Sample Representativeness: Implications for Administering and Testing Stated Preference Surveys

Proceedings of a Workshop held at Resources for the Future October 2, 2006

Alan Krupnick and David A. Evans
Resources For the Future

January 11, 2008

Prepared under EPA Cooperative Agreement CR83299-01
with the National Center for Environmental Economics,
U.S. Environmental Protection Agency, Washington, DC 20460

Project Officer Alan Carlin
National Center for Environmental Economics,
U.S. Environmental Protection Agency, Washington, DC 20460

Resources For the Future
1616 P Street, NW
Washington, DC 20036

DISCLAIMERS

Although the information in this report has been funded wholly or in part by the United States Environmental Protection Agency under Cooperative Agreement No. CR83299-01 with Resources For the Future, it does not necessarily reflect the views of the Agency, and no official endorsement should be inferred. Mention of trade names or commercial products does not constitute endorsement or recommendation for use.

Alan Krupnick is a Senior Fellow and Research Director at Resources for the Future. David A. Evans was a Research Associate at Resources for the Future when the workshop was planned and conducted and when this summary was written. He is now an Economist with the National Center for Environmental Economics at the U.S. Environmental Protection Agency. The preceding disclaimer applies in that the views expressed in this summary are those of the authors and do not necessarily represent those of the U.S. Environmental Protection Agency.

Furthermore, the workshop participants have not endorsed, either explicitly or implicitly, this summary as a wholly accurate representation of their views.

Sample Representativeness: Implications for Administering and Testing Stated Preference Surveys

Proceedings of a Workshop held at Resources for the Future,
October 2, 2006

Sponsored by the National Center for Environmental Economics,
U.S. Environmental Protection Agency

Acronyms iv
Opening Remarks 1
Agency Perspectives 1
Evidence of the Relation of Response Rates to Sample Representativeness 5
Meta-analysis of Response Rates to Contingent Valuation Studies Conducted by Mail . 13
Open Discussion Regarding First Three Sessions 14
Practitioners' Perspectives 15
Research Frontiers 19
Panel Session on Implications for OMB Review Procedures 23
Organizers' Concluding Thoughts 27
Appendix A: Workshop Précis and Agenda A-1
Appendix B: Workshop Participants B-1
Appendix C: Compendium of Practitioner Perspectives C-1
- Introduction C-2
- Environmental Risk to Children's Health C-3
- Assessing the Value of Improvements to Lakes and Rivers C-6
- Popular Support for Climate Change Policy C-14
- The Economics of Environmental Vulnerability: Epidemiology, Household Status, and Air Pollution C-18
- Valuation of Natural Resource Improvements in the Adirondacks C-20
- Improved Valuation of Ecological Benefits Associated with Aquatic Living Resources C-23
- Willingness to Pay Survey for §316(B) Regulations C-27
- The Nonconsumptive Value of Steller Sea Lion Protection Measures C-30
- Valuing Visibility in National Parks C-33
- Hawaiian Coral Reef Valuation Study C-36
- North Atlantic Right Whale Economic Benefit Study C-39
- Template for Description of Study and Techniques for Addressing Nonresponse . C-42

Appendix D: Workshop Presentations D-1
- Sample Representativeness: Implications for Administering and Testing Stated Preference Surveys (A. Krupnick) D-2
- OMB Guidance and Standards for Statistical Surveys (B. Harris-Kojetin) D-4
- Response Rates and Nonresponse Bias: What We Know and What We Don't Know (R. Groves) D-10

A Meta-analysis of Response Rates to Contingent Valuation Studies Conducted by Mail (K. Boyle) D-17
Practitioners' Perspectives (A. Krupnick) D-21
Should We Move to Panels? (J. Huber) D-23
Scavenging for Covariates to Use in Selectivity Assessment and Correction Strategies (T. Cameron) D-25
Correcting for Nonresponse Bias in Discrete Choice Models: A Two-Stage Mixed Logit Model (D. Hellerstein) D-32
Internet vs. Phone Survey: An Application to Alcohol Research (E. Edwards) D-37

Acronyms

CE: choice experiment (conjoint)
CV: contingent valuation
EPA U.S. Environmental Protection Agency
FCSM: Federal Committee on Statistical Methodology
HI : Harris Interactive
ICR: information collection request
IQA: Information Quality Act
KN: Knowledge Networks
MNL: multinomial logit
NESARC: National Epidemiologic Survey on Alcohol and Related Conditions
NOAA: National Oceanic & Atmospheric Administration
NPS: National Park Service
NSF: National Science Foundation
OMB: Office of Management and Budget
PRA: Paperwork Reduction Act
RDD: random-digit dialing
RV: random components (variables)
SP: stated preference
USFS: U.S. Forest Service
WTP: willingness to pay

Opening Remarks

Alan Krupnick opened the workshop by welcoming the participants and briefly introduced **Phil Sharp**, president of Resources for the Future. Phil provided introductory remarks and noted the value of methodological research for survey design. He believes that improving data collections, by making them effective and efficient, will in turn improve cost–benefit analyses and allow for more informed decisionmaking.

Alan then led a presentation that motivated the workshop and described the goals and plan for the day. In Alan's view credible cost–benefit analyses often require stated preference (SP) surveys to measure benefits. However, he noted that both agencies and researchers perceive that getting Office of Management and Budget (OMB) approval of these surveys is often difficult and time-consuming and noted a few reasons as to why this is the case. Recently approved OMB guidance, for example, does not specify sufficiently detailed criteria for survey approval. The guidance states that if a survey's expected response rate is less than 80 percent, then the investigator must plan to conduct a study demonstrating that the results are not biased due to nonresponse. Alan noted that the guidance does not specify how one should conduct such a study.

Alan also commented on the lack of clarity in the relationship between nonresponse rates and sample bias or representativeness, and asked whether the nonresponse rate is the best proxy variable for whether a study yields representative results. He also finds a lack of agreement among environmental economists as to how one should demonstrate sample representativeness.

Alan next laid out the three major goals of the workshop: to improve clarity about nonresponse bias and nonresponse rates, to identify protocols to reduce and test for bias, and to develop a research agenda to support such protocols and tests. While admitting that these are lofty goals, Alan emphasized that the workshop would be successful if there were any notable progress toward these goals. He added his belief that the goal of improving sample representativeness ought to be considered alongside achieving other desirable qualities of a survey design.

Agency Perspectives

Brian Harris-Kojetin of the Statistical and Science Policy Office at OMB began this session by introducing the audience to OMB's role in developing standards, policies, and guidelines associated with all steps of federal data collections and statistical programs (except the final data analysis). He explained that the Paperwork Reduction Act (PRA) provides the authority for, and requires, the OMB's review of agency data collections. Only 5–10 percent of the roughly 8,000 active federal data collections are surveys or statistical collections. The approval process includes reviews to assure sound statistical analyses are used and to determine if the data collection is even necessary, and if so, to

determine whether there are alternative methods to collect the data that are less expensive and burdensome on the survey population.

Brian then gave a brief overview of two guidance documents to assist agencies in conducting surveys or other statistical data collections. The drivers to update and develop these guidance documents include a decline in response rates over time and changes in the technologies for collecting survey responses. Efforts to call attention to the quality of data collections have also increased in recent years. The goals of these documents are to provide greater transparency about the OMB review process and to improve the quality of the data collected.

The first document, "Questions and Answers When Designing Surveys for Information Collections," is intended for a broad and nonexpert audience and covers topics such as basic process issues for OMB review and survey primer topics for collection requests. The second document, "Standards and Guidelines for Statistical Surveys," updates Directives 1 and 2, which outlined OMB's "Standards for Surveys and Publications of Statistics." Theses Directives were last updated in 1974. The OMB formed an interagency subcommittee of the Federal Committee on Statistical Methodology (FCSM) to review agency best practices and help write the Standards document. The document contains 20 standards supplemented by guidelines or best practices.

Brian noted that the OMB reviews collections before they have been fielded. They do not typically have continuing oversight of a collection once it is approved (unless it is a multiple-wave survey). He provided the analogy that reviewing the survey collection requests was like doing a peer review when one only has an introduction and methods section to work from.

The balance of Brian's presentation focused on those standards that are particularly salient to response rates and nonresponse bias. The OMB recognizes that response rates are merely an indicator of bias and are not a direct measure of nonresponse bias. Nonetheless, they believe that they can be used as a rough indicator of potential bias problems.

With respect to receiving approval for a collection request, Brian explained that the agency must guarantee that the data it collects is of suitable quality for its intended use. Influential surveys that collect information that may have a substantial impact on an agency's programs or policies should achieve high response rates, but OMB is sensitive to different uses and has taken a more flexible approach to response rates given the purposes of the study. Brian mentioned stories of OMB desk officers using the rule of thumb that if response rates were not expected to be above 60–70 percent, then a survey would not be approved. He emphasized that this is not what the guidance says. Even if desk officers had rejected a survey on these grounds, they should not do so anymore.

Agencies need to report the expected response rate in an information collection request (ICR). If the expected response rate is below 80%, an agency should plan to conduct a nonresponse bias analysis. OMB does not interpret the guidance as saying that response

bias is absent if the response rate is greater than 80%. Rather, it is a marker below which agencies should consider response bias. Furthermore, the nonresponse bias analysis must look at bias in the substance of the information being collected. For example, it is not enough to say that the demographic distribution of survey respondents is similar to what can be found in the *Census*.

Brian indicated that he has been encouraging the survey and establishment nonresponse subcommittees of the FCSM to provide a clearinghouse of examples of nonresponse studies so that agencies can refer to this information when developing their ICR requests.

Brian then briefly delved into how the two OMB documents address multistage surveys and how this information relates to panels. His main point was that the last-stage response rate for a survey conducted by Knowledge Networks (KN) (i.e., the number responding divided by the number of panel members given the opportunity to respond) is not representative of the actual response rate because this rate ignores the impaneling stages. An agency needs to report expected response rates at each stage, and attrition must be accounted for.

Finally, Brian described how the "Q&A" is a living document that OMB plans to continually update. The Guidance and Standards document will not be updated at this pace, however.

Nathalie Simon of the U.S. Environmental Protection Agency (EPA) began by highlighting the EPA's responsibility in performing benefit–cost analysis. While the EPA prefers using revealed preference data for monetizing regulatory impacts, this information is often not available for the types of goods being valued, and SP survey results are needed. Despite this, Nathalie noted that the EPA has conducted very few surveys and, to her knowledge, none that are for the purposes of a Regulatory Impact Analysis. She hypothesized that this is explained in part by the effort and time required to conduct a survey. However, she feels that SP work may become more prevalent at the EPA because of continuing controversies about mortality risk valuation, where the EPA may move away from hedonic approaches, and the need to value changes in ecosystem services.

Nathalie believes the OMB documents will be useful for conducting surveys and that surveys were improved as a result of external review. However, the standards for response rates and nonresponse analysis concern her, and she questions whether the response rate criteria can be met given the typical budgets for SP surveys. She is also concerned that some of the suggestions for improving response rates, such as using a mixed-mode approach, might lead to other biases.

She is most interested in learning what can be done to increase response rates (with specific mention of internet panel surveys) without introducing bias, and whether there are indicators of nonresponse bias other than response rates. She is also interested in what research should be conducted to improve understanding of survey nonresponse.

Bob Leeworthy of the National Ocean Service at the National Oceanic & Atmospheric Administration (NOAA) began by mentioning that his agency is responsible for damage assessment and coastal zone management, including National Marine Sanctuaries. He supports these tasks by conducting surveys to value changes in the quality of natural resources. When he started at NOAA in 1985 OMB approval of valuation surveys took up to 90 days. Now, it takes a minimum of 450 days if approval is required for focus groups, pretests, and the final survey (which includes the time it takes to get Commerce Department and NOAA clearance). Even with this amount of time, there are often delays. In Bob's view this suggests that OMB is likely overworked. He also pointed out that such a lengthy approval process discourages private partners from participating with NOAA in projects.

Bob believes there has been a creep of the Information Quality Act (IQA) into the PRA. NOAA is a scientific agency and wants to achieve high statistical standards. However, the appropriate stage for reviewing the quality of the responses and the analysis is after the survey has been conducted. Bob argues that it is the IQA that applies at this stage of the study, not the PRA.

NOAA wants quality analysis and has no problem with the Standards per se, but Bob does have a problem with the response rate requirements in the Guidelines document. He gave the example of his experience with a recent proposal for a socioeconomic research and monitoring program survey for the Florida Keys. A similar study was funded under the Sea Grant in 1995–1996. The goal was to do a decadal replication of this study, but OMB rejected the recently proposed survey because the response rate from members of local environmental groups was too low. Bob believes there is no scientific evidence that raising this response rate would meaningfully change the study results.

Bob is also concerned about unintended consequences of the approval process, such as contractors conducting unreported focus groups and survey testing to avoid the long time delays.

Dr. Linda Langner of the U.S. Forest Service (USFS) was the last discussant. She explained that the USFS is a management agency and does not do regulatory work, so the surveys they conduct are not for regulatory analysis. Rather, these surveys are in support of developing and evaluating management plans. She is in the research branch of the Service, which supports other branches and may conduct these surveys on their behalf. They have done SP surveys in the past and believe that there is considerable opportunity to use them more for developing land-management plans. However, they have limited funds and time to conduct such surveys. Also, it is difficult to work with those in academia collaboratively given the challenges of the PRA.

Scientists working at the USFS have told Linda they do not want to bother going through the lengthy and uncertain OMB review process. The USFS does field interception surveys, and it is unclear to Linda how OMB's suggestions for performing nonresponse analysis and increasing response rates are applicable to these types of surveys. She is concerned that while the OMB Guidelines say that the treatment of the data depends on

the use of the survey, it is silent on what the different uses of the surveys may be and what the standards are in those cases. She wonders if there should be different gradations based on the intent of the surveys.

Evidence of the Relation of Response Rates to Sample Representativeness

Robert (Bob) Groves, of the University of Michigan, led off this session. The purpose of his presentation, "Response Rates and Nonresponse Bias: What We Know and What We Don't Know," was to summarize his ongoing research on the effect of nonresponse on survey estimates. He began with four observations: (1) that the inferential paradigm based on probability sampling essentially ignores nonresponse, (2) that response rates in the developed world are declining, (3) that three recent well-designed studies have found no link between nonresponse bias and nonresponse rates, and yet (4) students in survey methodology are still being trained to achieve high response rates. Together these lead to a bit of confusion in understanding the state of the survey methodology field.

Stepping back, Bob made the point that nonresponse is only one of many issues that create error in measures and that there are two inferential processes under which all sources of error fall. The first is measurement error and the other representation error (see slide). Respondent representativeness falls into the latter category. The title of the workshop is thus wrong in his view. He would not call the issue at hand one of sample representativeness but rather one of respondent pool representativeness.

Bob next presented a meta-analysis of prior studies specifically designed to improve understanding of nonresponse bias. The surveys included in the meta-analysis typically are found in the health literature because records on both respondents and nonrespondents are collected in those studies. For any particular estimate, the relative nonresponse bias is measured as:

$$\left| \frac{\bar{y}_r - \bar{y}_n}{\bar{y}_n} \right|$$

where $\bar{y}_n$ is the estimate mean from the full sample and $\bar{y}_r$ is the estimate mean from the survey respondents.

Bob presented a graph plotting out this sample statistic for each estimate (i.e., particular question) from each survey relative to the nonresponse rate for that survey (all the points that trace out a vertical line above a particular nonresponse rate). Looking at just one particular study on this graph, he showed how the relative bias for the different estimates varies from 0–70 percent. This implies that the effect of nonresponse depends on the variable in question.

Bob then provided a few examples. The *Survey of Consumer Finances* includes one frame for the very wealthy. The wealthiest of the already very wealthy do not respond, leading to relative nonresponse bias in income of 55 percent. In an SP survey in his sample the nonresponse biases range between 60 percent and 70 percent. This is because residents had a higher willingness to pay (WTP) for aquatic plant management in a lake than visitors. These anecdotal findings imply that the nonresponse mechanism is related to the topic of measurement.

The meta-analysis yielded additional results. While the bias in estimates within a survey may vary considerably, Bob finds, using a naïve ordinary least squares regression, that the nonresponse rate is a poor predictor of nonresponse bias across surveys. Furthermore, he notes that it has generally been hoped in the survey methodology literature that if one were interested in a difference in means, say between males and females:

$$\bar{y}_n^m - \bar{y}_n^f$$

that somehow the estimate:

$$\bar{y}_r^m - \bar{y}_r^f$$

would have a lower relative nonresponse bias than either $\bar{y}_r^m$ or $\bar{y}_r^f$ evaluated separately. That is, it was hoped that somehow the biases would cancel each other out. The meta-analysis suggests that this is not the case. The relative nonresponse bias from the difference in the means is less than the average bias of the two subclass means less than half the time.

Reviewing the studies in the meta-analysis, Bob finds that most correct for potential nonresponse in some way. He finds that some use selection equations to adjust structural model estimates while others construct postsurvey weighting class adjustments from the frame variables. He noted that each approach typically uses demographic variables. He used the metadata to learn that the bias in the demographic variables cannot be used to predict the bias in the estimates of the substantive variables. He believes this suggests that naïve selection bias or propensity models are not particularly valuable.

Bob then provided a formula demonstrating the relationship between the sample mean of the variable of interest, the respondent mean, the nonrespondent mean, and the nonresponse rate (see slide). He reported that the survey methodology field has long assumed that as one reduces the rate of nonresponse, which is the ratio of number of nonrespondents to the total number of individuals surveyed, the difference in the estimate of the mean between the respondents and the nonrespondents stays the same. However, the field has come to recognize that this is wrong because there is a covariance between the nonresponse rate and the difference between the respondent and nonrespondent means. He then provided an alternative measure of bias that acknowledges this relationship between the response rate and the variable of interest (see slide).

Bob's current research focuses on the correlation between response propensity and bias in the estimates of a variable of interest because he finds (from his scatterplots, shown earlier) a variation in bias across estimates within a particular survey, but not across surveys. Thus, he argues that one needs theories and practices that focus on different

estimates in the same survey because each question in a survey is subject to different biases from nonresponse.

To try to determine which variable estimates are affected by mechanisms that produce nonresponse and to try to control for this effect, in one study Bob and his colleagues are trying to systematically and deliberately create nonresponse bias. This study proposes three candidate causes of nonresponse that may also be correlated with the variable of interest: topic, sponsorship, and incentives. Incentives are thought to dampen the effect of the correlation between both the topic and sponsorship with response propensity by encouraging those uninterested in the topic/sponsor to respond anyway. He described one study where he and his colleagues administered two surveys to three different frames. Two of the frames were expected to have high salience for the survey topic. They found an effect of the topic of the survey on the variable of interest (e.g., the surveys about birds were more apt to be responded to by birders), but that one needs to strongly signal what the topic is (e.g., by using stamps or writing on envelopes) in order for this effect to present itself.

Bob feels that previous efforts and studies about how one can increase response rates were mostly "misplaced energy" because this research ignores the variation in response bias within estimates in the same survey. In his view it is not clear that the additional respondents are improving the estimates of the variable of interest.

He then summarized, for the sake of the SP survey practitioners, where the field of survey methodology is at the moment. There is a reduced emphasis on the response rates per se and more on reducing nonresponse bias (error). He views this as consistent with the change in OMB focus as represented by the Guidelines and Standards. He actually thinks the Guidance is ahead of the field because few survey practitioners could conduct sophisticated nonresponse analyses and that survey methodology texts do not provide descriptions of methods for conducting such analyses. He also thinks that important contemporary research is rethinking what explanatory variables could be used in selection and propensity models.

