XXX xx, 2020

EPA-SAB-xx-xxx

The Honorable Andrew Wheeler
U.S. Environmental Protection Agency
1200 Pennsylvania Avenue, N.W.
Washington, D.C. 20460

Subject: Transmittal of the Science Advisory Board Report titled, "Review of EPA's Reduced Form Tools Evaluation."

Dear Administrator Wheeler,

Please find enclosed the final report from the Scientific Advisory Board (SAB). The EPA's Office of Air Quality Planning and Standards (OAQPS) requested that the SAB review EPA's draft report titled, *Evaluating Reduced-Form Tools for Estimating Air Quality Benefits (October 2019).* In response to the EPA's request, the SAB selected subject matter experts from the Science Advisory Board, Clean Air Scientific Advisory Committee and the SAB Chemical Assessment Advisory Committee and assembled the SAB Reduced Form Tools (RFT) Review Panel to conduct the review.

The SAB RFT Review Panel met in-person using a virtual meeting platform on May 28 and 29, 2020 to deliberate on the EPA's charge questions and held a second virtual meeting on September 10, 2020 to discuss their draft report. Oral and written public comments were considered throughout the advisory process. This report conveys the consensus advice of the SAB.

The SAB recognizes the attractiveness of reduced-form tools (RFTs) to support the Agency's goal to conduct streamlined air quality benefits analyses when time or resources constrain the ability to conduct full-form modeling. RFTs could also be useful for screening analyses. The SAB applauds the Agency's efforts to examine the opportunities and challenges presented by RFTs.

While the SAB provided several recommendations within this report, we would like to highlight the following. In general, the SAB agreed that the evaluation approach is organized in a reasonable fashion to derive certain initial insights about how RFTs perform in comparison to each other and to two full-form models (FFMs) that EPA relies on when estimating air quality inputs for benefits assessment. The SAB does recommend the EPA:

- Explicitly acknowledge, in the draft report, the assumption that is embedded in the choice to compare all RFT results to full-form Community Multiscale Air Quality (CMAQ) coupled with Benefits Mapping and Analysis Program-Community Edition (BenMAP-CE) results as it has a substantial potential to drive the conclusions that may come from EPA's evaluation exercise;
- Evaluate the ability of various RFTs to allow for evaluation of the sensitivity of their projected benefits to alternative concentration response relationships and potency assumptions;
- Provide more information that would allow reviewers to reproduce the results of EPA's evaluation, such as providing regional results directly, rather than as summary statistics;
- Increase the range of policy scenarios and provide more information to clarify the performance of the RFTs on regional scales; and
- Provide a discussion on the usefulness of RFTs in different parts of the regulatory decision process and clarify how they would be used for screening or prioritizing procedures.

As the EPA finalizes its draft evaluation of RFTs, the SAB encourages the EPA to address the SAB's concerns raised in the enclosed report and consider their advice and recommendations. The SAB appreciates this opportunity to review EPA's draft report titled, *Evaluating Reduced-Form Tools for Estimating Air Quality Benefits (October 2019)* and looks forward to the EPA's response to these recommendations.

Sincerely,

Chair EPA Science Advisory Board	Chair EPA SAB RFT Review Panel

Enclosure:

NOTICE

U.S. Environmental Protection Agency
Science Advisory Board
Reduced Form Tools Review Panel

CHAIR

Dr. Jay Turner, Professor and Vice Dean for Education, Department of Energy, Environmental and Chemical Engineering, McKelvey School of Engineering, Washington University, St. Louis, MO

MEMBERS

Dr. Richard Belzer, Independent Consultant, Mt. Vernon, VA

Dr. James Boylan, Program Manager, Planning & Support Program, Air Protection Branch, Georgia Department of Natural Resources, Atlanta, GA

Dr. Louis Anthony (Tony) Cox, Jr., President, Cox Associates, Denver, CO

Dr. Alison C. Cullen, Interim Dean and Professor, Daniel J. Evans School of Public Policy and Governance, University of Washington, Seattle, WA

Dr. Sabine Lange, Toxicology Section Manager, Toxicology, Risk Assessment, and Research Division, Texas Commission on Environmental Quality, Austin, TX

Dr. Anne Smith, Managing Director, NERA Economic Consulting, Washington, DC

Dr. Richard Smith, Professor, Department of Statistics and Operations Research, University of North Carolina, Chapel Hill, NC

SCIENCE ADVISORY BOARD STAFF

Dr. Suhair Shallal, Designated Federal Officer, U.S. Environmental Protection Agency, Science Advisory Board (1400R), 1200 Pennsylvania Avenue, NW, Washington, DC, Phone: 202-564-2057, Fax: 202-565-2098, (shallal.suhair@epa.gov)

U.S. Environmental Protection Agency
Science Advisory Board

[ROSTER TO BE ADDED]

TABLE OF CONTENTS

ACRONYMS AND ABBREVIATIONS

APX	Air Pollution Emission Experiment and Policy Analysis Model
BenMAP-CE	Benefits Mapping and Analysis Program-Community Edition
CAMx	Comprehensive Air Quality Model with Extensions
CMAQ	Community Multiscale Air Quality
CPP	Clean Power Plan
CRDM	Climatological Regional Dispersion Model
CRR	Concentration-Response Relationship
EASIUR	Estimating Air Pollution Social Impacts Using Regression
EC	Elemental Carbon
EGU	Electricity Generating Unit
FFM	Full Form Model
InMAP	Intervention Model for Air Pollution
MFB	Mean Fractional Bias
MFE	Mean Fractional Error
NCA	National Climate Assessment
NEI	National Emissions Inventory
NH_3	Ammonia
NMB	Normalized Mean Bias
NME	Normalized Mean Error
OAQPS	Office of Air Quality and Standards
OC	Organic Carbon
$PM_{2.5}$	particulate matter
$prPM_{2.5}$	primary particulate matter
PSAT	Particulate Matter Source Apportionment Technology
RFT	Reduced Form Tools
RIA	Regulatory Impact Analysis
RSM	Reduced Surface Model
SA-BPT	Source Apportionment Benefit-per-Ton
SA-Direct	Models applied directly to obtain monetized health benefit results from emissions inputs have “Direct” added to the model name
SAB	Science Advisory Board
SOA	Secondary Organic Aerosols
U.S. EPA	U.S. Environmental Protection Agency
VOC	Volatile Organic Compounds
VSL	Value of Statistical Life
VSLY	Value of Statistical Life Year
WRF-Chem	Weather Research and Forecasting model coupled with Chemistry

1. INTRODUCTION

The EPA Office of Air Quality Planning and Standards (OAQPS) conducted a study of reduced-form tools to develop and demonstrate a protocol for systematically comparing $PM_{2.5}$ monetized health benefits estimated using reduced-form tools with those generated using full-form air quality and health benefits models, in the specific context of using such tools to inform the economic impacts of regulatory analyses. The EPA's draft report first describes the analytical approach developed to compare the two types of approaches and then presents the evaluation results for several reduced-form tools (RFTs) across multiple policy scenarios. The tools evaluated include: 1) EPA's Source Apportionment Benefit-per-Ton (SA-BPT) approach based on the 2005 National Emissions Inventory (NEI); 2) Air Pollution Emission Experiment and Policy Analysis Model (APX); 3) Intervention Model for Air Pollution (InMAP); and 4) Estimating Air Pollution Social Impacts Using Regression (EASIUR). The EPA's draft report concludes with a description of the limitations of the evaluation approach and findings, with suggestions for future research. EPA representatives noted that they expect that RFTs will continue to evolve in the future. EPA also stated that they have already begun to update the 2005 National Emissions Inventory (NEI) BPT tool to reflect recent updates to the emissions inventories and plan to investigate other efficient modeling techniques that can also approximate full-form modeling (FFM) approaches. As a result, EPA requested a peer review to assess whether the evaluation framework developed in their draft report is appropriate, and to provide input regarding future design improvements to enhance the capabilities of reduced form tools.

The EPA's OAQPS requested that the SAB review EPA's draft report titled, *Evaluating Reduced-Form Tools for Estimating Air Quality Benefits (October 2019).* In response to the EPA's request, the SAB identified subject matter experts from the Science Advisory Board (SAB), Clean Air Scientific Advisory Committee (CASAC) and the SAB Chemical Assessment Advisory Committee (CAAC) and assembled the SAB Reduced Form Tools (RFT) Review Panel to conduct the review. The SAB RFT Review Panel met in-person using a virtual video meeting platform on May 28 and 29, 2020 to deliberate on the agency's charge questions and held one teleconference to discuss their draft report. Oral and written public comments were encouraged throughout the advisory process.

This report is organized to state each charge question raised by the agency followed by the SAB's consensus response and recommendations. The Panel provided key recommendations that are necessary to improve the critical scientific concepts, issues and/or narrative within the assessment. The Panel deemed these recommendations as important for improving the understanding of the suitability and reliability of reduced form tools (RFTs) as compared to full form models (FFMs) for estimating air quality benefits.

All editorial comments are presented within Appendix A. All materials and comments related to this report are available at:
https://yosemite.epa.gov/sab/sabproduct.nsf//LookupWebProjectsCurrentBOARD/46C3F741097CD634852585500048F4BA?OpenDocument

2. RESPONSE TO CHARGE QUESTIONS

2.1. Charge Question 1. Evaluation Approach

2.1.1 Charge Question 1a. *Please comment on the evaluation approach developed by EPA to compare reduced-form models to full-form equivalents.*

In general, the Panel agreed that the evaluation approach is organized in a reasonable fashion to derive certain initial insights about how reduced form tools (RFTs) perform in comparison to each other and to two full-form models (FFMs) that EPA relies on when estimating air quality inputs for benefits assessment (i.e., CMAQ and CAMx).[1] EPA's draft report follows the structure used in many other model comparison exercises, such as those of Stanford University's Energy Modeling Forums. That is, the evaluation establishes a set of scenarios to be run with each model under shared key assumptions. Shared assumptions usually focus on defining the baseline scenario against which policy alternatives are to be run, but they can also include making other key input parameters constant. Differences in model structure then drive differences in predicted outcomes. Results can be compared across models to understand which structural elements caused results to differ.

The interpretation of results from any model comparison, no matter how well structured, is inherently limited by the range of scenarios considered and how the standardization of assumptions narrowed the potential ways model results could differ. There are several attributes of this evaluation that limit the generalizability of insights it can produce. They are listed here and discussed in more detail, where appropriate, in the additional Charge Question 1 sub-sections below.

- Although they do reflect a diverse range of policies, the set of five scenarios are not representative of the array of regulatory applications for which EPA apparently intends to apply RFTs. Thus, no inferences can be made based on observed differences in outputs across the RFTs in the EPA's draft report with respect to the potential performance of these (or other) RFTs for other policy scenarios.
- The decision to standardize key benefits-related assumptions, if successfully completed, would constrain evaluation of the relative performance of these tools to differences in their air quality projections.[2] This decision would have made sense if the objective was to estimate differences among selected RFTs with respect to their air quality ***inputs to benefits assessment***, but not if the objective was to compare ***differences in estimated***

[1] The Panel notes that the EPA's draft report itself does not convey any hint of OAQPS staff contributions to the modeling, analysis, or exposition. These contributions were clarified during the May 28-29, 2020 public meeting.
[2] The EPA's draft report implies that EPA has standardized the key parameters that determine mortality and benefits per unit of air quality change. The Panel notes, however, that this standardization has failed in one or more key dimensions. For example, if the benefits formulas were identical, results from AP3 Direct would be insignificantly different from results from AP3 BenMAP. But as Exhibit C-1 clearly shows, in the nitrate component of the Pulp & Paper policy scenario, results from AP3 Direct and AP3 BenMAP differ by 47-fold. Given that AP3 BenMAP and AP3 Direct are said to have used the same air quality inputs, one must infer that this difference reflects divergent benefits-related assumptions. This discrepancy needs to be explained and addressed.

benefits. Very little can be inferred about performances of the selected RFTs with respect to the ***outputs*** of a benefits assessment.

- Contrary to assertions in the EPA's draft report, uncertainties in a key benefits-related assumption (the concentration-response relationship, or CRR) are not simply proportional in their effects on RFT outputs.[3] Unfortunately, the evaluation, by design, cannot be informative on this matter even though it is critical for ascertaining how accurately RFTs can estimate the health risk reductions and benefits of alternative regulatory policies.
- The evaluation focuses solely on how well RFT outputs match those of a single FFM, i.e., CMAQ. This decision to compare RFTs in terms of how well they match CMAQ is therefore biased in favor of RFTs that rely on CMAQ-compatible inputs. Although the EPA's draft report is not clear on this point, the primary RFT that benefits from this analytic structure is EPA's own SA-Direct.[4]
- The evaluation design can provide only point estimates of the output ratios for the selected RFTs' performance relative to CMAQ. Therefore, we cannot know whether reported departures are materially or statistically indistinguishable.
- An important analytic capability provided by FFMs is the flexibility to quickly conduct many types of sensitivity analyses in the benefits estimation step, including analyses of benefits under alternative benefits-related assumptions. Some RFTs (including SA-Direct) lack this capability. This results in lost analytic utility and transparency, both of which regulatory impact analysis must have. The EPA's draft report is silent concerning which of the RFTs (if any) have this essential flexibility.

