), lifetime cancer risk is correlated with the degree of racial residential segregation, with the highest risks accruing to non-Hispanic black, Asian and Pacific Islanders, and Hispanics in highly segregated neighborhoods. Poverty does not appear to explain differences in lifetime cancer risk by race and ethnicity.

Ringquist (2005) conducts a meta-analysis of both facility location and emissions across 49 studies published prior to 2002 and finds evidence that plant location and higher emissions are positively associated with non-white populations. He finds little evidence, however, that this is the case in communities with lower average household incomes or higher poverty rates. The finding for race holds across a wide variety of environmental risks (e.g., hazardous waste sites and air

---

406 Differences in exposures or health effects alone may not be representative of differences in total benefits and costs. As discussed in Serret and Johnstone (2006) and Fullerton (2011), for example, the full distributional effects of environmental policy could include differences in product prices, wage rates, employment effects, economic rents, etc. These impact categories and approaches to examine them are discussed in Chapter 9.

407 Others note the strength of this contemporaneous relationship but find that the direction and magnitude of the relationship between location and race or income at time of siting is less clear (e.g., Been 1994; Been and Gupta 1997; Wolverton 2009). See Shadbegian and Wolverton (2010) for a summary of the literature on firm location and environmental justice, including a discussion of whether plant location precedes changes in socioeconomic composition that result in higher percentages of non-white and poor households nearby or vice versa. Most of these studies examine partial correlations between pollution and household characteristics, using statistical techniques that control for other factors.

408 A common empirical challenge in this literature is the possibility of sorting (i.e., some poorer households may choose to move to neighborhoods with higher level of pollution). Many studies use demographic data that precedes the siting or emissions decision to control for the possibility of reverse causation. Gray and Shadbegian (2004) use a spatially lagged instrumental variable approach because they do not have demographic information that precedes plant siting.
pollution concentrations), levels of aggregation (e.g., zip codes, census tracts, and concentric circles around a facility), and controls (e.g., land value, population density, and percent employed in manufacturing). The finding for race appears sensitive, however, to the comparison group utilized (e.g., all communities versus a subset of communities).

A potential unintended consequence of improving environmental quality in some communities more than others is that rents may increase in the improved neighborhoods, making them potentially unaffordable for poorer households. For example, Grainger (2012) shows that about half of the increases in home prices due to the Clean Air Act Amendments are passed through to renters. Thus, the net health effect of improvements in environmental quality for renters depends on whether they move. Those who do not move experience higher rents, but also improved air quality (and potentially other neighborhood attributes). For those who do move the net effect depends on the quality of the neighborhood to which they relocate. If these households receive far less of the health benefit predicted from a static model and also face transaction costs from moving in addition to higher rent, they could be worse off. The literature refers to this phenomenon as “environmental gentrification” (see also Banzhaf and McCormick 2012).

Sieg et al. (2004) find that even with no moving costs, local households could be worse off because other households move into the clean neighborhood and bid up the rents.409 Earlier work by Banzhaf and Walsh (2008) shows that neighborhood income increased following cleanup, but more recent analysis (Banzhaf, et al. 2012) shows racial characteristics in the neighborhood may not change. The authors postulate that richer minorities may move back into neighborhoods following cleanup. Bento and Freedman (2014) find that lower-income home owners experienced relatively larger reductions in particulate matter emissions post 1990 Clean Air Act Amendments with little evidence of re-sorting after air quality improvements occurred. Using data on repeat residential location choices, Depro, et al. (2015) find that Hispanic households are less willing to give up other types of consumption to reduce cancer risk compared to white households, but that this result is likely driven by large income disparities between the two populations (i.e., the opportunity cost of avoiding cancer risk is higher for Hispanics because they are more income constrained).

10.2.2 Analyzing EJ Impacts in the Context of Regulatory Analysis

In the context of regulatory analysis, examining distributional effects of changes in human health and environmental outcomes or costs can be accomplished, when data are available, by comparing effects in the baseline to those under each regulatory option for minority, low-income, or indigenous populations.410 When evaluating human health and environmental outcomes, the following fundamental questions can guide consideration of potential analytical methods for assessing EJ.

- As a basis of comparison, what is the baseline distribution of health and environmental outcomes across population groups of concern for pollutants affected by the rulemaking?411
- What is the distribution of health and environmental outcomes for the options under consideration for the rulemaking effort?
- Under the options being considered, how do the health and environmental outcomes change for population groups of concern?412

---

409 The market dynamics associated with the relationship between household location decisions and pollution was first examined in a rigorous context in Been and Gupta (1997), and further explored by Banzhaf and Walsh (2008).

410 OMB (2003) defines the baseline as “the best assessment of the way the world would look absent the proposed action.” Section 10.2.6 describes the concept of baseline briefly. For a more detailed discussion on properly defining a baseline to measure the incremental effects of regulation, see Chapter 5 of these Guidelines.

411 The term “outcome” is used to indicate that these questions should be interpreted more broadly than just applying to health effects. EPA Program Offices have the flexibility to adapt the wording of these questions to reflect the realities of the particular endpoints under consideration for a rulemaking.

412 It would be useful to quantify the degree to which disparities change from baseline, so that one could rank in order of preference the relative merits of
Note that these questions ask the analyst to provide information on changes to the distribution of outcomes, but do not ask the analyst to determine whether differences across population groups constitute disproportionate impacts. The term disproportionate is not defined in EO 12898; nor does the academic literature provide clear guidance on what constitutes a disproportionate impact. The determination of whether an impact is disproportionate is ultimately a policy judgment, though one that may be informed by analysis.

This chapter presents a suite of methods for analyzing distributional effects across a variety of regulatory contexts. Because the data, time, and resource constraints will differ across programs and rules, these guidelines are intended to provide flexibility to the analyst while introducing greater rigor and transparency in how EJ is considered in a regulatory context.

10.2.2.1 Evaluating Changes in the Distribution of Health and Environmental Outcomes

The analysis of EJ should ideally consider how a regulation affects the distribution of relevant health and environmental outcomes (e.g., mortality risk from a regulated pollutant) across population groups of concern. If outcome data are unavailable, a second-best option is to consider how a regulation affects the distribution of ambient environmental quality indicators (e.g., pollutant concentrations). Such indicators are less informative than the outcomes themselves if, for example, population groups of concern vary in vulnerability to the pollutant. If projecting changes in ambient environmental quality is not feasible, then a third-best option is for analysts to examine to what extent regulated entities are within a certain proximity of populations of concern (which could be further refined based on characteristics that may correlate with emissions such as plant age or size). Evaluating proximity to emission sources is less desirable than evaluating effects on the distribution of ambient environmental quality or human health outcomes due to uncertainty and local variability in how emission changes affect the health of populations of concern.

As with other types of distributional analyses, it is important to characterize baseline conditions prior to evaluating how they are affected by each regulatory option. The baseline allows one to determine how the pollutant and its human health and environmental effects are distributed across population groups prior to any regulatory action. It is also the basis of comparison for understanding how a regulation affects the distribution of these health and environmental effects. Baseline assumptions used in a distributional analysis should be consistent with those used in the benefit-cost analysis.

Because an unequal distribution of environmental improvements across population groups may actually help alleviate existing disparities, analysts should consider how a regulatory option changes the overall distribution of human health and environmental outcomes not just how changes in human health and environmental outcomes are distributed across these population groups (Maguire and Sheriff 2011). For example, suppose a policy is expected to reduce a pollutant, causing a greater reduction in adverse health outcomes for non-minorities than for minorities. One might conclude that this change in the distribution of outcomes could pose an EJ concern. If, however, the non-minority population suffered

---

413 The CEQ NEPA Guidance for EJ (CEQ 1997) provides some guidance on the use of the term. A population group may be disproportionately affected if health effects are significant or “above generally accepted norms,” the risk or rate of exposure is significant or “appreciably exceeds or is likely to appreciably exceed the risk or rate to the general population or other appropriate comparison group,” or is subject to “cumulative or multiple adverse exposures from environmental hazards.”

414 A large epidemiological literature explores differences in health effects across various demographic groups. See, for example, Schwartz et al. (2011b).

415 OMB (2003) defines the baseline as “the best assessment of the way the world would look absent the proposed action.” See Chapter 5 for a more detailed discussion of baseline issues.
greater ill effects from the pollutant in the baseline relative to the minority population, then such a change in the distribution of outcomes may reduce, rather than increase, a pre-existing disparity in outcomes.

Benefit-cost analysis estimates society’s willingness to pay for a change in environmental quality.\textsuperscript{416} As an alternative to the change in willingness to pay, one could examine the distribution of physical indicators. Such an evaluation is relatively straightforward if there is only one outcome to consider. Analysis that evaluates multiple outcomes (e.g., asthma risk and fatal heart attack risk) raises the challenge of whether and how to aggregate these outcomes into a single measure, especially when their distributions are dissimilar across populations of concern. Combining several outcomes into a single aggregate measure may be desirable, but it entails normative value judgments regarding the weight to be given to each component. Absent clear guidance on how to make these judgements, analysts should present the distributional impacts for multiple outcomes separately.

10.2.2.2 Evaluating the Distribution of Costs

Activities to address EJ often focus on reducing disproportionate environmental and health outcomes in communities. However, certain directives (e.g., EO 13175 and OMB Circular A-4) specifically identify distribution of economic costs as an important consideration. The economic literature also typically considers both costs and benefits when evaluating distributional consequences of an environmental policy to understand its net effects on welfare. As discussed in Chapter 9, Tietenberg (2002) and Robinson, et al. (2016) describe how regulatory costs can affect product prices, labor compensation, and returns to capital, all of which may affect population groups of concern in different ways. Fullerton (2009, 2011) discusses distributional effects that may result from an environmental policy in terms of higher costs of production causing higher product prices, decreased production reducing revenues and affecting workers and investors, changes in scarcity rents, and transitional effects.

In the context of EJ, the distribution of health or environment effects alone might convey an incomplete – and potentially biased – picture of the overall burden faced by population groups of concern. For instance, if a regulation results in higher energy prices, low income households may be particularly hard hit because they spend a greater share of their income on energy compared to other households.

This chapter frames the discussion in terms of how a regulation affects environmental and health outcomes across population groups of concern (referred to as benefits, when monetized), but many of the methods can be applied to the evaluation of costs and other impacts as well. Whether to consider costs in an evaluation of EJ issues will depend on the relevance of the information for the regulatory decision at hand; the likelihood that economic costs of the regulatory action will be concentrated among particular types of households or, even if costs are not concentrated, whether their effects are expected to be more pronounced among low income households; as well as the availability of data and methods to conduct the analysis. (See 9.5.1 for more discussion of measuring the impacts of regulation on consumers.)\textsuperscript{417}

In many cases, analysis of how economic costs from the regulation are distributed are not expected to substantially alter the assessment of distributional impacts for population groups of concern. For example, this could be the case if regulatory costs result in a relatively small change in the prices of goods consumed by lower-income households or these households have a high elasticity of demand for these goods (i.e., can substitute away from them easily). When

\textsuperscript{416} The empirical techniques used to monetize health and environmental benefits for BCA estimate an individual’s marginal willingness to pay for a change in the outcome. Economic theory suggests that even if all individuals have identical preferences, the marginal willingness to pay to avoid a bad outcome should increase with the level of the outcome. However, marginal willingness-to-pay measures typically used in benefit-cost analysis are constant values. Thus, they cannot be used to evaluate the distribution of the change in welfare across groups may be of interest.

\textsuperscript{417} Note that there may be other impacts of a regulatory action (e.g., employment effects) beyond direct compliance costs, but understanding how all impacts vary across population groups of concern may not be feasible. For example, data on the distribution of changes in employment across low-income and minority populations may be difficult to assess. See Chapter 9.
costs are expected to differentially burden populations of concern, further exploration of the distribution of economic costs can offer substantial insight. Such cases may include situations when costs to comply with the regulatory action represent a noticeably higher proportion of income for population groups of concern; some population groups are less able to adapt to or substitute away from goods or services with now higher prices; costs to consumers are concentrated among particular types of households (e.g., renters); there are identifiable plant closures or facility relocations that could adversely affect certain communities; or when households may change their behavior in response to the imposition of costs in such a way that populations of concern are less protected than other groups. Also relevant is consideration of whether other government programs available to low-income households may mitigate some of these distributional effects.

While it is important to find ways to incorporate economic costs into an analysis of potential EJ concerns, detailed analyses may be challenging due to data or modeling constraints. A static analysis may be possible in some circumstances, but it is even more challenging is to anticipate and model the dynamic effects of a regulatory action on migratory patterns and other types of behavioral change. For example, while hedonic approaches (discussed in Chapter 7) may be useful for demonstrating how changes in environmental quality factor into housing prices, predicting the effect of such price changes on household migration by race or income may be infeasible. Likewise, spatial sorting models have been used in the literature to examine responses to regulation, but they typically limit their focus to a particular city or region.

In addition, incomplete data may prevent fully characterizing the distribution of costs across population groups of concern. Available data may only shed light on baseline distributions but be insufficient for purposes of modeling behavioral changes by household type in response to the costs of a regulatory action. Impacts that cannot be quantified can be qualitatively characterized, including a discussion of key methodological and data limitations and assumptions.

In addition, we may expect the distributional impacts across population groups of concern to differ in the short run relative to the long run even when all or almost all consumers face similar price changes due to a regulatory action. For instance, in the short run budget-constrained households may face more difficulties accommodating higher prices than in the long run. In contrast, if there is a robust used market for the regulated good, higher prices due to a regulatory action may initially affect higher income households who purchase new goods; over a longer period of time, however, these higher prices may also affect lower-income households due to higher prices for used goods.

When analyzing the distribution of costs, another consideration is the use of partial versus general equilibrium models for analysis. While general equilibrium models could be utilized to examine first and second-order costs and their implications for changes in wages and prices across households over time, such analyses are typically very resource- and time-intensive and are usually only utilized when a large number of sectors are expected to experience significant economic impacts. Such models also are generally focused on medium- to long-run impacts.

---

418 The regulatory analysis for the EPA’s Lead Renovation, Remodeling, and Painting Final Rule (U.S. EPA 2008a) provides an example of consideration of costs in the context of a rulemaking.

419 See Section 8.2.5.1 of the Handbook on the Benefits, Costs and Impacts of Land Cleanup and Reuse (U.S. EPA 2011b) for a more detailed discussion of EJ in the context of the potential effects of environmental policy on land values and household location decisions.

420 See Kuminoff, et al. (2013) for a discussion of equilibrium sorting models used to evaluate household housing choices.

421 Data for exploring differential consumption patterns in the baseline may be available from the Consumer Expenditure Survey, which provides information on the purchase of goods and services across different types of households. The baseline distribution of electricity and other energy prices by household type is also available from the Energy Information Administration. In addition, industry-specific data sources on baseline household consumption patterns may be available for certain types of products or services related to the regulatory scenarios under consideration.
10.2.3 Populations of Concern for EJ Analysis

EO 12898 identifies a number of relevant population groups of concern: minority populations, low-income populations, Native American populations and tribes, and “populations who principally rely on fish and/or wildlife for subsistence.” It may be useful to analyze these categories in combination — for example, low-income minority populations — or to include additional population groups of concern, but such analysis is not a substitute for examining populations explicitly mentioned in the Executive Order. In this section, we discuss existing Federal definitions for population groups of concern in the context of EJ. We also discuss credible options for defining these populations in the absence of a Federal definition.

10.2.3.1 Minority and Native American Populations

OMB (1997) specifies minimum standards for “maintaining, collecting, and presenting data on race and ethnicity for all Federal reporting purposes.... The standards have been developed to provide a common language for uniformity and comparability in the collection and use of data on race and ethnicity by Federal agencies.” In particular, it defines the following minimum race and ethnic categories:

- American Indian or Alaska Native
- Asian
- Black or African American
- Native Hawaiian or Other Pacific Islander
- White
- Hispanic or Latino

Statistical data collected by the Federal government, such as the U.S. Census Bureau, use this classification system. Beginning with the 2000 Census, individuals were given the option of selecting more than one race, resulting in 63 different categories. OMB (2000) provides guidance on how to aggregate these data in a way that retains the original minimum race categories (i.e., the first five categories listed above; note that the sixth category is an ethnicity and would therefore be tracked separately) and four double race categories that are most frequently reported by respondents. In addition, the U.S. Census Bureau collects data useful for identifying minority populations not completely captured by either the race or ethnicity categories, such as households that speak a language other than English at home or foreign-born populations.

CEQ’s NEPA Guidance for EJ (CEQ 1997) provides useful direction for defining minority and minority population based on these Federal classifications. Minority is defined as “individual(s) who are members of the following population groups: American Indian or Alaskan Native; Asian or Pacific Islander; Black, not of Hispanic origin; or Hispanic.” A minority population is identified if “either (a) the minority population of the affected area exceeds 50 percent or (b) the minority population percentage of the affected area is meaningfully greater than the minority population percentage in the general population or other appropriate unit of geographic analysis.” The term meaningfully greater is not defined, although the guidance notes that a minority population exists “if there is more than one minority group present and the

---

422 EO 12898 clarifies in Section 6 that the EO applies to Native Americans and also Indian Tribes, as specified in 6-606, as well as populations who principally rely on fish and/or wildlife for subsistence as specified in 4-401.

423 Analysts should refer to OMB (1997) for the specific definitions.

424 The U.S. Census Bureau releases two data products every five years with details on race and ethnicity the Selected Population Tables, and the American Indian and Alaska Native Tables. See: https://www.census.gov/programs-surveys/acs/data/race-aian.html (accessed on April 1, 2020).

425 See OMB (2000) for specific guidance on how to conduct this aggregation.

426 For example, see information on these data from the American Community Survey at: https://www.census.gov/topics/population/foreign-born/data/tables/acs-tables.html, and https://www.census.gov/topics/population/language-use.html (accessed on April 1, 2020).
minority percentage, as calculated by aggregating all minority persons, meets one of the above-stated thresholds.” Finally, the CEQ’s NEPA Guidance for EJ states that analysts “may consider as a community either a group of individuals living in geographic proximity to one another, or a geographically dispersed/transient set of individuals (such as migrant workers or Native Americans), where either type of group experiences common conditions of environmental exposure or effect.”

10.2.3.2 Low-Income Populations

CEQ’s NEPA Guidance for EJ (CEQ 1997) states that "low-income populations in an affected area should be identified with the annual statistical poverty thresholds from the Bureau of the Census’ Current Population Reports, Series P-60 on Income and Poverty. In identifying low-income populations, agencies may consider as a community either a group of individuals living in geographic proximity to one another, or a set of individuals (such as migrant workers or Native Americans), where either type of group experiences common conditions of environmental exposure or effect." The extent to which this same definition should be applied in a regulatory context will be context specific.

OMB has designated the U.S. Census Bureau’s annual poverty measure, produced since 1964, as the official metric for program planning and analytic work by all Executive branch agencies in Statistical Policy Directive No. 14 (OMB 1978), although it does not preclude the use of other measures. Many Federal programs use variants of this poverty measure for analytic or policy purposes, and the U.S. Census Bureau publishes data tables with several options, described below.

The U.S. Census Bureau measures poverty by using a set of money income thresholds that vary by family size and composition to determine which households live in poverty. If a family’s total income is less than the threshold, then that family and every individual in it is considered in poverty. The official poverty thresholds do not vary geographically, but they are updated for inflation using the national Consumer Price Index for All Urban Consumers (CPI-U). The official poverty definition uses money income before taxes and does not include capital gains, tax credits, or noncash benefits (such as public housing, Medicaid, and food stamps). The way poverty is defined has remained essentially unchanged — apart from relatively minor alterations in 1969 and 1981 — since its inception.

There is considerable debate regarding this poverty measure’s ability to capture differences in economic well-being. In particular the National Research Council (NRC) recommended that the official measure be revised because “it no longer provides an accurate picture of the differences in the extent of economic poverty among population groups or geographic areas of the country, nor an accurate picture of trends over time” (Citro and Michael 1995). OMB convened an interagency group in 2009 to define a supplemental poverty measure based on NRC recommendations. The U.S. Census Bureau released the Supplemental Poverty Measure (SPM) in November 2011 (Short 2011). This measure uses different measurement units to account for “co-resident unrelated children (such as foster children) and any co-habitors and their children,” a different poverty threshold, and modified resource measures (to account for in-kind benefits and medical expenses, for example). It also adjusts for differences in housing prices by metropolitan statistical area, as well as family size and composition.

The NRC recognized that annual income is not necessarily the most reliable measure of relative poverty as it does not account for differences in accumulated assets across households. Neither the SPM nor the official U.S. poverty thresholds take into account differences in wealth across families. However, the SPM examines whether a household is likely to fall below a specific poverty threshold as a function of inflows of income and outflows of expenses. The U.S.