Audience Discussion

> One participant wanted to hear more about the effect of sponsorship on response propensity. Bob suggested that little is known empirically about sponsorship effects on response, but provided a few examples of different effects of sponsorship. He noted that with customer satisfaction surveys the service provider often does not conduct the survey because there is a concern that only the satisfied would respond. His own research on Census surveys showed that low-income people responded at higher rates than high-income people, which is the opposite of findings from other studies he has conducted. His post hoc reasoning is that this is a sponsorship effect in that those who are dependent on transfer payments are more responsive to the *Census* survey. Despite these findings, in the birder study (described in his presentation) he and his colleagues really needed to emphasize the topic before they saw an effect on response. That said, he would

argue that the topic should still be deemphasized in the survey recruitment protocol if the topic is expected to be correlated with the variable of interest.

Another participant asked if Bob would encourage the use of incentives. He would not provide a blanket endorsement of incentives. He noted that, while in some cases incentives may dampen the effect, say, that the topic of the survey has on response propensity, the use of incentives may be detrimental if the reasons to respond were independent of the variable of interest. One example of when an incentive increased bias was in an exit poll where Democrats liked the pens being provided as an incentive more than Republicans did.

Bob was also asked if he felt his results would change if they were applied to survey panels. He did not know but felt that it was an excellent research question. He explained that clients of market research firms are concerned about the reliability of responses to "online open recruitment" surveys (volunteer opt-in surveys). This is because online surveys ignore all the sources of error that come from the recruitment process.

Another participant noted that Bob's presentation raised concerns regarding the use of panels like KN and Harris Interactive, but also that the potential for exploring the relationship between response propensity and the variables of interest suggests a benefit of using the results from these panels. Bob agreed that the use of rich data vectors on the propensity to respond to a particular request is very useful, but argued that the problem is that one is not interested in this propensity. In his view what is more important is what explains the propensity to enter the panel, which is a component of what one is interested in if one wishes to generalize the panel results to the target population. Bob further noted that if there is online open recruitment there is no way to know the process that yields this propensity of interest. If researchers start with a probability sample they are better off to the extent that they can maintain that probability sample.

The next presenter, Stanford University's **Jon Krosnick,** drew on a number of different studies he has participated in and tied Bob's presentation to environmental valuation studies.[1] While Bob's presentation focused on consequences of nonresponse, Jon's presentation focused more on the causes of nonresponse and whether strategies used to reduce nonresponse are successful. Jon also considered whether response rates influenced results of old contingent valuation (CV) surveys and exit-poll studies.

Jon noted survey methodologists are concerned about declining response rates and whether they are an indicator of survey quality. Jon first set out to ask whether these concerns were well founded. This is a difficult question to answer because response rates are not easy to find. Government survey response rates are buried in reports, and with news media (polling) surveys response rates are not published. Jon also noted that a decline in response rates may not signal changing behavior on the part of respondents. Rather, it may also be explained by less being spent to conduct surveys.

[1] At Jon's request his slides will not be distributed.

With colleagues, Jon conducted a survey of news media and government surveys to try to learn whether response rates for random-digit dialing (RDD) surveys were indeed declining. They asked for surveys conducted nearest particular dates (to avoid the incentive for the agency/firm to report on their "best" survey), disposition codes, demographic data, and data about the survey administration itself. Their sample contains 114 surveys. Using this sample, they find that the response rate of a survey is highly correlated with the contact rate, and even more so with the cooperation rate. The cooperation rate is defined as the percent of subjects contacted that yield a useful response. That the correlation with response rate is higher for the cooperation rate provides some notional sense that cooperation after contact is more important for increasing response rates. The sample shows that response, contact, and cooperation rates are declining over time. The decline over time in contact is bigger than the decline in cooperation. Caution must be made in this interpretation, however, because avoiding contact may be a form of noncooperation.

Jon next compared the techniques these surveys use to increase response rates and whether increasing response rates increases accuracy. His slides showed descriptive statistics of characteristics of the survey, such as the percentage of surveys that sent advanced letters, and a regression analysis of the survey response rate on these characteristics. Jon and his colleagues find higher response rates the longer the survey is fielded, the more call attempts are made, and when incentives are used. The regression also shows that any attempt to try to control for the household member who takes the survey, as opposed to allowing any adult at home to respond, lowers the response rate.

Jon and his colleagues also find that refusal conversions and advanced letters do not have a statistically significant effect. Despite this finding, Jon noted that in their sample many surveys have employed these two techniques for a long time, probably because experimental studies have shown that these techniques increase response rates. (Jon discussed other techniques that their study shows do not increase response rates, even though they are expected to from experimental studies.)

Going forward, Jon believes that focusing attention on the interviewer–interviewee relationship will help improve response rates. That is, research should focus on the effectiveness of different methods of *how* contact is conducted (i.e. the tenor and form of letters, calls, refusal conversions) and what the letter or interviewer says. He described a conference that brought together survey methodologists and psychologists studying communication to focus on the first 10 seconds of an RDD contact. Bob Groves had provided the psychologists a tape with 100 contacts between interviewer and interviewee. Upon hearing these conversations, the psychologists were horrified. With interviewers left to their own devices, psychologists found numerous violations of conventional conversation, suggesting room for improvement in these initial contacts.

Jon next considered whether increasing the response rate increases survey accuracy. The subsample in this analysis consists of the 81 surveys that are national in scope and are of the general population (of the 114 total). They tested for linear and nonlinear

relationships between response rates and the absolute difference in means between demographic variables from the sample and from the *Current Population Survey* (which serves as the "true" value of the estimate of interest). Accuracy here is thus how closely the sample looks like the general population based on observable characteristics. He defined the discrepancy between the sample and the actual demographic mix as the percentage point difference between these values. They found, for example, that increasing response rates would increase gender balance (at least in the range of response rates for which they had sufficient data [≤50 percent] to make such a claim). However, this is only from a discrepancy of 4.5 percentage points with a 5 percent response rate to a discrepancy of .5 percentage points with a 50 percent response rate. The lesson is that, at least with respect to observable demographics, increasing response rates does increase accuracy, but that the samples from the surveys with low response rates were fairly representative to begin with. However, as Bob Groves pointed out in his presentation, demographic accuracy is not the same as nondemographic accuracy.

To focus on the effects of response rates on nondemographic accuracy, Jon turned to a very recent analysis of extant data from two classic CV surveys: the Montrose (California) and Exxon Valdez studies. Both studies had very large budgets and achieved very high response rates. Using the data from these studies, he asks how increasing response rates affect median and mean WTP. Furthermore, he looks at the effect of response rates on tests of validity using regressions predicting the respondent's vote (validity tests), whereas up until this point the focus has been on biases in point estimates.

Looking at the mean and the median WTP in both studies, these estimates vary considerably up until the point where 50 percent of the survey responses used in the study were collected. With the following 50 percent of responses, the mean and median estimates generally fell, but, from Jon's perspective, not to an economically significant degree. Jon noted that one might argue that high response rates are preferred if the goal is to have a conservative estimate of WTP (i.e., have estimates that are more likely to be biased low rather than high). He also noted that the data show that the easiest respondents to contact (i.e., when very little sample had yet to be collected) have a much lower WTP than the rest of the sample.

Moving on to the regression equation predicting the vote of respondents, Jon showed how little the magnitude of the regression coefficients change from using the first 50 percent of responses collected to using 100 percent of the responses for the Montrose survey. More systematically, he compared the coefficients of the regression equation using the first X percent of the responses collected to coefficients for the remaining (100–X) percent responses for different values of X. There are only two cases where one can reject the hypothesis that the set of coefficients are the same, suggesting that increasing the sample size does not affect the parameters of this regression significantly.

The next question Jon posed was whether the significance of the regression coefficients changes by increasing the response rate. He finds that significance and numbers of significant variables increase with X.

The question then remains if significance itself is important to government surveys. Jon prefers coefficients to be statistically significant in such an equation because he finds significance reassuring, but he does not know if establishing significance itself is important for estimating median WTP.

The next part of his discussion focused on exit polls. A convenient attribute of exit polls is that the surveyor administrators quickly know the accuracy of votes reported in the survey relative to the actual vote count. Jon described a study of exit poll paper-and-pencil survey of voters in randomly selected precincts around the country. The distribution of response rates has a large range and is bell-shaped. He also showed the distribution of errors, defined as the absolute value of the difference between the predicted percentage voting for a candidate and the actual share. The study explored whether precincts with higher response rates had higher accuracy and found that there was no such relation. Again, this result shows that response rates do not increase accuracy.

Jon's discussion of the results of political polls concluded with an anecdote about the *Columbus Dispatch* preelection poll. The newspaper used a mail survey from 1980 to 1994 and on average had a response rate of about 25 percent. Two university polls used RDD telephone surveys and had an average response rate of 62 percent. However, the average error of the poll conducted by the newspaper was lower than for the two universities.

In conclusion, Jon argued that both more response and a higher response rate are valuable. However, higher response rates do not increase accuracy much, and methods designed and thought to increase response rates might not actually work. As such, one should be careful about discrediting surveys that have a low response rate. However, he acknowledged that this is the statistic consumers of surveys know, and may judge the credibility of the survey on this statistic despite the recent evidence that it is not very reliable in demonstrating bias.

Jon also argued that future survey administrations should use more than standard demographics for benchmarking response characteristics, as Brian Harris-Kojetin and Bob Groves also suggest. He suggested that survey administrators could buy questions on the American National Election Studies Survey and the General Social Survey, each of which are face-to-face surveys and have high response rates.

Audience Discussion

> One audience member noted that not all variables in the regression explaining the vote are of equal interest. Some of those with lower significance on their regression coefficients are those that were especially important to the NOAA Panel in judging survey validity. In his view, one may then want to be more selective about evaluating the significance with respect to response rate comparison. Jon does not think that this relationship between response rates and significance for some variables is coincidental.

Another audience member noted that Jon and Bob Groves both explored whether response rates increase accuracy and both concluded that while there is a benefit to increasing response rates, it is limited. He also noted that Bob thinks that the OMB guidance is consistent with his view and wondered if Jon felt the same and encouraged Bob to engage in the conversation. He wondered why, given Bob's and Jon's results, response rates are a focal point at all in OMB's Guidance and whether they could abandoned in favor of something more informative.

Jon responded that he likes the new OMB guidelines as they are much more informed than the old guidelines, which were based on scant evidence. He also likes that the OMB views are now clear. Plus, he thinks the new OMB guidelines are responsive to the data with the view that lowering response rates increases the risk of nonresponse bias but does not imply nonresponse bias. In Jon's view, the real value is encouraging studies of nonresponse bias. Seconding Bob, Jon commented that the real challenge is determining what types of nonresponse bias analyses are informative and affordable. He believes that the survey methodology discipline has little experience addressing this sort of problem and that it is an exciting direction for future work.

Bob followed by noting that it is the fault of the survey methodology field that there is such a focus on this one statistic (response rates) because those in the field thought the world was much less complex than it actually is. Bob commented that he is not sure he agrees with Jon on the point that increasing sample itself is a good thing. In part, that is because the field is good at increasing response rates, which would increase sample size; but as his work shows, increasing response rates may actually lead to greater bias. Bob also feels that there has been progress in moving away from the response rate as an indicator for bias. He described the work of other researchers who are looking for an alternative, and more informative, single indicator (a different sort of response rate). The hope is that one might eventually wean people off their reliance on the response rate as an indicator of bias in favor of a more informative single indicator. He also hopes that eventually researchers can even get to a point where there is a simple way to report nonresponse bias itself and not just a better indicator.

In Jon's opinion, approving surveys with greater than an 80 percent response rate likely implies a wise use of funds. Furthermore, his comparisons of response rates and accuracy across a large number of surveys have convinced him, at least with probability samples from firms conducting high-quality, large-sample surveys, that results are actually quite accurate at low response rates. This implies there would need to be a very large bias in order for the results of a survey to change significantly. Related to this observation, Jon commented that he was glad Bob discussed his failed attempts to create bias first because they showed that one needs to be almost unrealistically blatant to generate bias.

Meta-analysis of Response Rates to Contingent Valuation Studies Conducted by Mail

Kevin Boyle of Virginia Tech began his talk by sharing some observations relevant to the previous discussion. First, he noted that the effect of who is in and who is out of the sample is not always what one expects it to be. He once conducted a field study on a river and there were many folks from an environmental group there one day. The members of the group told people to take the survey, and they took the survey themselves. The respondents from this day looked no different from any other day.

He also expressed concern that the Guidelines might create incentives for SP studies to avoid this scrutiny, consequently be of lower quality, and then possibly be used for benefits transfer.

The study he discussed, which focused on responses to SP mail surveys, began around the time that the NOAA Panel report was released. Kevin believes the panel implicitly challenged mail administrations of CV by recommending the use of in-person interviews. He believes his study is also relevant to the OMB guidance with respect to describing expected response rates, approaches to increase response, and methods for evaluating nonresponse bias.

The sample of CV studies began with the 1994 compilation of studies by Carson et al. (at the time none were choice experiment [CE] surveys). While 124 surveys (with 1,672 citations) were referenced, few provided enough documentation of the design and implementation of the survey to include in their analysis. Kevin and his colleagues conducted a survey of the principal investigators of these studies to collect these data. The response rate of that survey was 74 percent. They also asked respondents if they knew of any other studies in the grey literature that the authors might want to track down. If a study did a split sample, then they treated each sample separately. Their final sample included 146 studies with 188 observations. Kevin noted that the average response rate of the surveys in the sample is 62 percent. From flipping through recent journals, he thinks that response rates are now about 20 percent lower.

Kevin then looked at which survey design variables have significant coefficients in four regression equations. One estimate uses (to the extent possible) only those variables that were used in a 1978 meta-analysis of the effect of survey design on response rates. The second equation included a few more variables thought to affect response rates from the survey methodology literature. The third equation included those design variables unique to SP surveys, such as resources type and payment vehicle. The final approach was to use a stepwise regression. Kevin noted that as Bob pointed out, the strength of the influence of the different design variables varies. Furthermore, few of the many variables studied significantly affected response rates.

Kevin and his coauthors find that surveys of the general population have lower response rates. This is an example of a design variable that the investigator has limited control over. There are variables that the investigator does have more control over. They do find

strong investigator-specific effects, but what the investigator is doing to affect response rates is unclear. They also find that the type of resource affects response rates, but it is not clear if this is attributable to the resource having more salience to respondents or due to greater experience with measuring WTP for improvements to these resources.

Audience Discussion

> One audience member asked whether Kevin and the previous presenters felt that there was a response rate below which no amount of nonresponse correction would help and the results should just be thrown out. Kevin could not provide a specific rate where that would be true. He noted, however, that response rates are not correlated with the magnitude of the WTP estimate for the surveys he and his colleagues analyzed, although there are a few studies with very low response rates in the sample they are using (perhaps due to self-selection).
>
> Another participant asked why one should think that an open-ended WTP question would matter to response rates, as Kevin's analysis showed. Would the question form actually affect the respondent's choice to mail in the survey, or was that choice perhaps correlated with other factors? Kevin was not sure, but he thinks that once people get to that question they may find it difficult to respond to.

Open Discussion Regarding First Three Sessions

One audience member expressed concern that the Guidelines allow surveys with over an 80 percent response rate to avoid conducting a nonresponse analysis because this might provide perverse incentives given that methods to increase response rates might themselves lead to bias.

Brian accepted this possible cost, but felt that the potential cost was outweighed by other considerations as to how the survey budget could be allocated. Furthermore, some national surveys have found that small increases in high response rates influence the results. Jim Laity of OMB followed up by noting that it is not true that an 80 percent response rate provides a free pass on an OMB review. It is a value that helps OMB approve the survey, but the response rate is not the only thing they focus on.

Another audience member seconded the concern of the first regarding the 80 percent threshold. He wondered whether the focus should be on the structure of the peer review process and defining what we mean by "influential" with respect to survey review. He was uncomfortable with the use of the requirement that only influential surveys are subject to OMB review because we do not know if survey response are influential until the survey is conducted. Brian responded by noting that the PRA requires the agency conducting the survey to demonstrate that the work has "practical utility" and that the information is useful. The OMB desk officer who is familiar with the working of the agency will review any survey proposal and if it is a CV survey, the review process will be supplemented with economists familiar with the technique as well as statisticians who

would evaluate the sampling design. Thus, the survey is reviewed from a number of different perspectives.

Practitioners' Perspectives

Alan Krupnick presented a summary of the responses provided by eleven research teams to a request for information on recent or in-progress high-profile SP surveys valuing ecological "commodities" or similar goods. These practitioner perspectives are provided in Appendix C. Five of these studies are completed; most of the others are awaiting OMB review. Sponsorship was primarily from the EPA and NOAA, with one study sponsored by the National Park Service.

Most of the studies are national in scope and were expecting large samples (1,500 or more). Many were using KN's internet/Web–TV panel methodology, although occasionally the samples were supplemented by alternative modes, such as mail or on a computer at a centralized facility. A few studies used a mail mode only. The response rates for completed or piloted studies varied considerably. Those using a panel frame typically report response rates as the number responding divided by the number of panel members given the opportunity to respond, rather than by the population contacted for panel recruitment. Rates from panel administrations range from 60 percent to 75 percent. Mail response rates were in the 20 percent range. Former panel members were also sampled in one study.

Alan then listed the ex ante strategies used by the studies to boost response rates. These included the following:

- Initial and followup contact letters following Dillman's Total Design Method.
- Nonresponse conversion (reminder phone calls).
- Oversampling of underrepresented groups.
- Incentives
  - One study found that increasing the incentive paid from $2 to $5 significantly raised the response rate.
- Signaling the importance of the survey by using priority mailing and individualizing contact letters.
- Identification of survey sponsor.
  - While the practitioners believe this helps to signal the importance of the survey, they recognize that it might bias responses as well as the sample of respondents.
- Data collection on nonresponders.
  - To check and correct for nonrepresentativeness of the sample, a few demographic and/or attitude questions are collected from nonrespondents, usually during initial contact or reminder phone calls. This is an ex ante technique in that it anticipates that these data will be valuable before the survey response rate is known.

Researchers supplement their ex ante strategies with strategies to test and correct for nonresponse bias. These include the following:

- Comparison of sample statistics with population statistics. All studies do this. Some go on to weight their sample observations to reflect differences in the population. However, most studies find very little difference in the estimates of the variables of interest by controlling for observable differences.
- Comparison of respondent characteristics and WTP across multiple modes of administration (or frames, if applicable). A new calibration approach was offered by Kevin Boyle that involved administering one low-cost mode to the majority of the sample and another mode, believed to be less susceptible to nonresponse bias, to a small share of the sample. The results from the small share of the sample can then be used to calibrate estimates from the larger share of the sample.
- Comparison of attitudinal and demographic characteristics of responders with those of nonresponders or of responders to a survey presumed or shown to have a low nonresponse bias.
- Comparison of WTP between current and former members of survey panels. Such a comparison helps address the concern that panel selection and attrition introduce nonresponse bias.
- Use of Heckman selection models to correct for differences in nonobservables that affect response propensity. This also includes using Trudy Cameron's approach, which involves using tract-level Census data matched to respondent location to reduce the number of nonobservable characteristics when correcting for differences in response propensities.
- Comparison of WTP based on various measures of engagement in the survey. Internet surveys in particular permit the collection of information that may be related to the degree to which respondents are engaged in the survey (e.g., the time they spent on various questions, how often they referred back to earlier parts of the survey, etc.). The hypothesis is that engaged respondents may make more informed decisions based on the content of the survey than the unengaged.