RECOMMENDATIONS:

The Panel has also identified many instances in which the EPA's draft report lacks sufficient background information or explanations concerning certain key aspects of the evaluation approach.[5] The Panel recommends that the EPA's draft report address the following areas where more information is critical to understanding the RFTs and their relationships with FFMs:

The Panel recommends that EPA provide details describing how CMAQ/CAMx and BenMAP-CE work independently and together, including the purpose of each model, governing equations, input data requirements, model outputs, and post-processing steps.

---

[3] Besides the CRR's slope, other important benefits-related assumptions include regional and sub-regional (e.g., county, grid) differences and the relative toxicity of different $PM_{2.5}$ constituents. Further, the assumption that emission reductions ***cause*** (and are not just associated with) modeled reductions in health effects has obvious effects on benefits assessment.

[4] For another example of bias favoring EPA's own RFTs, see footnote 9 and the accompanying text.

[5] For example, Panel members noted that the slides presented by OAQPS staff during the May 28-29, 2020 public meeting (U.S. Environmental Protection Agency Office of Air Quality Planning and Standards 2020) contained information that should have been in the EPA's draft report itself, particularly the information on slides 7 through 12.

The Panel recommends that EPA provide descriptions of how each RFT works, including an overview of each model, governing equations and algorithms, input data requirements, model outputs, and post-processing steps.[6] As noted above, a particularly important item of missing information is which, if any, RFTs provide users with ability to conduct sensitivity analyses on alternative economics-related assumptions, such as non-linearities and/or spatial variability in the CRR.

The Panel recommends that EPA provide information on how the air quality estimates of each RFT have been derived from underlying FFMs. For example, APX uses source-receptor matrices produced by the Climatological Regional Dispersion Model (CRDM); InMAP starts with source-receptor relationships from WRF-Chem, and estimates air quality surfaces with variable grid resolutions using its own dispersion-reaction algorithm; and EASIUR relies on statistical regression for emissions and benefits based on air quality fields derived as "averaged plume" out of randomly selected full-form modeling (i.e., CAMx PSAT) grid cells.

The Panel recommends that EPA provide information on which RFTs can produce air quality concentration surfaces and whether those projections have been "fused" with monitored values in the manner EPA does for its own full-form modeling (and apparently used for SA-Direct). The Panel recommends that EPA provide a clearer explanation of the methods and purpose of the primary $PM_{2.5}$ scaling (including when estimates are being presented that have been scaled up from EC-only, and when the estimates are still unscaled). More discussion of the potential errors this introduces as well as the methods should be provided in both the main body and in the Appendices.

The Panel recommends that EPA provide greater clarity regarding where and how ammonia (NH_3), volatile organic compounds (VOCs), and secondary organic aerosols (SOA) were accounted for, given that they are listed as RFT outputs in Exhibit 2-9 but are not listed as "precursors of interest" in Appendix A, Section 4 of the EPA's draft report when discussing the BenMAP model and methods.

The Panel recommends that EPA provide a discussion of the differences between reduced surface models (RSMs) and reduced form models (RFMs) and identify which models in the EPA's draft report are RSMs and which are RFMs.[7] This might provide insight when comparing the RFT results to those of FFMs.

The Panel recommends that EPA provide additional details concerning the errors that EPA's draft report discovered: 1)information on the baseline mortality rates in BenMAP-CE, 2) the basis for concluding these errors would result in "the overestimation of benefits by less than three percent for aggregate benefits values," and finally, 3) information concerning regional variability of this error.

[6] Panel members also identified a need for more information on the averaging times and forms of their concentration metrics; their population and health incidence data (including when they were not standardized for this study); and how each RFT accounts for the proximity of emissions changes to population centers.
[7] Reduced surface models (RSMs) estimate concentrations based on concentrations out of FFMs. Reduced form models (RFMs) use unique algorithms to estimate concentrations.

The Panel recommends that EPA provide a more complete referencing of sources of assumptions, such as the precise source for the SA-BPT estimates used in this analysis, the precise source for the Krewski CRR assumption, and the basis for the VSL assumption. The Panel recommends that EPA provide details on the FFM runs that were used, including for the clean power plan (CPP; which was apparently not used in that RIA) and for the three "hypothetical" industrial sector scenarios (i.e., those applied to cement kilns, pulp and paper facilities, and refineries). These appear to have been done solely for EPA's study and thus require more documentation.

2.1.2 Charge Question 1b. *Please comment on whether the emissions reduction scenarios used in the proposed evaluation approach provide enough diversity to adequately assess reduced-form performance over a range of possible applications (e.g., magnitude, type, and spatial variations of emissions reductions).*

This model comparison exercise covers five scenarios, a convenience sample intended to compare selected RFTs under a variety of conditions. While these scenarios do reflect diversity in the ways that EPA regulations may affect changes in ambient $PM_{2.5}$ precursor emissions, there is no evidence that they are representative of the range of possible policy scenarios. The EPA's draft report partially acknowledges this in the limitations section,[8] but it does not consistently reflect this limitation when discussing results.[9]

Panel members have noted that additional types of policies could produce very different RFT performance patterns and reveal more insight concerning the robustness of RFT performance. This would require additional consideration of other types of scenarios. Most notably, additional mobile and area source scenarios should be considered, as the current five scenarios include just one mobile source scenario, no area source scenarios, one electricity generating unit (EGU) point source scenario, and three industrial point source scenarios. Other types of sources that might produce materially different results are residential wood combustion, marine/aircraft/rail sources, and on-road diesel emissions. Further, even for industrial point source scenarios, the range of variation in RFT performance appears likely to have been greater if other sectors had been analyzed instead. For example, the Panel notes that, as described in EPA's overview presentation at the public meeting of the Panel of the 17 industrial sectors for which SA-BPT estimates are available, the three sectors selected for the EPA draft report's "hypothetical" policies do not have as much variation in their BPT values as other sectors (e.g., iron and steel). This suggests that greater diversity might have been achieved had other sectors been selected.

[8] Industrial Economics (2019, p. 4-4): "While the policies that were analyzed to demonstrate the abilities of each reduced-form tool compared with full-form model results are a thorough subset of policy types, ranging from mobile sources to industrial point sources to EGUs, it is not an exhaustive or fully representative set of policies."

[9] Compare Industrial Economics (2019, p. 4-4), quoted in footnote 8, with Industrial Economics (2019, p. 4-2) ("[T]he SA Direct and EASIUR Direct models … demonstrated consistent performance for total $PM_{2.5}$ and its components, which indicates that they would perform in a similarly reliable way for air quality policies beyond those considered in this analysis"). These statements are inconsistent; they do not accurately capture the limitations of the evaluation design; and they appear to impart a bias in favor of SA Direct and EASIUR Direct.

There is no evidence provided in the EPA's draft report that a structured approach was taken to maximize the variations of possible situations affecting RFT performance among the five scenarios analyzed.

Nevertheless, the five scenarios analyzed do reflect some of the diversity in the ways that EPA regulations may affect ambient $PM_{2.5}$ precursor emissions, including allowing for point estimates of comparisons across regions, magnitude of different emissions species, temporal patterns of emissions, and emission release heights. Given the extent to which critical benefits-related input assumptions have been standardized, it is interesting to see as much variability in results as were reported. Evidence has emerged that estimates of benefits from NOx emissions reductions are subject to the greatest inconsistencies, and that use of RFTs to value benefits of policies with such changes may be most questionable. However, there is also substantial variability in RFT performance even for the point sources scenarios, and for the other $PM_{2.5}$ constituents/precursors. This evidence of variability among RFTs, and between RFTs and FFMs, indicates that the choice of RFT for any particular future regulation could have a material effect on outputs of the benefits analysis and would likely be a source of considerable controversy given the absence of objective criteria for making such a choice. Finally, the Panel expressed concern that any reductions in the analytic burden during regulatory development resulting from using an RFT could be offset (or exceeded) by an increased burden to defend the validity of the results.

RECOMMENDATIONS:

The Panel recommends that a structured approach be followed to maximize the possible situations affecting RFT performance among the five scenarios analyzed. The Panel has identified alternative ways that the five scenarios might have been selected,[10] but the fundamental limitations of the EPA's draft report are that it is a convenience sample and a stratified random sample of five scenarios would be too small. Therefore, care must be taken not to generalize the EPA's draft report findings. Although this point is stated in the Limitations section, it is not always adhered to when results are discussed in other sections of the EPA's draft report. Convenience samples are useful for pilot studies, and that is how the current draft report should be understood and characterized. No inferences can be made with respect to other scenarios based on observed differences in outputs across RFTs in EPA's draft report.

2.1.3. Charge Question 1c. *Please discuss whether the specific assumptions that EPA made to apply each tool as consistently as possible (e.g., emissions, meteorology, use of direct vs. BenMAP estimates, etc.) are appropriate and clearly explained.*

The following assumptions were standardized across all the RFT runs: (a) to report benefits for chronic mortality only, using specifically the Krewski *et al.* (2009) CRR point estimate that

[10] Other ways that the selection of scenarios might have helped identify causes of differences in results across RFTs include: (1) using a more generic set of equal reductions for each of the multiple precursor emissions from each selected sector (while also avoiding the suggestion that the "hypothetical" control scenarios are actually indicative of potential real policies); and (2) comparing the Mercury and Air Toxics Standard (MATS) rule to the CPP rule, as they both apply to one sector but may have had very different spatial patterns of projected emissions changes.

BenMAP treats as its default values,[11] (b) to apply the same point estimate for VSL also relied upon in BenMAP,[12] (3) various other demographic inputs to the health impact function such as population and mortality rates, and (4) that the emissions reductions (quantity and geographic location) associated with each scenario are the same for each alternative model.[13] The EPA's draft report, however, does not clearly explain the extent to which this standardization affects the scope of the model comparison. RFTs reduce model details in two ways. First, they condense complex, nonlinear fate and transport models (i.e., FFMs) into a simpler summary format that is quicker to run but less accurate (such as a source-receptor matrix). Second, they apply an assumed CRR to the reduced-form air quality outcomes. The assumptions for the second step depend more on subjective judgment than scientifically-defined phenomena like air quality modeling, and the resulting benefits estimates vary enormously as a result of alternative assumptions – perhaps more widely than the variations in air quality changes that are associated with different choices of modeling and model-summarizing for the first step. By standardizing the CRR and VSL assumptions, the primary insights that can be extracted from this model comparison exercise are about the relative performance of the various models in predicting how air quality changes in different locations as a result of changes in emissions. Nonetheless, the EPA's draft report compares RFT outputs not as changes in air quality but in terms of benefits, as if the conversion of air quality changes to benefits is merely formulaic.

The EPA's draft report is not at all clear on this fact, yet it places significant limitations on the proper interpretation of the study's comparisons. And more worrisome than this lack of clarity is the incorrect assumption that, because these unaddressed sources of uncertainty are common across all the models compared, the inclusion of CRR uncertainty would not change the *relative* performance of the RFTs. This would be true if the only uncertainty in the CRR assumption were its slope, making differences across outputs simply multiplicative.[14] However, there is substantial evidence that the CRR is nonlinear in quantity, spatially variable (perhaps due to behavioral differences), and different across PM species independent of spatial differences. Each of the scenarios examined in the EPA's draft report has unequal regional and local emission

---

[11] Among the missing information in the EPA's draft report is the source of this single CRR out of the hundreds that are in Krewski *et al.* (2009). One Panel member surmises that it comes from Commentary Table 4 (p. 126) is the all causes random effects model using the 1999-2000 $PM_{2.5}$ exposure levels, because this is the CRR that BenMAP uses for its "default" CRR assumption.

[12] The source for this VSL value is also not referenced.

[13] The descriptions of other assumptions are less clear. For example, it is unclear why some RFTs were applied using different meteorological-year assumptions than the other RFTs, and only in some scenarios. Also, it is unclear why EASIUR used 36 km grids while the other models used 12 km grids. In addition, the population and health incidence data for the EASIUR runs were not consistent with inputs to other RFTs. A more detailed explanation and potential implications of these differences should be provided.

[14] It is notable that this multiplicative factor alone is large in the context of the other variations in RFT performance. For example, the upper confidence interval in the single Krewski *et al.* (2009) CRR is twice the lower confidence interval. And uncertainty about this CRR estimate is larger than this confidence interval implies, since this CRR is just one of a large number of CRR estimates produced in that one study under different statistical modeling assumptions without any clear-cut criterion for choosing which CRR is "best."

changes, so these CRR uncertainties, if they were to be considered, could result in distinctly different relative performance of each RFT under any given scenario.[15]

Thus, while standardizing the key benefits module assumptions enabled a comparative analysis of the air quality performance of RFTs, it significantly limited the interpretability of results beyond those related to the air quality outputs. That is, the EPA's study ***by design*** cannot illuminate the RFTs' relative performance with respect to key parameters of the benefits module. This limits the usefulness of the comparative analysis to evaluating RFTs' performance with respect to air quality ***inputs <u>to</u> a benefits assessment****,* but not ***outputs <u>of</u> a benefits assessment***.