428 The U.S. Census Bureau produces single-year estimates of median household income and poverty by state and county, and poverty by school district as part of its Small Area Income and Poverty Estimates. It also provides estimates of health insurance coverage by state and county as part of its Small Area Health Insurance Estimates. These data are broken down by race at the state level and by income categories at the county level.
Census Bureau asserts that this measure is therefore more likely to capture short-term poverty since many assets are not as easily convertible to cash in the short run (Short 2012).

The U.S. Census Bureau also includes several additional measures that may prove useful in characterizing low-income families. Unlike poverty, there is no official or standard definition of what constitutes “low-income,” though it is expected to vary similarly by region due to differences in cost-of-living as well as with family composition. It is therefore appropriate to examine several different low-income categories, including families that make some fixed amount above the poverty threshold (e.g., two times the poverty threshold) but still below the median household income for the United States or for a region.

Educational attainment or health insurance coverage may also be useful for characterizing low-income families relative to other populations, although we caution analysts that some measures may be hard to interpret and use in a regulatory context. It is also possible to examine the percent of people who are chronically poor versus those that experience poverty on a more episodic basis using the U.S. Census Bureau's Survey of Income and Program Participation, which provides information on labor force participation, income, and health insurance for a representative panel of households on a monthly basis over several years (see Iceland 2003). Finally, cross-tabulations often are available between many of these poverty measures and other socioeconomic characteristics of interest such as race, ethnicity, age, sex, education, and work experience.

10.2.3.3 Populations that Principally Subsist on Fish and Wildlife

EO 12898 directs agencies to analyze populations that principally subsist on fish and wildlife. CEQ's NEPA Guidance for EJ (CEQ 1997) defines subsistence on fish and wildlife as “dependence by a minority population, low-income population, Indian tribe or subgroup of such populations on indigenous fish, vegetation and/or wildlife, as the principal portion of their diet.” It also states that differential patterns of subsistence consumption are defined as “differences in rates and/or patterns of subsistence consumption by minority populations, low-income populations, and Indian tribes as compared to rates and patterns of consumption of the general population.”

Neither the U.S. Census Bureau nor other Federal statistical agencies collect nationally representative information on household consumption of fish and/or wildlife. However, the EPA has conducted consumption surveys in specific geographic areas. If fish and wildlife consumption is a substantial concern for a specific rulemaking, EPA’s guidance can provide useful information for collecting these data (see U.S. EPA 1998b). There may also be surveys conducted by state or local governments. It is important to verify that any survey used in an analysis of distributional impacts in the context of EJ adheres to the parameters and methodology set out in U.S. EPA (1998b).

10.2.4 Data Sources

Many data sources can be used for conducting analyses of EJ issues. In general, the type of analysis that can be conducted depends upon the type of data available and its quality. In some cases, spatially disaggregated individual-level data may be most appropriate and relevant for conducting an analysis of potential EJ concerns. In other cases, distance as a proxy for risk may be the best available relevant metric for conducting the analysis. At times qualitative information will be the best available information for the analysis. In all cases, analysts should use the highest quality and most relevant data and information.

Recognizing the importance of data quality, information needed to conduct an EJ analysis may include:

- Demographic characteristics (e.g., race, ethnicity, age, education, gender);
- Income data (e.g., median household income or percent below poverty level);
- Health data (e.g., hospital and emergency admissions, race/ethnicity-stratified mortality rates,
• race/ethnicity-stratified asthma, or other morbidity rates); 429
• Other triggers or co-stressors that may be confounders (e.g., low birth weight or asthma; exposure to indoor air pollution); 430
• Risk coefficients stratified by socio-economic variables (e.g., race/ethnicity, income);
• Location of pollution sources (e.g., latitude/longitude coordinates, zip code, county locator);
• Proximity to the nearest source(s) (e.g., distance in miles);
• Distribution of baseline emissions, exposure, and risk,
• Modeled changes in the distribution of emissions, exposure, and risk under different regulatory options, and
• Distribution of economic costs, when relevant (see Section 10.2.2.2).

Socioeconomic data is easily available in more aggregate form from the U.S. Census Bureau’s “Quick Facts” website, which contains frequently requested Census data for all states, counties, and urban areas with more than 25,000 people. 431 They include population, percent of population by race and ethnicity, and income (median household income, per-capita income, and percent below poverty line).

In 2010 the U.S. Census Bureau began to administer the decennial Census using a short form to collect basic socioeconomic information. More detailed socioeconomic information is now collected annually by the American Community Survey (ACS), which is sent to a smaller percentage of households than the decennial Census. 432 The ACS provides annual estimates of socioeconomic information for geographic areas with more than 65,000 people, three-year estimates for areas with 20,000 or more people, and five-year estimates for all areas. 433 The five-year estimates, which are based on the largest sample, are the most reliable and are available at the census tract and block group levels. Some of the Quick Facts data include estimates from the ACS.

The U.S. Census Bureau’s American Housing Survey (AHS), is a housing unit survey that provides data on a wide range of housing and demographic characteristics, including information on renters. 434 Unlike the ACS, which selects a random sample every year, the AHS returns to the same 50,000 to 60,000 housing units every two years.

10.2.5 Scope and Geographic Considerations

While most EPA rules are national in scope, there may be reasons to consider a rule’s distributional effects at a sub-national level. For instance, there may be differences in implementation at the state level (e.g., as with many waste rules under RCRA). A rule may also affect a limited part of the country (e.g., a single-sector regulation where affected facilities are geographically concentrated). In such cases the analyst may wish to evaluate the effects of the regulation at a regional level. For some regulations, such as those governing the use of a household chemical or as a product ingredient, geography may not be as relevant for determining how health and environmental outcomes vary across population

430 See the EJTG (U.S. EPA, 2016) for a discussion of possible co-triggers and stressors.
431 Quick Facts is available at: https://www.census.gov/quickfacts/table/US/PST045218 (accessed April 1, 2020). Note the year associated with a specific Quick Facts data element, as data are updated as new information becomes available. Not all data elements represent the same year.
432 The ACS is available at: https://www.census.gov/programs-surveys/acs (accessed April 1, 2020).
433 Because ACS variables change over time, caution should be used when comparing ACS estimates across samples and years. Guidance for comparing ACS data can be found at: https://www.census.gov/programs-surveys/acs/guidance/comparing-acs-data.html (accessed on April 1, 2020).
434 Information on owner-occupied homes versus renters may be useful when exploring issues of gentrification, where renters could be worse off due to rising housing costs.
groups of concern. Two main issues to consider when comparing impacts of a rulemaking on minority, low-income, or indigenous populations across geographic areas are:

- Unit of analysis (e.g., facilities or aggregate emissions to which a population group is exposed within a designated geographic area); and
- Geographic area of analysis used to characterize impacts (e.g., county or census tract).\(^{435}\)

The **unit of analysis** refers to how the environmental harm is characterized. For instance, in a proximity-based analysis the unit of analysis could be an individual facility or the total number of facilities within a specific geographic area (e.g., a county or census tract). In an exposure-based analysis the unit of analysis could be the emissions or ambient concentrations to which the population is exposed aggregated within a specific geographic area. The unit of analysis is often identical to the geographic scale used to aggregate and compare effects on minority, low-income, or indigenous populations in one area to another (see Section 10.2.7 regarding how to select an appropriate comparison group).\(^{436}\) The choice will vary depending on the nature of the pollutant (e.g., how far it disperses; whether it is possible to identify the specific source of emissions). In considering various units, an important consideration is whether the data are sufficiently disaggregated to pick up potential variation in impacts across socioeconomic characteristics. More aggregated units of analysis (e.g., metropolitan statistical area (MSA) or county) may mask variation in impacts across socioeconomic groups compared to more disaggregated levels (e.g., facility or census tract) for some types of pollutants.

The **geographic area of analysis** is the area used to characterize impacts (e.g., distance around a facility). Outcomes are aggregated by population groups within geographic areas to compare across groups. As with unit of analysis, the appropriate geographic area will vary depending on the pollutant and regulatory context. Some air pollutants, for example NO\(_x\), may travel hundreds of miles away from the source, making it appropriate to choose a large area for measuring impacts. In contrast, water pollutants or waste facilities may affect smaller areas, making it appropriate to consider a smaller area for analysis. Likewise, an assessment of outcomes from specific industrial point sources may require more spatially resolved air quality, demographic and health data than one that affects regional air quality, where coarser air quality, demographic and health data may suffice. Using more than one geographic area of analysis to compare effects across population groups may also be useful since outcomes are unlikely to be neatly contained within geographic boundaries. The literature has demonstrated that results are sensitive to the choice of the geographic area of analysis (Mohai and Bryant 1992; Baden et al. 2007).

Commonly used geographic areas of analysis include:

**Counties:** The United States has more than 3,000 counties according to the 2012 Census of Governments. Although counties are well-defined units of local government and provide complete coverage of the United States, they vary in size from a few to thousands of square miles and population density ranges from less than one person per square mile in some Alaskan counties to over 66,000 in New York County. In addition, spatial considerations associated with using counties present concerns for an analysis of distributional impacts in the context of EJ. A facility located in one corner of a county may have greater effects on neighboring counties than on residents of the county where the plant is located.\(^{437, 438}\)

**Metropolitan and Micropolitan Statistical Areas:** The U.S. Census Bureau publishes data on metropolitan and micropolitan statistical areas, as defined by OMB (OMB 2009). Metropolitan statistical areas include an urban core

---

\(^{435}\) This is often referred to in the literature as geographic scale.

\(^{436}\) In Fowlie et al. (2012), for example, the scale of the analysis varies between 0.5, 1 and 2 miles of the facility (which is the unit of analysis).

\(^{437}\) These same advantages and disadvantages can apply to other units of government.

\(^{438}\) For criteria pollutants, baseline health data may be available at the county level (e.g., baseline death rates, hospital admissions, and emergency department visits).
and adjacent counties that are highly integrated with the urban core. A micropolitan statistical area corresponds to the concept of a metropolitan statistical area but on a smaller scale. Metropolitan statistical areas have an urban core of at least 50,000 persons; micropolitan statistical areas have an urban core population between 10,000 and 50,000 persons. According to the U.S. Census Bureau, almost 94 percent of the U.S. population lived in a metro- or micropolitan statistical area in 2010. Rural areas of the United States are not covered by these statistical designations.

**Zip codes:** Zip codes are defined by the U.S. Post Office for purposes of mail delivery and may change over time. They also may cross state, county, and other more disaggregated Census statistical area definitions, making them difficult to use for analysis. Zip code tabulation areas are statistical designations first developed by the U.S. Census Bureau in 2000 to approximately characterize the zip code using available census block level data on population and housing characteristics. Data are readily available for the approximately 33,000 U.S. zip code tabulation areas. While smaller than counties, they also vary greatly in size and population. As a result, they may often be less preferable than other geographic areas for analyzing distributional effects across population groups of concern.

**Census tracts/block groups/blocks:** Census tracts are small statistical subdivisions of a county, typically containing from 1,500 to 8,000 persons. The area encompassed within a census tract may vary widely, depending on population density. Census tracts in denser areas cover smaller geographic areas. Census tract boundaries were intended to remain relatively fixed. However, they are divided or aggregated to reflect changes in population growth within an area over time. Although they were initially designed to be homogeneous with respect to population characteristics, economic status, and living conditions, they may have become less so over time as demographics have changed.

Analysts may also choose to use census blocks or block groups. A census block is a subdivision of a census tract and the smallest geographic unit for which the U.S. Census Bureau tabulates data, containing from 0 to 600 persons. Many blocks correspond to individual city blocks bounded by streets, but may include many square miles, especially in rural areas. And census blocks may have boundaries that are not streets, such as railroads, mountains or water bodies. The U.S. Census Bureau established blocks covering the entire nation for the first time in 1990. Census block groups are a combination of blocks that are within — and a subdivision of — a given census tract. Block groups typically contain 600 to 3,000 persons.439

**GIS-based approaches to defining geographic areas:** Because Census-based definitions often reflect topographical features such as rivers, highways, and railroads, they may exclude affected populations that, although separated by some physical feature, receive a large portion of the adverse impacts being evaluated. Since Census-based definitions vary in geographic size due to differences in population density, Geographic Information System (GIS) software and methods may enable the use of spatial buffers around an emissions source that are more uniform in size and easier to customize to reflect the appropriate scale and characteristics of emissions being analyzed for a given rulemaking. Dasymetric mapping techniques that use land cover information to more accurately distribute populations within selected Census-based boundaries while accounting for physical features (e.g., Mennis and Hultgren 2006). Analysts should be aware that there are sometimes challenges when working with geospatial data. Statistical techniques may rely on assumptions that often are violated by these types of data (Chakraborty and Maantay 2011). Analysts should follow best practices in the literature when using these types of data in order to address or minimize these challenges when possible.440

---

439 Other Census statistical area definitions (e.g., public use microdata areas or PUMAs) are also available.

440 For instance, spatial autocorrelation — when locations in closer proximity are more highly correlated than those further away from each other — violates the assumption that error terms are independently distributed (an assumption that underlies ordinary least squares). There are a variety of ways to test for spatial autocorrelation in the data, such as Moran’s I or a Mantel statistical test, as well as methods for addressing it in regression analysis.
10.2.6 Comparison groups

To evaluate impacts on population groups of concern, information needs to be presented in relation to another group, typically referred to as a comparison group. The choice of a relevant comparison group is important for evaluating changes in health, risk, or exposure effects across population groups relative to a baseline. Within-group comparisons involve comparing effects on the same demographic group across different areas in the state, region or nation, while across-group comparisons examine effects for different socioeconomic groups within an affected area. From the perspective of EO 12898, across-group comparisons may be most relevant.

It may make sense in some contexts to define the comparison group at a sub-national level to reflect differences in socioeconomic composition across geographic regions. For instance, because of their larger populations, effects in urban areas often dominate the results of an analysis. If a regulatory action primarily affects rural areas, inclusion of urban areas in the comparison group may not be valid. For such reasons, it is important to articulate clearly how the comparison group is defined in the EJ analysis. The literature suggests using more than one comparison group to discern the sensitivity of the results to these aspects of the analysis.\(^{441}\)

10.2.7 Measuring and estimating impacts

This section presents potentially useful approaches for describing EJ impacts in regulatory analysis. Basic summary statistics of a regulation’s impacts on relevant human health or environmental endpoints by race and income are recommended. When data permit, such statistics are straightforward to calculate, and they promote consistency across EPA analytical efforts.

It is important for analysts to conduct a screening analysis for determining when more in-depth analysis of the impacts of a regulatory action on population groups of concern is warranted. While there is no single prescribed screening method, the analyst should review the quality and availability of data, availability of defensible methods to analyze the data, and the peer-reviewed literature and stakeholder input that might be used to evaluate potential EJ concerns.\(^{442}\)

Such information may include the following:

- Proximity of regulated sources to minority populations, low-income populations, and/or indigenous peoples;
- Number of sources that may be impacting these populations in the baseline;
- Nature and amounts of different pollutants that may already be impacting these populations;
- Any unique exposure pathways associated with the pollutant(s) being regulated;
- Stakeholder concern(s) about the potential regulatory action; and
- History of EJ concerns associated with the pollutant(s) being regulated.

This review may enable the analyst to initially assess potential EJ concerns associated with the regulatory action, and to identify whether more detailed information is available for a more in-depth analysis. For economically significant actions, it is recommended that the results of the screening analysis be demonstrated through summary statistics.

---

\(^{441}\) Bowen (2001) also argues that restricting the comparison group to alternative locations within the same metropolitan area may be more defensible than a national level comparison in some instances, given heterogeneity across geographic regions in industrial development and economic growth over time and inherent differences in socioeconomic composition (e.g., relatively more Hispanics reside in the Southwest). Ringquist (2005), however, notes that placing restrictions on comparison groups in this way may “reduce the power of statistical tests by reducing sample sizes” or bias results against a finding of disproportionate impacts because such restrictions reduce variation in socioeconomic variables of interest.

\(^{442}\) A screening analysis is also recommended by the EJTG (U.S. EPA 2016). EJ SCREEN may be a useful starting point for a screening analysis in some cases, though it is not sufficient alone. In addition, analysts should be aware of the extent to which the information included in EJSCREEN overlaps with the affected sources and time frame for the regulation. EJSCREEN is not appropriate as a way of identifying an area as an EJ community or as a basis for agency decision-making regarding the existence or absence of EJ concerns (U.S. EPA 2019). See https://www.epa.gov/ejscreen for more information (accessed on March 31, 2020).
Summary statistics can be supplemented with other approaches described below when a screening analysis indicates that a more careful evaluation is needed. The health effects of exposure to pollution may vary across populations (likewise, with economic costs). One way to capture these effects is to use information regarding variation in risk and incidence by groups, when available, to characterize the baseline and projected response to a change in exposure (for example, see Fann et al. 2011).

Available scientific literature and data (which also often requires some level of spatial resolution) may not allow for a full characterization of potential EJ concerns. In these cases, it is recommended that the analyst qualitatively discuss limitations and sources of uncertainty in the risk and exposure characterization used to assess health effects for minority populations or low-income populations, as highlighted in the literature (U.S. EPA 2016). When data are available to approximate risk or exposure, for instance locations of emitting facilities, some level of quantitative analysis may be possible.

It may be desirable to identify which regulatory option best improves distributional outcomes. However, similar to the observation in Section 10.2.2 that it is a policy judgement as to whether an impact is disproportionate, there is no commonly accepted way to rank distributional outcomes. Text Box 10.1 discusses social welfare functions and inequality indices for potentially ranking distributions of environmental impacts. While these methods are useful for combining efficiency and equity considerations into one measure, they are not sufficiently developed for application to EPA regulatory analysis. Nor is there a consensus as to which one should be used. For discussion of the advantages and disadvantages of methods commonly used to rank environmental outcomes, see Maguire and Sheriff (2011).

### 10.2.7.1 Summary Statistics

Summary measures can characterize potential differences in baseline and regulatory options within and across populations of concern relative to the appropriate comparison group. However, summary statistics alone do not necessarily provide a complete description of differences across groups as they are not usually calculated for all variables of interest. In addition, summary statistics can mask important details about the tails of the distribution for outcomes of particular relevance for population groups of concern (see Gochfeld and Burger 2011). Nonetheless, such information can prove a useful starting point for understanding potential differences between population groups. After reviewing the available data and feasible methods for developing information on potential differences, the analyst should present information in a transparent and accessible manner such that the decision maker can consider:

- Population groups of concern for the regulatory action,
- Geographic scale and unit of analysis, when relevant,
- Primary conclusions regarding baseline conditions and, when possible, the extent to which different population groups may be affected by the rule (e.g., statistical differences),
- Sources of uncertainty across alternative results (e.g., comparison groups and geographic scale), and
- Data quality and limitations of the results.

A variety of measures can be used to characterize regulation’s distributional effects for population groups of concern. We discuss a few examples below.

**Means, medians and quantiles**

Reporting geometric mean or median outcomes by group at the baseline and for each regulatory option can be a straightforward way to display information. Tests for statistically significant differences across means or medians are often useful (see Been and Gupta 1997 and Wolverton 2009). It is important to be alert, however, to potential changes in the tails of the distribution. For example, the baseline outcomes could be uniformly distributed across the population but become concentrated around the mean or median for the regulatory scenario. Presenting data using different quantiles can provide additional information illuminating these effects.
**Text Box 10.1 - Social Welfare Functions and Inequality Indices**

The costs, benefits, and distributional effects of a regulation can be evaluated by using a social welfare function (SWF). A SWF provides a way to aggregate welfare or utility across individuals into a single value, thus allowing simple, direct comparisons in ranking alternative distributions of health outcomes, risk, or environmental quality. Such comparisons are potentially useful in evaluating whether a change from the baseline to a regulatory option makes society better off, and to understand which segments of society are negatively or positively affected. They can also facilitate comparisons between possible regulatory options (see Adler 2008, 2012 for a discussion). Sen (1970), Arrow (1977), and Just et al. (2004) provide theoretical discussions of SWFs, and Norland and Ninassi (1998) provide an example of an application to energy markets. Adler (2012) addresses practical issues of incorporating both health and income effects in a SWF.

Any ranking of alternative outcomes uses an implicit set of normative criteria; a SWF makes the criteria explicit regarding how society prefers to distribute resources across individuals. Since there is no consensus regarding those preferences, a universally accepted SWF does not exist. For example, suppose an increase in exposure to a specific pollutant results in an average loss of 0.1 IQ points across a population of 1,000 children. It is not obvious how society should rank alternative distributions of this total loss of 100 IQ points. Is it worse to have 250 individuals suffer a loss of 0.1 each, 250 suffer a 0.3 loss, and 500 suffer no loss; or 500 individuals suffer a loss of 0.01 and 500 suffer a loss of 0.19? A SWF would provide a ranking of the alternatives, determined by its specifications.