From this review, several issues emerged. These include:

- Lack of clarity as to whether members of probability-based panels were more representative based on demographics and overall knowledge than respondents to other modes/frames. The teams' experiences differed. Mark Dickie and Shelby Gerking found that, compared to a self-administered survey at a central location, KN panelists tended to skip key questions more frequently, complete the survey in much longer or much shorter times, were more likely to say they were unaware of skin cancer, and were less sensitive to the scope of the good. Other studies using multimodes do not find these results.
- Does engagement matter? Dickie-Gerking finds that KN panelists were poorly engaged. Kip Viscusi, Joel Huber, and Jason Bell find high levels of engagement from the KN internet panelists and speculate that the panelists are smarter, more aware of current events, more compliant, and an altogether better source of

information. Viscusi-Huber-Bell offers the concept of a distortion indicator, which is described in Joel Huber's presentation.

- Are nonrespondents different from respondents? Most of the studies found significant but small differences between respondents and nonrespondents. Consistent with this statement, they also found that corrections for unobservable characteristics in sample selection did not affect WTP estimates.
- An unaddressed issue was whether mode/frame effects are important. None of our SP studies tackled this issue of separating mode from frame effects.
- Does observation weighting based on demographics affect WTP estimates? Most of the studies found that weighting did not affect estimates given that the respondents did not differ much from the survey population.

Finally, Alan examined research needs that were either suggested by the practitioners or that arose from a comparison of the different studies. These include:

- More analyses of effort/engagement by mode.
- Further analysis of the notion that a KN panelist is a preferred/inferior respondent.
- Designing studies that disentangle mode from frame effects.
- Mounting analyses of selection effects associated with different modes of response as well as different stages of empanelment for probability-based panels.
- Development of sample selection (Heckman) correction algorithms for CE-type data (see summary of Dan Hellerstein's presentation below).
- More analyses of the effect of incentives on response rates.

Audience Discussion

One participant highlighted the multiple stages of possible selection bias. For instance, we can collect information from the people initially contacted, but that misses the people who do not answer the phone. He noted that there is a literature on this issue.

He also voiced concern about the incentives to respond that are presented by competing modes and frames. For example, there is an endogenous learning effect—people learn to answer in a way that minimizes their time commitment on panels, such as doing less fishing so they have less to report; or people's answer to an initial questionnaire with a known followup product-testing phase may answer the initial survey in a way to maximize their chance of getting into the product-testing phase.

Another participant mused about whether people respond more when the survey is about activities they are interested in (a salience effect) or more for the desire to learn something new. In one study she found that the latter does mitigate the former.

Joel Huber of Duke University gave a presentation, "Should We Move to Panels?" He first provided the perceived advantages of using panels of survey respondents:

- Lower cost per completed survey
- Much faster and more current
- Representative of the U.S. adult population
- Possibility of targeting subpopulations (by region, preferences, demographics, etc.)
- Can be used in longitudinal studies, which we do not generally do for SP studies, but which would be very valuable
- Can be used in experiments
- Willingness of respondents to answer difficult survey questions

These perceived advantages were contrasted with the possible problems with panels:

- Respondents become professional survey takers—either worn out or the best at answering questions
- Panelists are not ordinary people. They…
  - …like to answer questions
  - …have time to answer them
  - …value the rewards from answering them
  - …are at ease with computer interface

Whether being a typical member of the target population is good or not is unclear, according to Joel, since we would like survey responses from motivated, knowledgeable people.

Joel also considered what might make panelists (of the KN variety) different from the general public. In his view, they may be those more likely to use computers and the internet, have greater openness to alternative perspectives, and have more free time. Those less likely to be on the panel might have more geographic mobility than those on the panel or be less good at dealing with forms (such as those who get help with taxes). He also raised the concern that perhaps those who belong to panels are not typical members of their communities, and thus Trudy's approach to correcting for selection is more susceptible to error.

Joel described the results—in particular the selection tests for nonresponse bias—of a study he and colleagues conducted that valued the benefits of drinking water improvements using a KN panel. They found that:

- Many Census tract–level demographic characteristics predict a person's likelihood of reaching each stage in the KN panel recruitment process.
- A censored regression analysis predicting water quality value based upon tract-level data showed minimal differences between surveyed and nonsurveyed people.
- A Heckman selection model of water quality value did not produce a significant difference in the WTP estimate.

Joel then talked about their "distortion indicator," which provides a measure of the degree of nonresponse bias. This indicator is computed from information on panelists for which information from the target population is somehow known. If these qualities of the panelists are independent of the responses to the critical decision variables then presumably nonresponse bias is not a concern. If the qualities of the panelists are correlated with the decision variables then that correlation could serve as the basis for adjusting the outcome variable. In his work, they find very low correlation.

Audience Discussion

> In response to Joel's talk about developing a participation propensity index, one participant said that his proposed method is the Harris Interactive's propensity weighting approach! He noted that people who are on panels tend to be more inquisitive. He also noted that RDDs have similar problems as internet panels (time availability, interest in topic), as both have similar incentive issues.

Research Frontiers

Trudy Cameron of the University of Oregon presented next. Her presentation, "Scavenging for Covariates To Use in Selectivity Assessment and Correction Strategies," focused on strategies for finding covariates to correct for nonresponse bias due to observables and nonobservables, but touched on a number of other topics. She first discussed the availability of data for weighting, lamenting that the Bureau of the Census does not make available cross tabs on key variables (such as age, sex, race, income) at the tract level, and does not collect data that might really be associated with response biases, such as political views.

Next, Trudy discussed her work with KN to collect information on each empanelment stage, which included pulling together a huge number of Census tract variables to try to explain response propensities at each stage. She compared conditional models to marginal models of explaining the response process, where conditional models involve the process of going from any given stage to the next stage, while marginal models involve the process of going from (say) the RDD contact stage to any later stage. So the conditional models look at stage-by-stage attrition while the marginal models look at attrition over multiple stages. Trudy laid out five marginal and five conditional models and noted that most studies that use KN only examine the conditional model of those drawn from the panel for the survey to those present in the estimating sample for the model. Her study estimated all ten models. She finds that many variables explain participation at various stages, and those variables that explain participation for one model might not explain it in another or may even change sign. For instance, "well-to-do prime age" is negative and significant in the conditional participation models up until the final stage, where it becomes positive and significant.

Trudy then provided analysis on how attitudes toward government regulation are related to other variables in the context of participating in the KN panel. She and her colleagues

estimated a selection equation on this outcome variable, with the Census tract variables again significant, as are voting factors. For example, they find that Nader voters are more likely to want government involvement. The Heckman correction factor is not significant.

Trudy then turned to the problem that no model is available for selectivity correction for responses to a CE survey. She offered an interim solution that involves seeing whether the preference parameters within the estimating sample vary systematically with participation propensities. If these shift parameters are significant across the participants, then researchers should be concerned that preferences differ between participants and nonparticipants. She applied this approach to her Value of a Statistical Illness Profile study. Using 15 Census tract factors, she finds insignificant and small effects of participation propensity on the WTP to avoid future illness.

Trudy then turned to her Climate Policy Study, which used a mail survey to solicit preferences for climate policies. This study had a response rate below 20 percent. The selection equation included a variety of weather and disaster variables in addition to Census tract variables. She found that those who live in areas with natural disasters had a lower response propensity. Trudy thinks that this unexpected finding is explained by self-selection in location decisions.

Trudy closed by urging researchers to look for more variables that might influence response propensities given different survey topics. She wants government agencies to help companies set up cradle-to-grave selection studies with panels (despite her work suggesting that selection problems are minimal). Her last idea was to support the work of our next speaker, Dan Hellerstein.

Audience Discussion

> Joel mentioned that in the water quality study he and his colleagues conducted that used the KN panel, they employed Trudy's approach using Census tract data and had similar results.

Dan Hellerstein of the Economic Research Service, U.S. Department of Agriculture, offered a very technical presentation on "Correcting for Nonresponse Bias in Discrete Choice Models: A Two-Stage Mixed Logit Model." Such an effort to devise Heckman selection models for CE-type data was the most prominent research suggestion among our practitioners. Dan stressed that his work is still very preliminary.

His problem is that if respondents to a survey systematically differ from the general population, misleading inferences may result—conclusions based on information provided by survey respondents may not hold true for the population of interest.

Systematic differences may be observable or unobservable.

If observable, measurable factors can be identified for respondents and nonrespondents. For instance, the average income of the respondents to the survey may not be the same as

the general population. Such differences should not lead to biased estimates if the regression equation is correctly specified.

Unobservable characteristics are those that cannot be measured in either respondents or nonrespondents; for example, due to idiosyncratic personality factors, people who return a survey may be the most avid users of a resource.

For obvious reasons, direct use of measurable characteristics (of the sample and population) cannot be used to control for unobservable differences. However, indirect measures can be used by employing Heckman-like sample selection models.

Dan first explained the basic idea behind sample selection models. Respondents are "selected into" the sample: a first-stage model predicts who will be selected, and a second-stage model predicts the variable of interest. However, the random components (variables) (RV) in both stages are potentially correlated. This means that the first-stage RV of participants may not have an expected value of zero. Hence, if the two RVs are correlated, the second-stage RV may not have an expected value of zero. Hence, the second-stage model must control for such a situation.

Since econometric models for analyzing multiple-choice discrete choice models, such as the multinomial logit (MNL), are often used in environmental valuation, Dan asked whether sample selection models can be used to control for nonresponse bias in the MNL. The naïve approach is to estimate a first-stage probit model, compute the appropriate values of the selection coefficient λ, and then include λ in the MNL second stage. The problem is that it is not obvious what the first-stage RV are correlated with.

Dan then considered the two-stage mixed-logit model, and showed that it is the values of one (or several) of the varying parameters that will be correlated with the first-stage decision. That is, the RV in the first-stage model will be correlated with the RV used to generate the varying parameter. He then tested this model on artificial data. From these tests he concludes that:

- The two-stage mixed logit offers a somewhat theoretically appealing method of controlling for nonresponse bias in discrete-choice models.
- Under very limited testing, this approach has shown capacity to improve estimates by correcting for selection.
- More work is needed. In particular, the current Heckman-like two-stage estimator uses limited information when linking correlations to simulated draws. A more ambitious estimator could use a Gibbs sampler to estimate both stages nearly simultaneously. Also, it was unexpected that MNL outperforms the mixed-logit model.

Dan offered his program to everyone.

The final research speaker was **Erika Edwards** of Boston University. Erika presented research comparing the results of an internet and a phone survey on alcohol consumption.

Her team started with the *National Epidemiologic Survey on Alcohol and Related Conditions* (NESARC) study with in-home interviews of 43,093 adults ages 18+ conducted by the Bureau of the Census in 2001–2002 (with the proviso that the adults have at least one drink per month). Her team also interviewed KN panel members via the internet and on the phone and nonpanelists recruited through RDD who were administered the survey via the phone. They found that the sample of nonpanelists that was administered the phone survey yielded a very low response rate (38 percent).

Erika concluded that the panel using the internet mode is a better choice for targeted surveys than using the telephone with an RDD sample. They see minor mode effects between those on the panel who were administered the survey on the phone versus on the internet. In conclusion, she and her research team feel that KN is a good approximation to a national survey and much cheaper to administer.

Audience Discussion

One panel member asked if their results were driven by the sensitivity to the survey topic. Erika thought this was possible and that the internet may be a better choice in this case. Also, she noted that it may be impossible to find this population of drinkers over the phone.

A discussion developed about how the institutionalized population should be addressed. For example, are college students living in dorms, a group that presumably contains a large number of drinkers, included in the panel and in the RDD recruitment (and are they included in the NESARC survey)? Erika said that the NESARC study oversamples college-aged students, but she does not know if they are surveyed in their dorms. In the *Census*, dorms are "grouped quarters." KN excludes grouped quarters from the panel. Erika also mentioned that one university stopped providing land lines to dorm rooms because only 6 percent of students use them.

Another participant noted the importance of identifying criteria that will allow the surveyor to identify the population that is relevant.

Another participant asked if there were any studies analyzing the best mode for conducting SP surveys. In his view the controlling attribute for deciding on a mode is that well-designed SP surveys provide a large amount of information. Jon Krosnick responded that there are many studies of mode effects. The problem with these studies is that they often confound mode effects and factors such as sample frames so the mode effect comparison is not a clean one. Of those that are clean, face-to-face produces less measurement error and social desirability bias than telephone. On these criteria, paper-and-pencil administrations are generally worse than telephone, while the internet mode seems to perform better than a telephone administration. In Jon's view the real issue is quantity of information you can present, not on how the mode effects change results. Jon finds that even a phone survey performs well. Certainly one cannot provide a lot of information on

an SP survey if it is administered over the phone, but Jon thinks too many SP surveys are overly complex and wants to undo complexity.

Bob also mentioned that we would like to see work that bridges the surveyors' and practitioners' worlds. He thinks we need model diagnostics and critical evaluative steps. He reads the literature applying selection models to say that such corrections usually do not change the results much. There are two interpretations to this. The first is that selection effects are not very important. The other interpretation is that the selection models are poorly specified. Either is possible, so the question remains as to which is right. Moving beyond selection models, Bob feels we need diagnostic tools for rigorous sensitivity analysis of explanatory variables. There is now a kitchen sink approach to adding explanatory variables to selection models without much consistent rationale.

Panel Session on Implications for OMB Review Procedures

The final session of the day was a panel discussion moderated by Alan Krupnick. Each speaker was given about ten minutes to discuss the implications of the day for OMB's guidance to the agencies. At the end of the session, Brian Harris-Kojetin and Katherine Wallman of OMB responded to a number of the issues that were raised.

Norman Meade of NOAA praised OMB for being open to new ideas from the conference and stated his support for improving survey quality and reliability. He noted that the Data Quality Act creates an unfunded mandate, both to OMB, which has to review these surveys, and the agencies that do them. In addition to budgetary effects, meeting this mandate requires huge amounts of time. To make the review system more efficient he offered the following thoughts: license agencies to approve their own surveys, and expand the use of generic survey approval processes. Both of these would yield better use of agency resources. OMB's budget examiners could help in this regard. Norman also commented that given the apparent lack of correlation between nonresponse bias indicators and response rates, the 80 percent threshold does not make sense in its bluntness. He also thinks the response rate threshold leads to incentives to increase response rates even though techniques to increase response rates might have perverse effects on survey quality. Finally, Norman called for the National Science Foundation (NSF) and others to provide additional funds for funding research in this area.

Kerry Smith of Arizona State University provided numerous suggestions and observations:

- Given the concern about the burden for agencies, we could have an information budget and then a set of licenses to collect information that could be traded among agencies.
- Just because information is not collected, this does not mean that decisions will not be made. The implication is that the perfect data collection should not be the enemy of an informative, but perhaps not perfect, collection.

- The current OMB standards stifle learning. For example, this happens through requirements for separate approvals each time the focus group script is changed. The incentive is not to learn in a focus group because it will lead to a change in the focus group script. He therefore urged that focus group scripts be exempted from review or otherwise treated more leniently in review.
- Surveys should be reviewed through normal peer review channels.
- Recognizing that getting responses to a survey is a process, he counseled that more research be directed to understanding the economics of that process. For instance, he and Carol Mansfield found that people have a reservation price of $19–$25/hour to take a survey. Such research would involve understanding the mechanisms that lead to response and how the effects of these mechanisms are correlated with the variables one is trying to measure.
- He suggested that we use information from existing surveys to help identify distortion indexes. We can buy space for questions on existing major surveys or put questions from these surveys on our new surveys and compare the answers to our surveys with those of the well-regarded national surveys.
- With the stifling of survey research, he warned that survey research may move offshore with the consequence that American preferences may be proxied by benefit transfers based on preferences of people in other nations.
- Kerry felt that we should be adding behavioral questions in our surveys that correspond to questions asked in major surveys, such as the *Consumer Expenditure Survey*, to test for "behavioral correspondences" between surveys.

Michael Brick of the University of Maryland noted the cultural differences between practitioners and OMB. Mike cautioned that if the 80 percent rule is taken away, then every study would be highly scrutinized (although such studies even now do nonresponse bias analyses). In any event, he advocated that measures of survey quality be devised. He noted that many survey studies do not do acceptable nonresponse bias analyses. He discussed a new RDD study where the researchers expect less than a 50 percent response rate; the funding agency has authorized a followup study to examine response bias. Mike felt that agencies need to fund such efforts, as expensive as they are, for their most important studies. He also noted that the focus of the workshop has been on response rates, but that the focus should have been broader to include coverage and related issues (e.g., how to deal with cell phone use in surveys of college students and the representativeness of the 6 percent without them that could be reached by RDD studies). Finally, Mike said that this research is much harder than rocket science because it deals with people and, therefore, we need to step up our research efforts accordingly.

The floor was then opened for comments. Shelby Gerking returned to the issue of how respondent effort varies by mode, with more effort in a lab than on the internet because respondents had to expend so much effort to come to the lab and because there are no distractions in the lab. The issue, then, is how to provide incentives for similar behavior among panelists in probability-based panels (like KN).

Trudy Cameron suggested that researchers should think more systematically about how to model participant survey-taking behavior. She drew the parallel between an economic

model, where observed behavior is considered a consequence of the interactions between preferences and constraints, and a possible interpretation for survey purposes. She thinks of this model as the outcome of the inclination of a participant to take part in the survey (interest in the topic), the participant's opportunity cost to take part in the survey (available time versus income), and the participant's ability to complete the survey (cognitive capacity).

Kevin Boyle noted that the group should not forget that attention to response rates could result in giving short shrift to other questions, like how responses are affected by survey design. He thinks the focus on low response rates might lead to an inefficient allocation within a survey's budget and in particular away from things that affect the WTP estimate. Low response rates can occur because people are oversurveyed and because people do not think a survey will help them (as in the case of Katrina victims), which is the opposite effect of salience. He also wondered, like Joel, whether having experienced survey takers is a good thing.

Rich Bishop laid out three consequences of OMB's survey review policy: 1) fewer studies will be done because the OMB rules make them costly and time consuming; 2) therefore policies will be less well informed and may even rely on older, less credible values or on benefits-transfer values; 3) there is a risk of misallocation of funds to achieve high response rates at the expense of reducing quality in other areas.

Bob Groves agreed that in many cases the focus on achieving very high response rates tends to discourage the activities that would lead to higher-quality data (e.g., the time constraints on researchers focusing on response rates may lead to perverse incentives such as cutting back on survey development).

Jim Laity from OMB had a different view on the standards by which SP studies are reviewed by his agency. He sees it as appropriate that the surveys under review at OMB must be carefully scrutinized because their use for official government business potentially has great monetary and policy implications. Academic SP research not sponsored by federal agencies is not reviewed by OMB but is still important for policy. He said that response-rate requirements are a lightning rod for general frustration about the high standards set in the NOAA Panel report on CV. He further stated that neither he nor, to his knowledge, others at OMB look solely at response rates when evaluating collection proposals.