To elucidate this perspective on what this comparative analysis has accomplished, it would be helpful to see the analysis conducted in two separate steps with the first part comparing concentration fields generated by RFTs and FFMs, and the second part comparing monetized benefits estimated by each RFT and BenMAP using a consistent concentration field. This would help readers better understand which component (concentration fields or benefit estimation parameters) were responsible for differences between the RFTs and FFMs.

An additional feature of the EPA's draft report is that the BenMAP model itself was substituted for the original RFTs' internal ("direct") benefits calculations, where this was feasible to do.[16] That this step was done to create additional RFT variants is explained clearly enough, but the EPA's draft report is less clear about the purpose and merits of these variants, as well as, the implications of observed differences. Is this because the population and mortality rate assumptions still differ from those of BenMAP in the "RFT-Direct" calculations, even though the EPA's draft report indicates they were standardized? Is it because the geographical detail differs? Why would they be expected to differ at all? If differences were *not* expected, why were these variants important, given the complexity they have added to the EPA's draft report? The EPA's draft report is unclear on these points, and without clarification of the reasons for the differences, a misleading impression is created that BenMAP's computations of benefits per $\mu g/m^3$ of change in $PM_{2.5}$ are inherently superior to benefits calculations of the other RFTs.[17]

The EPA's draft report chose to compare every RFT's results to the benefits estimate predicted by the full-form model CMAQ. This imposes a strong assumption that CMAQ produces the "correct" benefits estimate. The EPA's draft report should clearly state this assumption, explain why EPA believes it is likely to be true, including providing references demonstrating the CMAQ and CAMx are better than other FFMs.

---

[15] There is evidence that the spatial nature of how CRR uncertainties could strongly affect RFTs' relative performance in the Results section of the EPA's draft report. This seems to suggest that RFTs produce results that differ at a regional level more markedly than at the national level even when the CRR has been assumed to be linear.

[16] This appears to have been feasible for all the RFTs evaluated except EASIUR.

[17] One possible explanation is that the differences are because some benefits estimates are computed for the locations in which the emissions reductions occur while others are computed for the locations where the air quality changes occur. If so, however, how this may be affecting the comparisons among the models is insufficiently explored in the EPA's draft report. The Panel seeks a more thorough discussion of this point, and a transparent illustrative numerical example could be very helpful.

The EPA's draft report includes a comparison of CMAQ and CAMx in addition to comparisons of RFT results to those of CMAQ. This shows that CAMx produces very consistent results to those of CMAQ, which presumably is important because EPA's SA-Direct BPT estimates are derived from CAMx. But other RFTs use different FFMs as their starting points. Thus, the decision to compare RFTs in terms of how well they match CMAQ outputs benefits estimates makes the analysis biased in favor of RFTs that are based on CMAQ or CAMx. The RFTs advantaged by this decision are SA-Direct and EASIUR (both of which were based on CAMx using PSAT). Although there are differences in how the original *full*-form air quality changes have been converted into a *reduced*-form estimate of air quality changes, it should not be surprising that SA-Direct and EASIUR tracked CMAQ-based full-form estimates better than the other RFTs.

RECOMMENDATION:

The Panel recommends that this design feature of comparing RFTs in terms of how well they match CMAQ outputs benefits estimates, and its implications for output comparisons, must be explicitly acknowledged because it likely drives the EPA draft report's results and conclusions. Insights about the importance (or not) of the choice of foundational FFM would be enhanced if the EPA's draft report were to compare the *air quality outputs* of all the RFTs for each scenario to those implied by the full-form runs.[18] This would further help clarify the extent to which this model comparison exercise has eliminated differences in *benefits* estimation.

2.1.4. Charge Question 1d. *Please assess whether the Report's description of its limitations is complete.*

Section 4.4, which summarizes limitations of the analysis, is clearly written but materially incomplete.

One major limitation was mentioned multiple times in the EPA's draft report but was not included in the discussion. Specifically, the limitation that BPT approaches assign health impacts to the county in which the emissions changes occur rather than where the health impacts occur should be added.

Section 4.4 also should provide a more thorough discussion of limitations resulting from none of the comparisons addressing uncertainties in the underlying CRR (and other benefits-related assumptions). The EPA's draft report incorrectly states that this uncertainty is simply multiplicative, which therefore would have no effect on relative performance. As discussed above, however, CRR uncertainty has multiple aspects (including a strong spatial dimension),

[18] Although SA-Direct may appear to be a set of BPT estimates with an air quality projection, the fact that they were updated for this study (see Industrial Economics 2019, p. 2-11) indicates that the underlying air quality grid to estimate the BPTs is available. That grid could be used to estimate the $\mu g/m^3$ per ton of each precursor that is implicit in its BPT estimates, which could in turn be compared to the outputted air quality changes of all the other models (except, apparently, EASIUR).

and alternative CRR specifications could strongly affect the relative performance of the RFTs. The importance of this uncertainty is thus unreasonably diminished in Section 4.4.

Another limitation not mentioned Section 4.4 concerns RFTs that do not allow users to *directly* test the sensitivity of projected results to alternative economics-related parameters, including the slope of the CRR and relative constituent potency assumptions. This limitation is hinted at on page ES-1 of the EPA's draft report, which states:

> *The study did not evaluate the ability of each approach to characterize the distribution of $PM_{2.5}$-related premature deaths according to the annual mean concentration at which they occurred.*

This is a vague way of saying the comparison of models in EPA's draft report has not addressed their respective abilities to provide information about the sensitivity of the estimated benefits to alternative CRR functional forms and cut points. Plenty of evidence exists in prior $PM_{2.5}$ benefits studies that this is a major source of uncertainty, one that is likely larger in magnitude than the uncertainty in projecting air quality changes resulting from emissions changes. Thus, choosing to rely on a benefits analysis method that eliminates the ability to perform this kind of sensitivity analysis implies a major limitation compared to FFMs (and a potentially fatal deficiency under Circular A-4 guidelines for Regulatory Impact Analyses of major rulemakings). If all RFTs are equally unable to perform such sensitivity analyses, then using an RFT instead ofFFM represents a significant deficiency, with highly controversial implications for proposed or final benefit-cost analyses of important regulatory decisions. That this has not been mentioned in the EPA's draft report as a significant limitation is disconcerting; it should be fully disclosed in Section 4.4, not just mentioned in passing.

The Panel also expressed concern with the possible inference that its efforts may be inappropriately represented as a peer review of these RFTs (or RFTs in general), pursuant to the Agency's Peer Review Policy (U.S. Environmental Protection Agency 2015, Section 1.3). This policy explicitly states that EPA utilizes peer review for the purpose of complying with pre-dissemination review requirements[19] under applicable information quality guidelines.[20] Our charge does not mention information quality, however, and the panel is generally unfamiliar with

[19] See U.S. Environmental Protection Agency (2015, p. 27), which says Agency pre-dissemination work products undergoing peer review should contain the following disclaimer: "*This information is distributed solely for the purpose of pre-dissemination peer review under applicable information quality guidelines. It has not been formally disseminated by EPA. It does not represent and should not be construed to represent any Agency determination or policy*." Further: "*In cases where the information is highly relevant to specific policy or regulatory deliberations, the disclaimer should appear on each page of the work product*." Such disclaimers are presently missing from the EPA's draft report.

[20] Office of Management and Budget (2002); U.S. Environmental Protection Agency (2002). For pre-dissemination review requirements, see Office of Management and Budget (2002, p. 8459): "As a matter of good and effective agency information resources management, agencies shall develop a process for reviewing the quality (including the objectivity, utility, and integrity) of information before it is disseminated. Agencies shall treat information quality as integral to every step of an agency's development of information, including creation, collection, maintenance, and dissemination. This process shall enable the agency to substantiate the quality of the information it has disseminated through documentation or other means appropriate to the information."

the requirements of applicable guidelines. Note that the key procedural information quality standard is reproducibility,[21] but Panel members who attempted to reproduce the EPA draft report's results were unable to do so. Therefore, the Panel cautions that their review must not be used to satisfy information quality pre-dissemination review requirements.[22] Thus, regardless of the potential merits of the EPA's draft report (even with limitations acknowledged), EPA cannot disseminate the draft report in a manner that conveys Agency endorsement. This limitation also should be acknowledged in Section 4.4.

RECOMMENDATION:

The Panel recommends that a more thorough discussion of limitations resulting from the fact that none of the comparisons addressed uncertainties in the underlying CRR be added.

Panel members were not able to reproduce some of the EPA draft report's results; therefore, a key component of the information quality pre-dissemination review requirements was not satisfied. The Panel recommends providing more information that would allow reviewers to reproduce the results of EPA's evaluation.

2.2. <u>Charge Question 2. Evaluation Results</u>

Charge Question 2- Please comment on the results of the reduced form tool evaluation in Section 3, considering both the quantitative and qualitative aspects of the model intercomparison.

2.2.1. Charge Question 2a - *Was the information clearly presented and informative?*

In general, the information was clearly presented and informative, nevertheless modifications would improve the presentation clarity. Presenting results with a log x-axis, ensuring that the differences between models are highlighted rather than the difference between scenarios, and inclusion of all study results for full transparency and to allow results to be reproduced would be beneficial.

Y-Axis Scale on Section 3 Exhibits

Exhibits 3-2 to 3-4 of EPA's draft report would be more easily interpreted (and less likely misinterpreted) with a logarithmic x-axis. The linear x-axis gives much greater visual attention to positive than negative biases of equal magnitude. That, in turn, reinforces the visual impression that RFTs may be upwardly biased in a systematic manner. This problem is especially acute in

[21] Office of Management and Budget (2002, p. 8460): "'Reproducibility' means that the information is capable of being substantially reproduced, subject to an acceptable degree of imprecision. For information judged to have more (less) important impacts, the degree of imprecision that is tolerated is reduced (increased)… With respect to analytic results, 'capable of being substantially reproduced' means that independent analysis of the original or supporting data using identical methods would generate similar analytic results, subject to an acceptable degree of imprecision or error."

[22] To be clear, for this (or any) Panel to conduct a pre-dissemination review of these RFTs, or RFTs in general, requires a very different charge. That, in turn, would require full disclosure of model data, code, and output files, and the panel would need much more review time.

Exhibit 3-4 because the x-axis spans a range of zero to 10 instead of zero to 4.5. Exhibit 3-4 also contains the biggest discrepancy between CMAQ and an RFT in the entire EPA's draft report, for the nitrate component AP3-BenMAP model and the "Pulp and Paper" policy, for which the ratio of CMAQ to RFT costs was 130/7=18.6 according to the table in Exhibit C1, easily greater than any of the RFT:CMAQ ratios, but the reader would easily miss this in a quick glance at Exhibit 3-4.

Highlighting Model Differences

Exhibit 3-1 of EPA's draft report does not match the purpose of the study, which was to characterize the *relative performance* of selected RFTs. In Exhibit 3-1, the projected total benefits for each of the 5 scenarios are compared on a by-scenario basis. This leaves readers with the first "insight" in the Results section being that some policy scenarios would, if implemented, have higher benefits than others. This inappropriately focuses the EPA's draft report away from comparisons *of the RFTs* and onto comparisons *of the scenarios*. It implicitly suggests an intent of the study may have been comparing policies rather models. The rest of the Results section uses comparisons relative to CMAQ's results, which avoids any sense of comparison of the 5 types of policy benefits. Also, the other exhibits provide all the information necessary to understand the model comparison insight, which effectively renders Exhibit 3-1 unnecessary.

Additionally, Exhibit 3-4 of EPA's draft report (reports the results for the $PM_{2.5}$ species) would be better formatted to look like Exhibit 3-3, rather than Exhibit 3-2. Exhibit 3-2 presents the comparisons of the RFTs within the different scenarios. Given that the purpose of this analysis is to compare the RFTs to the FFMs, it makes more sense to compare models within scenarios, rather than scenarios within models.

Present a Map of the Regions

Exhibit B-1 of EPA's draft report contains the states in each National Climate Assessment (NCA) region. The presumed source of these regional assignments is the 2017 report based on the reference cited on page 2-18 of EPA's draft report (https://www.epa.gov/cira). In the 2017 report, there is a regional map on page 17 that is drastically different than the breakdown of states listed in Exhibit B-1. The states in the different regions should be clarified, and the EPA's draft report should add a map to clearly show the groupings.

Transparency

Panel members who attempted to reproduce the EPA draft report's results were unable to do so (described more below). During the public meeting held on May 28-29, the Panel became aware that a substantial amount of information collected during the study was not included in the EPA's draft report. It is possible that if this information had been included the reproducibility defect noted above would go away. In any case, the Panel can evaluate the transparency of only the information that was disclosed.

Similarly, the lack of presentation of regional results (also discussed more below) is especially important because the overall tenor of the EPA's draft report seems to be favoring the SA-Direct and EASIUR models – if either of these doesn't work well in regions, it is important to know that.

Presentation of Results as Benefits Estimates
The results from the EPA's draft report should be presented as differences in air quality projections, not only as differences of benefits, particularly for RFTs that utilize BenMAP, because all relevant parameters for benefits estimation were held constant and are not part of the review. Reported biases reside somewhere in the emissions/air quality interface for those models and this should be discussed in addition to presenting the benefits results.

RECOMMENDATIONS:

The Panel recommends presenting the results with a log x-axis to allow for easier interpretation of results across different model comparisons.

The Panel recommends ensuring that the differences between models be highlighted rather than the difference between scenarios.