An inequality index is a related concept used to assign a numerical value to distributions of a single “good” or “bad” (e.g., income or pollution), independent of the total amount produced. A distribution with a higher index value is less “equal” than one with a lower number. Commonly used indices are based on simple SWFs and are subject to the same limitations (Blackorby and Donaldson 1978, 1980). However, unlike a SWF, an index number value has cardinal significance, i.e., the magnitudes, not just the rankings, contain information about how much society would be willing to give up in exchange for the rest to be equally distributed.

We do not recommend using inequality indices in regulatory analysis of distributional impacts. Inequality indices were originally developed for ranking “goods,” like income. Therefore, it is inappropriate simply to use positive values of a “bad” such as pollution in the index, since doing so would imply that the underlying SWF is increasing in pollution, i.e., it would rank scenarios with higher overall pollution as more desirable. Since indices cannot accommodate negative values, some commonly used income inequality measures, such as the Gini coefficient, and Atkinson index, are inappropriate for evaluating distributions of adverse outcomes. In contrast, the Kolm index (Kolm 1976a, 1976b) does not suffer from this problem (see Maguire and Sheriff 2011). Unfortunately, the peer-reviewed literature does not yet contain environmental applications of the Kolm Index, and the Atkinson Index is undefined for “bads.”

**Ratios**

A simple ratio can be calculated to determine whether certain groups are relatively more exposed to an environmental hazard. For instance, an analyst can use a ratio to make an across-group comparison. One can compare the number of individuals within a specific demographic group (e.g., minority or low-income) to the number of individuals outside of the demographic group living within a specified distance of a polluting facility (e.g., three miles). A value of one in this case indicates that there is no distinguishable difference between the number of potentially exposed individuals within the demographic group relative to individuals outside of the demographic group. An analyst also can use a ratio to conduct a within-demographic group comparison. One can compare the number of individuals within a specific demographic group that live close to a regulated facility to the number of individuals from the same demographic group living further away. In this case, a value of one indicates that there is no identifiable spatial pattern where individuals from that demographic group tend to live closer to (or further away from) a regulated facility.

Because ratios may magnify absolute differences, ratios should be used in conjunction with other statistics. For example, it is possible for a ratio may show a 100-fold difference between two groups’ potential pollution exposure, but the absolute difference could be small.

**Tests for Differences**

Statistical tests can determine whether a significant disparity exists across demographic groups. One of the simplest is a t-test of the difference in means (i.e., the null hypothesis is that the means between two groups are equal). However, a
t-test assumes an underlying normal distribution. For non-normal distributions, nonparametric methods may be used. In cases where comparisons are based on the difference in probabilities between two groups, tests such as the Kendall test and the Fisher Exact test (for small samples) or the Mann-Whitney-Wilcoxon test (for larger samples) may be useful. These tests compare standard errors of two separate and independent statistics to determine how likely it is that the calculated distribution is the actual one. More sophisticated tests (e.g., the Kruskal Wallis test) are needed when making comparisons across more than two groups or a more formal examination of the full distribution is desired.

**Correlation coefficients**

Simple pair-wise correlations between impacts and relevant demographic groups may be useful information for characterizing distributional effects (e.g., Brajer and Hall 2005). The value of a Pearson correlation coefficient, for example, is a measure of how closely the distribution of the relationship between two variables (e.g., percent minority population and ambient pollution concentrations) can be represented by a straight line. It does not provide information regarding the slope of the line, apart from being positive or negative. Similarly, a Spearman rank correlation coefficient measures how closely the relationship can be captured by a generic monotonically increasing or decreasing function. Determination of what constitutes a “strong” or “weak” correlation is somewhat arbitrary, and caution should be used when comparing coefficients across socio-economic variables of interest.

**Counts**

A count of geographic areas (e.g., counties) where the incidence of an environmental outcome affected by a rule, disaggregated by race/ethnicity and income, exceeds the overall average is a useful measure. For comparison, this count should be accompanied by a count of geographic areas where the incidence does not exceed the overall average. These counts do not account for magnitude of differences but can help identify the need for more detailed analysis.

10.2.7.2 Visual Displays

Using GIS software and built-in graphical functions in spreadsheet or statistical software, analysts can produce visual displays of EJ-related information (e.g., maps, charts, graphs). Such displays can illustrate baseline levels of pollutants or locations of certain facilities, and the distribution, demographic profile and baseline health status of population groups of concern.

There are several challenges with using GIS-based visual displays as the main approach to evaluate potential EJ concerns. These include possible spatial and data deficiencies as well as geographic considerations that can lead to misleading or inaccurate results in some cases. It may be difficult to discern differences that arise between baseline and regulatory options, unless such differences are stark. While the use of visual displays in an EJ analysis may help communicate the geographic distribution of impacts, this information may be more effective if it is accompanied by other analytical information (for example, 10.2.5 discusses using GIS to create buffers for analysis).

10.2.7.3 Proximity-Based Analysis

Proximity- or distance-based analysis is used when direct measures of changes in risk or exposure are not available. This approach examines demographic and socioeconomic characteristics in proximity to a specific location, typically a waste site, permitted facility, or some other polluting source subject to the regulation (e.g., Baden and Coursey 2002, Cameron et al. 2012, and Wolverton 2009). While a simplistic approach examines the population within a Census-defined geographic boundary, it is also possible to use GIS methods to draw a concentric buffer around an emission source, such

---

443 See Chakraborty and Maantay (2011) for further discussion of the limitations of using GIS for EJ analyses.
as a one-mile radius around a site to approximate exposure, or to allow the data to define the appropriate distance through statistical techniques, as in the case for evaluating property value effects.\textsuperscript{444,445}

Several analytical considerations are important for conducting a proximity-based analysis.\textsuperscript{446} First, accurate information is needed on the location of regulated sources. Addresses or latitude/longitude coordinates must reflect physical locations of these facilities, and not the location of a headquarters building, for example. Second, a decision regarding the appropriate distance from the facility is needed. A solid waste facility with strict monitoring and safety controls is likely to have a limited geographic impact, whereas a permitted air pollution source may have the potential for a more widespread geographic impact. In general, Census-defined geographic boundaries (e.g., county, MSA) are unlikely to provide an accurate portrayal of the potential effects on the relevant affected population when emission sources are located along a boundary and thus mostly affect a neighboring jurisdiction, when the affected population is not uniformly distributed within the geographic boundary, or when pollutant exposures do not conform to the boundary.\textsuperscript{447}

In addition, Census-defined areas often vary widely in size, implying that they may differ in how well they proxy for actual exposure. Using buffer-based approaches (e.g., through GIS or fate and transport modeling) around an emissions source has the potential to more closely approximate actual risk and exposure, but the appropriate distance measure can vary by situation. The literature has demonstrated that results in proximity-based analyses can vary substantially with the choice of the geographic area of analysis (see Sheppard, et al. 1999; Rinquist 2005; Mohai and Saha 2007; Mohai and Saha 2015a; Mohai and Saha 2015b). For this reason, it is recommended that the analyst explore the potential value of defining and applying more than one specification for distance or proximity.\textsuperscript{448}

When a proximity-based approach is used, analysts should explain why it is not feasible to pursue an exposure-based modeling approach. It also is important to discuss the biases and limitations introduced when proximity or distance is used as a substitute for risk and exposure modeling (see Chakraborty and Maantay 2011). For instance, regardless of how the boundary is defined, proximity-based approaches typically account for the effects of a stressor only within a designated boundary.

10.2.7.4 Exposure Assessment

Spatial patterns of health or environmental effects - and changes in those effects - are difficult to analyze when pollution is diffuse. Air and water pollution, for example, are typically dispersed widely and may undergo physical, chemical, and other changes once released to the receiving media, thus changing the nature of the risk posed. While identifying the “proximity” to the hazards via GIS analysis is difficult in these cases, monitoring and/or modeling data may still allow for an assessment of the distributional effects at a disaggregated level. Criteria air pollutants (i.e., carbon monoxide, lead, nitrogen dioxide, ozone, particulate matter and sulfur dioxide) are monitored nationally. EPA’s National Air Toxics Assessment (NATA) data provide an assessment of hazardous air pollutants across the U.S. at the census tract level. Data from these sources may be combined with demographic data and dispersion models to generate baseline and regulatory distributions of pollutants by population groups of concern.\textsuperscript{449}

\textsuperscript{444} See Linden and Rockoff (2008) and Muehlenbachs, et al. (2015) for examples of approaches for identifying an appropriate distance.

\textsuperscript{445} In some cases, it may be possible to use dispersion models to select a buffer that approximates the effect of atmospheric conditions (e.g., wind direction and weather patterns) on exposure, though these types of models are data-intensive (Chakraborty and Maantay 2011).

\textsuperscript{446} For an overview of proximity analysis, including a discussion of various spatial analysis techniques used in the literature, see Chakraborty and Maantay (2011), and Mohai and Saha (2007).

\textsuperscript{447} Mohai and Saha (2007) refer to this as the “unit-hazard coincidence” approach because the analyst uses the available geographic units and determines whether they are coincident with an environmental hazard instead of first identifying the exact location of the hazard and then examining effects within a particular distance.

\textsuperscript{448} The analysis of distributional impacts in the context of EJ completed for EPA’s Definition of Solid Waste is an example of this type of analysis in a rule-making context. See EPA (2014).

\textsuperscript{449} For examples of studies that have used this approach to evaluate ambient concentrations of particular matter, see Fann et al. (2011), Rosenbaum, et
While this approach is promising due to the spatial detail of the data, it is currently only available for certain air pollutants. In addition, the data measure emissions, not individual exposures or health effects associated with the pollutant under consideration. These data are therefore still a proxy for actual effects associated with a specific regulation. Actual exposures or health effects may differ across individuals for a variety of reasons discussed throughout this chapter.

10.2.7.5 Risk Considerations

Activities linked to a specific culture or socioeconomic status could expose some populations groups of concern to higher levels of pollution both in the baseline and after a regulation is put in place. For example, some indigenous peoples and immigrant populations rely on subsistence fishing which could result in higher mercury levels from consumption of fish or expose these populations to other forms of pollution if fishing occurs in contaminated waters (see Donatuto and Harper 2008).

In addition to the potential for greater exposure to environmental risk, certain pre-existing factors also make some populations more susceptible (i.e., experience a greater biological response) to a specific environmental stressor for a given level of exposure (see Adler and Rehkopf 2008, Sacks et al. 2011 and Schwartz et al. 2011a). These factors can be genetic or physiological (such as sex and age). They may also be acquired due to variation in factors such as health-care access, nutrition, fitness, stress, housing quality, other pollutant exposures, or drug and alcohol use. For instance, many populations face exposures from multiple pollutants or exposures that have accumulated in ways that may affect their susceptibility to a specific pollutant due to pre-existing disease and adverse health conditions. In such instances, addressing EJ concerns is complicated. See the EJTG (U.S. EPA 2016) for a more detailed discussion of risk considerations.

10.2.7.6 Identification and Analysis of Potential Community "Hot Spots"

"Hot spots" is a term that is often used to refer to geographic areas with the potential for a high level of exposure to pollution or contamination than occur within a larger geographic area of lower or more "normal" exposure. Populations and communities in these geographic areas may face potential EJ impacts if higher exposure results in more concentrated environmental risk or negative health outcomes (e.g., cancer) for population groups of concern. Relevant issues in a local setting may include exposure pathways and drivers of differential susceptibility. It is important to note that hot spots may result from pre-existing conditions (i.e., that exist prior to the regulatory action), such as other stressors within the community or may be created as a result of the regulatory action.

It may be possible to identify the areas that have the potential for a higher than normal level of pollution or contamination using quantitative proximity analyses. In addition, information received via public comments can yield insights into potential hot spots. In cases where they are a relatively small in number, in-depth qualitative analysis may

---

450 It is also worth considering conditions that reduce a community’s ability to participate fully in the decision-making process such as time and resource constraints, lack of trust, lack of information, language barriers, and difficulty in accessing and understanding complex scientific, technical, and legal resources (see Dietz and Stern 2008).

451 A special issue of the American Journal of Public Health (Volume 101, Issue S1, December 2011) provides a set of papers exploring these and other issues.

452 EPA’s Integrated Risk Information System (IRIS) defines susceptibility as “increased likelihood of an adverse effect, often discussed in terms of relationship to a factor that can be used to describe a human subpopulation (e.g., life stage, demographic feature, or genetic characteristic).” See http://www.epa.gov/iris/help_gloss.htm#s (accessed on April 1, 2020).

453 Sexton (1997) suggests that low-income families may be more susceptible to environmental stressors due to differences in quality of life and lifestyle. Centers for Disease Control data show higher incidences of asthma-related emergency room visits and asthma-related deaths among African-American populations. See http://minorityhealth.hhs.gov/templates/content.aspx?ID=6170 (accessed April 1, 2020).

454 EPA (2003a) may serve as a useful reference when assessing how prior exposures may affect the impacts of emission changes from the rule being analyzed.
be useful. More sophisticated approaches may be required (e.g., fate and transport modeling) when potential hotspots are more numerous or widespread.

10.3 Children’s Environmental Health

Distributional analysis may shed light on differential effects of regulation on children, a life-stage defined group characterized by a multitude of unique behavioral, physiological, and anatomical attributes. There are two sets of important differences between children and adults regarding health benefits. First, there are differences in exposure to pollutants and in the nature and magnitude of health effects resulting from the exposure. Children may be more vulnerable to environmental exposures than adults because their bodily systems are still developing; they eat, drink, and breathe more in proportion to their body size; their metabolism may be significantly different especially shortly after birth; and their behavior can expose them more to chemicals and organisms (e.g., crawling leads to greater contact with contaminated surfaces, while hand-to-mouth and object-to-mouth contact is much greater for toddler age children). In addition, since children are younger they have more time to suffer adverse health effects from exposure to contaminants. Second, individuals may systematically place a different economic value on reducing health risks to children than on reducing such risks to adults. In part this is because children cannot provide marginal willingness to pay values for their own risk reductions, unlike adults, so children’s health risk valuation necessarily requires some model, implicit or explicit, about household decision making. These models differ in their implications for valuation. The perceived or actual effects of a given health outcome, too, may differ across children and adults. Empirical evidence also suggests that parents value a given risk reduction to themselves differently than to their children, with WTP for own risks generally valued less than those for children.

EO 13045 requires that each federal agency address disproportionate health risks to children. In addition, EPA’s Children’s Health Policy (U.S. EPA 1995) requires the Agency to “consider the risks to infants and children consistently and explicitly as a part of risk assessments generated during its decision-making process, including the setting of standards to protect public health and the environment.”

Generally, many approaches described earlier in this chapter to characterize the distribution of impacts may be adapted to evaluate children’s environmental health risks. For example, when proximity-based analysis is appropriate for evaluating EJ impacts, it might also be used to examine whether children are disproportionately located near facilities of concern. In such a case, the considerations described earlier about geography, defining the baseline and comparison groups, and use of summary statistics would all apply.

10.3.1 Childhood as a Life Stage

Evaluating distributional impacts of regulatory actions on children differs in an important way from evaluating the same impacts on population groups of concern for EJ. When the EPA evaluates disproportionate health risk impacts from environmental contaminants, it views childhood as a sequence of life stages from conception through fetal development, infancy, and adolescence, rather than a distinct “subpopulation.”

---

455 See Grineski (2009), Rao et al. (2007), Arcury et al. (2014), and Schwartz et al. (2015) for examples.


458 In principle there is a potential distinction in distributional analysis to be made between factors that are fixed, such as race and sex, and those defined by lifestages. The latter raises the possibility, at least, of examining distribution concerns through the lens of differences in lifetime utility or well-being rather than focusing on a single life stage. See Adler (2008) for one proposal consistent with this approach.
Use of the term “subpopulation” is ingrained in both EPA’s past practices as well as various laws that the EPA administers such as the Safe Drinking Water Act Amendments. Prior to publication of revised risk assessment guidelines in 2005, the EPA described all groups of individuals as “subpopulations.” In the 2005 guidelines, the Agency recognizes the importance of distinguishing between groups that form a relatively fixed portion of the population, such as those described in section 10.2, and life stages or age groups that are dynamic groups drawing from the entire population.

The term “life stage” refers to a distinguishable time frame in an individual’s life characterized by unique and relatively stable behavioral and/or physiological characteristics associated with development and growth. Thus, since 2005 the EPA characterizes childhood as a sequence of life stages.459

10.3.2 Analytical Considerations

Assessing distributional consequences of policies that affect children’s health requires considerations that span risk assessment, action development, and economic analysis. In each case there are existing Agency documents that can assist in the evaluation.

10.3.2.1 Risk Assessment

Effects of pollution can differ depending upon age of childhood exposure. Analysis of disproportionate impacts to children or from childhood life stages begins with health risk assessment, but also includes exposure assessment. Many risk guidance and related documents address how to consider children and childhood life stages in risk assessment.

A general approach to considering children and childhood life stages in risk assessment is found in A Framework for Assessing Health Risks of Environmental Exposures to Children (U.S. EPA 2006a). The framework identifies existing guidance, guidelines and policy papers that relate to children’s health risk assessment. It emphasizes the importance of an iterative approach between hazard, dose response, and exposure analyses. In addition, it includes a discussion of principles for weight-of-evidence consideration -- that is, the critical evaluation of available and relevant data -- across life stages.

EPA’s 2005 Guidelines for Carcinogenic Risk Assessment (Cancer Guidelines) (U.S. EPA 2005a) explicitly call for consideration of possible sensitive subpopulations and/or lifestages such as childhood. The Cancer Guidelines were augmented by Supplemental Guidance for Assessing Susceptibility from Early-Life Exposure to Carcinogens (2005b). Recommendations from this supplement include calculating risks utilizing life stage-specific potency adjustments in addition to life stage-specific exposure values which should be considered for all risk assessments.

EPA’s Child-Specific Exposures Handbook (U.S. EPA 2008b) and Highlights of the Child-Specific Exposure Factors Handbook (U.S. EPA 2009a) help risk assessors understand children’s exposure to pollution. The handbook provides important information for answering questions about life stage specific exposure through drinking, breathing, and eating. EPA’s guidance to scientists on selecting age groups to consider when assessing childhood exposure and potential dose to environmental contaminants is identified in Guidance on Selecting Age Groups for Monitoring and Assessing Childhood Exposures to Environmental Contaminants (U.S. EPA 2005c).

10.3.2.2 Action Development

Disproportionate impacts during fetal development and childhood are considered in EPA guidance on action development, particularly the Guide to Considering Children’s Health When Developing EPA Actions: Implementing Executive Order 13045 and EPA’s Policy on Evaluating Health Risks to Children (U.S. EPA 2006b). The guide helps determine whether EO 13045 and/or EPA’s Children’s Health Policy applies to an EPA action and, if so, how to

459 The 2005 Risk Assessment Guidelines “view childhood as a sequence of lifestages rather than viewing children as a subpopulation, the distinction being that a subpopulation refers to a portion of the population, whereas a life stage is inclusive of the entire population.” (U.S. EPA 2005a).
implement the Executive Order and/or EPA’s Policy. The guide clearly integrates EPA’s Policy on Children’s Health with the Action Development Process and provides an updated listing of additional guidance documents.

10.3.2.3 Economic Analysis

While these Economic Guidelines provide general information on benefit-cost analyses of policies and programs, many issues concerning valuation of health benefits accruing to children are not covered. Information provided in the Children’s Health Valuation Handbook (U.S. EPA 2003b), when used in conjunction with the Guidelines, allows analysts to characterize benefits and impacts of Agency policies and programs that affect children.

The Handbook is a reference tool for analysts conducting economic analyses of EPA policies when those policies are expected to affect risks to children’s health. The Handbook emphasizes that regulations or policies fully consider the economic impacts on children, including incorporating children’s health considerations into benefit-cost analysis, as well as separate distributional analysis focused on children.

Economic factors may also play a role in other analyses that evaluate children’s environmental health impacts. For example, because a higher proportion of children than adults live in poverty, the ability of households with children to undertake averting behaviors might be compromised. This type of information could inform the exposure assessment.

10.3.3 Intersection Between Environmental Justice and Children’s Health

The burden of health problems and environmental exposures is often borne disproportionately by children from low-income communities and minority communities (e.g., Israel et al. 2005; Lanphear et al. 1996; Mielke et al. 1999; Pastor et al. 2006). The challenge for the EPA is to integrate both EJ and life stage susceptibility considerations for children where appropriate when conducting distributional analysis. This is especially true when short-term exposure to environmental contaminants such as lead or mercury early in life can lead to life-long health consequences.

10.4 Other Distributional Considerations

10.4.1 Older adults

While there are no standard procedures for including effects on older adults in a distributional analysis, the EPA stresses the importance of addressing environmental issues that may adversely impact them. Most of the Agency’s work in this area has been related to risk and exposure assessment.