To address the frustration of researchers over a perceived lack of transparency in the OMB review process, Mary Evans suggested that OMB provide information (similar to the Stevens report for Regulatory Impact Analyses) on the characteristics of successful proposals so that researchers could determine which study features and analyses lead to a higher propensity of approval and a shorter time to approval. This would provide transparency to researchers regarding the approval process. Alan Krupnick expanded on this idea by suggesting that OMB could set more specific criteria for estimating sample representativeness and other evaluation elements so that researchers could have more

certainty when initiating the review process. Alan also suggested that OMB create an external review board to be used in special appeals cases.

Jon Krosnick counterbalanced the conversation by stating that his sense is that the new OMB guidelines represent a step forward in flexibility and recognize changes in the field. He also noted that he has a sense that the problems of nonresponse rates are not well enough understood to provide specific guidelines as to how nonresponse bias analyses are to be carried out. He does not think issuing specific guidelines now will work well in practice anyway because rules created at this point in the development of the field will be outdated by the time they are published. He thinks a more effective way of providing guidance on nonresponse analysis is to work from the ground up within agencies to figure out what works; he sees this as a better approach because it is flexible to change as the field develops.

Kerry Smith revisited the comment that Jon made about the review process. In Kerry's view it is important to remember that the NOAA Panel's study was the product of a huge volume of research and that additional research is needed before certain questions can be answered. Researchers who would like to add to the necessary research have a hard time finding funding from nonmission-oriented agencies (such as the NSF). Where the funding is available, methodology development is not high on the agencies' priority list. He asks why a distinction cannot be made between the fielding of truly influential surveys that will play a dominant role in decisions and the activities that precede the development of these surveys (focus groups, pretests, etc.). These development activities might not need the same kind of OMB scrutiny and could be reviewed through a peer review process and evaluation. Bob Leeworthy agreed with a process revision such as this given the way that the IQA has crept into the PRA, which has led to a bottleneck at OMB.

In addition, Kerry believes that the PRA isn't applicable to survey participants because they are usually compensated and then by definition are not burdened. Others agreed with this comment.

Dan McCollum noted that the USFS is encouraged to solicit public participation in their planning processes, but at the same time the Service is told that they need to go through a lengthy review process, which results in the information coming after decisions are made. Agencies are also told that they need to evaluate their performance even though they do not have the long-term data to make those evaluations, so there are conflicting requirements. Surveys that receive funding from NSF do not require OMB approval, and Dan doesn't understand the difference between NSF-sponsored studies and studies sponsored by other agencies.

Julie Hewitt noted that there are Type I and Type II errors in survey research. There is a sense that OMB is doing a good job eliminating Type I errors, but there is uncertainty on how they are doing on Type II errors. She went on to say that environmental statutes leave considerable room for interpretation and that leads to lawsuits. In these lawsuits all the emphasis is put on the analysis EPA does. The EPA is already able to force firms to

provide information for the cost side of regulatory analyses, but they must find a way to get people to respond to surveys that inform the benefits side of the analysis.

Brian Harris-Kojetin of OMB then responded to points made by the various panelists in this session as well as the audience comments.

- Brian noted that pretests and generic survey development work can be evaluated by a number of other federal agencies besides OMB that have cognitive labs and a generic clearance to do so in about two weeks' time (examples include the Bureau of the Census and the Energy Information Agency).
- According to the PRA, there is a burden any time the public is asked for information. Thus, the use of incentives does not imply the absence of a burden as it is defined.
- Response rates are only one piece of information that OMB looks at in an ICR.
- There are a lot of errors in surveys and a checklist of sources of survey error would need to cover all of these errors even though the research has focused mostly on minimizing measurement error.

Katherine Wallman, the Chief Statistician of OMB, then spoke to some of the points made by the panelists and the audience. They include:

- OMB uniquely has a cross-agency perspective, which is why it is suitable that it is involved in the review process.
- OMB does make suggestions about questions to use in surveys that have already been tested.
- NSF-sponsored studies do not differ from studies sponsored by other agencies but there is a difference between contracts and grants. NSF largely provides grants to institutions, and unless the grant specifically calls for collection of information it is not subject to the OMB process. An NSF contract would be subject to the ICR process.
- A lot of the hold-up in the approval process is due to deliberations within the federal agencies and not with OMB review.
- The PRA permits delegating an external review process that is conducted by the federal agency. The Federal Reserve Board is the only agency that has taken on the task to date. This suggestion generated great interest among the agency representatives and practitioners present.
- OMB encourages agencies to communicate with OMB early in development of an ICR in order to save time and resources.

Organizers' Concluding Thoughts

The workshop was characterized by frank discussions and enlightening presentations, with improved understanding by all sides of the issues involved. Here is a brief summary of some of the take-home messages:

What is the problem?

The basic problem is that agencies believe their surveys, particularly internet panel surveys, are being held up by OMB to an unreasonable degree and length of time, without clear guidance from OMB on how to expedite their review. The concern is that decisions at agencies will be made with or without information from these surveys, and that these decisions will be the poorer without this information. In addition, researchers may cut corners in the survey process to avoid multiple OMB reviews of draft surveys, such as conducting fewer focus groups and pretests or making surveys shorter to raise response rates at the cost of increased measurement error. Similarly, researchers may overinvest in reducing nonresponse bias to the detriment of other components of the survey protocol. OMB, for its part, believes that there are problems with representativeness of the respondents to these surveys and that representativeness is getting too little attention from the agencies. Furthermore, OMB suggests that part of the delay in approvals is because of delays in requesting approval by the agencies. OMB also believes that the agencies could improve their navigation of the process by, for instance, getting generic clearance for focus groups, which will reduce review times.

Some respondents thought that some underlying causes of the problem are (1) OMB has too little staff to devote to survey review and is unnecessarily applying the PRA to survey development stages; (2) OMB is trying to apply the PRA to address concerns that are arguably the domain of the IQA, such as sample (respondent) representativeness; (3) the implications of changes in methods of communication, behavior, and technological advances in survey methods are not yet fully appreciated by both OMB and researchers. Examples of the third possible cause include the decline in landline phone use and increased resistance to taking surveys. Both of these phenomena make it harder to achieve representativeness using traditional sampling methods. Changes in technology have made it easier to conduct internet surveys, but the body of knowledge on sampling issues has not yet caught up to these advancements. Even with more traditional sampling methods there is a dearth of research on survey representativeness and response selection. A final cause cited may be that the guidance issued by OMB and its survey review processes need to be improved to provide clarity and flexibility for different survey settings.

Is the 80 percent threshold a standard or a guideline?

Before the workshop it was unclear to us whether the OMB's 80 percent response rate threshold was a standard or a guideline. Now it is clear that it is a guideline. OMB's position is that many of the surveys they review have not been designed with thoughtful consideration of response bias issues. Thus, surveys with response rates above this threshold are presumed to have low nonresponse bias, but OMB emphasized that even these surveys will still be scrutinized on other elements. Surveys below 80 percent have to make a showing that nonresponse bias is tolerably low. OMB was emphatic, however, that response rates for panel data must be calculated based on the panel recruitment stage, and not based on the number of panelists asked to respond.

Does OMB offer any specific guidance to make such a showing?

While there is some general guidance, it was agreed that there are no sampling protocols that OMB considers definitive.

What is the appropriate amount of effort for testing and reducing nonresponse bias?

There was broad recognition that researchers face tradeoffs in designing and implementing surveys, with minimizing response bias only one of the factors to be considered. Other factors include components of the protocol such as sample size and respondent burden. Yet there was no agreement on the weights that should be given to these various factors. Researchers want relatively low weight to be attached to response rate per se because they do not believe that it is a sound indicator of sample representativeness. However, they appreciate that sample representativeness itself deserves close attention, provided that a lack of clear guidance from the literature as to appropriate methods to avoid and test for nonresponse bias is acknowledged.

Is there research offering guidance on best practices to achieve and demonstrate sample (respondent) representativeness?

We learned from Jon Krosnick and Robert Groves that this area of research is in its nascent stages and the guidance that is available emphasizes the wrong thing (i.e., response over representativeness). Furthermore, from the evidence available, the relationship between response rate and nonresponse bias or sample representativeness is weak. They also indicated, as noted above, that there is little guidance as to what indicators could be used instead. What is clear is that efforts to increase response rates are wasted without considering directly their implications for nonresponse bias. OMB indicated that, although the new Guidance Document, which was updated from 1974, will not be updated any time soon, the Q&A document could be a living document to reflect the latest research results on the relationship between survey response and bias.

A useful suggestion was to form a panel in the spirit of the NOAA Panel (the expert panel that issued SP protocols after the Exxon Valdez spill compensation controversy) to develop protocols for avoiding and testing for bias.

Is there anything special about panels?

Yes and no. There is almost no research by survey methodology experts on panel-specific issues in nonresponse bias. However, as discussed above, this is a problem common to all methods of survey sampling. The panel approach is different in that it is relatively new, at least with respect to computer-aided methods, and is growing in popularity. For example, marketers are moving increasingly to panel-based approaches. This alone suggests that their biases are low because mistakes made here can translate into lower profits. However, as Bob Groves reported, some marketers are becoming less comfortable with this method, although there is an important distinction between probability- and nonprobability-based approaches. SP researchers have done a variety of studies on this issue, generally concluding that panels developed from RDD techniques (e.g., KN) generate low nonresponse bias despite low cumulative response rates.

Another provocative suggestion came from a challenging question: whether a representative sample is actually desirable, perhaps because informed and compliant respondents may deliver results that are more reliable and relevant for public policy than respondents drawn randomly from the general population.

What research needs exist?

The discussion at the workshop generated some suggestions regarding promising paths for future research. In our view the primary need is for rigorous diagnostics and behavioral models of the panel selection/withdrawal process. Extant methods for identifying and correcting for nonresponse bias include the development of propensity scores from a large number of demographic and behavioral variables. This approach is somewhat ad hoc. We agree with the suggestion that the OMB should support the funding of primary cross-agency research on methods to avoid and test for nonresponse bias.

What can the agencies do to improve the survey review process?

There were several intriguing ideas that surfaced during the workshop. The most compelling was for the agencies to develop their own survey review procedures, which at the time of this workshop was only being done by the Federal Reserve Board. Subsequent participants proposed that the agencies partner to develop a center for SP survey research that would serve all the partnering agencies' needs. In this case, OMB's reviews would focus on the Center's guidelines and less on any particular survey.

A less dramatic option is for an agency to develop a protocol for internally reviewing surveys and overseeing the survey development stages. With this approach, the OMB would then only need to review the entire development stage, and not have to approve an ICR for each focus group and pretest.

Another easy-to-implement option is for OMB to make researchers aware of on-going and previously administered surveys that have low response bias to help them design their sampling strategy and the survey itself.

What can OMB do to improve the survey review process?

Agency participants agreed that OMB should make the review process more transparent. This would include developing and distributing a list of studies that have been approved (as well as rejected) along with the approach each took to address and measure nonresponse bias. It was also suggested that OMB should permit agencies to increase their budgets for funding survey development, encourage greater funding of surveys by Congress, encourage the funding of primary research on survey response, and exempt survey development stages from ICR. One of the most dramatic options was for OMB to eliminate their review in favor of outside peer review.

APPENDIX A: WORKSHOP PRECIS AND AGENDA

Sample Representativeness:
Implications for Administering and Testing Stated Preference Surveys

Alan Krupnick and David A. Evans, RFF

Stated preference surveys are frequently being used in regulatory contexts to inform policy. Conducting a high quality stated preference survey may present a variety of methodological challenges. Assuring representative responses is one of these challenges. For stated preference surveys, the NOAA Panel on Contingent Valuation, formed in response to questions raised about projects to estimate nonuse values associated with the Exxon Valdez oil spill, suggested that in-person surveys with a high response rate were needed to help assure credibility of stated preference surveys. However, as the Panel acknowledges, many of its recommendations were made absent evidence from high quality surveys. The purpose of this EPA-funded workshop is to identify and discuss appropriate means of measuring sample representativeness and reducing sample biases in stated preference surveys.

Many surveys have experienced declining response rates in recent years, and researchers have devoted increasing attention to assessing the effects of declining response rates on data quality. Although surveys sponsored by Federal agencies have traditionally achieved higher response rates than those sponsored by commercial or academic organizations, many Federal surveys have also experienced declines in response rates and attendant increases in resources required to maintain high response rates. The Office of Management and Budget (OMB), which must approve all government sponsored surveys, has recently issued standards and guidance related to statistical surveys. One component that has received attention and interest concerns response rates and potential non-response bias. The OMB guidance requires Federal agencies to conduct extensive non-response analyses when expected response rates are below 80%, which the OMB uses as an indicator of sample representativeness.

The use of panel-based internet surveys (which has been growing in popularity) poses particular challenges to achieving high response rates due to the multiple stages of recruitment and participation to become a panel member. However, achieving representative samples and, further, samples that do not significantly bias outcomes of interest is not necessarily tied to response rates.

To address issues of sample representativeness in stated preference surveys, we are inviting 30 of the top experts in survey methodology, statisticians and stated preference surveys practitioners to share their knowledge, insights and experiences regarding this issue. In addition, we will be asking authors of recently performed stated preference surveys or other germane research to write a two-page summary describing their studies and results, how they tested for and addressed any biases in sample representativeness. The focus will lean towards administrations via internet with samples drawn from a probability-based panel. However, experience with more traditional modes that can also accommodate extensive information treatments, such as mail and in-person, will also be considered. These summaries from practitioners will be distributed to the authors of commissioned papers well in advance of the due date for such papers, so that they may be referenced or incorporated as appropriate. The workshop deliberations are designed to be informative for Federal agencies conducting stated preference surveys for regulatory purposes as well as for OMB in providing guidance to agencies in these areas.

Sample Representativeness:
Implications for Administering and Testing Stated Preference Surveys

October 2, 2006
Resources For the Future Conference Center
1616 P Street NW, Washington, DC

8:00-8:30	**Registration and Coffee**	
8:30-8:50	**Opening Remarks:** Philip Sharp, President of Resources for the Future **Purpose, Regulatory Context and Plan of the Day:** Alan Krupnick	
8:50-9:30	**Session I:**	**Agency Perspectives** Brian Harris-Kojetin, Statistical and Science Policy, Office of Management and Budget *OMB Guidance and Standards for Statistical Surveys* Nathalie Simon, Associate Director, National Center of Environmental Economics, U.S. Environmental Protection Agency Bob Leeworthy, Leader, Coastal and Ocean Resource Economics Program National Oceanic and Atmospheric Administration Linda Langner, Economics Program Leader, Resource Valuation & Use Research, U.S. Forest Service
9:30-11:00	**Session II:**	**Evidence on the Relation of Response Rates to Sample Representativeness** Robert Groves, University of Michigan *Response Rates and Nonresponse Bias - What We Know and What We Don't Know* Jon Krosnick, Stanford University *The Causes and Consequences of Survey Response Rates*
11:00-11:15	**Break**	
11:15-12:00	**Session III:**	**Meta-analysis of Response Rates to Contingent-Valuation Studies Conducted by Mail** Kevin Boyle, Virginia Tech
12:00-1:00	**Lunch (provided)**	
1:00-2:00	**Session IV:**	**Summarizing Analyses of Sample Representativeness in Recent Stated Preference Surveys-Practitioner Perspectives** Alan Krupnick and David A. Evans, Resources for the Future Discussant: Joel Huber, Duke University

2:00-3:30	**Session V:**	**Research Advances/Agenda** Trudy Cameron, University of Oregon *Scavenging for Covariates to Use in Selectivity Assessment and Correction Strategies* Dan Hellerstein, Economic Research Service, U.S. Department of Agriculture *Correcting for Non-response Bias in Discrete Choice Models: A Two-Stage Mixed Logit Model* Erika Edwards, Boston University *Internet vs. Phone Survey: An Application to Alcohol Research*
3:30-3:45	**Break**	
3:45-5:00	**Session VI:**	**Implications for Agency Guidance Panel Discussion** Alan Krupnick (Moderator), Resources for the Future Brian Harris-Kojetin, Statistical and Science Policy, Office of Management and Budget Kerry Smith, Arizona State University Norman Meade, Office of Response and Restoration, NOAA Fisheries Mike Brick, University of Maryland and Westat

APPENDIX B

Workshop Participants

Elena Y. Besedin, Ph.D.
Environment & Resources Division
Abt Associates Inc.

Kathryn Bisack, Ph.D.
Economist
Protected Species Branch
Northeast Fisheries Science Center

Richard Bishop, Ph.D.
Professor Emeritus
Department of Agricultural & Applied Economics
University of Wisconsin
Senior Consultant
Stratus Consulting Inc.

Kevin J. Boyle, Ph.D.
Professor and Head
Agricultural & Applied Economics
Virginia Tech University

Sarah Brabson
PRA Clearance Officer
NOAA

Michael Brick, Ph.D.
Director, Survey Methods Unit
Senior Statistician, Westat
Research Professor,
Joint Program in Survey Methodology

Trudy Ann Cameron, Ph.D.
R.F. Mikesell Professor of Environmental & Resource Economics
Department of Economics, 430 PLC
University of Oregon

Alan Carlin, Ph.D.
Senior Economist
National Center for Environmental Economics
U.S. Environmental Protection Agency

Bruce D. Carlson
Economist
U.S. Army Corps of Engineers

David Chapman, Ph.D.
Principle
Stratus Consulting

Michael Dennis, Ph.D.
Senior Vice President
Government & Academic Research
Knowledge Networks

Mark Dickie, Ph.D.
Professor
Department of Economics
University of Central Florida

Cheryl L. Eavey, Ph.D.
Program Director
Methodology, Measurement, and Statistics
National Science Foundation

Erika M. Edwards, MPH
Project Manager, School of Public Health
Boston University

David A. Evans
Research Associate
Resources for the Future
Ph.D. Candidate, Economics
University of Maryland

Mary Evans, Ph.D.
Assistant Professor
Department of Economics
University of Tennessee

Shelby Gerking, Ph.D.
Professor
Department of Economics
University of Central Florida

Robert M. Groves, Ph.D.
Director
Survey Research Center
University of Michigan

Brian Harris-Kojetin, Ph.D.
Statistical & Science Policy
Office of Information & Regulatory Affairs
Office of Management & Budget

Daniel M. Hellerstein, Ph.D.
Resource and Rural Economics Division
Economic Research Service
U.S. Department of Agriculture

Julie Hewitt, Ph.D.
Economist
Office of Policy, Economics, and Innovation
U.S. Environmental Protection Agency

Joel Huber, Ph.D.
Alan D. Schwartz Professor of Marketing
Fuqua School of Business
Duke University

F. Reed Johnson, Ph.D.
Economist
RTI International

Robert J. Johnston, Ph.D.
Associate Director, Connecticut Sea Grant
Associate Professor
Agricultural & Resource Economics
Connecticut Sea Grant Office
University of Connecticut at Avery Point

Jon A. Krosnick, Ph.D.
Frederick O. Glover Professor
Humanities & Social Sciences
Professor of Communication & Political Science
Director, Methods of Analysis Program in the Social Sciences
Stanford University

Alan Krupnick, Ph.D.
Senior Fellow & Director
Quality of the Environment
Resources for the Future

James A. Laity
Office of Information & Regulatory Affairs
Office of Management & Budget

Linda Langner, Ph.D.
Research & Development
Forest Service
U.S. Department of Agriculture

Amanda Lee
Office of Information & Regulatory Affairs
Office of Management & Budget

Bob Leeworthy, Ph.D.
Leader
Coastal & Ocean Resource Economics Program
Special Projects Office (N/MB7)
National Oceanic & Atmospheric Administration

Daniel K. Lew, Ph.D.
Economist
Alaska Fisheries Science Center
National Oceanic & Atmospheric Administration

Carol Mansfield, Ph.D.
Economist
RTI International

Daniel McCollum, Ph.D.
Rocky Mountain Research Station
Natural Resources Research Center

Norman Meade, Ph.D.
Economist
Office of Response and Restoration
National Oceanic & Atmospheric Administration

Chip Paterson, Ph.D.
Senior Associate
Industrial Economics, Inc.