The Panel also recommends that all study results be included for full transparency and to allow results to be reproduced.

2.2.2. Charge Question 2b - *Were EPA's conclusions reasonable*?

The EPA's draft report did not offer many conclusions, but rather mostly provided descriptions of what was done. As shown below, Panel members were unable to reproduce the results presented for SA Direct, and also found that the origin of the CPP data (used for 4 of the 5 scenarios) is unclear; therefore, it is difficult to be confident in the data used to draw conclusions. When drawing conclusions from these analyses, it is important to caveat the benefits estimates with considerations about the appropriateness of the underlying assumptions for the use of CRRs, and the flexibility of the models to respond to changes in CRRs. In general, Panel members did not find that the results could readily be generalized to other RFTs or to other policy scenarios.

Reproducibility of SA Direct Results
To confirm the reliability of the presented results, the Panel conducted a rough calculation of the benefits estimated using the SA Direct method. This calculation used the emissions changes for policy scenarios in tons from Exhibit 2-2 and the benefits per ton (BPT) for each $PM_{2.5}$ species from USEPA (2018), using the Krewski *et al.* estimates with a 3% discount rate for the year 2025 for the different sections (Tables 69 (cement kilns), 71 (pulp & paper), 73 (refineries), 100 (EGUs), 131 (2030, on-road mobile)). The emissions change for SO_2, NOx, or $PrPM_{2.5}$ were multiplied by the matched BPT (and by the appropriate mortality-only adjustment factor in Exhibit A-3) and compared to the data provided in Exhibit C-1. The full tables of data for these calculations for the Clean Power Plan (CPP) scenario are shown below in Table 1, and for the other scenarios are included in Appendix 1. In general, this calculation could very closely

recreate the SO_2 and NOx benefits estimates, but the $PrPM_{2.5}$ estimates were substantially different, being lower by a factor of 4 to 14 (depending on the scenario).

One possible source of the discrepancy in $PrPM_{2.5}$ values noted in Table 1 is the "scaling" of elemental carbon (EC)-only $PrPM_{2.5}$ to include organic carbon (OC) and crustal $PrPM_{2.5}$. This was investigated by comparing the $PM_{2.5}$ emission reduction estimates from the 2014 USEPA Proposed CPP Regulatory Impact Analysis (USEPA, 2014), which was the basis of the CPP emissions reductions (as stated on page 2-3), specifically Option 1 State estimates. Table 4-11 from the CPP RIA shows that nationally for 2025, 49,000 tons of crustal $PM_{2.5}$ and 6,000 tons of EC+OC $PM_{2.5}$ were projected to be reduced. So, as per the calculation specified on pp 2-16 to 2-17 ("We scaled the results by multiplying the $prPM_{2.5}$ benefit-per-ton based on EC only by the total amount of primary $PM_{2.5}$ emissions to generate an estimate of impacts for total primary $PM_{2.5}$ emissions."), we multiplied the $PrPM_{2.5}$ BPT ($170,000 x 0.973 mortality-only factor) by 55,000 tons (49,000+6,000) = $9,097 M, which does not match the value in Exhibit C-1 for SA Direct, $PrPM_{2.5}$, CPP ($5,800 M, also shown in Table 1 below). Therefore, the scaling from EC-only to EC+OC+crustal $PM_{2.5}$ does not readily explain the discrepancy shown in Table 1.

Similarly, there are inconsistencies in the ratios presented in Exhibits 3-2 and 3-3. Exhibits 3-2 and 3-3 are the same data in two different forms of display, presumably both derived from Exhibit C-1 in Appendix C. However, there seem to be some minor inconsistencies in the way the data in Exhibit C-1 were reduced to the two figures: for example, for the results of AP2-Direct *versus* CMAQ-BenMAP under total $PM_{2.5}$ for the Tier3 scenario, Exhibit C-1 shows a benefit of $4,100 (millions) under CMAQ-BenMAP and $11,000 under AP2-Direct, a ratio of 2.68, not 2.8 as reported in Exhibits 3-2 and 3-3. There are numerous minor inconsistencies like that, in addition to the major inconsistency shown here in Table 1.

Table 1. Calculation of benefits for CPP Rule via SA Direct method (using EGUs Benefits per Ton estimates and a mortality-only adjustment factor of 0.973 from Exhibit A-3).

Data Source:	**Exhibit 2-2**	**USEPA 2018 Table 100**	**Calculated**	**Exhibit C-1**
Pollutant	**Ton Reductions**	**Benefits per ton ($)**	**Total Benefit ($ Mill)** [1]	**Total Benefit ($ Mill)** [2]
Pri-$PM_{2.5}$	2,481	$170,000	$410	$5,800
NOx	414, 479	$6,700	$2,702	$2,700
SO2	422,670	$46,000	$18,918	$19,000
Total $PM_{2.5}$			$22,030	$28,000

Note: estimates marked in red show substantial differences between calculated and presented total benefits
[1] Calculated Total Benefit ($ Millions) = Ton Reductions x Benefits per Ton x 0.973 mortality-only factor
[2] Exhibit C-1 Total Benefits taken directly from the appropriate row of Exhibit C-1 in the SA Direct column

Modeling from Proposed CPP Rule (USEPA 2014)

Page 2-3 of the EPA's draft report states that the basis for the CPP scenario was the Option 1 State estimates from the Proposed CPP Regulatory Impact Analysis (RIA; USEPA 2014). Three other scenarios (Pulp & Paper, Refineries, and Cement Kilns) used the CPP modeling as their basis. The CPP SA Direct results and the benefits estimated in the Proposed CPP RIA should be very similar, because they used the same estimation method, although slightly different BPT estimates. Therefore, we compared the CPP scenario SA Direct results in Exhibit C-1 to the $PM_{2.5}$ benefits provided in the 2014 CPP RIA. Table 4-14 of the Proposed CPP RIA presents the Summary of Estimated Monetized Health Co-Benefits for the Proposed EGU GHG Existing Source Guidelines in 2025 (millions of 2011$). Using Option 1 – State, 3% Discount Rate, the lower end of the range provided (which represents the results from Krewski *et al.* 2009) and multiplying by 0.973 for mortality-only and by 1.05 to roughly convert to 2015$, the results in Table 2 were generated. The benefits presented in the CPP RIA could be largely recreated using the inputs from that document (emissions reductions tons and BPT) and these were converted to a comparable number for the current analysis (conversion to 2015$, mortality-only benefits). This generally produced estimates that were similar to SA Direct calculations of SO2 and NOx benefits and would be comparable for the $PrPM_{2.5}$ if the same $PM_{2.5}$ source were used (EC for SA Direct, EC+OC and crustal separately for the CPP RIA).

Table 2. Estimates of benefits for CPP emissions changes based on data from the CPP RIA (2014) and the SA Direct benefits calculated based on the RFT (IEc, 2019) analysis.

Pollutant	**CPP RIA (2014) (2015$)**			**SA Direct Benefits (IEC, 2019) (2015$)**		
	BPT (2011$)	**Tons**	**Calculated Benefits (millions)** [1]	**BPT (2015$)**	**Tons**	**Calculated Benefits (millions)** [2]
SO2	$41,000	425,000	$17,800	$46,000	422,670	$18,900
NOx	(for NOx as $PM_{2.5}$) – $6,000	436,000	$2,670	$6,700	414,479	$2,700
$PrPM_{2.5}$ (EC+OC)	$150,000	6,000	$920	$170,000		
$PrPM_{2.5}$ (Crustal)	$17,000	49,000	$850	Not provided		
$PrPM_{2.5}$ (EC)	Not provided	Not provided		Not provided	2,481	$410 [3]
Total $PM_{2.5}$			$22,500			$22,200

[1] Benefits = BPT x tons x 0.973 (mortality-only adjustment factor) x 1.05 (2011$ to 2015$ adjustment)
[2] Benefits = BPT x tons x 0.973 (mortality-only adjustment factor)
[3] Benefits for $PrPM_{2.5}$ (EC) calculated using the BPT estimate for EC+OC

Table 3 shows the results presented from the CPP RIA or from the EPA's draft report *versus* calculated benefits for the CPP RIA (2014) and for the SA Direct model. The $PrPM_{2.5}$ estimates, and therefore the total $PM_{2.5}$ estimates, were very different in the CPP RIA compared to Exhibit C-1. For comparison, the CMAQ-BenMAP estimate for total $PM_{2.5}$ was quite similar to the CPP RIA estimate, but this was generated in the CPP RIA by higher estimates of SO_2 and NOx benefits and lower estimates of $PrPM_{2.5}$.

Table 3. Estimates of benefits for CPP emissions changes based on the presented benefits from the CPP RIA (2014) and calculated from the inputs of the CPP RIA, and the SA Direct benefits presented in the EPA's draft report analysis and calculated based on the EPA's draft report inputs, and the CMAQ-BenMAP benefits presented in the EPA's draft report analysis.

Pollutant	**CPP RIA (2014) (Millions 2015$)**	**SA Direct Benefits (2019) (Millions 2015$)**	**CMAQ-BenMAP Benefits (2019) (Millions 2015$)**

	Presented (Table 4-14) [1]	Calculated in Table 2	Presented (Exhibit C-1)	Calculated in Table 2	Presented (Exhibit C-1)
SO_2	$18,400	$17,800	$19,000	$18,920	$15,000
NOx	$3,000	$2,670	$2,700	$2,700	$1,700
$PrPM_{2.5}$ (EC+OC)	$920	$920			
$PrPM_{2.5}$ (Crustal)	$850	$850			
$PrPM_{2.5}$ (EC)				$410	
$PrPM_{2.5}$ (EC+OC+crustal)			$5,800 [3]		$3,500 [3]
Total $PM_{2.5}$	22,500 [2]	$22,200	$28,000	$22,000	$21,000

Note: estimates marked in red show substantial differences between calculated and presented total benefits
[1] Benefits = Benefits value for CPP RIA 2014 Table 4-14 Option 1-State, 3% Discount Rate, lower end of presented range x 0.973 (mortality-only adjustment factor) x 1.05 (2011$ to 2015$ adjustment)
[2] Total $PM_{2.5}$ Benefits = Total – NOx (as Ozone) from Table 4-14, then calculated as in footnote 1
[3] Assumed to be the benefits from total primary $PM_{2.5}$ (EC+ OC+ crustal) based on language about scaling on pages 2-16 to 2-17

Therefore, the presented $PrPM_{2.5}$ and total $PM_{2.5}$ benefits do not match the benefits presented in the Proposed CPP RIA (2014) upon which the scenario is based. During the panel's public meeting, it became clear that a full form model run was completed on the Proposed CPP RIA after the document was published in 2014. Furthermore, the full form model run was used as the basis for the calculations in this document. However, that is not what is presented in the document as being the basis, and the authors need to clarify the data source as well as explain the discrepancies noted above.

Interpretation of Results from BenMAP Analyses

Even though the focus of this analysis is on the inputs into the BenMAP-type tools, and not the workings of those tools themselves, it is still important to note that all the reduced-form tools treat the BenMAP statistical regression equations for health impacts (representing statistical relationships with model specification errors, unmodeled errors in variables, omitted confounders, omitted interaction terms, etc.) as if they were valid causal models (Exhibit 2-10, note b, p. 2-16). As detailed in the CASAC's comments on the $PM_{2.5}$ NAAQS review, regression equations such as those in BenMap-CE do not in general give correct answers to causal questions, such as how changing a predictor (e.g., pollutant levels) would change health effects (Pearl 2009, https://projecteuclid.org/euclid.ssu/1255440554).

Similarly, the analysis has not discussed the ability of the various RFTs to allow for evaluation of the sensitivity of their projected benefits to alternative C-R slope and potency assumptions. Much evidence exists in prior $PM_{2.5}$ benefits studies that this is a major source of uncertainty in benefits estimates – likely larger in magnitude than the uncertainty in projecting air quality changes from given emissions changes. If all RFTs are equally unable to perform such sensitivity analyses, then this represents an important trade-off when deciding to use a quicker RFT approach over a complex full form benefits analysis, and should be given serious consideration in the decision process. However, if some of the RFTs under consideration do allow C-R sensitivity analyses to be conducted, that would be an important positive attribute for those RFTs compared to more rigid BPT-based approaches. Whether some of the RFTs have this greater flexibility is an important qualitative consideration that is presently lacking in the comparison and would be useful to include.

Extrapolation of Results

The small sample size of reduced-form models (N = 8 at most, and fewer if the AP models are not counted as independent observations) and the small number of policies analyzed makes it difficult to draw confident general conclusions from the results presented in the EPA's draft report. It is not possible to get a sense of the error surface for different policies from this small sample.