Older adults may be more susceptible to adverse effects of environmental contaminants due to differential exposures arising from physiological and behavioral changes with age, disease status, drug interactions, as well as the body’s decreased capacity to defend against toxic stressors. These considerations are highlighted in EPA’s Exposure Factors Handbook (U.S. EPA 2011c) and have led EPA’s Office of Research and Development to consider an exposure factors handbook specifically for the aging (see U.S. EPA 2007). Additionally, the toxicokinetic and toxicodynamic impacts of environmental agents in older adults have been considered in EPA’s document entitled Aging and Toxic Response: Issues Relevant to Risk Assessment (U.S. EPA 2005d).

10.4.2 Intergenerational Impacts

Concern for intergenerational impacts arises when those affected by a policy are not yet alive when the policy is developed. If a policy’s benefits, costs, and impacts primarily fall upon the current generation, or if policy decisions are reversible within this time frame, there is little need for explicit consideration of intergenerational impacts. However, in


\[61\] There is a lack of broad agreement about when this life stage begins. The U.S. and other countries typically define this life stage to begin at the traditional retirement age of 65, but, for example, the U.N. has it begin at age 60 (U.S. EPA 2005d).
other cases, benefits and/or costs of the policy will be borne by future generations, and it is important to consider impacts on these generations. One such case would be policies to reduce greenhouse gases, which are expected to result in benefits related to reduced changes in climate for future generations. Other examples may relate to toxic chemical exposures. Exposures to parents prior to their child’s conception can result in adverse health effects in the child, including effects that may not become apparent until the child reaches adulthood (for more information, see U.S. EPA 2006a and WHO 2007).

Social welfare functions, described in Text Box 10.1, and social discounting can be useful when examining intergenerational impacts. In both cases, normative judgments are required. Under the Ramsey approach to intergenerational discounting, the “pure rate of time preference” parameter can be used to weigh the welfare of current and future generations. However, there is no consensus on its value or whether it should be used in decision-making. See Section 6.3.1 for more information on intergenerational discounting and debate about the value of this parameter. One way to clarify distributional consequences if intergenerational impacts are important is to display time paths of benefits and costs without discounting, as recommended in Chapter 6 of these Guidelines.

**Chapter 10 References**


Chapter 11

Presentation of Analysis and Results

This chapter provides some general guidance for presenting analytical results to policy makers and others interested in environmental policy development. Economic analyses play an important role throughout the policy development process. From the initial, preliminary evaluation of potential options through the preparation of a final economic analysis document, economic analysts participate in an interactive process with policy makers. The fundamental goal of this process is to collect, analyze, and present information useful for policy makers.

Economic analysis is often motivated by a desire to find an optimal outcome, such as a degree of stringency in a regulation, or a level of provision of a public good that yields the largest possible net benefits. Environmental statutes sometimes mandate criteria other than economic efficiency, such as best available control technology or lowest achievable emission rate. Policy makers rely on quantitative analysis to promulgate these approaches. In particular they rely on analyses that delineate the costs, benefits, or other impacts of a wide range of control options.

This guidance for presenting inputs, analyses, and results applies at all stages of this process, not only for the final document embodying the completed economic analysis. Conveying uncertainty effectively and reporting critical assumptions and key unquantified effects to decision makers is critical at all points in the policy-making process.

This chapter begins by providing general guidance on how to present the results of economic analyses, with a particular emphasis on presenting benefits and costs, including those that cannot be quantified and/or put into dollar terms. The chapter then discusses the components, or inputs, of an economic analysis, and how their effect on the economic analysis can best be communicated.

11.1 Presenting Results of Economic Analyses

The presentation of the results of an economic analysis should be thorough and transparent. The reader should be able to understand:

- What the primary conclusions of the economic analysis are;
- How the benefits and costs were estimated;
- What the important non-quantified or non-monetized effects are;
- What key assumptions were made for the analysis;
- What the primary sources of uncertainty are in the analysis; and
- How those sources of uncertainty affect the results.

An economic analysis of regulatory or policy options should present all identifiable costs and benefits that are incremental to the regulation or policy under consideration.
Benefits and costs should be reported in monetary terms whenever possible. In reality, however, there are often effects that cannot be monetized, and the analysis needs to communicate the full richness of benefit and cost information beyond what can be put in dollar terms. Benefits and costs that cannot be monetized should, if possible, be quantified (e.g., expected number of adverse health effects avoided). Benefits and costs that cannot be quantified should be presented qualitatively (e.g., directional impacts on relevant variables). Section 11.1.2 contains more detailed guidance on presenting this information in EPA’s economic analyses.

Agencies are also required to provide OMB with an accounting statement reporting benefit and cost estimates when sending over each economically significant rule. Analysts should rely upon these Guidelines and Circular A-4 for developing these estimates. Circular A-4 describes the accounting statement on pages 44-46 and contains a suggested format for this accounting statement.\footnote{The accounting statement is on page 47 of Circular A-4, available at www.whitehouse.gov/sites/default/files/omb/assets/omb/circulars/a004/a-4.pdf (accessed on January 21, 2011).}

The results of economic analyses of environmental policies should generally be presented in three sections.

- **Results from BCA.** Estimates of the net social benefits should be presented based on the benefits and costs expressed in monetary terms. Non-monetized and unquantifiable benefits and costs should also be included and described in the presentation.

- **Results from cost-effectiveness analysis (CEA).** Under OMB Circular A-4, CEA should generally be performed for rules in which the primary effect is human health or safety. Results of these analyses should also be presented when they are conducted.\footnote{The Institute of Medicine (IOM) (2006) issued recommendations to regulatory agencies on how to perform health-based CEA. Examples of CEA can be found in appendices of several TRIAs including those for PM NAAQS [see Appendix G listed at http://www.epa.gov/ttn/ecas/ria.html (accessed March 13, 2011)] and the Ground Water Rule [see Appendix H listed at http://www.epa.gov/safewater/disinfection/gwr/regulation.html (accessed March 13, 2011)].}

- **Results from economic impact analysis (EIA) and distributional assessments.** Results of the EIA should be reported, including predicted effects on prices, profits, plant closures, employment, and any other effects. Distributional impacts for particular groups of concern, including small entities, governments, and environmental justice populations should also be presented.

The relative importance of these three sections will depend on the policy and statutory context of the analysis.

### 11.1.1 Presenting the Results of Benefit-Cost Analyses

When presenting the results of a BCA, the expected benefits and costs of the preferred regulatory option should be reported, together with the expected benefits and costs of alternative approaches. OMB’s Circular A-4 requires that at least one alternative be more stringent and one less stringent than the preferred option, and the incremental costs and benefits would be reported for each increasingly stringent option. Separate time streams of benefits and costs should be reported, in constant (inflation-adjusted), undiscounted dollars. Per the discussion in Chapter 6, appropriately discounted benefits and costs should be reported as well.

Ideally, all benefits and costs of a regulation would be expressed in monetary terms, but this is almost never possible because of data gaps, unquantifiable uncertainties, and other challenges. It is important not to exclude an important benefit or cost category from BCA even if it cannot be placed in dollar terms. Instead, such benefits and costs should be expressed quantitatively if possible (e.g., avoided adverse health impacts). If important benefit or cost categories cannot
be expressed quantitatively, they should be discussed qualitatively (e.g., a regulation’s effect on technological innovation).

Quantifiable benefits and costs, properly discounted, should be compared to determine a regulation’s net benefits, even if important benefits or costs cannot be monetized. However, an economic analysis should assess the likelihood that non-monetized benefits and costs would materially alter the net benefit calculation for a given regulation.

Incremental benefits, costs, and net benefits of moving from less to more stringent regulatory alternatives should also be presented. If a regulation has particularly significant impacts on certain groups or sub-populations, the various options’ incremental impacts on these sub-populations or source categories should be reported. This should include a discussion of incremental changes in quantified and qualitatively described benefits and costs.

Given the number of potential models presented in Chapters 7 and 8, the analyst should take care to clearly indicate the correspondence between the benefit and cost estimates. For example, the cost analysis may include results from a general equilibrium model but the benefit analysis may only include partial equilibrium effects. In this case, the cost side of the equation includes general equilibrium feedback effects while the benefit side does not. This difference should be clearly presented and explained.

The tables at the end of this chapter contain templates for presenting information on regulatory benefits and costs, including those benefits that cannot be quantified or put into dollar terms. The analyst’s primary goal, using these tables, is to communicate the full richness of benefit and cost information instead of focusing narrowly on what can be put in dollar terms. Some guiding principles for constructing these tables follow.

- All meaningful benefit and costs are included in all of the tables even if they cannot be quantified or monetized. Not only does this provide consistency for the reader, but it also maintains important information on the context of the quantified and monetized benefits.

- The types of benefits and costs are described briefly in plain terms to make them clearer to the public and to decision makers, and they should be well-defined and mutually exclusive, to the extent possible. Benefits should be grouped a manner consistent with the categories in Table 7.1 of Chapter 7, although the order and specific characterization can be expected to vary by rule as needed.

- The benefits are expressed first in natural or physical units (i.e., number) to provide a more complete picture of what the rule accomplishes. These units are not discounted as they would be in a CEA because the goal here is to describe what might be termed the “physical scope” of the rule’s benefits. It may be the case that physical or natural units are not relevant for presenting costs.

- Explanatory notes accompany each benefit and cost entry and can be used to describe whatever the most salient or important points are about scientific uncertainty, the type of benefit or cost, how it is estimated, or the presentation.

The benefit categories in these templates (e.g., improved human health, improved environment, and other benefits,) will need to be revised to reflect the benefits categories for the rule under consideration. Simpler analyses may need only the overview (Table 11.1) and the final summary (Table 11.4).

Table 11.1 is a quick-glance summary of regulatory benefits and costs, the extent to which they could be quantified and monetized, and a reference to where they are more fully characterized or estimated in the economic analysis. Some benefits may be described only qualitatively.

Table 11.2 reports benefits in non-monetary terms along with the units and additional explanatory notes. The goal of this table is to communicate the physical scope of the regulation’s benefits and costs rather than the dollar equivalent. Benefits here do not need to be discounted to present value, but the time associated with the quantities should be made clear (e.g., “annual” or “more than ten years”).
Table 11.3 reports benefits in monetary terms along with a total for dollar-valued benefits. Here it is important to specify the reference year for the dollars (i.e., real terms), the discount rate(s) used, and the unit value and/or source.

Table 11.4 contains a template for bringing all this information together in summary that includes the type of benefit or cost, how it is measured, its quantity, and dollar benefits. When multiple regulatory options are included in this table, it is appropriate for including in the regulatory preamble as requested by OMB.

Consistent with recommendations in these Guidelines for communicating uncertainty, quantitative entries should generally include a central or best estimate in addition to a range or confidence interval. The ability to do this, of course, may be limited by data availability.

**Table 11.1 - Template for Regulatory Benefits Checklist**

<table>
<thead>
<tr>
<th>Benefits</th>
<th>Effect can be Quantified? (put in numeric terms)</th>
<th>Effect can be Monetized? (put in dollar terms)</th>
<th>More Information (e.g., reference to section of the economic analysis)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Improved Human Health</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reduced incidence of adult premature mortality from exposure to PM$_{2.5}$</td>
<td>=</td>
<td>=</td>
<td>e.g., see Section 5.2 of the economic analysis</td>
</tr>
<tr>
<td>Reduced incidence of fetal loss from reduced exposure to disinfection byproducts</td>
<td>=</td>
<td>--</td>
<td>Notes and reference to section of the economic analysis</td>
</tr>
<tr>
<td>Unquantified human health benefit with a brief description</td>
<td>--</td>
<td>--</td>
<td>Notes and reference</td>
</tr>
<tr>
<td>Improved Environment</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fewer fish killed from reduced nutrient loadings into waterways</td>
<td>=</td>
<td>=</td>
<td>Notes and reference</td>
</tr>
<tr>
<td>Improved timber harvest from lower tropospheric ozone concentrations</td>
<td>=</td>
<td>=</td>
<td>Notes and reference</td>
</tr>
<tr>
<td>Other environmental benefit with a brief description</td>
<td>--</td>
<td>--</td>
<td>Notes and reference</td>
</tr>
<tr>
<td>Other Benefits</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fuel savings from improved efficiency in automobiles and light trucks</td>
<td>=</td>
<td>=</td>
<td>Notes and reference</td>
</tr>
<tr>
<td>Other benefit with a brief description</td>
<td>--</td>
<td>--</td>
<td>Notes and reference</td>
</tr>
</tbody>
</table>
Table 11.2 - Template for Quantified Regulatory Benefits

<table>
<thead>
<tr>
<th>Benefits</th>
<th>Quantified Benefits (confidence interval or range)</th>
<th>Units</th>
<th>More Information (w/possible reference to section of the economic analysis)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Improved Human Health</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reduced incidence of adult premature mortality from exposure to PM$_{2.5}$</td>
<td>estimate (range)</td>
<td>expected avoided premature deaths per year</td>
<td>e.g., range represents confidence interval</td>
</tr>
<tr>
<td>Reduced incidence of fetal loss from reduced exposure to disinfection byproducts</td>
<td>estimate (range)</td>
<td>expected avoided fetal losses per year</td>
<td>e.g., confidence interval cannot be estimated. Range based on alternative studies</td>
</tr>
<tr>
<td>Unquantified human health benefit with a brief description</td>
<td>*</td>
<td>*</td>
<td>e.g., data do not allow for quantification</td>
</tr>
<tr>
<td><strong>Improved Environment</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fewer fish killed from reduced nutrient loadings into waterways</td>
<td>estimate (range)</td>
<td>thousands of fish per year</td>
<td>Notes (reference)</td>
</tr>
<tr>
<td>Improved timber harvest from lower tropospheric ozone concentrations</td>
<td>estimate (range)</td>
<td>thousands of board feet per year</td>
<td>Notes (reference)</td>
</tr>
<tr>
<td>Other environmental benefit with a brief description</td>
<td>*</td>
<td>*</td>
<td>Notes (reference)</td>
</tr>
<tr>
<td><strong>Other Benefits</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fuel savings from improved efficiency in automobiles and light trucks</td>
<td>estimate (range)</td>
<td>millions of gallons of gasoline reduced per year</td>
<td>Notes (reference)</td>
</tr>
<tr>
<td>Other benefit with a brief description</td>
<td>*</td>
<td>*</td>
<td>Notes (reference)</td>
</tr>
</tbody>
</table>

Note: * indicates the benefit cannot be quantified with available information
Table 11.3 - Template for Dollar-Valued Regulatory Benefits

<table>
<thead>
<tr>
<th>Benefit</th>
<th>Dollar Benefits (millions per year)</th>
<th>Basis of Value</th>
<th>More Information (w/possible reference)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Improved Human Health</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reduced incidence of adult premature mortality from exposure to PM$_{2.5}$</td>
<td>$\text{estimate} ($ range)</td>
<td>e.g., $X$ based on Agency guidance</td>
<td>Notes (reference)</td>
</tr>
<tr>
<td>Reduced incidence of fetal loss from reduced exposure to disinfection byproducts</td>
<td>*</td>
<td>Not available</td>
<td>Notes (reference)</td>
</tr>
<tr>
<td>Unquantified human health benefit with a brief description</td>
<td>*</td>
<td>*</td>
<td>e.g., data insufficient to quantify (reference)</td>
</tr>
<tr>
<td><strong>Improved Environment</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fewer fish killed from reduced nutrient loadings into waterways</td>
<td>$\text{estimate} ($ range)</td>
<td>e.g., $X$ based on WTP for recreational fishing</td>
<td>e.g., range reflects two different valuation approaches (reference)</td>
</tr>
<tr>
<td>Improved timber harvest from lower tropospheric ozone concentrations</td>
<td>$\text{estimate} ($ range)</td>
<td>e.g., change in consumer and producer surplus</td>
<td>e.g., estimated from market model across several species (reference)</td>
</tr>
<tr>
<td>Other environmental benefit with a brief description</td>
<td>*</td>
<td>*</td>
<td>Notes (reference)</td>
</tr>
<tr>
<td><strong>Other Benefits</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fuel savings from improved efficiency in automobiles and light trucks</td>
<td>$\text{estimate} ($ range)</td>
<td>e.g., $X$, based on net-of-tax average per gallon price</td>
<td>e.g., there is debate on how well fuel savings represent consumer benefits (reference)</td>
</tr>
<tr>
<td>Other benefit with a brief description</td>
<td>*</td>
<td>Not available</td>
<td>Notes (reference)</td>
</tr>
<tr>
<td><strong>TOTAL Benefits that can be monetized</strong> ($millions per year)</td>
<td>$\text{estimate} ($ range)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: * indicates the benefit cannot be quantified with available information.
### Table 11.4 - Template for Summary of Benefits and Costs

**Benefits**

Notes: e.g., “annual average numbers; 2006 dollars annualized at 3% discount rate”
Best estimate, with range

<table>
<thead>
<tr>
<th>Option 1</th>
<th>Proposed Option</th>
<th>Option 3</th>
<th>Source, limitations, or other key notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number</td>
<td>Number</td>
<td>Number</td>
<td>Source, limitations, or other key notes</td>
</tr>
<tr>
<td>$ Millions</td>
<td>$ Millions</td>
<td>$ Millions</td>
<td>Source, limitations, or other key notes</td>
</tr>
</tbody>
</table>

11-7
**Improved Human Health**

<table>
<thead>
<tr>
<th>Effect Description</th>
<th>Estimate (range)</th>
<th>$ Estimate (range)</th>
<th>$ Estimate (range)</th>
<th>$ Estimate (range)</th>
<th>$ Estimate (range)</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reduced incidence of adult premature mortality from exposure to PM$_{2.5}$</td>
<td>estimate</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>highlight most important points, as needed</td>
</tr>
<tr>
<td>Reduced incidence of fetal loss from reduced exposure to disinfection byproducts</td>
<td>estimate</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>e.g., no valuation data exist. Effects are sensitive to dose-response model.</td>
</tr>
<tr>
<td>Unquantified human health benefit with a brief description</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>e.g., risk data insufficient for quantification</td>
</tr>
</tbody>
</table>

**Improved Environment**

<table>
<thead>
<tr>
<th>Effect Description</th>
<th>Estimate (range)</th>
<th>$ Estimate (range)</th>
<th>$ Estimate (range)</th>
<th>$ Estimate (range)</th>
<th>$ Estimate (range)</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fewer fish killed from reduced nutrient loadings into waterways</td>
<td>estimate</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>Notes</td>
</tr>
<tr>
<td>Improved timber harvest from lower tropospheric ozone concentrations</td>
<td>estimate</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>Notes</td>
</tr>
<tr>
<td>Other environmental benefit with a brief description</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>Notes</td>
</tr>
</tbody>
</table>

**Other Benefits**

<table>
<thead>
<tr>
<th>Effect Description</th>
<th>Estimate (range)</th>
<th>$ Estimate (range)</th>
<th>$ Estimate (range)</th>
<th>$ Estimate (range)</th>
<th>$ Estimate (range)</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fuel savings from improved efficiency in automobiles and light trucks</td>
<td>estimate</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>Notes</td>
</tr>
<tr>
<td>Other benefit with a brief description</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>Notes</td>
</tr>
</tbody>
</table>

**TOTAL Benefits that can be monetized (annualized, millions $2006)**

<table>
<thead>
<tr>
<th>Estimate (range)</th>
<th>$ Estimate (range)</th>
<th>$ Estimate (range)</th>
<th>$ Estimate (range)</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>e.g., total range may be overstated because of aggregation</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(See Section 8.1 of economic analysis)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Costs
2006 dollars annualized at 3% discount rate
Best estimate, with range

<table>
<thead>
<tr>
<th></th>
<th>Option 1</th>
<th>Proposed Option</th>
<th>Option 3</th>
<th>Source, limitations, or other key notes</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Initial capital costs</strong> with any brief description and units.</td>
<td>$ estimate (range)</td>
<td>$ estimate (range)</td>
<td>$ estimate (range)</td>
<td>e.g., estimated from engineering cost models</td>
</tr>
<tr>
<td><strong>Type of cost</strong> with a brief description and units. (This could include non-monetized costs.)</td>
<td>$ estimate (range)</td>
<td>$ estimate (range)</td>
<td>$ estimate (range)</td>
<td>Notes</td>
</tr>
<tr>
<td><strong>Type of cost</strong> with a brief description and units. (This could include non-monetized costs.)</td>
<td>$ estimate (range)</td>
<td>$ estimate (range)</td>
<td>$ estimate (range)</td>
<td>Notes</td>
</tr>
</tbody>
</table>

**TOTAL Costs that can be monetized** (annualized, millions $2006)

<table>
<thead>
<tr>
<th></th>
<th>$ estimate (range)</th>
<th>$ estimate (range)</th>
<th>$ estimate (range)</th>
</tr>
</thead>
</table>

**TOTAL Net Benefits that can be monetized** (annualized, millions $2006)

<table>
<thead>
<tr>
<th></th>
<th>$ estimate (range)</th>
<th>$ estimate (range)</th>
<th>$ estimate (range)</th>
</tr>
</thead>
</table>

Note: * indicates the benefit cannot be quantified with available information.