Nathalie Simon, Ph.D.
National Center for Environmental Economics
U.S. Environmental Protection Agency
Ariel Rios Building
1200 Pennsylvania Avenue NW
Washington, DC 20460

V. Kerry Smith, Ph.D.
W.P. Carey Professor of Economics
Department of Economics
Arizona State University

Randall K. Thomas
Senior Research Scientist
Harris Interactive

Katherine Wallman, Ph.D.
Chief Statistician
U.S. Office of Management & Budget

Kristy Wallmo, Ph.D.
Acting Division Chief, Economics & Social Analysis
National Oceanic & Atmospheric Administration

APPENDIX C

Compendium of Practitioner Perspectives

Prepared for:

Sample Representativeness:
Implications for Administering and Testing
Stated Preference Surveys

October 2, 2006

Resources for the Future

Compiled by David A. Evans and Alan Krupnick
Resources for the Future

Practitioner Survey Template Written by David A. Evans and Alan Krupnick

Introduction

We asked practitioners of recently performed or pending high profile stated preference surveys to summarize and describe their studies and results, and how they tested for and addressed any biases in their sample. To maintain consistency in the descriptions of the surveys, the practitioners were provided a template along with instructions for completing the template (the template is provided on page C-42). We received full cooperation from the participants that we asked to complete these templates. None of the summaries have been subject to any external review. Please contact the practitioners with any questions about these studies, including their status.

Summary of Studies

Survey Study	Type	Primary Investigators	Page
Environmental Risk to Children's Health	CV	Dickie, Gerking	C-3
Assessing the Value of Improvements to Lakes and Rivers	CE, Ref	Viscusi, Huber, Bell	C-6
Popular Support for Climate Change Policy	CE	Cameron, DeShazo	C-14
The Economics of Environmental Vulnerability: Epidemiology, Household Status, and Air Pollution	CV	M.F. Evans, Smith, Poulos	C-18
Valuation of Natural Resource Improvements in the Adirondacks	CV	Banzhaf, Burtraw, D. Evans, Krupnick	C-20
Improved Valuation of Ecological Benefits Associated with Aquatic Living Resources	CE	Johnston, Besedin	C-23
Willingness to Pay Survey for §316(B) Regulations	CE	Besedin, Johnston	C-27
The Nonconsumptive Value of Steller Sea Lion Protection Measures	CE	Lew, Rowe, Layton	C-30
Valuing Visibility in National Parks	CE	Boyle, Paterson	C-33
Hawaiian Coral Reef Valuation Study	CE	Bishop et al.	C-36
North Atlantic Right Whale Economic Benefit Study	Undecided	Bishop, Chapman, Tourangeau, Bisack	C-39

Environmental Risks to Children's Health: Parents' Risk Beliefs, Protective Actions, and Willingness to Pay

I. Basic Information

1. Title of project: Environmental Risks to Children's Health: Parents' Risk Beliefs, Protective Actions, and Willingness to Pay.

2. Sponsor(s): U.S. EPA, ORD/NCER.

3. Date work was (is expected to be) finished: Dec 31, 2005.

4. Timeline (if any) for review of this survey by OMB: N/A.

5. Team members and affiliations (including survey administrator, e.g. KN):
   Survey #1 & #2: Mark Dickie & Shelby Gerking, University of Central Florida.
   Survey #2: KN, survey administrator

6. Author of this summary: Mark Dickie & Shelby Gerking.

7. Purpose of the study: To estimate parents' marginal rates of substitution between risks to their own and to their children's health. USEPA provided funding for two surveys to examine the extent to which results from a local sample would mirror results from a national.

8. Type of stated preference survey (e.g. CV, conjoint): CV.

9. Where are descriptions/results of the study published/available? An OECD report and three working papers:

Dickie, Mark and Shelby Gerking. (2006). "Valuing children's health: parental perspectives," in *Economic Valuation of Environmental Health Risks to Children*, (ed.) Pascale Scarpecci, OECD: Paris.

Dickie, Mark and Shelby Gerking. (2006). "Parental altruism and environmental risks to health of parents and the children," Department of Economics, University of Central Florida, http://www.bus.ucf.edu/wp/.

Dickie, Mark and Shelby Gerking. (2006). "Parental altruism, protective behavior, and effectiveness of policies to reduce children's environmental risk," Department of Economics, University of Central Florida.

Dickie, Mark, Shelby Gerking and William L. Goffe. (2006). "Valuation of non-market goods using computer assisted surveys: A comparison of data quality from Internet and RDD samples," Department of Economics, University of Central Florida.

10. Sampling description

i. Population of interest (e.g. voters, adults, households): Parents with biological child/children aged 3-12 years living at home.

ii. Frame(s) (e.g. RDD): Survey #1 (summer 2002): RDD in Hattiesburg, Mississippi. Survey #2 (summer 2005): KN.

iii. Mode(s) of administration (e.g. internet, mail, in-person; please indicate if multiple waves)**:** Survey #1: computer-administered in-person. Survey #2: via KN Internet/WebTV.

iv. Location of administration (central facility, home): Survey #1: central facility. Survey #2: home.

v. Total number of unique cases fielded (i.e. the total number of unique contacts attempted to survey): Survey #1: Approximately 26680 working residential numbers. Survey #2: Survey fielded to 1199 KN panelists.

vi. Response and Contact Rates[1]: Note eligibility requirements in 10(i) above. Survey #1: Approximately 30% of calls to presumed working residential numbers yielded no contact with an adult after three attempts at different times of the day and days of the week. In 64% of cases in which a call reached an adult, the adult stated that the household did not meet eligibility requirements (no biological children aged 3-12 living at home). Parents agreeing to participate in the study constituted 3.5% of working residential numbers, 5% of contacts with adults, and 14.3% of contacts with adults who did not declare the household ineligible. Finally, 68% of persons agreeing to participate completed the instrument. Survey #2: Transmitted to 1199 panelists and 755 (63%) panelists agreed to complete the survey. Of those who agreed to complete the survey, 644 panelists were eligible to participate (had one or more biological children aged 3-12 living at home. (Twenty-eight percent of these parents completed the survey by WebTV©.)

II. Ex Ante Strategies for Reducing and Testing for Sampling Biases

Survey #1: respondents paid $25 for completing 25-min survey at central location. In recruiting sample, up to three calls made to working numbers – weekday, weekday evening, weekend. Call-backs to no-shows.

III. Ex Post Analyses of Sample Representativeness and Consequences

[1] Please use APPOR Response Rate 1. If available, please provide Response Rate 3 and Contact Rate 1. Please provide all variables used in the calculation (sample size, contacts, complete interviews, etc.). The APPOR standard response rate and contact rate definitions can be found here http://www.aapor.org/pdfs/newstandarddefinitions.pdf . If you do not use the AAPOR standard definitions, please describe how you calculated a response rate.

Analysis focused on non-black parents because African-American parents perceived little risk of skin cancer and therefore had little incentive to think about precautions that might be taken to avoid this disease or about how their own risk might differ from that of their children. Survey #1: 93% of the 488 non-black parents were white and 7% were members of other races. Their mean age was 35 years, 75% were female, 83% were married, 58% were college graduates, mean household income (in 2005 dollars) was $64,000 per year, and the mean number of persons per household was 4.02. Survey #2: 77% of the 592 non-black KN parents were white, the mean age was 37 years, 61% were female, 82% were married, 30% were college graduates, mean household income was $58,000 per year and the mean number of persons per household was 4.18. Thus, at the 1% level, a significantly greater percentage of males participated in the KN survey than participated in the Hattiesburg survey. A possible explanation for this outcome is that in the KN survey, respondents could complete the survey at home and did not need to travel to a central location. Additionally, at the 1% level, Hattiesburg respondents were more likely to be college graduates. As discussed by Chang and Krosnick (2004), respondents with greater levels of education tend to be over-represented in RDD telephone samples. Moreover, local graduates of the University of Southern Mississippi may have been more likely to respond positively to a request from their alma mater to return to the campus and participate in a research project.

Results of a comparison of the two surveys indicate that in comparison to the RDD/Hattiesburg Survey #1 respondents, relatively more Internet/KN Survey #2 respondents: (1) failed to answer key questions, (2) took either a very long time or a very short time to complete the survey, (3) indicated that they were unaware of persons who had been afflicted with skin cancer, (4) took perhaps unwarranted short-cuts in estimating skin cancer risk, and (5) provided stated preference values indicating no greater willingness to pay for large risk reductions than for small risk reductions, thus failing an important external scope test. Split-sample regressions and dummy-interactive regressions were used to assess whether observable differences in characteristics of respondents in the two samples could explain the divergence in scope test outcomes. With few exceptions, RDD/Hattiesburg parents seem to have considered the size of risk changes as well as the cost in responding to the contingent valuation question, regardless of their demographic characteristics, knowledge of skin cancer, or initial risk perception. Internet/KN parents in contrast seem to have considered the cost but ignored the risk changes, regardless of their observed characteristics. Thus the difference in results does not appear to be explained by sample differences in education, gender or other demographic characteristics, knowledge about skin cancer or initial perceived risk levels, or in the time taken or the interface used to complete the survey. In any case, the contrasting results of the stated preference experiment imply different conclusions about resource allocation within the family and about the effectiveness of public policies designed to protect children against health risks.

IV. Research Needs/Additional Observations

Obtaining a representative sample of willingness to pay (or other values/behaviors/relationships of interest) depends on more than sampling. If respondents give less than their best effort in completing a survey, data quality will inevitably suffer.

Assessing the Value of Improvements to Lakes and Rivers

I. Basic Information

1. Title of project: U.S. Environmental Protection Agency. "Assessing the Value of Improvements to Lakes and Rivers," With Kip Viscusi,

2. Sponsor(s): USEPA--Office of Policy, Economics, and Innovation (National Center for Environmental Economics) and Office of Water.

3. Date work was (is expected to be) finished: August 2006

4. Timeline (if any) for review of this survey by OMB: 2001-2003 (Basically it took about two years to get approval as "research")

5. Team members and affiliations (including survey administrator, e.g. KN):

Joel Huber—Duke University
Kip Viscusi—Vanderbilt
Jason Bell—Duke University
Knowledge Networks, survey administrator

6. Author of this summary: Joel Huber, Kip Viscusi, Jason Bell and Alan Carlin

7. Purpose of the study: To use iterative choice to value the benefits of cleaner lakes and rivers

8. Type of stated preference survey (e.g. CV, conjoint): Choice and referendum

9. Where are descriptions/results of the study published/available? Economics of Environmental Improvement

10. Sampling description

i. Population of interest (e.g. voters, adults, households): US national household survey

ii. Frame RDD then selection from Knowledge Network's Panel.

iii. Mode(s) of administration (internet)

iv. Location of administration: Home.

v. Total number of unique cases fielded (i.e. the total number of unique contacts attempted to survey): 5655 invited to participate, 4257 completed interviews

vi. Response and Contact Rates:
-Main KN panel: *4257/5655= 75%*

II. Ex Ante Strategies for Reducing and Testing for Sampling Biases

The base and scope versions of the survey were fielded to a representative panel of US adults from KN's internet panel. To examine whether the KN panelists are representative of the general population and to increase the total number of responses, we compared the final sample characteristics with those of the US.

To boost the response rate from the KN panel the survey was open for at least three weeks and at least three reminder emails were sent.

III. Ex Post Analyses of Sample Representativeness and Consequences

A. Demographic match to target population.

The demographic profile corresponded closely to the US population as shown in Table 1.

TABLE 1

Comparison of KN Sample to the National Adult Population[2]

Demographic Variable	Survey Participants (n=4257) Percent	US Adult Population Percent
Employment Status (16 years or older)		
Employed	61.3	62.3
Age		
18 - 24 years old	13.4	13.3
25 - 34 years old	20.1	18.3
35 - 44 years old	19.4	20.4
45 - 54 years old	18.6	18.7
55 - 64 years old	11.9	12.2
64 - 74 years old	11.7	8.4
75 years old or older	4.9	8.1
Educational Attainment		
Less than HS	18.5	15.4
HS Diploma or higher	59.4	57.4
Bachelor or higher	22.2	27.2
Race / Ethnicity		
White	80.3	81.9
Black/African-American	13.3	11.8

[2] *Statistical Abstract of the United States, 2004-5*. 2003 adult population (18 years+), unless otherwise noted.

American Indian or Alaska Native	1.6	0.9
Asian/Pacific Islander/Other	4.8	5.5
Race / Ethnicity of Household		
Hispanic	10.6	12.1
Gender		
Male	51.0	48.5
Female	49.0	51.5
Marital Status		
Married	58.4	58.8
Single (never married)	25.6	24.4
Divorced	10.9	10.2
Widowed	5.1	6.6
Household Income (2002)		
Less than $15,000	15.0	16.1
$15,000 to $24,999	11.6	13.2
$25,000 to $34,999	12.5	12.3
$35,000 to $49,999	18.8	15.1
$50,000 to $74,999	18.2	18.3
$75,000 or more	23.8	25.1

Comparison of those KN panel members compared with those who declined the survey

Table 2 reports the logit regression for whether the individual in the panel chose to participate in the survey after having been invited to do so. The number of observations for this regression is slightly lower than that reported in Table 2 due to missing data for certain explanatory variables. The dprobit results reported are probit regression estimates for which the coefficients have been transformed to equal the marginal effects.

Older panel members were more likely to participate in the survey, as were those who were married.

Less likely to participate were black panel members, Hispanic panel members, panel members from large households, and panel members who owned their residence. Panel members reporting a high level of stress were less likely to participate in the survey, as were those panel members who did not provide an answer to that question. Also, panel members who retired from the KN panel within 5 months of the date their survey round closed were less likely to participate.

Table 2

Did the Invited Panel Member Participate in the Survey

Variable	Logistic Odds Ratio	Logistic Standard Error	DProbit dF / dx	DProbit Standard Error
Log (Income)	1.0172	0.0472	0.0026	0.0070
Years of education	1.0069	0.0167	0.0013	0.0025
Age	1.0083 ***	0.0030	0.0013 ***	0.0005
Race: Black	0.6408 ***	0.0684	-0.0748 ***	0.0184
Race: Non-black, Non-white	0.8162	0.1245	-0.0361	0.0254
Hispanic	0.7302 ***	0.0843	-0.0521 ***	0.0197
Gender: Female	0.9605	0.0751	-0.0073	0.0118
Household Size	0.8688 ***	0.0252	-0.0219 ***	0.0044
Region: Northeast	0.8703	0.1059	-0.0226	0.0192
Region: South	0.9471	0.0985	-0.0102	0.0159
Region: West	0.9697	0.1155	-0.0069	0.0182
Currently Employed	0.8874	0.0780	-0.0173	0.0131
Not living in 150 largest MSA's	0.9506	0.0984	-0.0073	0.0159
Owner of Residence	0.7732 ***	0.0713	-0.0398 ***	0.0135
Marital Status: Married	1.2441 **	0.1129	0.0340 **	0.0138
Dual Income Household	0.9987	0.0863	0.0023	0.0130
Head of Household	0.9044	0.0923	-0.0147	0.0151
Time as Panel Member, in Months	0.9955 *	0.0027	-0.0007 *	0.0004
Stress Level	0.7011 ***	0.0636	-0.0546 ***	0.0138
Stress Data Missing	0.3115 ***	0.0369	-0.2131 ***	0.0238
Respondent Retired from KN Panel	0.0804 ***	0.0064	-0.4774 ***	0.0146

N	Complete= 1	Complete= 0
5639	4246	1393

Notes: * significant at .10 level, ** significant at .05 level, *** significant at .01 level, all two-tailed tests.

B. Impact of respondent characteristics on valuation estimates.

We consider these effects above predictable. The same kinds of people who are difficult to recruit on to a panel (minorities, unmarried, high stress) will be those who are going to be more likely to not participate. The next section considers whether respondent panel behavior altered our critical variable, the dollar value of a 1% improvement in water quality. These results are shown in Table 3.

The first variable describes whether the respondent stopped the survey and then continued taking the survey at a later time. Conceivably such respondents might be less engaged in the survey task. However, there was no significant effect of this variable on benefit values.

The second variable of interest is the time the person has been a member of the Knowledge Networks panel. Length of time in the panel may affect attentiveness to the surveys, and potentially could be correlated with other personal characteristics that influence water quality valuations. The results fail to indicate any significant effect of this variable as well.

Third, the days the respondent took to complete the survey after first being offered the opportunity to participate could reflect a lack of interest in the survey topic or in taking surveys generally. This variable did have a significant but small negative effect. Given that respondents were told the general subject of the survey before taking it, this result could demonstrate a lower concern or interest in the subject matter exhibited in a delay to complete the survey.

Finally, if the respondent retired from the Knowledge Networks panel within five months of taking the survey (this duration was used so that all rounds could be analyzed for the same amount of time), that might be associated with a diminishing interest in taking surveys. The results did not show any significant effect on our results from this variable.

Overall, this analysis provides evidence that these key aspects of the Knowledge Networks methodology had minimal impact on the survey responses. The number of days a panel member took to complete the survey after being invited had a small but significantly negative effect on regional water quality value, but when demographic characteristics were included, the significance of that effect was lost.

Table 3
Impact of Respondent Characteristics on Water Valuation

Variable	Log (Dollar Value for 1% Better Water Quality) Parameter Estimate	Standard Error	Log (Dollar Value for 1% Better Water Quality) Parameter Estimate	Standard Error
Log (Income)	0.1216 ***	0.0234	-	-
Years of education	0.0394 ***	0.0082	-	-
Age	0.0069 ***	0.0013	-	-
Environmental Organization Membership	0.5298 ***	0.0897	-	-
Visited a Lake or River, last 12 Months	0.1940 ***	0.0440	-	-
Race: Black	-0.1425 **	0.0615	-	-
Race: Non-black, Non-white	0.0102	0.0838	-	-
Hispanic	0.1095	0.0670	-	-
Gender: Female	-0.0498	0.0400	-	-
Household Size	-0.0304 *	0.0156	-	-
Region: Northeast	0.0214	0.0621	-	-
Region: South	-0.0524	0.0566	-	-
Region: West	-0.0315	0.0619	-	-
State Lake Quality	0.0003	0.0008	-	-
Lake Acres per State Square Mile	0.0044 **	0.0022	-	-
Subject Stop and Continue Survey Later	-0.0938	0.0753	-0.0919	0.0772
Time as Panel Member, in Months	-0.0021 *	0.0012	-0.0001	0.0012
Days from Invitation to Completion	-0.0037	0.0025	-0.0067 ***	0.0025
Respondent Retired from KN Panel	-0.0115	0.0595	-0.0887	0.0607
INTERCEPT	0.5210 **	0.2624	2.6951 ***	0.0382
N	Pseudo R-square		N	Pseudo R-square
4033	0.0172		4033	0.0010
(376 left censored)			(376 left censored)	
(403 right censored)			(403 right censored)	

Notes: * significant at .10 level, ** significant at .05 level, *** significant at .01 level, all two-tailed tests.