It appears from the text of the EPA's draft report that the main conclusion is that certain RFTs produce results sufficiently close to FFMs that their prior use in Regulatory Impact Analyses (RIAs) was reasonable and they are nearly ready to be applied to one or more unspecified future RIAs. It would be inappropriate for EPA to rely upon the report of this Panel as an external validation of these conclusions. Looking backward, this Panel has not reviewed prior RIAs. Looking forward, members of this Panel clearly concluded that the scenarios considered in the EPA's draft report should not be deemed representative, which makes extrapolation to other scenarios inherently dubious. The EPA's draft report seems to concur, but that concurrence is not as clear as it should be.[23] Elsewhere in the EPA's draft report, RFT outputs are described as "a quicker approach to generating ballpark estimates" (pp. ES-7, 5-1) – a much lower level of practical utility than what is expected of an RIA. Members of the Panel concluded that none of the RFTs examined produced results so obviously reliable that extrapolation to other scenarios is justified. This is especially so for RFTs that predict benefits directly from emissions changes. The Panel came to this conclusion because the EPA's draft report seems to implicitly state that a 2:1 relative error (in either direction) is acceptable when comparing RFT results to the full form model results. However, many benefit-cost analyses result in costs and benefits estimates that are quite close (less than a factor of 2 apart), so a 2-fold error in the benefits estimate could result in a different conclusion about the cost-benefit comparison and therefore potentially a different policy decision.

Comparison to CMAQ or CAMx Models

The EPA's draft report includes a comparison of two full-form models – CMAQ and CAMx - to each other in addition to comparisons of the RFT results to those of CMAQ. This shows that CAMx produces very consistent results to those of CMAQ, which in turn is important because EPA's SADirect BPT estimates are derived from CAMx-based model runs. However, other RFTs have used different full-form models as the starting point for their projections of air quality changes. Because the standardization process has eliminated any differences due to benefits-related assumptions, the RFT's results in the EPA's draft report will differ only due to their methods of estimating air quality changes. Thus, the decision to compare RFTs in terms of how well they match the full-form CMAQ benefits estimates has implicitly created an initial

[23] Industrial Economics (2019, p. 4-4) seems to minimize the representativeness problem by describing the scenarios as "not an *exhaustive* or *fully* representative set of policies" (italics added). A convenience sample is never representative; there is no such thing as a "*partially* representative set of policies." Further, no representative sample is *exhaustive*; if it were, it would be a census, not a sample.

comparative advantage for RFTs that have been developed based on either CMAQ or CAMx. The RFTs with this implicit initial advantage are SA-Direct and EASIUR (both of which were based on CAMx using PSAT). Although there are differences in how the original full-form air quality changes have been converted into a *reduced*-form estimate of air quality changes, it should not be surprising if SA-Direct and EASIUR were to match CMAQ-based full-form estimates better than the other RFTs evaluated in the EPA's draft report. In addition, the errors of the CMAQ-based RFTs may be underestimated, because we are comparing a CMAQ-generated model directly to CMAQ. The assumption that is embedded in the choice to compare all RFT results to full-form CMAQ results is thus important to acknowledge explicitly, as it has a substantial potential to drive the conclusions that may come from this particular evaluation exercise.

RECOMMENDATIONS:

The Panel reiterates its concern regarding the reproducibility of the results presented for SA Direct. The Panel recommends EPA clarify the origin of the CPP data used for 4 of the 5 scenarios.

The Panel recommends that the ability of various RFTs to allow for evaluation of the sensitivity of their projected benefits to alternative C-R slope and potency assumptions be investigated.

The Panel recommends that the EPA's draft report should explicitly acknowledge the assumption that is embedded in the choice to compare all RFT results to full-form CMAQ results as it has a substantial potential to drive the conclusions that may come from this particular evaluation exercise.

2.2.3. Charge Question 2c - *Are there other results which would be useful to include in the comparison?*

The Panel concluded that regional results would substantially improve their ability to determine how to interpret the differences between the models and whether the RFTs can or should be used at the regional level. Even unreliable regional results would provide useful information. Perhaps most importantly, comparisons of air quality surfaces would get to the root of the differences between the models and would best inform the use of the models for various scenarios. Additional model evaluation methods and summary statistics would help to further evaluate the RFTs compared to the full form models.

Regional effects and other forms of disaggregation
In contrast to the national results, which overall were well explained, the regional parts of the analysis were not thoroughly presented or explained. According to Appendix B of EPA's draft report, the continental US was divided into 7 regions defined by states. County-level results for each modeling approach were aggregated into the 7 regions, but instead of presenting separate results for all 7 regions, the authors calculated summary statistics (principally R^2, normalized

mean bias and normalized mean error). These summary statistics were hard to interpret. Presenting individual results for the 7 regions would have been far better. Exhibits 2-3 through 2-7 show substantial variability among the policy scenarios in terms of which parts of the country they affect – it is therefore entirely plausible that the results for different scenarios will be quite different in different parts of the country. Also, given the different results that were obtained nationally for total $PM_{2.5}$ and for different components, it would be informative to show those regional results as well. In principle, the authors could present results from 9 models (the same ones as in Exhibit C-1), 8 regions (counting all-US as the eighth region), 4 pollutants (total $PM_{2.5}$, $prPM_{2.5}$, sulfate, nitrate) and 5 policy scenarios – a total of 1,440 numbers. It should have been possible to present that information in the EPA's draft report without overwhelming the reader, and it would allow others to make more comparisons to supplement those presented in the EPA's draft report.

The distribution of regional benefits and costs is always important, as are distributional differences across other margins, such as income (e.g. Fullerton, 2017). Air quality regulations have highly variable regional and local impacts that the EPA is likely to consider when setting and administering national policy. RFTs that cannot accurately identify the geography where benefits are projected to be realized have limited practical utility for regulatory decision-making. In general, the EPA should strongly prefer models and tools that estimate benefits and costs at the lowest possible level of aggregation.

In addition, only some of the regional results were presented even as summary statistics (Exhibits 3-5 to 3-7), but in other parts of the EPA's draft report, it seems that summary statistics were in fact completed for the other RFTs, but not presented. For example, on page 4-2 the EPA's draft report states, "EASIUR Direct also did a reasonable job capturing variation in benefits across large regions of the US (0.88 R^2value on average)." This information is not present in the Results chapter or in the appendices. Similarly, on page 5-2 the EPA's draft report states, "In our analysis we saw differences in how the tools performed at different geographical scales and locations." The EPA's draft report should offer more information about what data this conclusion was based on.

Information about Speciation of $PM_{2.5}$

The method and interpretation of scaling of $PrPM_{2.5}$ (EC) to $PrM_{2.5}$ (EC+OC+crustal) is poorly described in the EPA's draft report. One method for clarifying the $PrPM_{2.5}$ benefits would be to include the raw and scaled versions of the $PrPM_{2.5}$ and total $PM_{2.5}$ benefits, as well as a better explanation for how and why they were generated. Exhibit 3-4 of the EPA's draft report shows the ratio of benefits from $PrPM_{2.5}$ (labeled as EC only) for the RFTs compared to CMAQ BenMAP. The ratios are as expected based on the values shown in Exhibit C-1, but Exhibit C-1 presents the scaled estimates (defined on page ES-4 as $prPM_{2.5}$ BPT based on EC multiplied by the total amount of primary $PM_{2.5}$ emissions EC scaled to OC + crustal) - would these ratios be expected to be the same for EC only? This is another example of a lack of clarity in the estimates due to $PrPM_{2.5}$ scaling.

Another consideration for $PM_{2.5}$ speciation is the contribution of ammonia (NH_3), volatile organic compounds (VOCs), and secondary organic aerosols (SOA). Some of the RFTs consider changes in NH_3 and VOC emissions (noted in Exhibit 2-9) whereas others do not. The authors should discuss how much these emissions contribute to the total $PM_{2.5}$ benefits (they are not included in Appendix C Exhibits or discussed in the Appendix A methods) and how those may impact the relative outputs of the RFTs *versus* the full form tools. For example, as shown in Exhibit 2-2 a substantial amount of the reductions from the Tier 3 rule were from VOCs (33% of the change) – the authors should address whether those models that don't capture VOCs (SA Direct, AP2 BenMAP, AP3 BenMAP, EASIUR Direct) will capture this aspect of the benefits.

In general, the EPA's draft report does not indicate what fractions of total benefits were attributable to each $PM_{2.5}$ species. A small bias with respect to estimates from one species could translate into greater effects than a large bias in estimating another species. Relative contributions could be calculated if Exhibit 3-4 provided these proportions for CMAQ and CAMx.

Comparisons of RFT Air Quality Surfaces

Insights about the importance (or not) of the choice of RFTs would be enhanced if the EPA's draft report compared the air quality outputs of all the RFTs for each scenario to those produced by the full-form runs.[24] This would help clarify the extent to which this model comparison exercise has eliminated differences in the way benefits themselves are calculated, once the air quality changes have been estimated. In addition, transparency requires maximum disaggregation to fingerprint where RFTs lack accuracy and need to be revised.

Summary Statistics and Model Evaluation

Using model evaluation methods and additional summary statistics would improve the ability to compare the RFTs to the full form models. While some of these methods can be used on the existing data and models, others require more scenarios and many more model runs to fully answer the question: for what kinds of policies and scenarios do the RFTs work relatively well or badly? Therefore, which methods to use (from those discussed below) will depend on the EPA's time and the amount of additional work they are willing to do.

For model evaluation, Verification, Validation, and Uncertainty Quantification (VVUQ) methods could be used (https://asmedigitalcollection.asme.org/verification, https://www.nap.edu/catalog/13395/assessing-the-reliability-of-complex-models-mathematical-and-statistical-foundations). It would also be more informative to test whether distributions of observed and model-predicted values (or full- and reduced-form analysis results) are significantly different from each other; and to use visualizations such as regression diagnostics to understand when and how the different reduced-form model predictions differ significantly from each other and from full-form results. Using optimization to identify scenarios that maximize error metrics, similar to Extreme Bounds Analysis (EBA) for regression models (https://cran.r-

[24] Although SA-Direct may appear to be a set of BPT estimates with an air quality projection, the fact that they were updated for this study (per p. 2-11 of the EPA'S DRAFT report) indicates that the underlying air quality grids to estimate the BPTs is available. That grid could be used to estimate the μg/m3per ton of each precursor that is implicit in its BPT estimates, which could in turn be compared to the outputted air quality changes of all the other models (except apparently EASIUR).

project.org/web/packages/ExtremeBounds/vignettes/ExtremeBounds.pdf) could help reveal how large the errors from reduced form model could possibly be and under what conditions relatively large errors occur. It would be helpful to use sensitivity analysis techniques (some of which have also been discussed in connection with Info-Gap robust design methods) to understand the types of scenarios that lead to relatively large or small prediction errors for some or all of the reduced form models.

For the summary statistics, mean squared error (MSE) is not reported (See Exhibit 2-11, p. 2-18), but could add useful information to the mean absolute error metrics. In addition, the presented statistics require careful interpretation: the coefficient of determination is insensitive to many types of errors (e.g. If each predicted value is 1000 times greater than the observed value, $r^2 = 1$, the same as for a perfect fit). The mean bias and normalized mean bias likewise can have 0 values (the same as for a perfect model) even if all predicted values are extremely wrong (e.g., way too high for all small values and way too low for all high values). Also, it would be more informative to show entire error distributions instead of just summary statistics.

Complexity and Level of Effort

Exhibit 3-8 is informative. This table describes the format of each of the modeling tools and qualitatively evaluates them according to their pre- and post-processing requirements, time requirements, and level of skill and software required. The APX tools require MATLAB, which (unlike R) is not a free package, but this should not deter an agency responsible for national policy. MATLAB is a very well-established package and its mathematical routines are widely applied across many scientific disciplines.

RECOMMENDATIONS:

The Panel recommends that the description of the method and interpretation of scaling of $PrPM_{2.5}$ (EC) to $PrM_{2.5}$ (EC+OC+crustal) be enhanced in the EPA's draft report.

The Panel recommends that entire error distributions be shown instead of just summary statistics when comparing RFTs to FFMs.

The Panel suggests testing whether distributions of observed and model-predicted values (or full- and reduced-form analysis results) are significantly different from each other; and using visualizations such as regression diagnostics to understand when and how the different reduced-form model predictions differ significantly from each other and from full-form results.

2.3. Charge Question 3. Suitability of RFTs

2.3.1. ***Charge Question 3a****. Does the report provide a clear and thorough explanation for why some tools under- or over-estimated $PM_{2.5}$ health benefits as compared to the full-scale air quality modeling? Please add any additional explanations for the pattern of results observed.*

Exhibit ES-3 of the EPA's draft report and the corresponding Exhibits 3-3 and 3-2 (which present the same data in different format) and C-1 (the raw data from which the figures are derived) generally present a useful picture of how the different RFTs perform with respect to the five specific emissions control scenarios evaluated in the EPA's draft report. Exhibit 3-4 provides further assistance in understanding how well the different components of $PM_{2.5}$ (primary $PM_{2.5}$ or $prPM_{2.5}$, sulfates and nitrates) are reproduced by the RFTs. Two specific suggestions to improve the plots (also made in response to Charge Question 2) are to use a logarithmic scale on the horizontal axis, and to use the same scale for all the plots. Those changes would make it easier to compare cases where RFTs underestimate CMAQ outputs with cases where they overestimate and would ensure that the ratios for different $PM_{2.5}$ components are comparable. One specific example is for the nitrate component of the AP3-BenMAP model on the Pulp and Paper policy scenario, where the RFT underestimates the CMAQ estimate by a factor of 18 (the largest relative error of any comparison in the EPA's draft report) but this in no way stands out from Exhibit 3-4. The Panel also noted some minor discrepancies between the ratios plotted in Exhibits 3-2 through 3-4 and the raw numbers derived from Exhibit C-1 – these are not big enough to affect any of the recommendations, but care should have been taken to ensure the results are internally consistent.