### 11.1.2 Presenting the Results of Cost-Effectiveness Analyses

When BCA is not possible, CEA may be the best available option. The cost-effectiveness of a policy option is calculated by dividing the annualized cost of the option by non-monetary benefit measures. Options for such measures range from quantities of pollutant emissions reduced, measured in physical terms, to a specific improvement in human health or the environment, measured in reductions in illnesses or changes in ecological services rendered.

In the context of RIA, or other analyses of specific regulatory or policy options, CEA is most informative when several different options are analyzed. The analysis should include at least one option that is less stringent and at least one option that is more stringent than the preferred option. The incremental costs and non-monetary benefit yield of each option, in order of increasing stringency, should be reported.

The non-monetary measure of benefits used in a CEA must be chosen with great care to facilitate valid comparisons across options. The closer the chosen measure is to the variable that directly impacts social welfare, the more robust a CEA will be. Consider the following steps that a typical environmental economic assessment follows:

- Changes in emissions are estimated (e.g., tons of emissions); then
- Changes in environmental quality (e.g., changes in ambient concentrations of a given air pollutant) are estimated; then
- Changes in human health or welfare (e.g., changes in illness or visibility) are estimated.
Each successive step in this sequence yields a better measure for CEA.

To illustrate, consider a typical air pollution scenario. Depending on where and when air pollutants are released into the atmosphere, a given ton of a particular pollutant can have widely divergent impacts on ambient air quality. Similarly, depending on when and where air quality changes, widely different levels of human health impacts may result. Particularly when different regulatory approaches are under consideration (e.g., regulation of different source categories in different locations), failing to standardize the analyses on the benefit measure that directly affects human health or welfare will significantly reduce the value of the analysis to decision makers (and the public).

When presenting the results of a CEA, the rationale for the selection of the non-monetary benefit measure must be described in detail. The presentation of results should also include a discussion of the limitations of the analysis, especially if an inferior measure, such as cost per ton of pollutant, must be used.

CEA is most useful when the policy or regulation in question affects a single endpoint. When multiple endpoints are affected (e.g., cancer and kidney failures), combining endpoints into a single effectiveness measure is impossible unless appropriate weighting factors exist for the multiple endpoints. The theoretically correct weights to apply are the dollar values associated with each endpoint, but generally it is the absence of these values that necessitates CEA. Therefore, it is not possible to compare a policy or regulation that reduces relatively more expected cancers, but fewer expected cases of kidney failure, with one that has the opposite relative effects. When this occurs, the effects of each option for each endpoint should be reported. A single endpoint may be selected for calculating cost-effectiveness, while other endpoints can be listed as ancillary benefits (or, if possible, their monetary value should be subtracted from the option’s cost prior to calculating its cost-effectiveness) (OMB 2003).

The most cost-effective option — i.e., the option with the lowest cost per unit of benefit — is not necessarily the most economically efficient. Moreover, other criteria, such as statutory requirements, enforcement problems, technological feasibility, or quantity and location of total emissions abated may preclude selecting the least-cost solution in a regulatory decision. However, where not prohibited by statute, CEA can indicate which control measures or policies are inferior options.

11.1.3 Presenting the Results of EIA and Distributional Analyses

EIA and distributional outcomes focus on disaggregating effects to show impacts separately for the groups and sectors of interest. If costs and/or benefits vary significantly among the sectors affected by the policy, then both costs and benefits should be shown separately for the different sectors. Presenting results in disaggregated form will provide important information to policy makers that may help them tailor the rule to improve its efficiency and distributional outcomes.

The results of the EIA should also be reported for important sectors within the affected population — identifying specific segments of industries, regions of the country, or types of firms that may experience significant impacts or plant closures and losses in employment.

Reporting the results in distributional assessments may include the expected allocation of benefits, costs, or both for specific subpopulations including those highlighted in the various mandates. These include minorities, low-income populations, small businesses, governments, not-for-profit organizations, and sensitive and vulnerable populations (including children). Where these mandates specify requirements that depend on the outcomes of the distributional analyses, such as the Regulatory Flexibility Act, the presentation of the results should conform to the criteria specified by the mandate.
11.1.4 Reporting the Effects of Uncertainty on Results of Economic Analyses

Estimates of costs, benefits and other economic impacts should be accompanied by indications of the most important sources of uncertainty embodied in the estimates, and, if possible, a quantitative assessment of their importance. OMB requires formal quantitative analysis of uncertainties for rules with annual economic effects of $1 billion or more.

In economic analysis, uncertainty encompasses two different concepts:

- Statistical variability of key parameters; and
- Incomplete understanding of important relationships.

Economic analyses of environmental policies and regulatory options will frequently have to accommodate both concepts. The importance of statistical variability is commonly assessed using Monte Carlo analyses. Delphic panels, or expert elicitation techniques, can help close knowledge gaps surrounding key relationships (see IEc 2004).

Ideally, an economic analysis would present results in the form of probability distributions that reflect the cumulative impact of all underlying sources of uncertainty. When this is impossible, due to time or resource constraints, results should be qualified with descriptions of major sources of uncertainty. If at all possible, information about the underlying probability distribution should be conveyed.

As recommended in Chapter 6, many EPA analyses will employ more than one discount rate to reflect different underlying approaches to discounting. When the choice of discount rate affects the outcome of the analysis, analysts should take extra care to convey the underlying theory and assumptions to decision makers. See Chapter 6 for more information.

11.2 Communicating Data, Model Choices and Assumptions, and Related Uncertainty

An economic analysis of an environmental regulation should carefully describe the data used in the analysis, the models it relies on, major assumptions that were made in running the models, and all major areas of uncertainty in each of these elements. Presentations of economic analyses should strive for clarity and transparency. An analysis whose conclusions can withstand close scrutiny is more likely to provide policy makers with the information they need to develop robust environmental policies.

11.2.1 Data

An economic analysis should clearly describe all important data sources and references used. Unless the data are confidential business information or some other form of private data, they should be available to policy makers, other researchers, policy analysts and the public. Providing documentation and access to the data used in an analysis is crucial to the credibility and reproducibility of the analysis.

EPA Order CIO 2105.0 (U.S. EPA 2000a) and the applicable federal regulations established a mandatory quality system for EPA. As required by the quality system, all EPA offices have developed quality management plans to ensure the quality of their data and information products.

At one time federal quality assurance (QA) requirements only applied to measurement and collection of primary environmental data. This meant that QA requirements often did not apply to economic analyses, which usually rely on the use of secondary data. However, this changed with the introduction of QA requirements regarding use of secondary data. In 2002 the Agency released QA guidelines regarding use of secondary data, and released Agency guidance, Guidance for Quality Assurance Project Plans, that includes procedures for documenting secondary data (U.S. EPA 2002f).
In any economic analysis, there should be a clear presentation of how data are used and a concise explanation of why the data are suitable for the selected purpose. The data’s accuracy, precision, representativeness, completeness, and comparability should be discussed when applicable. When data are available from more than one source, a rationale for choosing the source of the data should be provided.

### 11.2.2 Model Choices and Assumptions

An economic analysis of an environmental regulation should carefully describe the models it relies on, the major assumptions made in running the models (to be discussed more fully below), and any areas of outstanding uncertainty. The analyst should take particular care to explain any results that might be viewed as counter-intuitive. In particular, analysts should be careful not to accept model output blindly. Any model that is used without proper thought given to both its input and output may become a “black box” insofar as nonsensical results may result from a misspecified scenario, a coding error, or any of a number of other causes.

In the process of conducting an economic analysis, it is sometimes necessary to bridge an information gap by making an assumption. Analysts should not simply note the information gap, but should also justify the chosen assumption and provide a rationale for choosing one assumption over other plausible options. The analyst should take care not to overlook information gaps that are filled with a piece of information that is only slightly related to the desired information. Analysts are advised to keep a running list of assumptions. This will make it easier to identify “key assumptions” for the final report. The likely impact of errors in assumptions should be characterized both in terms of direction and magnitude of effect when feasible.

Maintaining a list of assumptions can benefit the analysis in several ways. In the short run, a list can serve to focus analysts’ attention on those assumptions with the greatest potential to affect net benefits, possibly leading to new approaches to bridging an information gap. In the long run, highlighting information gaps may encourage EPA or others to devote attention and resources to generating that information.

Whenever the likely errors in a particular assumption can be characterized numerically or statistically, the factor is a good candidate for sensitivity analysis or uncertainty analysis, respectively. In many cases, only a narrative description of the impact of errors in assumptions is possible. The analyst should include a table that clearly lays out all of the key assumptions and the potential magnitude and direction of likely errors in assumptions in the summary of results.

### 11.2.3 Addressing Uncertainty Driven by Assumptions and Model Choice

Every analysis should address uncertainties resulting from the choices the analyst has made. For example, many economic analyses performed at EPA include assessments of economic impacts expected to occur decades into the future. Estimates of the future costs and benefits of a regulation will be sensitive to assumptions about growth rates for populations, source categories, economic activity, and technological change, as well as many other factors. Sensitivity analyses on key variables in the baseline scenario should be performed and reported when possible. This allows the reader to assess the importance of the assumptions made for the central case. Some of these variables may be affected by a regulation, particularly the assumed rate of technological innovation (see Chapter 5 for additional guidance on specifying baselines).

The impact of using alternative assumptions or alternative models can be assessed quantitatively in many cases.

- **Alternative analysis.** An analysis of alternative assumptions or “alternative analysis” is the substitution of one of the key assumptions with another. In presenting the results, the alternative analysis is presented with equal weight as the primary analysis and is presented alongside of the primary analysis, even if the probability of the alternative assumption differs from that of the primary analysis. Because performing an alternative analysis on all the assumptions in an analysis is prohibitively resource intensive, the analyst should focus on
the assumptions that have the largest impact on the final results of the particular analysis. Thus, keeping a running list of the “key assumptions” in an analysis is recommended.

- **Sensitivity analysis.** A sensitivity analysis is used to assess how the final results or other aspects of the analysis change as input parameters change, particularly when only point estimates of parameters are available. A regulatory impact analysis benefits from knowing how the cost-effectiveness of a particular technology changes as fuel prices change, or how the net benefits of a BCA change as one of the model coefficients change. Typically, a sensitivity analysis measures how the model’s output changes as one of the input parameters change. Joint sensitivity analysis (varying more than one parameter at a time) is sometimes useful as well.

- **Model uncertainty.** In addition to explaining the uncertainty in a model’s parameters, analysts should discuss the uncertainty generated by the choice of model. Multiple models are often available to the analyst, and choosing among them is similar to making an assumption. Implicit in the choice of a model are many factors. For example, one model may take long-run effects into account while another model does not. When possible, presenting results of an alternate model can inform the reader. When resource limitations prevent the use of an alternative model, it is still often possible to predict the direction and likely magnitude of the use of an alternate model, and the analyst should present this information to the reader.

### 11.3 Use of Economic Analyses

The primary purpose of conducting economic analysis is to provide policy makers and others with detailed information on a wide variety of consequences of environmental policies. One important element these analyses have traditionally provided to the policy-making process is estimates of social benefits and costs — the economic efficiency of a policy. For this reason, these Guidelines reflect updated information associated with procedures for calculating benefits and costs, monetizing benefits estimates, and selecting particular inputs and assumptions.

Determining which regulatory options are best even on the restrictive terms of economic efficiency is often made difficult by uncertainties in data and by the presence of benefits and costs that can be quantified but not monetized, or that can only be qualitatively assessed. Even if the criterion of economic efficiency were the sole guide to policy decisions, social benefit and costs estimates alone would not be sufficient to define the best policies.

A large number of social goals and statutory and judicial mandates motivate and shape environmental policy. For this and other reasons, these Guidelines contain information concerning procedures for conducting analyses of other consequences of environmental policies, such as economic impacts and equity effects. This is consistent with the fact that economic efficiency is not the sole criterion for developing good public policies.

Even the most comprehensive economic analyses are but part of a larger policy development process, one in which no individual analytical feature or empirical finding dominates. The role of economic analysis is to organize information and comprehensively assess the economic consequences of alternative actions — benefits, costs, economic impacts, and equity effects — and the trade-offs among them. Ultimately statutory requirements dictate if and how the analytic results are used in standard setting. In any case, these results, along with other analyses and considerations, serve as important inputs for the broader policy-making process and serve as important resources for the public.

### Chapter 11 References


Appendix A

Economic Theory

This appendix provides a brief overview of the fundamental theory underlying the approaches to economic analysis discussed in Chapters 3 through 9. The first section summarizes the basic concepts of the forces governing a market economy in the absence of government intervention. Section A.2 describes why markets may behave inefficiently. If the preconditions for market efficiency are not met, government intervention can be justified. The usefulness of benefit-cost analysis (BCA) as a tool to help policy makers determine the appropriate policy response is discussed in Section A.3. Sections A.4 and A.5 explain how economists measure the economic impacts of a policy and set the optimal level of regulation. Section A.6 concludes and provides a list of additional references.

A.1 Market Economy

The economic concept of a market is used to describe any situation where exchange takes place between consumers and producers. Economists assume that consumers purchase the combination of goods that maximizes their well-being, or “utility,” given market prices and subject to their household budget constraint. Economists also assume that producers (firms) act to maximize their profits. Economic theory posits that consumers and producers are rational agents who make decisions taking into account all of the costs — the full opportunity costs — of their choices, given their own resource constraints. The purpose of economic analysis is to understand how the agents interact and how their interactions add up to determine the allocation of society’s resources: what is produced, how it is produced, for whom it is produced, and how these decisions are made. The simplest tool economists use to illustrate consumers’ and producers’ behavior is a market diagram with supply and demand curves.

The demand curve for a single individual shows the quantity of a good or service that the individual will purchase at any given price. This quantity demanded assumes the condition of holding all else constant, i.e., assuming the budget constraint, information about the good, expected future prices, prices of other goods, etc. remain constant. The height of the demand curve in Figure A.1 indicates the maximum price, $P$, an individual with $Q_d$ units of a good or service would be willing to pay to acquire an additional unit of a good or service. This amount reflects the satisfaction (or utility) the individual receives from an additional unit, known as the marginal benefit of consuming the good. Economists generally assume that the marginal benefit of an additional unit is slightly less than that realized by the previous unit. The amount an individual is willing to pay for one more unit of a good is less than the amount she paid for the last unit; hence, the individual demand curve slopes downward. A market demand curve shows the total quantity that consumers are willing to purchase at different price levels, i.e., their collective willingness to pay (WTP) for the good or service. In other words, the market demand curve is the horizontal sum of all of the individual demand curves.

---

465 EPA’s mandates frequently rely on criteria other than economic efficiency, so policies that are not justified due to a lack of efficiency are sometimes adopted.

Opportunity cost is the next best alternative use of a resource. The full opportunity cost of producing (consuming) a good or service consists of the maximum value of other goods and services that could have been produced (consumed) had one not used the limited resources to produce (purchase) the good or service in question. For example, the full cost of driving to the store includes not only the price of gas but also the value of the time required to make the trip.
The concept of an individual’s WTP is one of the fundamental concepts used in economic analyses, and it is important to distinguish between total and marginal WTP. Marginal WTP is the additional amount the individual would pay for one additional unit of the good. The total WTP is the aggregate amount the individual is willing to pay for the total quantity demanded ($Q_d$). Figure A.1 illustrates the difference between the marginal and total WTP. The height of the demand curve at a quantity $Q_{d-1}$ gives the marginal WTP for the $Q_{d-1}$th unit. The height of the demand curve at a quantity $Q_d$ gives the marginal WTP for the $Q_{d}$th unit. Note that the marginal WTP is greater for the $Q_{d-1}$th unit. The total WTP is equal to the sum of the marginal WTP for each unit up to $Q_d$. The shaded area under the demand curve from the origin up to $Q_d$ shows total WTP.

An individual producer’s supply curve shows the quantity of a good or service that an individual or firm is willing to sell ($Q_s$) at a given price. As a profit-maximizing agent, a producer will only be willing to sell another unit of the good if the market price is greater than or equal to the cost of producing that unit. The cost of producing the additional unit is known as the marginal cost. Therefore, the individual supply curve traces out the marginal cost of production and is also the marginal cost curve. Economists generally assume that the cost of producing one additional unit is greater than the cost of producing the previous unit because resources are scarce. Therefore, the supply curve is assumed to slope upward. In Figure A.2, the marginal cost of producing the $Q_s$th unit of the good is given by the height of the supply curve at $Q_s$. The marginal cost of producing the $Q_{s+1}$th unit of the good is given by the height of the supply curve at $Q_{s+1}$, which greater than the cost of producing the $Q_s$th unit, and greater than the price, $P$. The total cost of producing $Q_s$ units is equal to the shaded area under the supply curve from the origin to the quantity $Q_s$. The market supply curve is simply the horizontal summation of the individual producers’ marginal cost curves for the good or service in question.

In a competitive market economy, the intersection of the market demand and market supply curves determines the equilibrium price and quantity of a good or service sold. The demand curve reflects the marginal benefit consumers receive from purchasing an extra unit of the good (i.e., it reflects their marginal WTP for an extra unit). The supply curve reflects the marginal cost to the firm of producing an extra unit. Therefore, at the competitive equilibrium, the price is where the marginal benefit equals the marginal cost. This is illustrated in Figure A.3, where the supply curve intersects the demand curve at equilibrium price $P_m$ and equilibrium quantity $Q_m$.

A counter-example illustrates why the equilibrium price and quantity occur at the intersection of the market demand and supply curves. In Figure A.3, consider some price greater than $P_m$ where $Q_s$ is greater than $Q_d$ (i.e., there is excess supply). As producers discover that they cannot sell off their inventories, some will reduce prices slightly, hoping to attract more customers. At lower prices consumers will purchase more of the good ($Q_d$ increases) although firms will be willing to sell less ($Q_s$ decreases). This adjustment continues until $Q_d$ equals $Q_s$. The reverse situation occurs if the price becomes lower than $P_m$. In that case, $Q_d$ will exceed $Q_s$ (i.e., there is excess demand) and consumers who cannot purchase as much as they would like are willing to pay higher prices. Therefore, firms will begin to increase prices, causing some reduction in the $Q_d$ but also...
increasing $Q_s$. Prices will continue to rise until $Q_s$ equals $Q_d$. At this point no purchaser or supplier will have an incentive to change the price or quantity; hence, the market is said to be in equilibrium.

Economists measure a consumer’s net benefit from consuming a good or service as the excess amount that she is willing to spend on the good or service over and above the market price. The net benefit of all consumers is the sum of individual consumer’s net benefits — i.e., what consumers are willing to spend on a good or service over and above that required by the market. This is called the consumer surplus. In Figure A.3, the market demands price $P_m$ for the purchase of quantity $Q_m$. However, the demand curve shows that there are consumers willing to pay more than price $P_m$ for all units prior to $Q_m$. Therefore, the consumer surplus is the area under the market demand (marginal benefit) curve but above the market price. Policies that affect market conditions in ways that decrease prices by decreasing costs of production (i.e., that shift the marginal cost curve to the right) will generally increase consumer surplus. This increase can be used to measure the benefits that consumers receive from the policy.\footnote{Section A.4.2 provides a more technical discussion of how consumer surplus serves as a measure of benefits.}

On the supply side, a producer can be thought to receive a benefit if he can sell a good or service for more than the cost of producing an additional unit — i.e., its marginal cost. Figure A.3 shows that there are producers willing to sell up to $Q_m$ units of the good for less than the market price $P_m$. Hence, the net benefit to producers in this market, known as producer surplus, can be measured as the area above the market supply (marginal cost) curve but below the market price. Policies that increase prices by increasing market demand for a good (i.e., that shift the marginal benefit curve to the right) will generally increase producer surplus. This increase can be used to measure the benefits that producers receive from the policy.

\textit{Economic efficiency} is defined as the maximization of social welfare. In other words, the efficient level of production is one that allows society to derive the largest possible net benefit from the market. This condition occurs where the (positive) difference between the total WTP and total costs is the largest. In the absence of externalities and other market failures (explained below), this occurs precisely at the intersection of the market demand and supply curves where the marginal benefit equals the marginal cost. This is also the point where total surplus (consumer surplus plus producer surplus) is maximized. There is no way to rearrange production or reallocate goods so that someone is made better off without making someone else worse off — a condition known as Pareto optimality. Notice that economic efficiency requires only that net benefits be maximized, \textit{irrespective of to whom those net benefits accrue}. It does not guarantee an “equitable” or “fair” distribution of these surpluses among consumers and producers, or between subgroups of consumers or producers.