IV. Research Needs/Additional Observations

Our research showed that the KN panel matches the desired universe and that within panel selection has a relatively minor part in our estimates. We also performed reweighting giving the weights supplied by KN to adjust for panel biases, but got virtually no differences in our estimates. Additionally, a Heckman adjustment for non-response found minimal impact.

That said, there is no doubt in our mind that the KN panelists are different from other sampled populations. Compared with our studies in malls and with central location the KN panelists are more conscientious, in the sense of being less likely to stop in the middle of the survey; they are more careful, in the sense of being less likely to mistakenly choose a dominated alternative, and they had less difficulty with the survey as measured by how hard or easy they found the survey. They were, in short, more ideal respondents.

In hindsight, it is perhaps not surprising that the KN panelists do not differ substantially from a representative sample with respect to the value of water quality. We find that demographics, such as income, education and age have the greatest impact on that valuation, and for those demographics the panel matches the desired universe.

However for different policy variables one might not expect the same null results. The fact that KN panel members are by self selection more compliant, rule following and computer literate means that for a number of variables projected from the panel may not correspond to the nation. For example, one could reasonably hypothesize KN panelists being different from the nation for the following kind of ways.

- KN respondents are more likely to be honest about their tax returns
- KN respondents are likely to be more knowledgeable about current events
- KN respondents are more likely to be in favor of new electronic versions of communication
- KN respondents are more likely to be 'balanced' with respect to highly polarized issues, since they are accustomed to seeing both sides of issues

Notice, however, that these are hypotheses. Research is needed to characterize the kinds of questions for which recruited panel members are consistently different from the population. It would be then possible for KN to sample and recruit to minimize such biases. Further, future research could establish a series of standard 'distortion indicator' questions that could be used adjust any survey to the population at large.

Developing and using such distortion indicator questions would be a valuable project for future research. What is needed is an assessment of the biases in panels with respect to a large series of possible questions. This large battery would be factor analyzed to determine relatively small number of distortion indicators that correspond to each factor. Then when one ran a survey, the analyst could ask whether the key variable is correlated with any of these distortion indicators. If the answer is no, then the researcher can be confident that the bias is relatively minor. If yes, then Heckman-like adjustments are available provided one knows the value of the indicator

questions. This adjustment is possible for panels because each member's response to the distortion indicator questions could be part of the panel's background information.

Popular Support for Climate Change Policy

I. Basic Information

1. Title of project: NSF "Popular Support for Climate Change Policy: Regional, Intertemporal, and Sociodemographic Heterogeneity in Risk Perception, Receptiveness to 'Expert' Opinions, and Willingness to Incur Costs of Mitigation"

2. Sponsor(s): NSF (Trudy Ann Cameron, SBR 98-18875, 4/99-3/01, $179,985)

3. Date work was (is expected to be) finished: Some papers complete and published, some still under review/revision

4. Timeline (if any) for review of this survey by OMB: none

5. Team members and affiliations (including survey administrator, e.g. KN): Online survey: Geoffrey R. Gerdes, UCLA Ph.D. student, now at the Federal Reserve Board of Governors in Washington DC (and surveying commercial banks from time to time); Mail survey: Jae-Seung Jason Lee, UCLA Ph.D. student, now with ICF International, in Los Angeles.

6. Author of this summary: Trudy Ann Cameron

7. Purpose of the study: To assess the determinants of heterogeneity in stated preferences towards climate change mitigation policies, given the uncertainty and long time horizons, as well as up-front costs and far future benefits of mitigation.

8. Type of stated preference survey (e.g. CV, conjoint): Format for elicitation was randomized across respondents in the online version of the survey. Generally, this is a conjoint-type survey, but variants that were presented as a binary-choice CV were also used. Each respondent answered only a single choice question. Mail survey used a binary choice question, but listed the attributes of the alternatives (costs and first-round incidence of domestic and international costs) along with respondents' subjective assessments of the consequence of business-as-usual.

9. Where are descriptions/results of the study published/available?
For the preliminary surveys, two papers have been published:

- Trudy Ann Cameron, "Individual option prices for climate change mitigation," *Journal of Public Economics*, 2005, 89, 283-301.
- Trudy Ann Cameron, "Updating Subjective Risks in the Presence of Conflicting Information: An Application to Climate Change," *Journal of Risk and Uncertainty*, 2005, 30(1) 63-97.

For the main mail survey:

- J. Jason Lee, "Subjective Benefits from Climate Change Mitigation: Results from a Household Survey," Ph.D. dissertation, UCLA, Department of Economics, 2002.

A paper on response/non-response bias assessment for this sample is:

- J. Jason Lee and Trudy Ann Cameron, "Thorough Non-response Modeling as an Alternative to Minimum Survey Response Rate Requirements: Popular Support for Climate Change Mitigation," undergoing final line edits as of 09/08/06.

For the online survey:

- Trudy Ann Cameron and Geoffrey R. Gerdes, "Individual Subjective Discounting: Form, Context, Format, and Noise" (revise-and-resubmit at JRU)
- Daniel Burghart, Trudy Ann Cameron, and Geoffrey R. Gerdes, "Questioning some maintained hypotheses in stated preference estimation: allowing for failure, substitutes, adjustments, and inattention" (submitted to JEEM)

More papers are in the pipeline.

10. Sampling description

i. Population of interest (e.g. voters, adults, households):
For the preliminary surveys: a convenience sample of students at one university;
For the online survey: a convenience sample of students at over around 100 universities
For the mail survey: the general population

ii. Frame(s) (e.g. RDD): best we could do for the very limited budget available. Drawn as a stratified random sample, in proportion to populations of individual states (or groups of small states) from the universe of households offered by Households USA this inventory of households is a product of InfoUSA Inc., 5711 S. 86th Circle, Omaha, NE 68127. (888)999-0580. The database is used extensive for direct mail marketing.

iii. Mode(s) of administration (e.g. internet, mail, in-person; please indicate if multiple waves)**:**
Mail survey: surveys went out in waves every week (Tuesdays) for a full year (started on Tuesday, February 20, 2001 and ended on February 17, 2002, in order to permit current weather conditions to affect survey topic salience and concern about climate change).
Online survey: individual faculty at a wide range of colleges and universities were recruited to introduce the survey to their students. Survey remains online at http://globalpolicysurvey.ucla.edu (but no longer collects data)

iv. Location of administration (central facility, home):
Mail survey: completed by household recipients (adult in dwelling with most recent birthday) at the location of their choice
Online survey: completed wherever student had internet access

v. Total number of unique cases fielded (i.e. the total number of unique contacts attempted to survey):
Mail survey: Survey mailed to 8576 U.S. addresses; there was also a Canadian sample, but targeted households could not be geocoded and comparable geocoded Census data was not readily available.

vi. Response and Contact Rates[3]:

Mail survey: 1651 sufficiently complete responses for thorough analysis (19.25% response rate if absolutely no adjustments are made to the denominator)

II. Ex Ante Strategies for Reducing and Testing for Sampling Biases

In one or two paragraphs describe any steps you took before the survey was completed to boost response rates or otherwise try to make the sample more representative of the population of interest (e.g., paying incentives, non-response conversion techniques such as follow-up phone calls). Also indicate any steps taken to prepare for later tests of sample representativeness, such as asking basic questions of non-responders.

Mail survey: We used a $1 bill as an incentive in each mail survey, except for the Canadian sample, where only dollar coins remain in circulation. We followed Dillman's method by sending an advance notification first, then a cover letter with survey on the second week, and then a reminder postcard on the third week. For the first 1000 survey recipients, we also sent replacement survey packets, one month after sending the original surveys, to households who did not respond. We subsequently decided to cancel this step due to the extremely low response rate (only about 5%.) and high cost for these replacement surveys. We elected instead to send out a larger number of original surveys and rely upon our formal response/non-response correction process.

III. Ex Post Analyses of Sample Representativeness and Consequences

Describe any techniques you used to adjust for potential observable bias from non-response (or coverage bias) such as weighting. Further, describe any tests believed to *suggest* the presence or absence of non-observable non-response bias in the sample, such as comparing respondents who answered the survey quickly to those that answered it later or testing for convergent validity across different modes/frames. Finally, report any attempts to correct for selection on unobservables, such as a Heckman selection correction. For any of these analyses, report any results that suggest the consequences of non-response on the outcome of interest (e.g., willingness to pay). This section should be several paragraphs long.

Mail survey: We geocoded each survey recipient's address to a precision of six decimal degrees of latitude and longitude, and mapped these to the census tract and county containing the dwelling. Using these geo-data, we merged the recipient sample with data from a wide variety of sources to construct variables in several categories: The explanatory variables we use can be classified into several categories: features of the survey instrument and its implementation, current versus historical temperatures, potential climate change vulnerability, seasonality, major discrete events, the county-level political mix, attributes of the address or addressee, and census tract sociodemographic characteristics. In all, our response/non-response specification considers

[3] Please use APPOR Response Rate 1. If available, please provide Response Rate 3 and Contact Rate 1. Please provide all variables used in the calculation (sample size, contacts, complete interviews, etc.). The APPOR standard response rate and contact rate definitions can be found here http://www.aapor.org/pdfs/newstandarddefinitions.pdf . If you do not use the AAPOR standard definitions, please describe how you calculated a response rate.

67 potential explanatory variables, finding nineteen of them to be statistically significant determinants of response propensities.

IV. Research Needs/Additional Observations

Briefly describe any research needs for analyzing sample representativeness.

Further, please briefly provide any additional observations you believe are salient to the question of sample representativeness in these surveys. For example, you might to comment on the relative importance of addressing concerns of sample bias versus other concerns associated with stated preference surveys.

Mail survey: We feel we have done just about everything we could have done to capture heterogeneity in response propensities among the targeted households for the mail survey. It was a relief to find that despite the heterogeneity in response propensities, a Heckman selectivity correction makes only negligible differences to the parameter estimates.

The Economics of Environmental Vulnerability: Epidemiology, Household Status, and Air Pollution

I. Basic Information

1. Title of project: The Economics of Environmental Vulnerability: Epidemiology, Household Status, and Air Pollution

2. Sponsor(s): Environmental Protection Agency, STAR

3. Date work was (is expected to be) finished: November 2006

4. Timeline (if any) for review of this survey by OMB: not relevant

5. Team members and affiliations (including survey administrator, e.g. KN): V. Kerry Smith, Arizona State University; Mary F. Evans, University of Tennessee; Christine Poulos, RTI International

6. Author of this summary: Mary F. Evans and V. Kerry Smith

7. Purpose of the study: Use pilot data collection effort to illustrate a strategy for examining the potential importance of dependency relationships and altruism using stated choice questions. This strategy offers a method for using pilot questions to help frame preference restrictions that subsequently condition the structure used for choice questions intended to recover valuation-related tradeoffs. This approach offers an alternative to focus groups and cognitive interviews, both of which are often used to inform the development of surveys.

8. Type of stated preference survey (e.g. CV, conjoint): CV

9. Where are descriptions/results of the study published/available? Work in progress

10. Sampling description

i. Population of interest (e.g. voters, adults, households): Adults

ii. Frame(s) (e.g. RDD): internet panel composed from RDD initial contacts

iii. Mode(s) of administration (e.g. internet, mail, in-person; please indicate if multiple waves)**:** Internet

iv. Location of administration (central facility, home): Home as far as we know; since internet survey could in principle be answered anywhere

v. Total number of unique cases fielded (i.e. the total number of unique contacts attempted to survey): 2670 based on information provided by KN

vi. Response and Contact Rates[4]: 2110/2670 = 79%

II. Ex Ante Strategies for Reducing and Testing for Sampling Biases

This was a pilot to evaluate a set of hypotheses that could not be addressed within focus groups. As a result we did not we did not include incentives beyond those provided by KN for participation. We did negotiate to obtain information on demographic attributes of those KN panelists who were invited to participate and did not. We also acquired geographic information to allow a link to census demographics. Finally we included questions that have been asked on other non-internet surveys that could be used with a sample from those surveys that was matched based on demographic attributes. The responses to these common questions could then be compared to gauge whether the KN sample had different behavioral responses in a related area.

III. Ex Post Analyses of Sample Representativeness and Consequences

We have not estimated selection models with this survey. In another survey conducted with Carol Mansfield using KN we did undertake such an analysis; introducing inverse Mills ratios into behavioral choice models. We also compared responses to common questions with other surveys.

IV. Research Needs/Additional Observations

Judgments about how representative a sample may be depend on the objectives of the research. As a result they should be based on comparing response to related behaviors in matched random samples taken from existing surveys to confirm that the features of any new survey do not diminish its ability to detect behaviors consistent with its primary objective.

[4] Please use APPOR Response Rate 1. If available, please provide Response Rate 3 and Contact Rate 1. Please provide all variables used in the calculation (sample size, contacts, complete interviews, etc.). The APPOR standard response rate and contact rate definitions can be found here http://www.aapor.org/pdfs/newstandarddefinitions.pdf . If you do not use the AAPOR standard definitions, please describe how you calculated a response rate.

Valuation of Natural Resource Improvements in the Adirondacks

I. Basic Information

1. Title of project: Valuation of Natural Resource Improvements in the Adirondacks

2. Sponsor(s): Environmental Protection Agency, Clean Air Markets Division

3. Date work was (is expected to be) finished: August 2004

4. Timeline (if any) for review of this survey by OMB: NR

5. Team members and affiliations (including survey administrator, e.g. KN):

   Dallas Burtraw, RFF
   Alan Krupnick, RFF
   Spencer Banzhaf, RFF
   David Evans, RFF/University of Maryland
   Knowledge Networks, survey administrator

6. Author of this summary: Spencer Banzhaf and David Evans

7. Purpose of the study: To estimate WTP of New York State residents for ecological improvements in the Adirondacks State Park, such as would be expected to result from reductions in acid deposition associated with likely national air quality policies (e.g. CAIR).

8. Type of stated preference survey (e.g. CV, conjoint): Contingent valuation

9. Where are descriptions/results of the study published/available? Report to EPA, Resources for the Future report (available at www.rff.org) in forthcoming *Land Economics* article.

10. Sampling description

   i. Population of interest (e.g. voters, adults, households): New York State adults (one per household)

   ii. Frame(s) (e.g. RDD): RDD. Three waves, one of the Knowledge Networks (KN) panel, one of former panel members, and the other of a fresh RDD sample.

   iii. Mode(s) of administration (e.g. internet, mail, in-person; please indicate if multiple waves)**:** Three waves. First two waves internet/web TV. Third wave mail.

   iv. Location of administration (central facility, home): Home.

v. Total number of unique cases fielded (i.e. the total number of unique contacts attempted to survey): 8868 (Panel: 1,351 individuals/1,284 households, Former Panel: 3,612 individuals/2,120 households, Mail: 3,905)

vi. Response and Contact Rates[5]:
-Main KN panel: 951/1,284 = 74%
-Former KN panel: 293/2,120 = 14%
-Mail: 570/2,372 = 24%
I: 570, R: 1533, NC: 117, UO: 152

II. Ex Ante Strategies for Reducing and Testing for Sampling Biases

The base and scope versions of the survey were fielded to residents of New York in KN's internet panel. To examine whether the KN panelists are representative of the general population and to increase the total number of responses, the base and scope surveys were also fielded by internet to former panel members and the base survey in paper form to an RDD mail sample.

To boost the response rate from the KN panel the survey was open for at least three weeks and at least three reminder emails were sent. Furthermore, reminder phone calls were made to those who did not complete the survey. Where possible, the former panel members received reminder emails and about half of them received reminder phone calls. The former panel members received $10 for completing the survey.

For the mail survey there was an initial contact letter prior to delivery of the survey and two reminder letters. The final reminder letter allowed respondents to take the survey on the internet; few of the respondents took this option. Furthermore, $10 was paid to each respondent to the mail survey. The mail survey required two administrations to achieve the targeted number of responses. Up to five call attempts were made to remind mail recipients to take the survey. Little process data was collected from the mail administration and none was used in our statistical analysis.

III. Ex Post Analyses of Sample Representativeness and Consequences

First, we weighted all respondents by their observable characteristics, including location of residence, so that the sample would reflect NY State's demographic distribution.

Second, as noted above, we surveyed KN's panel as well as a new RDD sample (via mail). Ignoring the possibility of an effect introduced by the mail (vs. internet) mode, the latter survey provides a check on the selectivity of KN's panel. After weighting the data to account for the differing demographics of the sample, we could not reject the hypothesis of equal WTP from the differing survey modes. Thus the internet mode/panel administration and the mail mode/RDD administration were either both unbiased or else identically so.

[5] Please use APPOR Response Rate 1. If available, please provide Response Rate 3 and Contact Rate 1. Please provide all variables used in the calculation (sample size, contacts, complete interviews, etc.). The APPOR standard response rate and contact rate definitions can be found here http://www.aapor.org/pdfs/newstandarddefinitions.pdf . If you do not use the AAPOR standard definitions, please describe how you calculated a response rate.

Third, we used process data to analyze whether more responsive or experienced KN panel members responded differently. We analyzed whether WTP was correlated with the lag between when the invitation to take the survey was sent and when it was completed and found the lag to be insignificant. Similarly, we found that the tenure of the respondent on the panel and the number of surveys they had taken previously were not significant in explaining WTP.

Fourth, we employed a Heckman selection equation to test and correct for bias resulting from selection on unobservables. We can do this because KN elicited initial background demographic and attitudinal questions for all its panel members. Thus, we have individual-level details about nonrespondents. This information provides a unique opportunity to identify factors leading to non-response and, more importantly, to test whether non-response biases our WTP estimates.

We found that whites, middle-aged people, members of smaller households, and environmentalists were all more likely to respond to the survey, as were households in certain regions of the state. Households who were not given or did not answer questions about political and environmental attitudes in previously administered panel profile surveys were more likely to be non-responders. Importantly, we could not reject the hypothesis that the correlation between the unobserved components of the selection and WTP equations is zero. This suggests that there is no selection bias. Moreover, mean WTP estimates from models with and without the Heckman correction for bias differ only by about 5 percent, and are not significantly different.

IV. Research Needs/Additional Observations

First, further research into the differential effects of mode and administration is warranted. Our comparison of the KN panel using their internet mode with a new RDD sample using a mail mode conflated these two effects. A research design that differentiated mode and administration would be useful.

Second, additional research into the selection effects of KN's (or other panel's) recruitment and retention would be helpful. Aggressive non-response conversion could be used on a sample of people who did not join the panel or who left the panel early. Their responses could be compared to those on the panel.

Improved Valuation of Ecological Benefits Associated with Aquatic Living Resources

I. Basic Information

1. Title of project: **Improved Valuation of Ecological Benefits Associated with Aquatic Living Resources: Development and Testing of Indicator-Based Stated Preference Valuation and Transfer**

2. Sponsor(s): **EPA/STAR**

3. Date work was (is expected to be) finished: **September 30, 2008**

4. Timeline (if any) for review of this survey by OMB: **Not required.**

5. Team members and affiliations (including survey administrator, e.g. KN):
Robert J. Johnston, Agricultural and Resource Economics, University of Connecticut
Eric Schultz, Department of Ecology and Evolutionary Biology, University of Connecticut
Kathleen Segerson, Department of Economics, University of Connecticut
Elena Besedin, Abt Associates.