Considering Exhibit 3-4, it seems clear that the biggest discrepancies, between benefits calculations for FFMs and RFTs, are for the nitrate components of the models. The problems are less severe for the SA-Direct and EASIUR models than for the APX class of models or for InMAP, though even for SA-Direct and EASIUR, the discrepancies are large enough to cause concern. The discrepancies are less severe for the sulfate component, except for the APX models applied to the Tier 3 scenario.

While the report generally does a good job of explaining HOW the model results differ, it generally fails to explain WHY. It is not clear whether this question was included in the scope the EPA's draft report, since the authors explicitly noted that they were not expected to change any of the basic model parameters (which would typically be needed to do a causal analysis).

Reasons why some models outperformed others could be better understood if the detailed surfaces (of $PM_{2.5}$ and its constituents) that are produced by some of the models were provided. The Panel's understanding is that this should be possible for each of the models whose intermediate air quality outputs could be input to BenMAP, but it may not be possible for the various "Direct" implementations of the RFTs.

The comparisons effectively treat the CMAQ-BenMAP approach as "ground truth." The justification for this is not clear, though the Panel acknowledges that CMAQ is well-established and generally believed to perform well in most scenarios. The good agreement between CMAQ and CAMx is further evidence that CMAQ is performing well in the cases where both models were run, but the Panel notes that the one case for which CAMx was not run (Tier 3 – this is the only scenario examined that involved mobile sources) is also the scenario that produced the biggest overall discrepancies between the FFM and RFTs.

Two parameters that are critical to the overall benefit estimates, but were not varied for this exercise, are the value of a statistical life/life-year (VSL/VSLY) and the concentration response relationship (CRR). The EPA's draft report makes the argument that changing either of these would not affect the comparisons between FFMs and RFTs, but matters are not so simple if the VSL and/or CRR are dependent on location, age, and in the case of CRR, the level of $PM_{2.5}$ (i.e., a nonlinear response curve). For example, the plots in Chapter 2 show that there are substantial variations in how different policy scenarios affect different regions of the country, so if the CRR or realization of VSL/VSLY are spatially dependent as well, that could materially affect the comparisons.

For the VSL, EPA has taken the value $8.7 million (in 2015 dollars) but has made no attempt to assess the uncertainty of that estimate. There is a substantial economics literature on this topic, which also addresses how the VSL varies by region, age, co-morbidities, competing risks, and other factors, that have not been addressed.

For the CRR, the EPA's draft report states that it uses an estimate from a 2009 report by Krewski *et al*. (2009), a Health Effects Institute study based on the American Cancer Society dataset but does not specify which of the numerous hazard rate estimates in that report is used. There is evidence that the CRR varies regionally, and is a nonlinear function of $PM_{2.5}$, both of which could affect the comparisons in the EPA's draft report. Moreover, the Krewski 2009 report did not address whether the regression relations they derived were causal; if they were not, their translation into estimated benefits would not be appropriate.

Insights from atmospheric chemistry

Although the EPA's draft report does not discuss the root causes of the discrepancies between CMAQ and the RFTs, the review panel feels that some explanation may be possible based on the atmospheric chemistry involved.

The relatively minor differences observed in $prPM_{2.5}$ concentration fields are likely because $prPM_{2.5}$ results are driven more by transport (advection and diffusion) rather than chemistry. There are added complexities associated with secondary $PM_{2.5}$ formation due to photochemistry and aerosol dynamics. For example, production of sulfate and nitrate is related to ozone formation and the presence of OH·radicals. When photochemical activity is diminished (e.g., during nights and winters) or under high NOx conditions (e.g., in inner cities with high vehicular emissions), NOx can titrate ozone and slow the secondary formation of sulfate and nitrate. Under certain atmospheric conditions, reductions in NOx emissions can actually increase nitrate and sulfate formation. In addition, free ammonia in the atmosphere has a significant impact on the formation of nitrate PM since the nitrate must be fully neutralized with ammonium (ammonium nitrate, NH_4NO_3). However, the amount of free ammonia will have a smaller impact on the formation of sulfate since sulfate can exist as ammonium sulfate ($(NH_4)_2SO_4$ which is fully neutralized), ammonium bisulfate ($(NH_4)HSO_4$ which is half neutralized), or sulfuric acid mist (H_2SO_4 which is not neutralized). If the EPA's draft report had performed the analysis in two separate steps with the first part comparing concentration fields generated by RFMs and FFMs and the second part comparing monetized benefits estimated by each RFM and BenMAP using a consistent concentration field, it would be much easier to distinguish estimated benefit differences due to air quality concentration fields compared to the estimated monetized benefits step.

The Panel was also uncertain whether the air quality surfaces generated by a RFT would be altered/normalized (as discussed below) before use, or if they are directly applied to the benefits assessment. In most regulatory applications, full-form model results are not used directly. EPA's "Modeling Guidance for Demonstrating Air Quality Goals for Ozone, $PM_{2.5}$ and Regional Haze" (EPA, 2018) recommends that when air quality models are used for regulatory application to predict future year concentrations and future year control scenarios, the models be used in a "relative" sense rather than an "absolute" sense, as noted below:

> *Air agencies should determine whether a control program scenario will provide sufficient emission reductions to demonstrate attainment of the NAAQS using the modeled attainment test. The modeled attainment test is a technical procedure in which an air quality model is used to simulate base year and future air pollutant concentrations for the purpose of demonstrating attainment of the relevant NAAQS. The recommended test uses model estimates in a "relative" rather than "absolute" sense to estimate future year design values.*
>
> *…this approach has the effect of anchoring the future concentrations to a "real" measured ambient value, which is important given model bias and error in the base year simulation(s). It is reasoned that factors causing bias (either under or over-predictions) in the base case will also affect the future case.*
>
> *The EPA has developed the Software for Modeled Attainment Test-Community Edition (SMAT-CE) tool to enable completion of the modeled attainment tests for $PM_{2.5}$ and ozone, as well as for calculating changes in visibility in Class I areas.*
>
> *The modeled attainment test is primarily a monitor-based test. As such, the focus of the attainment test is whether attainment can be reached at existing monitors. An additional "unmonitored area analysis" can also be performed to examine ozone and/or $PM_{2.5}$ concentrations in unmonitored areas.*

Many times, absolute modeled nitrate concentrations are significantly over-predicted by the full-form photochemical models. The approach described above reduces biases in future year projections and policy scenarios by using the model in a "relative" sense rather than an "absolute" sense. Therefore, the future nitrate concentrations calculated with SMAT-CE can be significantly lower than the absolute nitrate concentrations directly from the model.

The Panel observed that this approach was not used for the analysis described in the EPA's draft report. This raises the issue of whether the use of the RFTs in the relative sense described could mitigate some of the discrepancies noted, especially in the nitrate component.

RECOMMENDATIONS:

The Panel recommends the use a logarithmic scale on the horizontal axis, and to use the same scale for all the plots.

The Panel recommends that the EPA clarify if the air quality surfaces generated by a RFT would be altered/normalized before use, or if they will be directly applied to the benefits assessment.

The Panel recommends that the EPA clarify if the RFTs should be used in a "relative" sense rather than an "absolute" sense.

2.3.2. Charge Question 3b. *How do the results of this study inform our understanding of the suitability of these tools for regulatory economic analyses in their current form?*

In general, the results show that SA-Direct and EASIUR provide better agreement with CMAQ than InMAP or the APX class of models, with SA-Direct generally overestimating and EASIUR underestimating benefits (except for the Tier-3 scenario). However, the Panel does not support replacement of CMAQ with RFTs based on the limited information provided in the EPA's draft report. Some of the concerns are outlined below:

1. A wider range of policy scenarios is needed to assess the robustness of the RFTs under realistic conditions. For example, only one of the scenarios considered (Tier 3) involved mobile sources. Furthermore, policies that involved larger changes in emissions could well imply worse behavior of the RFTs because of nonlinearities in the underlying dynamics.
2. The acceptability of a model could depend on what it is used for. It may be acceptable to use an RFT in an initial scoping exercise, when EPA is considering several versions of a new rule prior to recommending one for public consideration, but not for a Regulatory Impact Analysis conducted pursuant to Executive Order 12,866 and related requirements.
3. Measures that would assess the agreement between an RFT and the FFM with which it is being compared need to be better defined. The EPA's draft report used five measures of agreement, including Normalized Mean Bias (NMB) and Normalized Mean Error (NME). Instead of NMB and NME, Boylan and Russell (2006) suggested Mean Fractional Bias (MFB) and Mean Fractional Error (MFE). The difference among these measures is in the denominator: for NMB and NME, the denominator is mean observation, but for MFB and MFE, it is the average of mean observation and mean model value. MFB and MFE are more symmetrical when comparing models that overestimate or underestimate the true value by the same fraction. For example, consider a case where the observation is 1 and the model value is 0.5. The NME is (1-0.5)/1=0.5 and the MFE is (1-0.5)/0.75=0.667. Now let the model value be 2 instead (still a 2:1 ratio

between the two values). The NME is now (2-1)/1=1 and the MFE is (2-1)/1.5=0.667. The MFE is the same in both cases but the NME is different. Guidelines for what is an acceptable model error were proposed by Boylan and Russell (2006) and Emery (2017) and could be considered by EPA as general criteria. Other measures that could be considered include the distribution of the difference between RFT and FFM results for a large set of randomly sampled scenarios, or the maximum possible difference for a given set of scenarios.

4. The Panel was critical of the fact that the EPA's draft report focused entirely on point estimates and did not give any consideration to the variability or uncertainty of those estimates. In principle, testing the agreement of one model with another could be viewed as a hypothesis testing problem with carefully defined null and alternative hypotheses, and type I and type II error rates. In the climate modeling literature, ensembles (collections of model run with variations in initial conditions and model parameters) are increasingly used as a means of assessing the natural variability of model predictions. Use of ensembles provides a more rigorous separation of bias and variance and could be valuable in the context of air quality models as well.
5. All the comparisons were based on treating CMAQ + BenMAP as ground truth, though in four of the five scenarios (the exception being Tier-3) there was also a comparison with the CAMx model, with good but not perfect agreement. Some assessment should have been made of the uncertainty in CMAQ + BenMAP as well.
6. Exhibit 3-8 provides helpful information about the time requirements and ease of implementation for each RFT. Given applicable information quality requirements (OMB, 2002; Environmental Protection Agency, 2002) to ensure the transparency and reproducibility of information it relies upon and/or disseminates (including data, models, and analyses thereof), this should also be taken into account in assessing which RFT (if any) to use. In this exercise, the two models that performed best on the benefit comparisons (SA-Direct and EASIUR) were also the ones that were judged quickest and easiest to run, but that may reflect exogenous factors such as the analysts' baseline familiarity.
7. Finally, the Panel urges EPA to consider overall costs to the Agency and the public, and not focus exclusively on the costs to EPA of producing the estimates themselves. The differences in running times of RFTs *versus* FFMs may be significant relative to the total modeling effort, but they are still relatively minor when compared with all the Agency and social costs of introducing a new rule. Just within EPA, this includes the costs of receiving and responding to public comments, and even the possibility of having to respond to litigation should EPA's modeling efforts be challenged by an outside group.

2.3.3. Charge Question 3c. *Can any of the reduced-form tools explored in this report easily be modified to allow quantifying the extent to which the total health benefits accrue to specific geographic areas (e.g., by state, or where ambient concentrations are above or below the NAAQS)?*

It should be straightforward to modify the SA-Direct and EASIUR methods to produce results at a regional/state/county level, and to use the APX and InMAP models without modification by simply aggregating results at the desired spatial level. However, the Panel questioned the value of doing this, given spatial variability in emissions and human population characteristics. Concerning the accuracy of RFTs on smaller scales than the national level, the Panel concluded that the EPA's draft report provides inadequate evidence in support of such applications.

In the EPA's draft report, Exhibit C-2 covered results for seven regions that were defined in Appendix B. No information was provided that would allow an assessment of the RFTs at a state or county level. However, even Exhibit C-2 is extremely limited in its usefulness. Consider the following table:

Table 4

NITRATE COMPONENT: Comparison of National RFT Costs by Exhibit C-1, C-2							
		National Estimate			Regional Estimate		
	CMAQ	AP2-B	AP3-B	InMAP	AP2-B	AP3-B	InMAP
CPP	1700	3400	720	11000	3463	761	10966
CK	600	990	350	3200	987	352	3232
P&P	130	250	7	740	243	2	740
Ref.	160	640	470	1500	639	472	1521
T-3	1900	7300	4600	11000	7274	4627	10602

Table 4, above, compares (for the nitrate component only) the national estimates for three of the RFTs, computed by two different methods. The "National Estimates" are direct quotes from Exhibit C-1. The "Regional Estimates" are computed by taking the mean biases from Exhibit C-2, converted to millions of dollars, multiplying by 7 to convert from a mean bias to a total bias, and adding the CMAQ + BenMAP benefit estimate. Ideally, the two ways of calculating the national RFT benefit estimate should be the same. One might expect small discrepancies because of rounding errors, possible missing values in some of the cells, and similar features. Most of the discrepancies between national and regional estimates are within the range that could plausibly be accounted for in this way.