Economists maintain that \textit{if the economic conditions are such that there are no market imperfections} (as discussed in Section A.2), then this condition of Pareto-optimal economic efficiency occurs automatically.\footnote{Technically, there are two types of efficiency. Allocative efficiency means that resources are used for the production of goods and services most efficiently within the market. In other words, resources are allocated to their most valued use. Productive efficiency occurs when resources are allocated in such a way that the marginal cost is minimized. Notice that productive efficiency can also be considered allocative efficiency since the resource of marginal cost is being minimized. Economic efficiency, however, is the maximization of social welfare and occurs where the marginal benefit equals the marginal cost, i.e., the point where total surplus is maximized.} That is, no government
intervention is necessary to maximize the sum of consumer surplus and producer surplus. This theory is summarized in the two Fundamental Theorems of Welfare Economics, which originate with Pareto (1906) and Barone (1908):

1. **First Fundamental Welfare Theorem.** Every competitive equilibrium is Pareto-optimal.
2. **Second Fundamental Welfare Theorem.** Every Pareto-optimal allocation can be achieved as a competitive equilibrium after a suitable redistribution of initial endowments.

One graphical representation of these results is given in Figure A.4, which shows utility (welfare) levels in a two-person economy. The curve shown is the utility possibility frontier (UPF) curve; the area within it represents the set of all possible welfare outcomes. Each point on the negatively sloped UPF curve is Pareto optimal since it is not possible to increase the utility of one person without decreasing the utility of the other. If the initial allocation is at point A, then the set of Pareto-superior (welfare-enhancing) outcomes include all points in the shaded area, bordered by \( H \), \( V \), and the UPF curve. If trading is permitted, the First Welfare Theorem applies and the market will move the economy to a superior, more efficient point such as \( B \). Then the Second Welfare Theorem simply says that for any chosen point along the UPF curve, given a set of lump sum taxes and transfers, an initial allocation can be determined inside the UPF from which the market will achieve the desired outcome.

### A.2 Reasons for Market or Institutional Failure

If the market supply and demand curves reflect society’s true marginal social cost and WTP, then a laissez-faire market (i.e., one governed by individual decisions and not government authority) will produce a socially efficient result. However, when markets do not fully represent social values, the private market will not achieve the efficient outcome (see Mankiw 2004, or any basic economics text); this is known as a *market failure*. Market failure is primarily the result of externalities, market power, and inadequate or asymmetric information. Externalities are the most likely cause of the failure of private and public sector institutions to account for environmental damages.

**Externalities** occur when markets do not account for the effect of one individual’s decisions on another individual’s well-being. In a free market producers make their decisions about what and how much to produce, taking into account the wanted by society. Productive efficiency implies that the least costly production techniques are used to produce any mix of goods and services. Allocative efficiency requires that there be productive efficiency, but productive efficiency can occur without allocative efficiency. Goods can be produced at the least-costly method without being most wanted by society. Perfectly competitive markets in the long run will achieve both of these conditions, producing the “right” goods (allocative efficiency) in the “right” way (productive efficiency). These two conditions imply Pareto-optimal economic efficiency. (See Varian 1992 or any basic economics text for a more detailed discussion.)

Another, perhaps more commonly used, graphical tool to explain the First and Second Welfare Theorems is an Edgeworth box. See Varian (1992) or other basic economic textbook for a detailed discussion.

Note that efficiency could be obtained by moving along the vertical line \( V \), which keeps utility of person 1 \((U_1)\) constant while increasing utility of person 2 \((U_2)\), or by moving along the horizontal line \( H \), which only shows improvements in utility for person 1. Moving to point \( B \) improves the utility for both individuals.

Note that outcomes on the frontier such as \( C \) and \( D \), although efficient, may not be desired on equity, or fairness, grounds.

More formally, an externality occurs when the production or consumption decision of one party has an unintended negative (positive) impact on the profit or utility of a third party. Even if one party compensates the other party, an externality still exists (Perman et al. 2003). See Baumol and Oates (1988) or any basic economics textbook for similar definitions and more detailed discussion.
cost of the required inputs — labor, raw materials, machinery, energy. Consumers purchase goods and services taking into account their income and their own tastes and preferences. This means that decisions are based on the private costs and private benefits to market participants. If the consumption or production of these goods and services poses an external cost or benefit on those not participating in the market, however, then the market demand and supply curves no longer reflect the true marginal social benefit and marginal social cost. Hence, the market equilibrium will no longer be the socially (Pareto) efficient outcome.

Externalities can arise for many reasons. Transactions costs or poorly defined property rights can make it difficult for injured parties to bargain or use legal means to ensure that the costs of the damages caused by polluters are internalized into their decision making. Activities that pose environmental risks may also be difficult to link to the resulting damages and often occur over long periods of time. Externalities involve goods that people care about but are not sold in markets. Air pollution causes ill health, ecological damage, and visibility impacts over a long time period, and the damage is often far from the source(s) of the pollution. The additional social costs of air pollution are not included in firms’ profit maximization decisions and so are not considered when firms decide how much pollution to emit. The lack of a market for clean air causes problems and provides the impetus for government intervention in markets involving polluting industries.

Figure A.5 illustrates a negative externality associated with the production of a good. For example, a firm producing some product might also be generating pollution as a by-product. The pollution may impose significant costs — in the form of adverse health effects, for example — on households living downwind or downstream of the firm. Because those costs are not borne by the firm, the firm typically does not consider them in its production decisions. Society considers the pollution a cost of production, but the firm typically will not. In this figure:

- $D$ is the market demand (marginal benefit) curve for the product;
- $MPC$ is the firm’s marginal private real-resource cost of production, excluding the cost of the firm’s pollution on households;
- $MSD$ is the marginal social damage of pollution (or the marginal external cost) that the firm is not considering; and
- $MSC$ is society’s marginal social cost associated with production, including the cost of pollution ($MSC = MPC + MSD$).

In an incomplete market, producers pay no attention to external costs, and production occurs where market demand ($D$) and the marginal private real-resource cost ($MPC$) curves intersect — at a price $P_m$ and a quantity $Q_m$. In this case, net social welfare (total WTP minus total social costs) is equal to the area of the triangle $P_0P_m$X less the area of triangle $XYZ$. If the full social cost of production, including the cost of pollution, is taken into consideration, then the marginal cost of the required inputs — labor, raw materials, machinery, energy. Consumers purchase goods and services taking into account their income and their own tastes and preferences. This means that decisions are based on the private costs and private benefits to market participants. If the consumption or production of these goods and services poses an external cost or benefit on those not participating in the market, however, then the market demand and supply curves no longer reflect the true marginal social benefit and marginal social cost. Hence, the market equilibrium will no longer be the socially (Pareto) efficient outcome.

Externalities can arise for many reasons. Transactions costs or poorly defined property rights can make it difficult for injured parties to bargain or use legal means to ensure that the costs of the damages caused by polluters are internalized into their decision making. Activities that pose environmental risks may also be difficult to link to the resulting damages and often occur over long periods of time. Externalities involve goods that people care about but are not sold in markets. Air pollution causes ill health, ecological damage, and visibility impacts over a long time period, and the damage is often far from the source(s) of the pollution. The additional social costs of air pollution are not included in firms’ profit maximization decisions and so are not considered when firms decide how much pollution to emit. The lack of a market for clean air causes problems and provides the impetus for government intervention in markets involving polluting industries.

Figure A.5 illustrates a negative externality associated with the production of a good. For example, a firm producing some product might also be generating pollution as a by-product. The pollution may impose significant costs — in the form of adverse health effects, for example — on households living downwind or downstream of the firm. Because those costs are not borne by the firm, the firm typically does not consider them in its production decisions. Society considers the pollution a cost of production, but the firm typically will not. In this figure:

- $D$ is the market demand (marginal benefit) curve for the product;
- $MPC$ is the firm’s marginal private real-resource cost of production, excluding the cost of the firm’s pollution on households;
- $MSD$ is the marginal social damage of pollution (or the marginal external cost) that the firm is not considering; and
- $MSC$ is society’s marginal social cost associated with production, including the cost of pollution ($MSC = MPC + MSD$).

In an incomplete market, producers pay no attention to external costs, and production occurs where market demand ($D$) and the marginal private real-resource cost ($MPC$) curves intersect — at a price $P_m$ and a quantity $Q_m$. In this case, net social welfare (total WTP minus total social costs) is equal to the area of the triangle $P_0P_m$X less the area of triangle $XYZ$. If the full social cost of production, including the cost of pollution, is taken into consideration, then the marginal cost of the required inputs — labor, raw materials, machinery, energy. Consumers purchase goods and services taking into account their income and their own tastes and preferences. This means that decisions are based on the private costs and private benefits to market participants. If the consumption or production of these goods and services poses an external cost or benefit on those not participating in the market, however, then the market demand and supply curves no longer reflect the true marginal social benefit and marginal social cost. Hence, the market equilibrium will no longer be the socially (Pareto) efficient outcome.

Externalities can arise for many reasons. Transactions costs or poorly defined property rights can make it difficult for injured parties to bargain or use legal means to ensure that the costs of the damages caused by polluters are internalized into their decision making. Activities that pose environmental risks may also be difficult to link to the resulting damages and often occur over long periods of time. Externalities involve goods that people care about but are not sold in markets. Air pollution causes ill health, ecological damage, and visibility impacts over a long time period, and the damage is often far from the source(s) of the pollution. The additional social costs of air pollution are not included in firms’ profit maximization decisions and so are not considered when firms decide how much pollution to emit. The lack of a market for clean air causes problems and provides the impetus for government intervention in markets involving polluting industries.
cost curve should be increased by the amount of the marginal social damage (MSD) of pollution. Production will now occur where the demand and marginal social cost (MSC) curves intersect — at a price $P^*$ and a quantity $Q^*$. At this point net social welfare (now equal to the area of the triangle, $P_0P_1X$, alone) is maximized, and therefore the market is at the socially efficient point of production. This example shows that when there is a negative externality such as pollution, and the social damage (external cost) of that pollution is not taken into consideration, the producer will oversupply the polluting good. The shaded triangle $(XYZ)$, referred to as the deadweight loss (DWL), represents the amount that society loses by producing too much of the good.

### A.3 Benefit-Cost Analysis

If a negative externality such as pollution exists, an unregulated market will not account for its cost to society, and the result will be an inefficient outcome. In this case, there may be a need for government intervention to correct the market failure. A correction may take the form of dictating the allowable level of pollution or introducing a market mechanism to induce the optimal level of pollution. Figure A.5 neatly summarizes this in a single market diagram. To estimate the total costs and benefits to society of an activity or program, the costs and benefits in each affected market, as well as any non-market costs or benefits, are added up. This is done through BCA.

BCA can be thought of as an accounting framework of the overall social welfare of a program, which illuminates the trade-offs involved in making different social investments (Arrow et al. 1996). It is used to evaluate the favorable effects of a policy action and the associated opportunity costs. The favorable effects of a regulation are the benefits, and the foregone opportunities or losses in utility are the costs. Subtracting the total costs from the total monetized benefits provides an estimate of the regulation’s net benefits to society. An efficient regulation is one that yields the maximum net benefit, assuming that the benefits can be measured in monetary terms.

BCA can also be seen as a type of market test for environmental protection. In the private market, a commodity is supplied if the benefits that society gains from its provision, measured by what consumers are willing to pay, outweigh the private costs of producing the commodity. Economic efficiency is measured in a private market as the difference between what consumers are willing to pay for a good and what it costs to produce it. Since clean air and clean water are public goods, private suppliers cannot capture their value and sell it. The government determines their provision through environmental protection regulation. BCA quantifies the benefits and costs of producing this environmental protection in the same way as the private market, by quantifying the WTP for the environmental commodity. As with private markets, the efficient outcome is the option that maximizes net benefits.

---

475 When conducting BCA related to resource stocks, the MSD or marginal external cost is the present value of future net benefits that are lost due to the use of the resource at present. That is, exhaustible resources used today will not be available for future use. These foregone future benefits are called user costs in natural resource economics (see Scott 1953, 1955). The marginal user cost is the user cost of one additional unit consumed in the present and is added together with the marginal extraction cost to determine the MSC of resource use.

476 Similarly, the private market will undersupply goods for which there are positive externalities, such as parks and open space.

477 Chapter 4 discusses the various regulatory techniques and some non-regulatory means of achieving pollution control.
The key to performing BCA lies in the ability to measure both benefits and costs in monetary terms so that they are comparable. Consumers and producers in regulated industries and the governmental agencies responsible for implementing and enforcing the regulation (and by extension, taxpayers in general) typically pay the costs. The total cost of the regulation is found by summing the costs to these individual sectors. (An example of this, excluding the costs to the government, is given in Section A.4.3.) Since environmental regulation usually addresses some externality, the benefits of a regulation often occur outside of markets. For example, the primary benefits of drinking water regulations are improvements in human health. Once the expected reduction in illness and premature mortality associated with the regulation is calculated, economists use a number of techniques to estimate the value that society places on these health improvements. These monetized benefits can then be summed to obtain the total benefits from the regulation.

Note that in BCA gains and losses are weighted equally regardless of to whom they accrue. Evaluation of the fairness, or the equity, of the net gains cannot be made without specifying a social welfare function. However, there is no generally agreed-upon social welfare function and assigning relative weights to the utility of different individuals is an ethical matter that economists strive to avoid. Given this dilemma, economists have tried to develop criteria for comparing alternative allocations where there are winners and losers without involving explicit reference to a social welfare function. According to the Kaldor-Hicks compensation test, named after its originators Nicholas Kaldor and J.R. Hicks, a reallocation is a welfare-enhancing improvement to society if:

1. The winners could theoretically compensate the losers and still be better off; and
2. The losers could not, in turn, pay the winners to not have this reallocation and still be as well off as they would have been if it did occur (Perman et al. 2003).

While these conditions sound complex, they are met in practice by assessing the net benefits of a regulation through BCA. The policy that yields the highest positive net benefit is considered welfare enhancing according to the Kaldor-Hicks criterion. Note that the compensation test is stated in terms of potential compensation and does not solve the problem of evaluating the fairness of the distribution of well-being in society. Whether and how the beneficiaries of a regulation should compensate the losers involves a value judgment and is a separate decision for government to make.

Finally, BCA may not provide the only criterion used to decide if a regulation is in society’s best interest. There are often other, overriding considerations for promulgating regulation. Statutory instructions, political concerns, institutional and technical feasibility, enforceability, and sustainability are all important considerations in environmental regulation. In some cases a policy may be considered desirable even if the benefits to society do not outweigh its costs, particularly if there are ethical or equity concerns. There are also practical limitations to BCA. Most importantly, this type of analysis requires assigning monetized values to non-market benefits and costs. In practice it can be very difficult or even impossible to quantify gains and losses in monetary terms (e.g., the loss of a species, intangible effects). In general, however, economists believe that BCA provides a systematic framework for comparing the social costs and benefits of proposed regulations, and that it contributes useful information to the decision-making process about how scarce resources can be put to the best social use.

### A.4 Measuring Economic Impacts

#### A.4.1 Elasticities

The net change in social welfare brought about by a new environmental regulation is the sum of the negative effects (i.e., loss of producer and consumer surplus) and the positive effects (or social benefits) of the improved environmental quality. This is shown graphically for a single market in Figure A.5 above. The use of demand and supply curves highlights

---

478 Chapter 7 discusses a variety of methods economists use to value environmental improvements.

479 Chapter 9 addresses equity assessment and describes the methods available for examining the distributional effects of a regulation.

480 Kelman (1981) argues that it is even unethical to try to assign quantitative values to non-marketed benefits.
the importance of assessing how individuals will respond to changes in market conditions. The net benefits of a policy will depend on how responsively producers and consumers react to a change in price. Economists measure this responsiveness by the supply and demand elasticities.

The term “elasticity” refers to the sensitivity of one variable to changes in another variable. The price elasticity of demand (or supply) for a good or service is equal to the percentage change in the quantity demanded (or supplied) that would result from a 1 percent increase in the price of that good or service. For example, a price elasticity of demand for tuna equal to -1 means that a 1 percent increase in the price of tuna results in a 1 percent decrease in the quantity demanded. Changes are measured assuming all other things, such as incomes and tastes, remain constant. Demand and supply elasticities are rarely constant and often change depending on the quantity of the good consumed or produced. For example, according to the demand curve for tuna shown in Figure A.6, at a price of $1 per pound, a 10 percent increase in price would reduce quantity demanded by 2.5 percent (from 8 lbs to 7.8 lbs). At a price of $4 per pound, a 10 percent increase in price would result in a 40 percent decrease in quantity demanded (from 2 to 1.2 lbs). This implies that the price elasticity of demand is -0.25 when tuna costs $1/lb but -4 when the price is $4/lb. When calculating elasticities it is important realize where one is on the supply or demand curve, and the price or quantity should be stated when reporting an elasticity estimate.

Elasticities are important in measuring economic impacts because they determine how much of a price increase will be passed on to the consumer. For example, if a pollution control policy leads to an increase in the price of a good, multiplying the price increase by current quantity sold generally will not provide an accurate measure of impact of the policy. Some of the impact will take the form of higher prices for the consumer, but some of the impact will be a decrease in the quantity sold. The amount of the price increase that is passed on to consumers is determined by the elasticity of demand relative to supply (as well as existing price controls). “Elastic” demand (or supply) indicates that a small percentage increase in price results in a larger percentage decrease (increase) in quantity demanded (supplied). All else equal, an industry facing a relatively elastic demand is less likely to pass on costs to the consumer because increasing prices will result in reduced revenues. In determining the economic impacts of a rule, supply characteristics in the industries affected by a regulation can be as important as demand characteristics. For highly elastic supply curves relative to the demand curves, it is likely that cost increases or decreases will be passed on to consumers.

The many variables that affect the elasticity of demand include:

- The cost and availability of close substitutes;
- The percentage of income a consumer spends on the good;
- How necessary the good is for the consumer;
- The amount of time available to the consumer to locate substitutes;
- The expected future price of the good; and
- The level of aggregation used in the study to estimate the elasticity.

Demand (or supply) is said to be “elastic” if the absolute value of the price elasticity of demand (supply) is greater than one and “inelastic” if the absolute value of the elasticity is less than one. If a percentage change in price leads to an equal percentage change in quantity demanded (supplied) (i.e., if the absolute value of elasticity equals one), demand (supply) is “unit elastic.”
The availability of close substitutes is one of the most important factors that determine demand elasticity. A product with close substitutes at similar prices tends to have an elastic demand, because consumers can readily switch to substitutes rather than paying a higher price. Therefore, a company is less likely to be able to pass through costs if there are many close substitutes for its product. Narrowly defined markets (e.g., salmon) will have more elastic demands than broadly defined markets (e.g., food) since there are more substitutes for narrow goods.

Another factor that affects demand elasticities is whether the affected product represents a substantial or necessary portion of customers’ costs or budgets. Goods that account for a substantial portion of consumers’ budgets or disposable income tend to be relatively price elastic. This is because consumers are more aware of small changes in the price of expensive goods compared to small changes in the price of inexpensive goods, and therefore may be more likely to seek alternatives. A similar issue concerns the type of final good involved. Reductions in demand may be more likely to occur when prices increase for “luxuries” or optional purchases. If the good is a necessity item, the quantity demanded is unlikely to change drastically for a given change in price. Demand will be relatively inelastic.

Elasticities tend to increase over time, as firms and customers have more time to respond to changes in prices. Although a company may face an inelastic demand curve in the short run, it could experience greater losses in sales from a price increase in the long run. Over time customers begin to find substitutes or new substitutes are developed. However, temporary price changes may affect consumers’ decisions differently than permanent ones. The response of quantity demanded during a one-day sale, for example, will be much greater than the response of quantity demanded when prices are expected to decrease permanently. Finally, it is important to keep in mind that elasticities differ at the firm versus the industry level. It is not appropriate to use an industry-level elasticity to estimate the ability of only one firm to pass on compliance costs when its competitors are not subject to the same cost.

Characteristics of supply in the industries affected by a regulation can be as important as demand characteristics in determining the economic impacts of a rule. For relatively elastic supply curves, it is likely that cost increases or decreases will be passed on to consumers. The elasticity of supply depends, in part, on how quickly per unit costs rise as firms increase their output. Among the many variables that influence this rise in cost are:

- The cost and availability of close input substitutes;
- The amount of time available to adjust production to changing conditions;
- The degree of market concentration among producers;
- The expected future price of the product;
- The price of related inputs and related outputs; and
- The speed of technological advances in production that can lower costs.