6. Author of this summary: **Robert J. Johnston**

7. Purpose of the study: **This project will develop and test Indicator-Based Stated Preference Valuation for aquatic living resources and assess implications for benefits transfer. As a case study, we will estimate preferences and WTP for multi-attribute diadromous fish restoration projects in New England states. Among areas of focus are the use of established and tested ecological indicators in choice experiment surveys to characterize ecosystem changes, and to isolate the specific aspects of restoration that are valued by the public.**

8. Type of stated preference survey (e.g. CV, conjoint): **choice experiment (conjoint)**

9. Where are descriptions/results of the study published/available? **None currently published.**

10. Sampling description

i. Population of interest (e.g. voters, adults, households): **Households (most likely Rhode Island households).**

ii. Frame(s) (e.g. RDD): **Expected RDD telephone solicitation with Rhode Island households for initial contact, followed by mail survey for households that agree to participate. Phone contact will ask for agreement to participate in mail survey, collect minimal demographic and attitudinal data for sample selectivity testing, and screen for households with a DVD player required to play DVD containing information and**

instructions. Subsequent mail survey will be sent to those agreeing to participate and who pass screening criteria (DVD player in household).

iii. Mode(s) of administration (e.g. internet, mail, in-person; please indicate if multiple waves)**: Anticipated dual mode phone-mail. Initial contact via phone to obtain agreement to participate in survey and to screen for households with CD or DVD equipment (see above). Second contact is a mail survey to households agreeing to participate. Mail survey will incorporate multiple mailings (likely 3-4) to increase response rates (e.g., following Dillman 2000). We may also consider dual phone-internet survey following similar mechanisms.**

iv. Location of administration (central facility, home): **Home**

v. Total number of unique cases fielded (i.e. the total number of unique contacts attempted to survey): **We hope to obtain a final usable sample of at least 1,500 completed mail surveys.**

vi. Response and Contact Rates[6]: **Not yet known.**

II. Ex Ante Strategies for Reducing and Testing for Sampling Biases

In one or two paragraphs describe any steps you took before the survey was completed to boost response rates or otherwise try to make the sample more representative of the population of interest (e.g., paying incentives, non-response conversion techniques such as follow-up phone calls). Also indicate any steps taken to prepare for later tests of sample representativeness, such as asking basic questions of non-responders.

These are currently being developed. We hope that the workshop will provide insights into particularly effective techniques. See discussion of concerns and challenges below in part IV. Anticipated methods include follow-up mailings following Dillman (2000) and perhaps follow-up telephone contacts. Small incentives will be paid ($1) if the budget allows. The initial phone solicitation will attempt to collect very basic demographic information (and perhaps responses for very simple attitudinal questions) from non-responders (those who do not agree to participate in the mail survey) to allow simple tests of differences between mail survey responders and non-responders. The mail survey will include questions requesting more complete demographic and other data to characterize respondents and allow comparisons to average RI households (e.g., using census data).

III. Ex Post Analyses of Sample Representativeness and Consequences

[6] Please use APPOR Response Rate 1. If available, please provide Response Rate 3 and Contact Rate 1. Please provide all variables used in the calculation (sample size, contacts, complete interviews, etc.). The APPOR standard response rate and contact rate definitions can be found here http://www.aapor.org/pdfs/newstandarddefinitions.pdf . If you do not use the AAPOR standard definitions, please describe how you calculated a response rate.

Describe any techniques you used to adjust for potential observable bias from non-response (or coverage bias) such as weighting. Further, describe any tests believed to *suggest* the presence or absence of non-observable non-response bias in the sample, such as comparing respondents who answered the survey quickly to those that answered it later or testing for convergent validity across different modes/frames. Finally, report any attempts to correct for selection on unobservables, such as a Heckman selection correction. For any of these analyses, report any results that suggest the consequences of non-response on the outcome of interest (e.g., willingness to pay). This section should be several paragraphs long.

The survey has not yet been conducted. However, anticipated methods include the comparison of the surveyed sample to the RI population using demographic indicators, with weighting used to adjust estimation results. These will be compared to unweighted results to assess potential implications of sampling bias for model results and WTP estimates. We will also contrast the basic demographic (and perhaps attitudinal) information collected from the initial phone contact with that from the final mail survey sample—similarly allowing a comparison of responders and non-responders. Heckman-type selection correction models are currently not well developed for conditional logit (choice experiment) estimation models—and particularly mixed logit models—so it is not anticipated that these will be used. However, if these models have been sufficiently developed by the time the survey data have been collected, the use of such models will be considered.

IV. Research Needs/Additional Observations

Briefly describe any research needs for analyzing sample representativeness.

Further, please briefly provide any additional observations you believe are salient to the question of sample representativeness in these surveys. For example, you might to comment on the relative importance of addressing concerns of sample bias versus other concerns associated with stated preference surveys.

We are one year in to the research effort and are currently designing our survey and implementation methods. Our primary sampling concern is the desire to obtain a representative sample of Rhode Island households. A second concern is the need to obtain a sufficiently large sample to conduct required statistical tests. The primary difficulties are related to the need to present substantial information in the survey instrument. Focus groups have indicated that this will likely require the inclusion of the DVD or similar media to present information (i.e., during an 8-12 minute video presentation), given respondents typical unwillingness to read long text passages in stated preference surveys. This will require an additional screening during the initial RDD telephone contact, and may further depress response rates due to the additional requirement of watching the DVD. (The survey will include a "test" question (e.g., a unique code number) that can only be answered after watching the DVD, to ensure that respondents watch the material and to identify those who do not.) Our desire is to ensure that these survey design elements do not contribute to a biased and/or overly small usable sample of responses. We currently anticipate a phone-mail survey but are also considering the possibility of a phone-internet

survey of similar format. We are also interested in researcher's comparative experiences with mail versus internet implementation of choice experiment surveys in terms of response rates and non-response bias. We also want to ensure that survey response rates are not correlated with environmental attitudes (e.g., that those with greater concern for environmental issues respond at higher rates than others).

Willingness to Pay Survey for §316(B) Regulations

I. Basic Information

1. Title of project: **Willingness to Pay Survey for §316(B) Regulations**

2. Sponsor(s): **EPA/OW/EAD**

3. Date work was (is expected to be) finished: **To be determined. The survey was planned to support benefits analysis of the §316(B) regulations for Phase III facilities. However, EPA decided to withdraw its ICR for the 316b rule due to time constraints and to finish this study as a research project.**
4. Timeline (if any) for review of this survey by OMB: **January – March 2007 (tentative)**

5. Team members and affiliations (including survey administrator, e.g.):
Elena Besedin, Abt Associates Inc.,
Robert J. Johnston, Agricultural and Resource Economics, University of Connecticut
Knowledge Networks, Survey Administrator
.

6. Author of this summary: **Elena Besedin**

7. Purpose of the study: **This study will estimate WTP for reductions in fish losses at cooling water intake structures (CWIS) due to the 316(b) regulations. The study would focus on a broad range of aquatic species, including forage fish and a variety of fish species harvested by commercial and recreational fishermen.**

8. Type of stated preference survey (e.g. CV, conjoint): **choice experiment (conjoint)**

9. Where are descriptions/results of the study published/available? **None currently published.**

10. Sampling description

i. Population of interest (e.g. voters, adults, households): **U.S. Households**

ii. Frame(s) (e.g. RDD): **RDD. Three waves. (1) Survey subjects will be randomly selected from a representative national panel of respondents maintained by Knowledge Networks. (2) To assess non-response bias subjects will be randomly selected from former KN panel members. (3) A fresh RDD sample will be selected.**

iii. Mode(s) of administration (e.g. Internet, mail, in-person; please indicate if multiple waves)**: (1) Internet/web TV; (2)/(3) phone**

iv. Location of administration (central facility, home): **Home**

v. Total number of unique cases fielded (i.e. the total number of unique contacts attempted to survey): **The target number of completed interviews is 3,900 for the main web survey and 600 for the non-response study.**

vi. Response and Contact Rates[7]: **Not yet known.**

II. Ex Ante Strategies for Reducing and Testing for Sampling Biases

In one or two paragraphs describe any steps you took before the survey was completed to boost response rates or otherwise try to make the sample more representative of the population of interest (e.g., paying incentives, non-response conversion techniques such as follow-up phone calls). Also indicate any steps taken to prepare for later tests of sample representativeness, such as asking basic questions of non-responders.

The survey administrator, KN will use its standard procedures to boost the response rates: (1) reminder emails will be sent to KN panel members who participate in the survey and reminder phone calls will be made to those who did not complete the survey within a specified time period; and (2) survey participants will receive incentives for completing the survey.

To avoid potential sampling biases associated with the web-based survey methodology, the survey sample will be stratified by demographic variables including age, education, Hispanic ethnicity, race, gender, and household income. The number of respondents in each demographic stratification group will be inversely proportional to the historical response rates of individuals in that group for similar types of surveys. By over-sampling groups that tend to have lower response and consistency rates, the demographic characteristics of respondents who provide valid completed surveys will mirror U.S. Census Bureau demographic benchmarks more closely.

We will assess the extent of nonresponse bias in key outcome measures in the Knowledge Networks sample of 3,900 respondents by obtaining national estimates for the key attitudinal and socio-demographic variables from a high quality random-digit-dialing survey of approximately 500 adults.

III. Ex Post Analyses of Sample Representativeness and Consequences

Describe any techniques you used to adjust for potential observable bias from non-response (or coverage bias) such as weighting. Further, describe any tests believed to *suggest* the presence or absence of non-observable non-response bias in the sample, such as comparing respondents who answered the survey quickly to those that answered it later or testing for convergent validity

[7] Please use APPOR Response Rate 1. If available, please provide Response Rate 3 and Contact Rate 1. Please provide all variables used in the calculation (sample size, contacts, complete interviews, etc.). The APPOR standard response rate and contact rate definitions can be found here http://www.aapor.org/pdfs/newstandarddefinitions.pdf . If you do not use the AAPOR standard definitions, please describe how you calculated a response rate.

across different modes/frames. Finally, report any attempts to correct for selection on unobservables, such as a Heckman selection correction. For any of these analyses, report any results that suggest the consequences of non-response on the outcome of interest (e.g., willingness to pay). This section should be several paragraphs long.

The survey has not yet been conducted. However, anticipated methods include the comparison of the surveyed sample to the U.S. population using demographic indicators, with weighting used to adjust estimation results. These will be compared to unweighted results to assess potential implications of sampling bias for model results and WTP estimates. We will also contrast the basic demographic and attitudinal information collected from the phone survey with the former KN panel members and a new sample of respondents (waves 2 & 3) with that from the main survey sample—similarly allowing a comparison of responders and non-responders.

Heckman-type selection correction models are currently not well developed for conditional logit (choice experiment) estimation models—and particularly mixed logit models—so it is not anticipated that these will be used.

IV. Research Needs/Additional Observations

Briefly describe any research needs for analyzing sample representativeness.

Further, please briefly provide any additional observations you believe are salient to the question of sample representativeness in these surveys. For example, you might to comment on the relative importance of addressing concerns of sample bias versus other concerns associated with stated preference surveys.

First, comprehensive selectivity assessment for the Knowledge Networks panel at different stages would be useful, including: 1) RDD panel recruitment, 2) connection of web-survey equipment, 3) completion of initial profile survey, 4) panel retention, and 5) survey completion.

Second, development of statistical methodology (Heckman-type procedure) for selection correction models for conditional logit (choice experiment) estimation models—and particularly mixed logit models would be helpful.

The Nonconsumptive Value of Steller Sea Lion Protection Measures

I. Basic Information

1. Title of project: The Nonconsumptive Value of Steller Sea Lion Protection Measures

2. Sponsor(s): NOAA Fisheries

3. Date work was (is expected to be) finished: December 2007

4. Timeline (if any) for review of this survey by OMB: Expected review beginning Fall 2006

5. Team members and affiliations (including survey administrator, e.g. KN):

Robert D. Rowe, Stratus Consulting
David F. Layton, University of Washington
Daniel K. Lew, NOAA Fisheries
PA Consulting, survey administrator

6. Author of this summary: Dan Lew

7. Purpose of the study: To measure the public's preferences and willingness to pay (WTP) for various levels of protection being considered for the endangered Western stock and the threatened Eastern stock of Steller sea lions.

8. Type of stated preference survey (e.g. CV, conjoint): Stated preference choice experiment (attribute-based stated choice method)

9. Where are descriptions/results of the study published/available? Brief description of the study is available at http://www.afsc.noaa.gov/REFM/Socioeconomics/Projects/NCVSSLPM.htm. No publications or results are currently available.

10. Sampling description

i. Population of interest (e.g. voters, adults, households): U.S. adult residents (one per household)

ii. Frame(s) (e.g. RDD): Purchased sample from Survey Sampling, Intl.

iii. Mode(s) of administration (e.g. internet, mail, in-person; please indicate if multiple waves): Modified Dillman mail-telephone survey w/ five contacts (advanced letter, initial mailing with monetary incentive, postcard reminder, telephone follow-up and nudge, and second full-mailing).

iv. Location of administration (central facility, home): home

v. Total number of unique cases fielded (i.e. the total number of unique contacts attempted to survey): 424 in pretest, 4000 (expected) in full survey. Alaska will be oversampled.

vi. Response and Contact Rates[8]:

Pretest response rates for a Spring/Summer 2006 formal pretest are the following:

	Total sample		**Alaska subsample**		**Rest of U.S. subsample**	
	$2.00	**$5.00**	**$2.00**	**$5.00**	**$2.00**	**$5.00**
Total mailed	212	212	47	24	165	188
Undeliverable (Mail 1)	18	8	3	1	15	7
Undeliverable (Priority)	1	1	1	0	0	1
Refused						
- Refused entire survey	1	6	0	0	1	6
- Partial survey	0	0	0	0	0	0
Deceased	1	3	1	0	0	3
Language barrier/R incapable	0	1	0	0	0	1
Total successfully delivered (eligible)	192	200	42	23	150	177
Completed before calls	55	81	17	15	38	66
Completed after calls	12	6	4	1	8	5
Completed after 2nd mailing	0	11	0	0	0	11
Total completes	67	98	21	16	46	82
Response Rate (total completes/eligible)	**34.9%**	**49.0%**	**50.0%**	**69.6%**	**30.7%**	**46.3%**

II. Ex Ante Strategies for Reducing and Testing for Sampling Biases

Monetary incentives: To encourage participation in the mail survey, an honorarium of $5 was given to half the participants in the initial mailing, while the other half received $2.

Follow-up phone calls: To better understand why non-respondents did not return the survey and to determine if there are systematic differences between respondents and non-respondents, those contacted in follow-up phone call and identified as non-respondents were asked a few questions to gauge their reasons for not responding to the mail survey. These included select socioeconomic and demographic classification questions and a few attitudinal questions. Although not used for the analysis of the pretest data, the information collected from non-respondents will be used to correct for non-response bias where necessary in the full survey (e.g., Heckman method).

[8] Please use APPOR Response Rate 1. If available, please provide Response Rate 3 and Contact Rate 1. Please provide all variables used in the calculation (sample size, contacts, complete interviews, etc.). The APPOR standard response rate and contact rate definitions can be found here http://www.aapor.org/pdfs/newstandarddefinitions.pdf . If you do not use the AAPOR standard definitions, please describe how you calculated a response rate.

Priority Mail mailing: For our second full mailing, we used Priority Mail to emphasize the importance of the materials to the respondent.

III. Ex Post Analyses of Sample Representativeness and Consequences

No ex post analysis of sample representativeness was conducted for the pretest.

IV. Research Needs/Additional Observations

The outside Alaska response rates for this pretest were lower-than-expected, particularly given the amount of qualitative testing done to develop the instrument. We are currently investigating whether higher incentives will improve response rates, but would benefit from research that identifies other cost-effective approaches to boosting response rates in these types of surveys.

Valuing Visibility in National Parks

I. Basic Information

1. Title of project: Valuing Visibility in National Parks

2. Sponsor(s): National Park Service

3. Date work was (is expected to be) finished: Unknown

4. Timeline (if any) for review of this survey by OMB: Focus group materials to be reviewed this fall.

5. Team members and affiliations (including survey administrator, e.g. KN):

Kevin Boyle, Virginia Tech
Richard Carson, UCSD
Susan Johnson, NPS
Robert Paterson, Industrial Economics, Inc.
Robert Unsworth, Industrial Economics, Inc.
John Molenar, Air Resource Specialists
Robert Mitchell, Clark University
PA Consulting (Pam Rathbun and Robert Baumgartner), survey administrator

6. Author of this summary: Kevin Boyle and Robert Paterson

7. Purpose of the study: To estimate the benefits of visibility improvements in national parks and Class I wilderness areas. The study will focus in particular on the visibility improvements expected to result from the implementation of the Regional Haze Rule.

8. Type of stated preference survey (e.g. CV, conjoint): Attribute-based choice questions.

9. Where are descriptions/results of the study published/available? Not completed.

10. Sampling description

i. Population of interest (e.g. voters, adults, households): All households within the continental U.S.

ii. Frame(s) (e.g. RDD): All households with listed phone numbers.

iii. Mode(s) of administration (e.g. internet, mail, in-person; please indicate if multiple waves): Mail with telephone follow-up.

iv. Location of administration (central facility, home): Self-administered at home.

v. Total number of unique cases fielded (i.e. the total number of unique contacts attempted to survey): Four to five.

vi. Response and Contact Rates: Not implemented yet.

II. Ex Ante Strategies for Reducing and Testing for Sampling Biases

Generally following Dillman's Total Design Method, the main survey will be designed and implemented to maximize response rates. This will include a respondent-friendly questionnaire, high quality initial contact, personalization of correspondence and branding of survey materials. Implementation will involve up to five separate mailings, with the fourth and fifth mailings described below sent only to non-respondents. The mailings will consist of: (1) a pre-survey notification, (2) the initial mailing of the survey instrument with a $5 incentive, (3) a thank you/reminder postcard, (4) a replacement survey and reminder letter for non-respondents, and (5) a second replacement survey (via Priority Mail) and reminder letter for non-respondents. In addition to the mailings, nonrespondents will receive a telephone call urging them to complete the survey. This call will also be used to collect data on some key items in the survey in order to determine the characteristics of nonresponders. This call will commence one week after the first replacement survey is mailed to non-respondents and continue until one week after the second replacement survey is mailed to non-respondents.

III. Ex Post Analyses of Sample Representativeness and Consequences

A Response Rate Pilot Study will be conducted involving a split-sample comparison of the data obtained from self-administered questionnaires collected through mail and in-person survey modes. The study will be designed to develop a calibration equation that would be used to adjust responses to the nationwide mail survey data for potential non-response bias.

The Response Rate Pilot Study would be conducted by selecting a number of neighborhoods in each of two metropolitan areas, enumerating all owner-occupied households in each neighborhood, taking a simple random sample of owner-occupied households from each neighborhood, then randomly assigning one-half of each neighborhood sample to a group of mail survey households and a group of in-person survey households.

The mail pilot survey would be implemented using a methodology identical to the methodology used in the nationwide mail survey. The in-person pilot survey would be conducted in a manner that would minimize the differences between the two modes. That is, rather than administering the survey through a personal interview, the interviewer would hand the questionnaire to the respondent and ask the respondent to fill out the questionnaire. The questionnaire would be identical to the one mailed to respondents. The respondent would place the completed survey in a "ballot box" carried by the interviewer, thus minimizing the potential for social desirability bias and mimicking the "self-administered" nature of the mail survey, as much as possible.