The "Mean Error" values in Exhibit C-2 are generally similar to the Mean Bias values, but sometimes larger when there is presumably cancellation among bias terms of opposite signs. The "Normalized Mean Bias" and "Normalized Mean Error" terms are essentially calculated from the Mean Bias and Mean Error by dividing by the total benefit, and the Panel has already argued in response to Charge Question 3b that it would have been better to use Mean Fractional Bias or Mean Fractional Error. However, none of these measures adds new information in the regional results that was not already implicit in the national results. The only part of Exhibit C-2 that contains new information is the R^2 values, but for the sulfate and nitrate components, nearly all of the R^2 values are less than 0.9, in some cases very much less, so they don't provide reassurance about the performance of the RFTs at the regional level. In addition, a high R^2 does not guarantee good performance, as the estimates may still be biased by location or scale without affecting R^2.

It would be better if the EPA's draft report had provided regional results directly, rather than as summary statistics in Exhibit C-2. There could be regions where the RFTs do much better than others, and this might be associated with the different regional impacts of the policy scenarios, as documented in Chapter 2. However, the information provided in the EPA's draft report does not allow the review panel to make that assessment.

Another aspect that the EPA's draft report does not explain is why the regional results were only presented for the models that use BenMAP, though it appears that some evaluation was also done for the DIRECT models (e.g. line 12, page 4-2, reports an R^2 value for EASIUR, but this does not correspond to anything in Exhibit C-2).

Without having better information about the regional results, it is impossible to assess performance on smaller scales such as states or counties. (The Charge Question does not ask directly about county level estimates, but it does ask about comparing sites that were or were not in compliance with the NAAQS; but compliance is typically assessed at county level, so one would need county-level estimates in order to answer this question.) The EPA's draft report does make the important point that for each of the "Direct" RFTs, the benefit attributed to an emission reduction is associated with the location where the emission reduction occurs, and not where benefits presumably would be realized (and after what lag), which is typically different because of transport of pollutants through the atmosphere. This could create a bias in the results even at regional level, and almost certainly at a state or county level.

Some RFTs develop fixed estimates of benefits-per-ton which are then used to estimate total benefits of an emissions reduction scenario by multiplying those benefit-per-ton values by the total emissions reduction of the scenario.[25] The problem with such an RFT is that its benefit-per-ton estimates are invariant to *where* the tons of reduction occur. For example, if such an RFT has been developed with national benefit-per-ton estimates for each pollutant, that RFT would produce the same national benefit estimate for a 1,000 ton reduction of pollutant A from Sector X occurring entirely in Maine as it would for a 1,000 ton reduction of pollutant A from Sector X occurring uniformly across the U.S. Clearly the location of those benefits would differ significantly, but the RFT would not be able to inform this question at all.[26] While such an RFT could be enhanced to produce benefit-per-ton estimates that differ for discrete regions, errors in

[25] SA-Direct is an RFT that works in this manner, although it is not clear from the EPA's draft report which other RFTs produce benefit-per-ton values that cannot be altered without returning to more runs of the original full-form model. The report should be clearer on this matter because of how much the reliance on benefit-per-ton estimates constrains the ability of an RFT to disaggregate total benefits to states, counties or areas with certain air pollutant levels.

[26] And because emissions do produce different impacts depending on their source/location, the two national total estimates will have different degrees of error, also unknowable. The use in the report of only one type of emissions reduction scenario for each of the sector-specific comparisons prevents the report from showing this fact, and the extent to which such variance may differ among RFTs that use fixed benefit-per-ton values versus those that do not (if any). As mentioned elsewhere in the Panel's report, the lack of evidence of the variance in errors for different types of sector-specific reduction scenarios makes it impossible to assess whether the report's point estimates of differences in RFT performance are indicative of systematic differences in performance or merely one random set of outcomes.

both regional and total RFT benefit estimates will still depend on the spatial match between the emissions reductions assumed in the original full-form model run and those of the policy scenario being evaluated using the regionally-disaggregated RFT.

This problem is exacerbated by the benefit-per-ton estimates being based on specific assumptions about a single concentration response relationship (CRR) and, for whatever form assumed, this CRR assumption makes the benefits estimates invariant to the concentration of pollutants in each location. Thus, a benefit-per-ton RFT approach will not be able to provide information on how much of its estimated benefits occur in areas above or below the NAAQS (or in areas with other concentration ranges). This is an important issue for understanding the sensitivity of regional (and hence also total) benefits to potential alternative slopes of the CRR, given that a number of recent epidemiology papers have argued for a nonlinear CRR, and this could affect the results differentially in different regions. This is another uncertainty in the estimates of RFTs at finer spatial scales that is of concern for RFTs that are characterized by their benefit-per-ton estimates.

Whether any of the RFTs could be modified to produce better results seems impossible to answer based on the information provided in the EPA's draft report. The Panel understands that EPA is in the process of updating the SA-Direct model to incorporate more up to date weather and emission scenarios. The Panel encourages this update and recommends EPA continue to work with other model developers to address the discrepancies between RFTs and the FFMs revealed by the EPA's draft report. The Panel advises against making some simple adjustment, such as an overall rescaling of results from any of the RFTs; the range of policy scenarios is too limited, and the performance of the RFTs on regional scales is too unclear, to recommend any such adjustment with confidence.

RECOMMENDATIONS:

The Panel recommends that the EPA's draft report provide regional results directly, rather than as summary statistics.

The Panel recommends that the EPA's draft report should address concerns related to RFTs that develop fixed benefits-per-ton estimates which constrain their ability to disaggregate total benefits to states, counties or areas with certain air pollutant levels.

The Panel recommends that the EPA increase the range of policy scenarios and provide more information to clarify the performance of the RFTs on regional scales.

2.4. Charge Question 4. BPT approaches

Question 4. *Since 2008 EPA has used SA-BPT to estimate the health impacts of numerous regulations. Under the scenarios examined in this report, EPA's SA-BPT approach over-estimated* $PM_{2.5}$*-related health benefits by between 10 and 30 percent, depending on the sector. To ensure BPT estimates correspond to full-form results as closely as possible, the report* *recommends updating the underlying emissions inventories and air quality modeling used to inform the EPA SA-BPT approach over time.*

2.4.1. Charge Question 4a- *In the interim, how might EPA improve its characterization of results derived from the 2005 SA-BPT approach, specifically the potential degree of over- or underestimation in BPT-based results for a particular regulatory scenario?*

The Panel supports the recommendation that the SA model should be updated to reflect more recent emissions inventories and air quality modeling. There are several areas in which additional modeling and evaluation is extremely important.

- Additional policy scenarios should be run for one or more select RFTs, and the results compared against the full form models to refine understanding about the degree of over- or under-estimation of the RFTs.

- Additional sensitivity analyses are needed to discern which inputs are playing the largest roles with respect to divergence of the RFT results from the full form models, and under what conditions or policy scenarios these divergences are greatest. The EPA's draft report's disaggregation of differences for the national benefits ratios (Exhibit 3-4), suggests that for all RFTs, SA-BPT included, the greatest divergence is for nitrate. This raises a question as to whether nitrate is playing an outsized role in overestimation of total $PM_{2.5}$, and if so, additional explanation is warranted.

- Additional uncertainty analysis is needed – aimed at characterizing and representing all the key forms of uncertainty in the estimate, and not limited to those that have been quantified in the EPA's draft report. Examples of assumptions that should be examined in such an analysis include the following:

    - Assumptions about the C-R relationship (CRR) and VSL/VSLY have been incorporated in this set of model comparisons. These assumptions may be more influenced by analyst or policy judgment than scientifically defined phenomena. The benefits estimates vary much more widely with changes in these assumptions than with variation in air quality changes associated with different choices of models.

    - Assumptions about the (uncertain) slope of the CRR and relative toxicities of $PM_{2.5}$ constituents. These assumptions affect the *relative* performance of each model at the regional level, which is the level at which the RFTs differ most markedly, as opposed to the national level. The claim in the Limitations section of the EPA's draft report that changing assumptions about CRR is not expected to change the relative performance of the models is almost certainly incorrect. This might be true if the only uncertainty in the CRR assumption were related to slope – but this is not the case, as non-linearity, differential toxicities, and issues about causality are more important and unexplored forms of CRR uncertainty.

Overall, the Panel notes that the EPA's draft report should reference earlier work constituting critical reviews of CMAQ and/or CAMx – with respect to how well they represent reality. The EPA's draft report is concerned solely with how closely CAMx and the selected RFTs approximate CMAQ. The EPA's draft report does not analyze how closely each model would be expected to align with observations.

The Agency should be clearer with respect to how it intends to use RFT results – i.e., as screening tools to produce "ballpark" estimates (see EPA's draft report at ES-7 and 5-1), or as substitutes for FFMs like CMAQ and CAMx (implied by Question 3(b) of the Charge). The Panel does not believe any of the RFTs are appropriate replacements for FFMs at this time, although they may be useful for screening purposes that do not substitute for or displace FFMs in regulatory impact analysis.

Finally, it would be misleading for EPA to rely solely on the results of this comparative analysis to "improve its characterization" of potential over- or underestimation of benefits using the SA-BPT (or SA-Direct) approach because that characterization would presume that all relevant potential sources of uncertainty in those estimates were evaluated in the EPA's draft report. As explained above, this range of uncertainty would be misleading because it would lack representation of the additional uncertainties not explored in the EPA's draft report, such as potential CRR non-linearities and the potential for $PM_{2.5}$ constituents to have non-equal toxicities. Further, an additional unexplored source of uncertainty relates to the choice of policy scenarios, as the five scenarios evaluated here are too narrow a group to allow analysts to make general statements.

RECOMMENDATIONS:

The Panel recommends that additional policy scenarios be run for one or more select RFTs and the results compared against the full form models to refine understanding about the degree of over- or under-estimation of the RFTs.

The Panel recommends that additional sensitivity analyses be conducted to discern which inputs are playing the largest roles with respect to divergence of the RFT results from the full form models, and under what conditions or policy scenarios these divergences are greatest.

The Panel recommends that additional uncertainty analysis to characterize the key forms of uncertainty in the model estimates.

2.4.2. Charge Question 4b- *What criteria (e.g., geographical scale, regulated sector, pollutants/precursors) should EPA examine to determine the potential for divergence between SA-BPT results versus full-form air quality modeling results (resulting in over- or under-estimation)?*

The Panel sees an opportunity to use criteria such as those outlined in the charge question to gain insight. Additional analyses based on geographical scale, demographics, regulated sector and different pollutants/precursors could be evaluated to further test the agreement between the SA and CMAQ models. If EPA wishes to understand what contributes to differences across models, it can compare air quality changes in the underlying air quality grid to those of CMAQ and identify where such *air quality* projections appear to have the largest error, rather than focusing on dollar value or mortality differences. Quite a bit of insight could be gained by comparison of air quality surfaces.

Regarding geographic scale - a separate set of BPT values should be generated for several geographic area subsets, and for each regulated sector, to allow additional sensitivity analyses. There is currently only one set of BPT values for each sector, originally derived from a BenMAP analysis using full form air quality projections from assumed emissions changes in that sector within the contiguous US.

The EPA should assess performance for a more recent mobile source scenario encompassing more contemporary aspects. For example, the EPA's draft report states that the SA BPT values that were applied do not vary with emission height. This is acknowledged as a potential contributing factor as to why RFT results deviate for the Tier 3 scenario. Also, on page 4-4 the authors note that "the fact that the Tier 3 scenario is exclusively comprised of ground-level emissions may be a secondary contributing factor, as may the use of a different base year emissions inventory (2005) than the other policies."

RECOMMENDATION:

The Panel recommends that EPA examine model performance for more scenarios such as area sources (e.g., residential wood combustion), marine/aircraft/rail (MAR) sources, additional industrial point sources (e.g., iron/steel), and on-road diesel emission reductions, in addition to examining $PM_{2.5}$ components ($prPM_{2.5}$, sulfate, nitrate, ammonium, SOA).

2.4.3. Charge Question 4c. *Based on the results of this study, does the panel have any additional recommendations about BPT-based approaches?*

Overall, the question of the "suitability" of using RFTs must be tied to the question, "For what purpose would they be used?" The charge question appears to mean "for use in final regulatory impact analyses," but elsewhere the EPA's draft report suggests using them only to produce "ballpark" estimates (e.g., see pp. ES-7 and 5-1). These alternative uses are not compatible. While RFT estimates can play a useful role in screening analyses, the EPA's draft report does not provide insight concerning how well RFT-based benefit estimates can meet the appropriate degrees of accuracy and precision needed for various types of purposes. BPT-based approaches face challenges with handling non-linear atmospheric processes that affect the spatial patterns of secondary forms of $PM_{2.5}$. Because these spatial uncertainties may average out over larger regions, errors in BPT-based benefits estimates might be less pronounced when aggregated over very large geographic scales (e.g., national scale). However even this may not be the case because population densities can differ substantially over the same geographic scale as the $PM_{2.5}$ change uncertainty. Also, BPT-based approaches that focus on aggregate national values by design diminish the policy relevance of variability across space, time, age, and a host of other important factors. Given these concerns, the EPA should clarify where and under what conditions it envisions using RFTs in lieu of full form approaches.