Similar to the determinants of demand elasticity, the factors influencing the price elasticity of supply all relate to a firm’s degree of flexibility in adjusting production decisions in response to changing market conditions. The more easily a firm can adjust production levels, find input substitutes, or adopt new production technologies, the more elastic is supply. Supply elasticities tend to increase over time as firms have more opportunities to renegotiate contracts and change production technologies. When production takes time, the quantity supplied may be more responsive to expected future price changes than to current price changes.

Demand and supply elasticities are available for the aggregate output of final goods in most industries. They are usually published in journal articles on research pertaining to a particular industry. When such information is unavailable, as is

---

482 Another useful source of elasticity estimates is the recently developed EPA Elasticity Databank (U.S. EPA 2007d). In the absence of an encyclopedic “Book of Elasticities” the Elasticity Databank serves as a searchable database of elasticity parameters across a variety of types (i.e., demand and supply elasticities, substitution elasticities, income elasticities, and trade elasticities) and economic sectors/product markets. The database is populated with EPA-generated estimates used in Environmental Impact Assessment studies conducted by the Agency since 1990, as well as estimates found in
often the case for intermediate goods, elasticities may be quantitatively or qualitatively assessed.\textsuperscript{483} Econometric tools are frequently used to estimate supply and demand equations (thereby the elasticities) and the factors that influence them.

### A.4.2 Measuring the Welfare Effect of a Change in Environmental Goods

As introduced in Section A.1 changes in consumer surplus are measured by the trapezoidal region below the ordinary, or Marshallian, demand curve as price changes. This region reflects the benefit a consumer receives by being able to consume more of a good at a lower price. If the price of a good decreases, some of the consumer’s satisfaction comes from being able to consume more of a commodity when its price falls, but some of it comes from the fact that the lower price means that the consumer has more income to spend. However, the change in (Marshallian) consumer surplus only serves as a monetary measure of the welfare gain or loss experienced by the consumer under the strict assumption that the marginal utility of income is constant.\textsuperscript{484} This assumption is almost never true in reality. Luckily, there are alternative, less demanding monetary measures of consumer welfare that prove useful in treatments of BCA. Intuitively, these measures determine the size of payment that would be necessary to compensate the consumer for the price change. In other words, they estimate the consumer’s WTP for a price change.

As mentioned above, a price decline results in two effects on consumption. The change in relative prices will increase consumption of the cheaper good (the substitution effect), and consumption will be affected by the change in overall purchasing power (the income effect). A Marshallian demand curve reflects both substitution and income effects. Movements along it show how the quantity demanded changes as price changes (holding all other prices and income constant), so it reflects both the substitution and the income effects. The Hicksian (or “compensated”) demand curve, on the other hand, shows the relationship between quantity demanded of a commodity and its price, holding all other prices and utility (rather than income) constant. This is the correct measure of a consumer’s WTP for a price change. The Hicksian demand curve is constructed by adjusting income as the price changes so as to keep the consumer’s utility the same at each point on the curve. In this way, the income effect of a price change is eliminated and the substitution effect can be considered alone. Movements along the Hicksian demand function can be used to determine the monetary change that would compensate the consumer for the price change.

Hicks (1941) developed two correct monetary measures of utility change associated with a price change: compensating variation and equivalent variation. \textit{Compensating variation} (CV) assesses how much money must be taken away from consumers after a price decrease occurred to return them to the original utility level. It is equal to the amount of money that would ‘compensate’ the consumer for the price decrease. \textit{Equivalent variation} (EV) measures how much money would need to be given to the consumer to bring her to the higher utility level instead of introducing the price change. In other words, it is the monetary change that would be ‘equivalent’ to the proposed price change.

Before examining the implications of these measures for valuing environmental changes, it is useful to understand CV and EV in the case of a reduction in the price of some normal, private good, \( C_1 \).\textsuperscript{485} This is shown with indifference curves and a budget line, as seen in Figure A.7.

\textsuperscript{483} Final goods are those that are available for direct use by consumers and are not utilized as inputs by firms in the process of production. Goods that contribute to the production of a final good are called intermediate goods. It is of course possible for a good to be final from one perspective and intermediate from another (Pearce 1992).

\textsuperscript{484} See Perman et al. (2003), Just et al. (2005) or any graduate level text for a more thorough exposition of this issue.

\textsuperscript{485} The notation and discussion in this section follow Chapter 12 of Perman et al. (2003).
Assume that the consumer is considering the trade-off between $C_1$ and all other goods, denoted by a composite good, $C_2$. The indifference curve, $U_0$, depicts the different combinations of the two goods that yield the same level of utility. Because of diminishing marginal utility, the curve is concave, where increasing amounts of $C_1$ must be offered for each unit of $C_2$ given up to keep the consumer indifferent. The budget line on the graph reflects what the consumer is able to purchase given her income, $Y_0$, and the prices of the two goods — $P_1'$ and $P_2'$, respectively. A utility-maximizing consumer will choose quantities $C_1'$ and $C_2'$, the point where the indifference curve is tangent to the budget constraint.

Figure A.8 shows the change in the optimal consumption bundle resulting from a reduction in the price of $C_1$. If the price of $C_1$ falls, the budget line shifts out on the $C_1$ axis because more $C_1$ can be purchased for a given amount of money. The consumer now chooses $C_1''$ and $C_2''$ at point $b$ and moves to a new, higher utility curve, $U_1$. CV then measures how much money must be taken away at the new prices to return the consumer to the old utility level. That is, starting at point $b$ and keeping the slope of the budget line fixed at the new level, by how much must it be shifted downward to make it tangent to the initial indifference curve, $U_0$? It is, therefore, the maximum amount the consumer would be willing to pay to have the price fall occur — i.e., the precise monetary measure of the welfare change. In Figure A.8, CV is simply given by the amount $Y_0 - Y_1$. EV, on the other hand, measures how much income must be given to the individual at the old price set to maintain the same level of well-being as if the price change did occur. That is, keeping the slope of the budget line fixed at the old level, by how much must it be shifted upwards to make it tangent to $U_1$? EV is, then, the minimum amount of money the consumer would accept in lieu of the price fall. This too is a proper monetary measure of the utility change resulting from the price decrease. In Figure A.8 then EV is the amount $Y_2 - Y_0$, leaving the individual at point $f$.

CV and EV are simply measures of the distance between the two indifference curves. However, the amount of money associated with CV, EV, and Marshallian consumer surplus (MCS) is generally not the same. For a price fall, it can be

---

486 In Figure A.7, $C_2$ is considered the numeraire good (i.e., prices are adjusted so that $P_2'$ is equal to 1).

487 For a review of the utility maximizing behavior of consumers, see any general microeconomics textbook.

488 In Figure A.8, this would result in a shift from $C_1'$ to $C_1''$. This is known as the income effect of the price change. The shift from $C_1'$ to $C_1''$ is considered the substitution effect.
shown that $CV < MCS < EV$, and for a price increase, $CV > MCS > EV$.\footnote{This can be seen by redrawing Figure A.8 using a graph of Marshallian and Hicksian demand curves. See Perman et al. (2003) for a detailed explanation.} Notice that in the case of a price decrease, the CV measures the consumer’s willingness to pay (WTP) to receive the price reduction and EV measures the consumer’s willingness to accept (WTA) to forgo the lower price. If the price of $C_1$ were to increase, then the relationships between WTP/WTA and CV/EV would be reversed. CV would measure the consumer’s WTA to suffer the price increase and EV would be the individual’s WTP to avoid the increase in price.

In order to examine the implications of these measures for valuing changes in environmental conditions, one can think of $C_1$ in the above discussion as an environmental commodity, henceforth denoted by $E$. Then an improvement in environmental quality (or an increase in an environmental public good) resulting from some policy is reflected by an increase in the amount of $E$. Holding all else constant, such an increase is equivalent to a decrease in the price of $E$ and can be depicted as a shifting outward of the budget line along the $E$ axis.

Welfare changes due to an increase in $E$ follow along the lines of the previous discussion. However, because $E$ is generally non-exclusive and non-divisible, the consumer consumption level cannot be adjusted. Therefore, the associated monetary measures of the welfare change are not technically CV and EV, but are referred to as compensating surplus (CS) and equivalent surplus (ES). In practice, however, the process is the same; a Hicksian demand curve is estimated for the unpriced environmental good. Analogous to the preceding discussion, if there is an environmental improvement, then CS measures the amount of money the consumer would be willing to pay for the improvement that would result in the pre-improvement level of utility. For the purposes of environmental valuation, this is the primary measure of concern when considering environmental improvements. ES measures how much society would have to pay the consumer to give him the same utility as if the improvement had occurred. In other words, this is how much he would be willing to accept to not experience the gain in environmental quality. If valuing an environmental degradation, then CS measures the WTA and ES measures WTP.

Whereas statements can be made about the relative size of CV, EV, and MCS for price changes of normal goods, Bockstael and McConnell (1993) find that it is not possible to make similar statements about CS, ES, and MCS for a change in environmental quality.\footnote{Willig (1976) shows that ordinary, or Marshallian, demand curves can provide an approximate measure of welfare changes resulting from a price change. In most cases the error associated with using MCS, with respect to CV or EV, will be less than 5 percent (see Perman et al. 2003).} Given that environmental quality is generally an unpriced public good, ordinary Marshallian demand functions cannot be estimated, so it may seem irrelevant that one cannot say anything about how MCS approximates the proper measure. However, Bockstael and McConnell’s results are important in relation to indirect methods for environmental valuation. However, most indirect valuation studies are based on Marshallian demand functions in practice, in the hope of keeping the associated error small.

**A.4.3 Single Market, Multi-Market, and General Equilibrium Analysis**

Both supply and demand elasticities are affected by the availability of close complements and substitutes. This highlights the fact that regulating one industry can have an impact on other, non-regulated markets. However, this does not necessarily imply that all of these other markets must be modeled. Changes due to government regulation can be captured using only the equilibrium supply and demand curves for the affected market, assuming: (1) there are small, competitive adjustments in all other markets; and (2) there are no distortions in other markets. This is referred to as partial equilibrium analysis.

For example, suppose a new environmental regulation increases per unit production costs. The benefits and costs of abatement in a partial equilibrium setting are illustrated in Figure A.9 where the market produces the quantity $Q_m$ in equilibrium without intervention. The external costs of production are shown by the marginal external costs (MEC) curve.
Figure A.9 - Benefits and Costs of Abatement

without any abatement. Total external costs are given by the area under the MEC curve up to the market output, $Q_m$, or the area of triangle $Q_mE0$.

With required abatement production, costs are the total of supply plus marginal abatement costs (MAC), shown as the new, higher supply curve in the figure. These higher costs result in a new market equilibrium quantity shown as $Q^*$. The social cost of the requirement is the resulting change in consumer and supplier surplus, shown here as the total observed abatement costs (parallelogram $P_0P_1AC$) plus the area of triangle $ABC$, which can be described as deadweight loss.

Abatement also produces benefits by shifting the MEC curve downward, reflecting the fact that each unit of production now results in less pollution and social costs. Additionally, the reduced quantity of the output good results in reduced external costs. The reduced external costs, i.e., the benefits, are given by the difference between triangle $Q_mE0$ and triangle $Q^*D0$, represented by the shaded area in the figure.

The net benefits of abatement are the benefits (the reduced external costs) minus the costs (the loss in consumer and producer surplus). In the figure this would equal the shaded area (the benefits) minus total abatement costs and deadweight loss as described above.

While the single market analysis is theoretically possible, it is generally impractical for rulemaking. As mentioned in Section A.3, this is often because the gains occur outside of markets and cannot be linked directly to the output of the regulated market. Therefore BCA is frequently done as two separate analyses: a benefits analysis and a cost analysis.

When a regulation is expected to have a large impact outside of the regulated market, then the analysis should be extended beyond that market. If the effects are significant but not anticipated to be widespread, one potential improvement is to use multi-market modeling in which vertically or horizontally integrated markets are incorporated into the analysis. The analysis begins with the relationship of input markets to output markets. A multi-market analysis extends the partial equilibrium analysis to measuring the losses in other related markets.491

In some cases, a regulation can have such a significant impact on the economy that a general equilibrium modeling framework is required.492 This may be because regulation in one industry has broad indirect effects on other sectors, households may alter their consumption patterns when they encounter increases in the price of a regulated good, or there may be interaction effects between the new regulation and pre-existing distortions, such as taxes on labor. In these cases, partial equilibrium analyses are likely to result in an inaccurate estimation of total social costs. Using a general equilibrium framework accounts for linkages between all sectors of the economy and all feedback effects, and can measure total costs comprehensively.493

A.5 Optimal Level of Regulation

Following from the definition in Section A.1, the most economically efficient policy is the one that allows for society to derive the largest possible social benefit at the lowest social cost. This occurs when the net benefits to society (i.e., total...

---

491 An example of the use of multi-market model for environmental policy analysis is contained in a report prepared for EPA on the regulatory impact of control on asbestos and asbestos products (U.S. EPA 1989).

492 General equilibrium analysis is built around the assumption that, for some discrete period of time, an economy can be characterized by a set of equilibrium conditions in which supply equals demand in all markets. When this equilibrium is “shocked” through a change in policy or a change in some exogenous variable, prices and quantities adjust until a new equilibrium is reached. The prices and quantities from the post-shock equilibrium can then be compared with their pre-shock values to determine the expected impacts of the policy or change in exogenous variables.

493 Chapter 8 provides a more detailed discussion of partial equilibrium, multi-market, and general equilibrium analysis.
Conceptually, net social benefits will be maximized if regulation is set such that emissions are reduced up to the point where the benefit of abating one more unit of pollution (i.e., marginal social benefit) is equal to the cost of abating an additional unit (i.e., marginal abatement cost). If the marginal benefits are greater than the marginal costs, then additional reductions in pollution will offer greater benefits than costs, and society will be better off. If the marginal benefits are less than marginal costs, then additional reductions in pollution will cost society more than they provide in benefits, and will

If the regulation were designed to maximize benefits, then any policy, no matter how expensive, would be justified if it produced any benefit, no matter how small. Similarly, minimizing costs would, in most cases, simply justify no action at all. A benefit-cost ratio equal to one is equivalent to saying that the benefits to society would be exactly offset by the cost of implementing the policy. This implies that society is indifferent between no regulation and being regulated; hence, there would be no net benefit from adopting the policy. Maximizing the benefit-cost ratio is not optimal either. Two policy options could yield equivalent benefit-cost ratios but have vastly different net benefits. For example, a policy that cost $100 million per year but produced $200 million in benefits has the same benefit-cost ratio as a policy that cost $100,000 but produced $200,000 in benefits, even though the first policy produces substantially more net benefit for society. Finally, finding the most cost-effective policy has similar problems because the cost-effectiveness ratio can be seen as the inverse of the benefit-cost ratio. A policy is cost effective if it meets a given goal at least cost — i.e., minimizes the cost per unit of benefit achieved. Cost-effectiveness analysis (CEA) can provide useful information to supplement existing BCA and may be appropriate to rank policy options when the benefits are fixed and cannot be monetized, but it provides no guidance in setting an environmental standard or goal.

If the benefits of pollution reduction are the reduced damages from being exposed to pollution. Therefore, the marginal social benefit of abatement is measured as the additional reduction in damages from abating one more unit of pollution.

The benefits of pollution reduction are the reduced damages from being exposed to pollution. Therefore, the marginal social benefit of abatement is measured as the additional reduction in damages from abating one more unit of pollution.

494 Benefit-cost ratios are useful when choosing one or more policy options subject to a budget constraint. For example, consider a case where five options are available and the budget is $1,000. The first option will cost $1,000 and will deliver benefits of $2,000. Each of the other four will cost $250 and deliver benefits of $750. If options are selected according to the net benefits criterion, the first option will be selected, because its net benefits are $1,000, while the net benefits of each of the other options are $500. However, if options are selected by the benefit-cost ratio criterion, the other four options will be selected, as each of their benefit-cost ratios equal 3, versus a benefit-cost ratio of 2 for the first option. In this case, choosing options by the benefit-cost ratio will yield $1,000 in total net benefits, while choosing options by the benefit-cost ratio will yield $1,000 in total net benefits. In most cases, choosing options in decreasing order of benefit-cost ratios will yield the largest possible net benefits given a fixed budget. This method will guarantee the optimal solution if the benefits and costs of each option are independent, and if each option can be infinitely subdivided: simply select the options in decreasing order of their benefit-cost ratios and once the budget is exceeded, subdivide the last option selected such that the budget constraint is met exactly (see Dantzig 1957). Also note that this strategy does not require measuring benefits and costs in the same units, which means that it is directly useful for CEA (Hyman and Leibowitz 2000), while the net-benefit criterion is not.

495 The idea that a given level of abatement is efficient — as opposed to abating until pollution is equal to zero — is based on the economic concept of diminishing returns. For each additional unit of abatement, marginal social benefits decrease while marginal social costs of that abatement increase. Thus, it only makes sense to continue to increase abatement until the point where marginal abatement benefits and marginal costs are just equal. Any abatement beyond that point will incur more additional costs than benefits. (Alternatively, one can understand the efficient level of abatement as the amount of regulation that achieves the efficient level of pollution. If one considers a market for pollution, the socially-efficient outcome would be the point where the marginal WTP for pollution equals the marginal social cost of polluting.)
make society worse off. When the marginal cost of abatement is equal to society’s marginal benefit, no gains can be made from changing the level of pollution reduction, and an efficient aggregate level of emissions is achieved. In other words, a pollution reduction policy is at its optimal, most economically efficient point when the marginal benefits equal the marginal costs of the rule.\(^{497}\)

The condition that marginal benefits must equal marginal costs assumes that the initial pollution reduction produces the largest benefits for the lowest costs. As pollution reduction is increased (i.e., regulatory stringency is increased), the additional benefits decline and the additional costs rise. While it is not always true, a case can be made that the benefits of pollution reduction follow this behavior. The behavior of total abatement costs, however, will depend on how the pollution reduction is distributed among the polluters since firms may differ in their ability to reduce emissions. The aggregate marginal abatement cost function shows the least costly way of achieving reductions in emissions. It is equal to the horizontal sum of the marginal abatement cost curves for the individual polluters. Although each firm faces increasing costs of abatement, marginal cost functions still vary across sources. Some firms may abate pollution relatively cheaply, while others require great expense. To achieve economic efficiency, the lowest marginal cost of abatement must be achieved first, and then the next lowest. Pollution reduction is achieved at lowest cost only if firms are required to make equiproportionate cutbacks in emissions. That is, at the optimal level of regulation, the cost of abating one more unit of pollution is equal across all polluters.\(^{498}\)

Figure A.11 illustrates why the level of pollution that sets the marginal benefits and marginal costs of abatement equal to each other is efficient.\(^{499}\) Emissions are drawn on the horizontal axis and increase from left to right. The damages from emissions are represented by the marginal damage (MD) curve. Damages may include the costs of worsened human health, reduced visibility, lower property values, and loss of crop yields or biodiversity. As emissions rise, the marginal damages increase. \(E_j\) represents the amount of emissions in the absence of regulation on firms. The costs of controlling emissions are represented by the marginal abatement cost curve (MAC). As emissions are reduced below \(E_j\), the marginal cost of abatement rises.

The total damages associated with emissions level \(E^*\) are represented by the area of the triangle \(AE_0E^*\), while the total abatement costs are represented by area \(AE_1E^*.\) The total burden on society of this level is equal to the total abatement costs of reducing emissions from \(E_1\) to \(E^*\) plus the total damages of the remaining emissions, \(E^*.\) That is, the total burden is the darkly shaded triangle, \(E_0AE_1.\)

Now assume that emissions are something other than \(E^*.\) For example, suppose emissions were \(E_0\), which is greater than \(E^*.\) Total damages for this level of emissions are equal to the area of the triangle \(BE_0E_0\) while total costs of abatement to

---

497 It is important to reemphasize the word “marginal” in this statement. Marginal, in economic parlance, means the extra or next unit of the item being measured. If regulatory options could be ranked in order of regulatory stringency, then marginal benefits equal to marginal costs means that the additional benefits of increasing the regulation to the next degree of stringency is equal to the additional cost of that change.

498 Thus a regulation that requires all firms to achieve the same level of reduction will probably result in different marginal costs for each firm and not be efficient. (See Field and Field 2005 or any other environmental economics text for a detailed explanation and example.)

499 Figure A.11 illustrates the simplest possible case, where the pollutant is a flow (i.e., it does not accumulate over time) and marginal damages are independent of location. When pollution levels and damages vary by location, then the efficient level of pollution is reached when marginal abatement costs adjusted by individual transfer coefficients are equal across all polluters. Temporal variability also implies an adjustment to this equilibrium condition. In the case of a stock pollutant, marginal abatement costs are equal across the discounted sum of damages from today’s emissions in all future time periods. In the case of a flow pollutant, this condition should be adjusted to reflect seasonal or daily variations (see Sterner 2003).
this level is equal to the area $CE_1E_x$. The total burden on society of this level is the sum of the areas of the darkly shaded and the lightly shaded triangles. This means that the excess social cost of choosing emissions $E_x$ rather than $E^*$ is equal to the area of the lightly shaded triangle, $ABC$. A similar analysis could be done if emissions levels were below level, $E^*$. Here, the additional abatement costs would be greater than the decrease in damages, resulting in excess social costs.