More exploration should target whether selection and/or use of RFTs should depend on the specific characteristics of the policy scenario of interest or on other factors. The EPA's draft report suggests that point source emissions are generally better approximated by RFTs than are mobile source emissions for example – although this finding is based on only one policy scenario for EGU, and only one for mobile sources thus far. Additional policy scenario modeling will be needed before the suitability of RFTs can be fully understood.

A limitation in using BPTs is that they do not allow users to test the sensitivity of projected benefits to alternative slope of the CRR and potency assumptions. The question of nonlinear CRRs is unresolved at the present time (see EPA-CASAC-19-002). If all RFTs are equally unable to perform sensitivity analyses related to CRR, then this represents an important trade-off when deciding to use a RFT approach over a complex full-form benefits analysis. However, if some of the RFTs under consideration do allow CRR sensitivity analyses to be conducted, that would be an important positive attribute for those RFTs compared to more rigid BPT-based approaches. Whether some of the RFTs have this greater flexibility has not been evaluated in the EPA's draft report but should be incorporated into decisions about whether to use any particular RFT in a BPT-based approach instead of a full-form analysis.

Any BPT and/or RFT approach that utilizes a source apportionment approach to underlying source-receptor relationships may suffer performance issues when direct/indirect NH_3 effects are involved in secondary inorganic $PM_{2.5}$ formation. For policy applications, it might be more appropriate to use a sensitivity approach, i.e., associating change in concentrations with change in specific emissions such as with the Brute Force method (Hwang *et al.*, 1997 and Clappier *et al.*, 2017) or High-Order Decoupled Direct Method in Three Dimensions (Zhang *et al.*, 2012 and Huang *et al.*, 2017). For an area where nitrate formation is limited by available NH_3, source apportionment may indicate that NOx emissions from the EGU sector contribute 50% of the nitrate concentration. This implies that the removal of all EGU NOx would result in a 50% reduction of nitrate in the area; however, a brute force sensitivity analysis may show by contrast that the nitrate concentration is unchanged when all EGU NOx is removed. Specifically, EPA should review Clappier *et al.*, 2017 and discuss the potential impacts of excluding brute-force runs when accounting for complex PM chemistry. Finally, EPA should discuss the computation benefits of source apportionment approaches, which can generate multiple scenario contribution tags in a single model run *versus* brute force approaches which require a new full-form model run for each sensitivity scenario.

Treatment of uncertainty should be augmented as the RFTs are assessed and their results are compared to full form models. Although uncertainty in some factors is found to be fairly consistent across the board for all RFTs, there are particular sources of uncertainty that make highly variable contributions to overall outputs depending on both model particulars and policy scenario. With the limited set of policy scenarios available in the EPA's draft report, it is difficult to gauge uncertainty structure and contribution related to scenario context.

Finally, the absence of CAMx-based full form Tier 3 results for comparison with CMAQ raises questions about how to interpret comparisons between full form models and reduced-form models. Whether CMAQ and CAMx align well with each other for point source scenarios does not contribute insight into how comparisons across mobile source scenarios might align.

RECOMMENDATIONS:

The Panel recommends that the EPA clarify where and under what conditions it envisions using RFTs in lieu of full form approaches.

The Panel recommends that EPA investigate which of the RFTs under consideration allow CRR sensitivity analyses to be conducted as this would be an important positive attribute for those RFTs compared to more rigid BPT-based approaches and preference given to incorporate RFTs displaying greater flexibility in a BPT-based approach when a full-form analysis is not feasible.

The Panel recommends that EPA provide a fuller discussion of uncertainties associated with the use of RFTs when compared to FFMs.

The Panel recommends that EPA include of CAMx-based full form Tier 3 results to determine how they may differ from those of CMAQ results in comparison with RFTs.

2.5. <u>Charge Question 5. Reliability of RFTs</u>

<u>Charge Question 5</u>. *How do the results of this study inform the future development of reduced-form tools that are capable of providing reliable estimates of impacts associated with different sectors, across a variety of spatial scales, and for different portions of the air quality distribution? Are there other, less resource intensive approaches than full-scale air quality modeling for informing the public about the size and distribution of PM health benefits associated with alternative regulatory scenarios?*

The results of the EPA's draft report suggest that none of the RFTs that were evaluated consistently reproduce the FFM. However, some RFTs might be useful for certain sectors when used for screening purposes at the national scale. For example, SA-Direct and EASIUR reduced-form models, which require less time and technical expertise than the other RFTs, can produce results that are within a factor of two of the FFMs. The EPA's draft report highlights several reasons for the deviations between the RFTs and the FFMs. Addressing the reasons for these differences by evaluating concentration fields and benefits estimates separately can inform the future development of new RFTs. Specifically, ground-level emissions and non-linear nitrate formation are not well characterized by the RFTs and should be further investigated.

Since the performance of the RFTs varies with policy scenario, additional policy scenarios will need to be modeled in order to better understand the differences. Along with additional scenarios, model performance should also be evaluated for different levels of emissions changes (e.g., 20%, 40%, 60%) within the same type of scenario since regulatory analyses many times use the differences between modeled alternative scenarios rather than the absolute numbers from the benefits analysis. Model results from additional scenarios with different levels of emission changes could be used to help provide guidance on when it might be appropriate to apply specific RFTs.

In some situations, using multiple models to produce an average result can lead to better performance compared to the individual models. Also, ensemble modeling can sometimes be used to produce a general range of benefits, noting that the actual benefit may be outside the upper/lower bounds of the range. However, additional research involving more models and more scenarios is required before it would be appropriate to combine RFTs in these ways. The development of performance guidelines for acceptable model performance would also help to guide model choice and improvement. In order to obtain reliable RFTs that are tailored to a variety of emission reduction scenarios, many FFM and RFT runs would have to be performed. Even doing that leaves many important sources of uncertainty in the RFTs uncharacterized.

It is important to understand when RFT estimates can be helpful to guide policy decisions and when they are too unreliable to be used to inform a policy decision. BPT estimates may be useful for screening out or refining potential regulatory options before reaching the proposed rule stage, even if they are deemed too unreliable to be used to inform the public about the benefits of a proposed or final regulatory option. The EPA's draft report should have included a discussion on the usefulness of RFTs in different parts of the regulatory decision process. In the future, the concept of data quality objectives and performance criteria may be useful to determine when and where these models should be used.

The Panel recommends that the RFTs be updated each time updates are made to CMAQ or BenMAP. In addition, RFTs that rely on concentration fields from FFMs other than CMAQ/CAMx might benefit by switching to CMAQ/CAMx.

A discussion of additional RFTs and less resource intensive approaches than full-scale air quality modeling should have been included in the EPA's draft report. On page 2-9, the EPA's draft report states, "We conducted an extensive literature review to identify reduced-form approaches for predicting policy-related air quality changes and associated benefits.[10] Based on this review, we selected four reduced-form tools for this analysis." The detailed literature review to identify all reduced-form approaches and the selection of the four reduced-form tools are a critical part of this report. Footnote "10" refers to a personal communication memorandum (November 17, 2017). The reference to a single personal communication memorandum does not capture the scope of "an extensive literature review." The EPA's draft report should include a copy of the personal communication memorandum in the Appendix. Also, the EPA's draft report should list all references that were reviewed and list all the RFTs that were considered for selection, including but not limited to ABaCAS (http://www.abacas-dss.com/abacas/Default.aspx).

Finally, the EPA's draft report should explain why the four RFTs were ultimately selected for this report while others were not selected.

RECOMMENDATIONS:

The Panel recommends that ground-level emissions and non-linear nitrate formation be better characterized by the RFTs and be further investigated.

The Panel recommends that model performance also be evaluated for different levels of emissions changes (e.g., 20%, 40%, 60%) within the same type of scenario.

The Panel recommends that EPA provide a discussion on the usefulness of RFTs in different parts of the regulatory decision process.

The Panel recommends that the RFTs be updated each time updates are made to CMAQ or BenMAP.

REFERENCES

Boylan, J.W. and Russell, A.G., 2006. PM and light extinction model performance metrics, goals, and criteria for three-dimensional air quality models. Atmospheric Environment 40 (26), 4946-4959. DOI: 10.1016/j.atmosenv.2005.09.087

Clappier, A., Belis, C.A., Pernigotti, D., and P. Thunis. 2017. Source apportionment and sensitivity analysis: two methodologies with two different purposes. Geosci. Model Dev. 10, 4245–4256.

Emery, C., Liu, Z., Russell, A.G., Odman, M.T., Yarwood, G., Kumar, N., 2017. Recommendations on statistics and benchmarks to assess photochemical model performance. Journal of the Air & Waste Management Association 67, 582-598.

Hwang, D., Byun, D. W., and M.T. Odman. 1997. An automatic differentiation technique for sensitivity analysis of numerical advection schemes in air quality models. Atmospheric Environment 31(6): 879-888

Huang, Z., Hu, Y., Zheng, J., Yuan, Z., Russell, A.G., Ou, J., and Z. Zhong. 2017. A New Combined Stepwise-Based High-Order Decoupled Direct and Reduced-Form Method to Improve Uncertainty Analysis in PM2.5 Simulations. *Environmental Science & Technology 51* (7), 3852-3859. DOI: 10.1021/acs.est.6b05479

Fullerton, D. (2017). *Distributional effects of environmental and energy policy*. Taylor and Francis. https://doi.org/10.4324/9781315257570.

Industrial Economics. 2019. "Evaluating Reduced-Form Tools for Estimating Air Quality Benefits; Final Report." In. Cambridge MA: IEc.

Krewski, Daniel, Michael Jerrett, Richard T. Burnett, Renjun Ma, Edward Hughes, Yuanli Shi, Michelle C. Turner, C. Arden Pope III, George Thurston, Eugenia E. Calle, and Michael J. Thun. 2009. "Extended Follow-Up and Spatial Analysis of the American Cancer Society Study Linking Particulate Air Pollution and Mortality." In *Research Report 140*. Boston MA: Health Effects Institute.

Office of Management and Budget. 2002. 'Guidelines for Ensuring and Maximizing the Quality, Objectivity, Utility, and Integrity of Information Disseminated by Federal Agencies; Notice; Republication', *Federal Register*, 67: 8452-60.

OMB Circular A-4, "Regulatory Analysis," September 17, 2003. The circular is available at http://www.whitehouse.gov/omb/assets/regulatory_matters_pdf/a-4.pdf. The circular took effect for economically significant proposed rules on January 1, 2004, and for economically significant final rules on January 1, 2005.

U.S. Environmental Protection Agency. 2002. "Guidelines for Ensuring and Maximizing the Quality, Objectivity, Utility, and Integrity of Information Disseminated by the Environmental Protection Agency " In *EPA/260R-02-008*. Washington DC: USEPA.

U.S. Environmental Protection Agency. 2015. "Peer Review Handbook." 4th Edition, In *EPA/100/B-15/001*. Washington DC: USEPA/Science and Technology Policy Council.

U.S. Environmental Protection Agency Office of Air Quality Planning and Standards. 2020. "Evaluating Reduced-Form Tools for Estimating Air Quality Benefits: Presentation to the SAB reduced form modeling peer review panel." In. Research Triangle Park NC: USEPA/OAQPS.

USEPA (2014). Regulatory Impact Analysis for the Proposed Carbon Pollution Guidelines for Existing Power Plants and Emission Standards for Modified and Reconstructed Power Plants.

USEPA. (2018). Technical Support Document: Estimating the benefit per ton of reducing $PM_{2.5}$ precursors from 17 sectors. Research Triangle Park, NC, Office of Air Quality Planning and Standards. US Environmental Protection Agency. February 2018.

Zhang S, Liu Z, Rosati A, Delworth TL (2012) A Study of enhancive parameter correction with coupled data assimilation for climate estimation and prediction using a simple coupled model. Tellus 64:10963

APPENDIX A: EDITIORIAL CORRECTIONS

On page 3-2, the report states "Some reduced-form tools tend to consistently underestimate CMAQ benefits, while others tend to overestimate." The report should list the tools that fit into each category. It looks like SA Direct and InMAP consistently overestimate benefits; however, none of the tools considered seems to consistently underestimate benefits.

On page 3-4, the report states "AP2 BenMAP, AP2 Direct, and EASIUR Direct all underestimate CMAQ benefits except for Tier 3, while SA Direct, AP3 BenMAP, AP3 Direct, and InMAP BenMAP all overestimate CMAQ results to varying degrees." AP3 BenMAP and AP3 Direct do not overestimate CMAQ results for Pulp and Paper.

On page 3-4, the report states "Of all the models, AP3 BenMAP and AP3 Direct estimates of health benefits are within 10% of CMAQ benefits estimates for more scenarios (3: CPP Proposal, Cement Kilns, and Pulp and Paper) than any of the other reduced form tools." AP3 Direct is within 10% for two scenarios, not three.

On page 3-8, the "Nitrate" chart shows a ratio of 0.0 (in fact 7/130=0.053) for AP3 BenMAP with Pulp and Paper compared to a ratio of 1.8 for AP2 BenMAP with Pulp and Paper and a ratio of 2.4 for AP3 Direct with Pulp and Paper. This large discrepancy between similar models should be examined and explained in the report.