The policy that sets the emissions level at $E^*$ — at the point where marginal benefits of pollution reduction (represented by the MD curve) and the MAC curve intersect — is economically efficient because it imposes the least net cost on, and yields the highest net benefits for, society. That is, the triangle $E_1AE_x$ is the smallest shaded region that can be obtained.

This section has focused on first-best optimal regulation when there are no pre-existing market distortions. However, it is important to note that realizable policy outcomes will often be “second best” due to information constraints, political constraints, imperfect competition, and market distortions created by tax and other government interventions. For example, many of the emissions-based policies emphasized in these Guidelines may be less feasible for addressing nonpoint source pollution, such as agriculture, which is less observable and more stochastic than emissions from point sources. Agriculture is also subject to multiple non-environmental policy distortions that must be considered in the measurement of the social benefits and costs of regulating agriculture.

A.6 Conclusion

The purpose of this appendix is to present a brief explanation of some of the fundamental economics relevant to Chapters 3 through 9. It is not intended to provide a comprehensive discussion of all microeconomic theory and its application to environmental issues. The interested reader can turn to undergraduate or graduate level textbooks for a more thorough exposition of the topics covered here. At the undergraduate level, Field and Field (2005) provide an introduction to the basic principles of environmental economics. Tietenberg’s (2002) and Perman et al.’s (2003) presentations are more technical but still used primarily for undergraduate courses. Freeman (2003) is the standard text for graduate courses in environmental economics and deals with the methodology of non-market valuation. Supplemental texts that provide a good handle on environmental economics with less technical detail include Stavins (2000a), and Portney and Stavins (2000). Finally, general microeconomics textbooks (Mankiw 2004, and Varian 2005 at the undergraduate level; and Mas-Colell et al. 1995, Kreps 1990, and Varian 2005 at the graduate level), and applied welfare economics textbooks (Just et al. 2005) are useful references as well.

Appendix A References


Appendix B

Mortality Risk Valuation Estimates

Some EPA policies are designed to reduce the risk of contracting a potentially fatal health effect such as cancer. Reducing these risks of premature death provides welfare increases to those individuals affected by the policy. These policies generally provide marginal changes in relatively small risks. That is, these policies do not provide assurance that an individual will not die prematurely from environmental exposures; rather, they marginally reduce the probability of such an event. For BCA, analysts generally aggregate these small risks over the affected population to derive the number of statistical lives saved (or the number of statistical deaths avoided) and then use a “value of statistical life” (VSL) to express these benefits in monetary terms.

The risk reductions themselves can generally be classified according to the characteristics of the risk in question (e.g., voluntariness or controllability) and the characteristics of the affected population (e.g., age and health status). These dimensions may affect the value of reducing mortality risks. Ideally the VSL would account for all possible risk and demographic characteristics that matter. It would be derived from the preferences of the population affected by the policy, based on the type of risk that the policy is expected to reduce. For example, if a policy were designed to remove carcinogens at a suburban hazardous waste site, the ideal measure would represent the preferences for reduced cancer risks for the exposed population in the area and would reflect the changes in life expectancy that would result. Unfortunately, time and resource constraints make it difficult if not impossible to obtain such unique valuation estimates for each EPA policy. Instead, analysts need to draw from existing VSL estimates obtained using well-established methods (see Chapter 7).

This appendix describes the default VSL estimate currently used by the Agency and its derivation, as well as how analysts should characterize and assess benefit transfer issues that may arise in its application. Benefit transfer considerations that are common to all valuation applications, including the effect of most demographic characteristics of the study and policy populations, are described in Chapter 7 Section 7.3 and will not be repeated here.

B.1 Central Estimate of VSL

Table B.1 contains the VSL estimates that currently form the basis of the Agency’s recommended central VSL estimate. Fitting a Weibull distribution to these estimates yields a central estimate (mean) of $7.4 million ($2006) with a standard
Table B.1 - Value of Statistical Life Estimates (mean values in millions of 2006 dollars)

<table>
<thead>
<tr>
<th>Study</th>
<th>Method</th>
<th>Value of Statistical Life</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kniesner and Leeth (1991 - US)</td>
<td>Labor Market</td>
<td>$0.85</td>
</tr>
<tr>
<td>Smith and Gilbert (1984)</td>
<td>Labor Market</td>
<td>$0.97</td>
</tr>
<tr>
<td>Dillingham (1985)</td>
<td>Labor Market</td>
<td>$1.34</td>
</tr>
<tr>
<td>Butler (1983)</td>
<td>Labor Market</td>
<td>$1.58</td>
</tr>
<tr>
<td>Miller and Guria (1991)</td>
<td>Contingent Valuation</td>
<td>$1.82</td>
</tr>
<tr>
<td>Moore and Viscusi (1988)</td>
<td>Labor Market</td>
<td>$3.64</td>
</tr>
<tr>
<td>Viscusi, Magat, and Huber (1991)</td>
<td>Contingent Valuation</td>
<td>$4.01</td>
</tr>
<tr>
<td>Gegax et al. (1985)</td>
<td>Contingent Valuation</td>
<td>$4.86</td>
</tr>
<tr>
<td>Kniesner and Leeth (1991 - Australia)</td>
<td>Labor Market</td>
<td>$4.86</td>
</tr>
<tr>
<td>Cousineau, Lecroix, and Girard (1988)</td>
<td>Labor Market</td>
<td>$5.34</td>
</tr>
<tr>
<td>Jones-Lee (1989)</td>
<td>Contingent Valuation</td>
<td>$5.59</td>
</tr>
<tr>
<td>Dillingham (1985)</td>
<td>Labor Market</td>
<td>$5.71</td>
</tr>
<tr>
<td>Viscusi (1978)</td>
<td>Labor Market</td>
<td>$6.07</td>
</tr>
<tr>
<td>R.S. Smith (1976)</td>
<td>Labor Market</td>
<td>$6.80</td>
</tr>
<tr>
<td>Olson (1981)</td>
<td>Labor Market</td>
<td>$7.65</td>
</tr>
<tr>
<td>Kniesner and Leeth (1991 - Japan)</td>
<td>Labor Market</td>
<td>$11.18</td>
</tr>
<tr>
<td>Leigh (1987)</td>
<td>Labor Market</td>
<td>$15.31</td>
</tr>
</tbody>
</table>

benefit transfer exercise is to account for all of the factors that significantly affect the value of mortality risk reductions between the policy and study scenarios. Policy analysts valuing mortality risk reductions should account for differences in risk and population characteristics. Until updated guidance is available, the SAB should be applied in relevant analyses while the Agency continues its efforts to update its guidance on this issue.

B.2 Other VSL Information

For most of mortality risk reductions EPA uniformly applies the VSL estimate discussed above. For a period of time (2004-2008), the Office of Air and Radiation (OAR) valued mortality risk reductions using a VSL estimate derived from a limited analysis of some of the available studies. OAR arrived at a VSL using a range of $1 million to $10 million (2000$) consistent with two meta-analyses of the wage-risk literature. The $1 million value represented the lower end of the interquartile range from the Mrozek and Taylor (2002) meta-analysis of 33 studies. The $10 million value represented the upper end of the interquartile range from the Viscusi and Aldy (2003) meta-analysis of 43 studies. The mean estimate of $5.5 million (2000$) was also consistent with the mean VSL of $5.4 million estimated in the Kochi et al. (2006) meta-analysis. However, the Agency neither changed its official guidance on the use of VSL in rulemakings nor subjected the interim estimate to a scientific peer-review process through the Science Advisory Board (SAB) or other peer-review group.

During this time, the Agency continued work to update its guidance on valuing mortality risk reductions. EPA commissioned a report from meta-analytic experts to evaluate methodological questions raised by EPA and the SAB on combining estimates from the various data sources. In addition, the Agency consulted several times with the SAB Environmental Economics Advisory Committee (SAB-EEAC) on the issue. With input from the meta-analytic experts, the SAB-EEAC advised the Agency to update its guidance using specific, appropriate meta-analytic techniques to combine estimates from unique data sources and different studies, including those using different methodologies such as wage-risk and stated preference (U.S. EPA 2007g).

Until updated guidance is available, the Agency determined that a single, peer-reviewed estimate applied consistently best reflects the SAB-EEAC advice received to date. Therefore, the VSL described above that was vetted and endorsed by the SAB should be applied in relevant analyses while the Agency continues its efforts to update its guidance on this issue.

B.3 Benefit Transfer Considerations

Policy analysts valuing mortality risk reductions should account for differences in risk and population characteristics between the policy and study scenarios and their potential effect on the overall results. The ultimate objective of the benefit transfer exercise is to account for all of the factors that significantly affect the value of mortality risk reduction in

---

500 The VSL was updated from the $4.8 million ($1990) estimate referenced in the 2000 Guidelines by adjusting the individual study estimates for inflation using a GDP deflator and then fitting a Weibull distribution to the estimates. The updated Weibull parameters are: location = 0, scale = 7.75, shape = 1.51 (updated from location = 0, scale = 5.32, shape = 1.51). The Weibull distribution was determined to provide the best fit for this set of estimates. See U.S. EPA 1997a for more details.

501 This VSL estimate was produced using the GDP deflator inflation index. Some economists prefer using the Consumer Price Index (CPI) in some applications. The key issue for EPA analysts is to ensure that the chosen index is used consistently throughout the analysis.

502 The studies listed in Table B.1 were published between 1974 and 1991, and most are hedonic wage estimates that may be subject to considerable measurement error (Black et al. 2003, and Black and Knesner 2003). Although these were the best available data at the time, they are sufficiently dated and may rely on obsolete preferences for risk and income. The Agency is currently considering more recent studies as it evaluates approaches to revise its guidance.

503 EPA is in the process of revisiting this guidance and has recently engaged the SAB-EEAC on several issues including the use of meta-analysis as a means of combining estimates and approaches for assessing mortality benefits when changes in longevity may vary widely (U.S. EPA 2006d). The Agency is committed to using the best available science in its analyses and will revise this guidance in response to SAB recommendations (see U.S. EPA 2007g for recent SAB recommendations).
the context of the policy. Analysts should carefully consider the implications of correcting for some relevant factors, but not for others, recognizing that it may not be feasible to account for all factors.

**B.4 Adjustments Associated with Risk Characteristics**

Risk characteristics appear to affect the value that people place on risk reduction. A large body of work identifies eight dimensions of risk that affect human risk perception:

- voluntary/involuntary
- ordinary/catastrophic
- delayed/immediate
- natural/man-made
- old/new
- controllable/uncontrollable
- necessary/unnecessary
- occasional/continuous

Transferring VSL estimates among these categories may introduce bias. There have been some recent efforts attempting to quantitatively assess these sources of bias. These studies generally conclude that voluntariness, control and responsibility affect individual values for safety, although there is no consensus on the direction and magnitude of these effects.

In addition, environmental risks may differ from those that form the basis of VSL estimates in many of these dimensions. Occupational risks, for example, are generally considered to be more voluntary in nature than are environmental risks, and may be more controllable. As part of the Agency’s review of our mortality risk guidance we are evaluating the literature from which the studies are drawn.

Support for quantitative adjustments in the empirical literature is lacking for most of these factors. The SAB reviewed an Agency summary of the available empirical literature on the effects of risk and population characteristics on WTP for mortality risk reductions (U.S. EPA 2000d). The SAB review concludes that among the demographic and risk factors that might affect VSL estimates, the current literature can only support empirical adjustments related to the timing of the risk. The review supports making the following adjustments to primary benefits estimates: (1) adjusting WTP estimates to account for higher future income levels, though not for cross-sectional differences in income; and (2) discounting risk reductions that are brought about in the future by current policy initiatives (that is, after a cessation lag), using the same rates used to discount other future benefits and costs. All other adjustments, if made, should be relegated to sensitivity analyses.

**Increases in income over time.** The economics literature shows that the income elasticity of WTP to reduce mortality risk is positive, based on cross-sectional data. As a result, benefits estimates of reduced mortality risk accruing in future years may be adjusted to reflect anticipated income growth, using the range of income elasticities (0.08, 0.40 and 1.0) employed in *The Benefits and Costs of the Clean Air Act, 1990-2010*. Recent EPA analyses have assumed a triangular

---


506 For details see Kleckner and Neuman (2000).
distribution from these values and used the results in a probabilistic assessment of benefits. At the time of this writing, EPA is engaged in a consultation with the SAB-EEAC on the appropriate range of income elasticities and will update this guidance as needed.

Timing of reduced exposure and reduced risk. Many environmental policies are targeted at reducing the risk of effects such as cancer, where there may be an extended period of time between the reduced exposure and the reduction in the risk of death from the disease. This delay between the change in exposure and realization of the reduced risk may affect the value of that risk reduction. Most existing VSL estimates are based on risks of relatively immediate fatalities making them an imperfect fit for a benefits analysis of many environmental policies. Economic theory suggests that reducing the risk of a delayed health effect will be valued less than reducing the risk of a more immediate one, when controlling for other factors.

B.5 Effects on WTP Associated with Demographic Characteristics

Two population characteristics are particularly noteworthy for their potential effect on mortality risk valuation estimates: age and health status of the exposed population. In September 2006, the Agency requested an additional advisory from the SAB-EEAC on issues related to valuing changes in life expectancy for which age and baseline health status are close correlates. Because the outcome of this review is not yet available, we focus here on previous advice received from the SAB on related questions.

Age. It has sometimes been posited that older individuals should have a lower WTP for changes in mortality risk given the fewer years of life expectancy remaining compared to younger individuals. This hypothesis may be confounded, however, by the finding that older persons reveal a greater demand for reducing mortality risks and hence have a greater implicit value of a life year (Ehrlich and Chuma 1990). Several authors have attempted to explore potential differences in mortality risk valuation estimates associated with differences in the average age of the affected population using theoretical models of life-cycle consumption. In general this literature has shown that the relationship between age and WTP for mortality risk changes is ambiguous, requiring strong assumptions to even sign the relationship. Empirical evidence is also mixed. A number of empirical studies (mostly hedonic wage studies) suggest that the VSL follows a consistent “inverted-U” life-cycle, peaking in the region of mean age. Others find no such statistically significant relationship and still others show WTP increasing with age. Stated preference results are also mixed, with some studies showing declining WTP for older age groups and others finding no statistically significant relationship between age and WTP.

In spite of the ambiguous relationship between age and WTP, two alternative adjustment techniques have been derived from this literature. The first technique, value of statistical life-years (VSLY), is derived by dividing the

507 See, for example, pp. 6-84 of the Final Economic Analysis for the Stage 2 DBPR (U.S. EPA 2005a).
508 Although latency is defined here as the time between exposure and fatality from illness, alternative definitions may be used in other contexts. For example “latency” may refer to the time between exposure and the onset of symptoms. These symptoms may be experienced for an extended period of time before ultimately resulting in fatality.
509 U.S. EPA (2006d) summarizes much of the literature related to the effects of age and health status on WTP for changes in mortality risk and includes the charge questions put to the SAB-EEAC on these issues.
510 See, for example, Shepard and Zeckhauser (1982), Rosen (1988), Cropper and Sussman (1988, 1990), and Johannson (2002).
511 See Evans and Smith (2006) for a recent summary.
513 Viscusi and Aldy (2003) review more than 60 studies of mortality risk estimates from 10 countries and discuss eight hedonic wage studies that explicitly examine the age-WTP relationship. Only five of the eight studies found a statistically significant, negative relationship between age and the return to risk. Smith et al. (2004) and Kniesner et al. (2006) find that WTP increases with age.
514 Krupnick et al. (2002) report that WTP for mortality risk reductions changes significantly with age after age 70. Alberini et al. (2004) find no difference in the WTP for younger age groups and find a 20 percent reduction for those aged 70 and older. However, this difference was not statistically significant.
estimated VSL by expected remaining life expectancy. This is by far the most common approach and presumes that: (1) the VSL equals the sum of discounted values for each life year; and (2) each life year has the same value. This method was applied as an alternative case in an effort to evaluate the sensitivity of the benefits estimates prepared for EPA’s retrospective and prospective studies of the costs and benefits of the Clean Air Act (U.S. EPA 1997a, and U.S. EPA 1999).

A second technique is to apply a distinct value or suite of values for mortality risk reduction depending on the age of incidence. However, there is relatively little available literature upon which to base such adjustments.515 Neither approach enjoys general acceptance in the literature as they both require large assumptions to be made, some of which have been contradicted in empirical studies. Since published support is lacking, neither approach is recommended at this time.

Analysts are advised to note the age distribution of the affected population when possible, especially when children are found to be a significant portion of the affected population.516 Although the literature on the valuation of children’s health risks is growing, there is still not enough information currently to derive age-specific valuation estimates.

**Health status.** Individual health status may also affect WTP for mortality risk reduction. This is an especially relevant factor for valuation of environmental risks because individuals with impaired health are often the most vulnerable to mortality risks from environmental causes. For example, particulate air pollution appears to disproportionately affect individuals in an already impaired state of health. Health status is distinct from age (a “quality versus quantity” distinction) but the two factors are clearly correlated and therefore must be addressed jointly when considering the need for an adjustment. Again, both the theoretical and empirical literatures on this point are mixed with some studies showing a declining WTP for increased longevity with a declining baseline health state (Desvousges et al. 1996) and other studies showing no statistically significant effects (Krupnick et al. 2002).

Application of existing VSLY approaches implicitly assumes a linear relationship in which each discounted life year is valued equally. As OMB (1996) notes “current research does not provide a definitive way of developing estimates of VSLY that are sensitive to such factors as current age, latency of effect, life years remaining, and social valuation of different risk reductions.” The second alternative, applying a suite of values for these risks, lacks broad empirical support in the economics literature. However, the potential importance of this benefit transfer factor suggests that analysts consider sensitivity analysis when risk data — essentially risk estimates for specific age groups — are available. An emerging literature on the value of life expectancy extensions, based primarily on stated preference techniques, is beginning to help establish a basis for valuation in cases where the mortality risk reduction involves relatively short extensions of life.518

515 This second approach was illustrated in one EPA study (U.S. EPA, 2002d) for valuation of air pollution mortality risks, drawing upon adjustments measured in Jones-Lee et al. (1985).

516 See U.S. EPA (2003a) for more information on the valuation of children’s health risks. OMB’s Circular A-4 advises agencies to use estimates of mortality risk valuation for children that are at least as large as those used for adult populations (OMB 2003).

517 The fields of health economics and public health often account for health status through the use of quality-adjusted life years (QALYs) or disability adjusted life years (DALYs). These measures have their place in evaluating the cost-effectiveness of medical interventions and other policy contexts, but have not been fully integrated into the welfare economic literature on risk valuation. More information on QALYs can be found in Gold et al. (1996) and additional information on DALYs can be found in Murray (1994).

518 It should be noted that many observers have expressed reservations over adjusting the value of mortality risk reduction on the basis of population characteristics such as age. One of the ethical bases for these reservations is a concern that adjustments for population characteristics imply support for variation in protection from environmental risks. Another consideration is that existing economic methods may not capture social WTP to reduce health risks. Chapter 9 details how some of these considerations may be informed by a separate assessment of equity.
B.6 Conclusion

Due to current limitations in the existing economic literature, these Guidelines conclude that, for the present time, the appropriate default approach for valuing these benefits is provided by the central VSL estimate described earlier. However, analysts should carefully present the limitations of this estimate. Economic analyses should also fully characterize the nature of the risk and populations affected by the policy action and should confirm that these parameters are within the scope of the situations considered in these Guidelines. While a qualitative discussion of these issues is generally warranted in EPA economic analyses, analysts should also consider a variety of quantitative sensitivity analyses on a case-by-case basis as data allow. The analytical goal is to characterize the impact of key attributes that differ between the policy and study cases. These attributes, and the degree to which they affect the value of risk reduction, may vary with each benefit transfer exercise, but analysts should consider the characteristics described above (e.g., age, health status, voluntariness of risk, and latency) and values arising from altruism.

As the economic literature in this area evolves, WTP estimates for mortality risk reductions that more closely resemble those from environmental hazards may support more precise benefit transfers. Literature on the specific methods available to account for individual benefit-transfer considerations will also continue to develop. In addition, EPA will continue to conduct periodic reviews of the risk valuation literature and will reconsider and revise the recommendations in these Guidelines accordingly. EPA will seek advice from the SAB as guidance recommendations are revised.

Appendix B References