Collected data (e.g., valuation information and various response rates from sampled neighborhoods) would be used to estimate the parameters of a calibration relationship such as:

$$MeanWTP = f(Response\ Rate_i, Survey\ Mode)$$

where *i* indicates neighborhood.

Mean WTP from the nationwide mail survey could then be calibrated using the response rate actually obtained and the estimated parameters of the above relationship.

IV. Research Needs/Additional Observations

Awaiting funding for full implementation.

Hawaiian Coral Reef Valuation Study

I. Basic Information

1. Title of project: **Hawaiian Coral Reef Valuation Study**

2. Sponsor(s): NOAA / National Ocean Service

3. Date work was (is expected to be) finished: 2007

4. Timeline (if any) for review of this survey by OMB: Fall 2006 / Winter 2007

5. Team members and affiliations (including survey administrator, e.g. KN):
    a. Richard Bishop, University of Wisconsin, Madison
    b. Robert Rowe, Stratus Consulting Inc.
    c. David Chapman, Stratus Consulting Inc.
    d. Norman Meade, NOAA
    e. Steve Thur, NOAA
    f. Bob Leeworthy, NOAA
    g. Tony Penn, NOAA
    h. Roger Tourangeau, University of Maryland, College Park

6. Author of this summary: David Chapman

7. Purpose of the study: Develop benefit estimates of protection of Hawaiian coral reef ecosystems. Develop estimates for the relative benefits of protecting the Northwestern Hawaiian Islands over the main Hawaiian islands.

8. Type of stated preference survey (e.g. CV, conjoint): Conjoint (Stated Choice)

9. Where are descriptions/results of the study published/available? Study is not complete, no results available at this time.

10. Sampling description

    i. Population of interest (e.g. voters, adults, households): Civilian Non-institutionalized population of the United States over the age of 18, as defined by the universe of U.S. households

    ii. Frame(s) (e.g. RDD): KN panel

    iii. Mode(s) of administration (e.g. internet, mail, in-person; please indicate if multiple waves)**:** Internet - KN panel

    iv. Location of administration (central facility, home): Home

v. Total number of unique cases fielded (i.e. the total number of unique contacts attempted to survey): Pilot – 216 completes out of 312 attempts ; Main – 1500 complete expected

vi. Response and Contact Rates[9]: 69% Response rate 216 completes out of 312 attempts

II. Ex Ante Strategies for Reducing and Testing for Sampling Biases

We will employ the following practices for the survey:

- field period of 3 to 4 weeks
- respondent incentives of $10 for participation
- use of the federal agency name in the e-mail invitation
- e-mail reminders
- telephone reminder calls to nonresponders.

Two types of respondent incentives are provided: survey-specific and nonsurvey-specific incentives:

Non-survey-specific incentives are used to maintain a high degree of panel loyalty and to prevent attrition from the panel. For the households provided Internet appliances and an Internet connection, their incentive is the hardware and Internet service. For households using their own personal computers and Internet service, Knowledge Networks enrolls the panelists into a points program that is analogous to a "frequent flyer" card in that respondents are credited with points in proportion to their regular participation in surveys. Panelists receive cash-equivalent checks approximately every four to six months in amounts reflecting their level of participation in the panel, which commonly results in distributions of $4 to $6 per month.

The survey-specific incentives are provided to respondents as a result of one of two conditions: 1) the survey is expected to require more than 20 minutes of time to complete, or 2) there is an unusual request being made of the respondent, such as specimen collection, the viewing of a specific television program, or completion of a daily diary. In these kinds of circumstances, panelists are being asked to participate in ways that are more burdensome than initially described to respondents during the panel recruitment stage. For the Coral Reef Protection Survey, an incentive will be provided because the survey is expected to require 20 or more minutes to complete. Extra encouragement will be required because the survey will require participants to read and digest more information than is the case in other types of surveys and because the topic of the survey will be of limited salience for significant numbers of people. Respondents who participate in the survey will be credited with 10,000 points, which equates to $10, which will be mailed to them at a later date.

In one or two paragraphs describe any steps you took before the survey was completed to boost response rates or otherwise try to make the sample more representative of the population of interest (e.g., paying incentives, non-response conversion techniques such as follow-up phone

[9] Please use APPOR Response Rate 1. If available, please provide Response Rate 3 and Contact Rate 1. Please provide all variables used in the calculation (sample size, contacts, complete interviews, etc.). The APPOR standard response rate and contact rate definitions can be found here http://www.aapor.org/pdfs/newstandarddefinitions.pdf . If you do not use the AAPOR standard definitions, please describe how you calculated a response rate.

calls). Also indicate any steps taken to prepare for later tests of sample representativeness, such as asking basic questions of non-responders.

One option for consideration is a nonresponse followup study (NRFUS) for the Coral Reef project. KN has experience conducting nonresponse followup surveys for measuring nonresponse bias and for improving the cumulative response rate. The method is to subsample randomly research subjects from defined groups of nonresponders, and then expend substantial resources to gain their cooperation. This is sometimes called a “double sampling” method. Persons, for instance, who were sampled to join the panel but did not join the panel would be eligible for sampling for this project. The same survey is used for these cases.

III. Ex Post Analyses of Sample Representativeness and Consequences

Describe any techniques you used to adjust for potential observable bias from non-response (or coverage bias) such as weighting. Further, describe any tests believed to *suggest* the presence or absence of non-observable non-response bias in the sample, such as comparing respondents who answered the survey quickly to those that answered it later or testing for convergent validity across different modes/frames. Finally, report any attempts to correct for selection on unobservables, such as a Heckman selection correction. For any of these analyses, report any results that suggest the consequences of non-response on the outcome of interest (e.g., willingness to pay). This section should be several paragraphs long.

Since the sampling has not been completed yet, we do not yet have any ex post analysis of sample representativness.

IV. Research Needs/Additional Observations

Briefly describe any research needs for analyzing sample representativeness.

Further, please briefly provide any additional observations you believe are salient to the question of sample representativeness in these surveys. For example, you might to comment on the relative importance of addressing concerns of sample bias versus other concerns associated with stated preference surveys.

North Atlantic Right Whale Economic Benefit Study

I. Basic Information

1. Title of project: **North Atlantic Right Whale Economic Benefit Study**

2. Sponsor(s): NOAA Northeast Fisheries Science Center

3. Date work was (is expected to be) finished: 2008

4. Timeline (if any) for review of this survey by OMB: Fall 2006 / Winter 2007

6. Team members and affiliations (including survey administrator, e.g. KN):
   a. Richard Bishop, University of Wisconsin, Madison
   b. David Chapman, Stratus Consulting Inc.
   c. Roger Tourangeau, University of Maryland, College Park
   d. Kathryn Bisack, NOAA

6. Author of this summary: David Chapman

7. Purpose of the study: Develop benefit estimates of protection of north Atlantic right whale for policy analysis.

8. Type of stated preference survey (e.g. CV, conjoint): Conjoint (Stated Choice)

9. Where are descriptions/results of the study published/available? Study is not complete, no results available at this time.

10. Sampling description

i. Population of interest (e.g. voters, adults, households): Civilian Non-institutionalized population of the United States over the age of 18, as defined by the universe of U.S. households

ii. Frame(s) (e.g. RDD): TBD – most likely either KN panel or Survey Sampling Inc. matched list.

iii. Mode(s) of administration (e.g. internet, mail, in-person; please indicate if multiple waves)**:** Either Mail or KN panel (Internet)

iv. Location of administration (central facility, home): Home

v. Total number of unique cases fielded (i.e. the total number of unique contacts attempted to survey): Approximately 1200-1500

vi. Response and Contact Rates[10]:

II. Ex Ante Strategies for Reducing and Testing for Sampling Biases

Anticipated efforts to increase response rates include:
For all modes:
Survey materials will prominently display NOAA sponsorship.
Incentive (5$ - $10) will be paid to increase response rates.

Mode dependent:
Mail:
Introductory letter on NOAA letterhead
Dillman mail design efforts – follow-up postcard, resending of instrument (phone calls if phone/address matched list used)
Internet:
Reminder emails

In one or two paragraphs describe any steps you took before the survey was completed to boost response rates or otherwise try to make the sample more representative of the population of interest (e.g., paying incentives, non-response conversion techniques such as follow-up phone calls). Also indicate any steps taken to prepare for later tests of sample representativeness, such as asking basic questions of non-responders.

III. Ex Post Analyses of Sample Representativeness and Consequences

Describe any techniques you used to adjust for potential observable bias from non-response (or coverage bias) such as weighting. Further, describe any tests believed to *suggest* the presence or absence of non-observable non-response bias in the sample, such as comparing respondents who answered the survey quickly to those that answered it later or testing for convergent validity across different modes/frames. Finally, report any attempts to correct for selection on unobservables, such as a Heckman selection correction. For any of these analyses, report any results that suggest the consequences of non-response on the outcome of interest (e.g., willingness to pay). This section should be several paragraphs long.

IV. Research Needs/Additional Observations

Briefly describe any research needs for analyzing sample representativeness.

Better understanding of the overall method for calculating response rates from matched (address/phone) survey sample lists.

[10] Please use APPOR Response Rate 1. If available, please provide Response Rate 3 and Contact Rate 1. Please provide all variables used in the calculation (sample size, contacts, complete interviews, etc.). The APPOR standard response rate and contact rate definitions can be found here http://www.aapor.org/pdfs/newstandarddefinitions.pdf . If you do not use the AAPOR standard definitions, please describe how you calculated a response rate.

Methods for non-response tests when using matched mail sample lists.

Further, please briefly provide any additional observations you believe are salient to the question of sample representativeness in these surveys. For example, you might to comment on the relative importance of addressing concerns of sample bias versus other concerns associated with stated preference surveys.

Template for Description of Study and Techniques for Addressing Nonresponse

For October 2, 2006 Sample Representativeness Workshop

Description: We, David A. Evans and Alan Krupnick, asked practitioners who attended the workshop to describe their survey and the techniques they were using or planning to use to address sample representativeness. To facilitate their response, we wrote this template and provided it to the practitioners. We also provided a copy of the completed template for the Adirondacks Valuation Survey (page C-22) as an example of the level of effort and information we were requesting. A copy of the template is provided here so that the reader may better understand what was asked of the practitioners.

I. Basic Information

1. Title of project:

2. Sponsor(s):

3. Date work was (is expected to be) finished:

4. Timeline (if any) for review of this survey by OMB:

5. Team members and affiliations (including survey administrator, e.g. KN):

6. Author of this summary:

7. Purpose of the study:

8. Type of stated preference survey (e.g. CV, conjoint):

9. Where are descriptions/results of the study published/available?

10. Sampling description

    i. Population of interest (e.g. voters, adults, households):

    ii. Frame(s) (e.g. RDD):

    iii. Mode(s) of administration (e.g. internet, mail, in-person; please indicate if multiple waves)**:**

    iv. Location of administration (central facility, home):

v. Total number of unique cases fielded (i.e. the total number of unique contacts attempted to survey):

vi. Response and Contact Rates[11]:

II. Ex Ante Strategies for Reducing and Testing for Sampling Biases

In one or two paragraphs describe any steps you took before the survey was completed to boost response rates or otherwise try to make the sample more representative of the population of interest (e.g., paying incentives, non-response conversion techniques such as follow-up phone calls). Also indicate any steps taken to prepare for later tests of sample representativeness, such as asking basic questions of non-responders.

III. Ex Post Analyses of Sample Representativeness and Consequences

Describe any techniques you used to adjust for potential observable bias from non-response (or coverage bias) such as weighting. Further, describe any tests believed to *suggest* the presence or absence of non-observable non-response bias in the sample, such as comparing respondents who answered the survey quickly to those that answered it later or testing for convergent validity across different modes/frames. Finally, report any attempts to correct for selection on unobservables, such as a Heckman selection correction. For any of these analyses, report any results that suggest the consequences of non-response on the outcome of interest (e.g., willingness to pay). This section should be several paragraphs long.

IV. Research Needs/Additional Observations

Briefly describe any research needs for analyzing sample representativeness.

Further, please briefly provide any additional observations you believe are salient to the question of sample representativeness in these surveys. For example, you might to comment on the relative importance of addressing concerns of sample bias versus other concerns associated with stated preference surveys.

[11] Please use APPOR Response Rate 1. If available, please provide Response Rate 3 and Contact Rate 1. Please provide all variables used in the calculation (sample size, contacts, complete interviews, etc.). The APPOR standard response rate and contact rate definitions can be found here http://www.aapor.org/pdfs/newstandarddefinitions.pdf . If you do not use the AAPOR standard definitions, please describe how you calculated a response rate.

APPENDIX D

Workshop Presentations

Sample Representativeness: Implications for Administering and Testing Stated Preference Surveys

RFF, October 2, 2006

Sponsored by NCEE, EPA

RESOURCES FOR THE FUTURE

Goals for the workshop

- Improve clarity about relationship between non-response bias and non-response rates
- Move towards consensus on appropriate protocols to reduce and test for bias
- Begin to develop research agenda

RESOURCES FOR THE FUTURE

Why this Workshop?

- We all want credible, low-cost cost-benefit analyses
- SP surveys are needed for estimating many types of benefits

BUT:

- "Difficult" review process by OMB

CAUSES

- OMB Guidance lacks clear rules (except 80% threshold).
- Lack of clarity of relationship between sample representativeness/non-response bias and non-response rates
- Lack of consensus on appropriate protocols to increase and test for representativeness/bias

RESOURCES FOR THE FUTURE

Trade-offs in Allocating Survey Budget

- Survey quality
- Sample size
- Sample representativeness
  - Is WTP affected?

"Agencies must design surveys to achieve the highest practical rates of response, commensurate with the importance of the survey's use, respondent burden and data collection costs to insure that the results are representative of the target population so that they can be used with confidence to inform decisions." (OMB Guidance, 9/06)

RESOURCES FOR THE FUTURE

Plan of the Day

- Government Views
- The Experts
- The Practitioners
- Research Agenda
- Towards Consensus Protocols

RESOURCES FOR THE FUTURE

OMB Guidance and Standards for Statistical Surveys

Brian A. Harris-Kojetin, Ph.D.

Statistical and Science Policy
Office of Management and Budget

October 2, 2006

OMB Guidance and Standards

- Questions and Answers When Designing Surveys for Information Collections
  - Issued in January 2006
- Standards and Guidelines for Statistical Surveys
  - Issued in final in September 2006
- Both available at ***www.whitehouse.gov/omb/***
  - Go to *Statistical Programs and Standards*

Background

- OMB is charged to develop and oversee the implementation of Government-wide polices, principles, standards and guidelines concerning
  - statistical collection procedures and methods
  - statistical data classifications
  - statistical information presentation and dissemination
  - timely release of statistical data

  44 USC 3504(e)

OMB Q&A Guidance

- Audience: all Federal agencies conducting and sponsoring collections of information that use statistical methods (broadly defined)
- Assumes little knowledge of clearance process, survey methodology, and statistics
- 81 Q&A's

Q&A Development Process

- Revision and update of FASQs
  - Solicited additional issues from OMB desk officers
- Reviewed by the Federal Committee on Statistical Methodology (FCSM) and the Interagency Council on Statistical Policy
- Circulated for Agency Comments
- External Peer Review

Standards Development Process

- Interagency team formed as subcommittee of the FCSM
  - Charged with reviewing and making recommendations to OMB on updates and revisions to standards
  - Solicited representatives from all ICSP agencies
- Subcommittee reviewed current standards as well as standards at statistical agencies
  - also some other National Statistical Institutes

Standards for Statistical Surveys

- Revision and Update of Statistical Policy Directives 1 and 2, Standards for Statistical Surveys and Publication of Statistics
  - last updated in 1974
  - all surveys were paper & pencil
  - references to punch cards

Standards Development Process

- Draft Subcommittee recommendations were reviewed by FCSM
- Then reviewed by ICSP agencies
- Proposed standards submitted to OMB
- OMB issued Federal Register Notice
  - 60-day public comment period
- OMB reviewed and addressed the public comments

Standards and Guidelines for Statistical Surveys

- 20 Standards
- Supplemented with Guidelines or Best Practices
  - help agencies interpret and fulfill the goals of the Standard

Why are response rates important?

- Common data quality and field performance indicator
- Nonresponse can occur for a number of reasons with different implications
  - refusals
  - noncontacts
- Response rates are a useful indicator for the risk of nonresponse bias

Response Rates and Nonresponse Bias

- Why are response rates important?
- How should response rates be calculated?
- What are acceptable response rates for different kinds of collections?
- How can agencies examine potential nonresponse bias?
- What response rate issues are involved with using samples derived from pre-existing multipurpose panels, such as Internet or consumer panels?

What are acceptable response rates for different kinds of collections?

- Surveys collecting "influential information" (see Information Quality Guidelines) should achieve high response rates
- Agencies need to consider how they will use the data and how the methods chosen will achieve acceptable response rates and data quality

What are acceptable response rates?

In their Information Collection Requests, Agencies need to:

- Provide expected response rates and
  - Description of how response rate determined
  - Description of steps to maximize the response rate
- If expected response rate is less than 80%
  - include plans to evaluate potential nonresponse bias

Survey Response Rates Guidelines

- **Guideline 1.3.3:** Prior to data collection, identify expected unit response rates at each stage of data collection, based on content, use, mode, and type of survey.
- **Guideline 1.3.4:** Plan for a nonresponse bias analysis if the expected unit response rate is below 80 percent.
- **Guideline 1.3.5:** Plan for a nonresponse bias analysis if the expected item response rate is below 70 percent for any items used in a report.

Standards for Statistical Surveys

Standard 1.3 Survey Response Rates

- Agencies must design the survey to achieve the highest practical rates of response, commensurate with the importance of survey uses, respondent burden, and data collection costs, to ensure that survey results are representative of the target population so that they can be used with confidence to inform decisions. Nonresponse bias analyses must be conducted when unit or item response rates suggest the potential for bias to occur.

Standards for Statistical Surveys

Standard 3.2 Nonresponse Analysis and Response Rate Calculation

- Agencies must appropriately measure, adjust for, report, and analyze unit and item nonresponse to assess their effects on data quality and to inform users. Response rates must be computed using standard formulas to measure the proportion of the eligible sample that is represented by the responding units in each study, as an indicator of potential nonresponse bias.

How can agencies examine potential nonresponse bias?

- Nonresponse bias analyses can include:
  - Response rates by different subgroups
  - Comparing nonrespondents and respondents on frame variables
  - Comparing initial refusers with initial respondents
- Subsample nonrespondents for more extensive follow-up efforts
  - Use information to estimate characteristics of nonrespondents

Nonresponse Analysis

- For a sample mean, an estimate of the bias of the sample respondent mean is given by:

$$B(\bar{y}_r) = \bar{y}_r - \bar{y}_t = \left(\frac{n_{nr}}{n}\right)(\bar{y}_r - \bar{y}_{nr})$$

Where:

- $\bar{y}_t$ = the mean based on all sample cases;
- $\bar{y}_r$ = the mean based only on respondent cases;
- $\bar{y}_{nr}$ = the mean based only on the nonrespondent cases;
- n = the number of cases in the sample;
- nnr= the number of nonrespondent cases.

Nonresponse Analysis

- **Guideline 3.2.9:** Given a survey with an overall unit response rate of less than 80 percent using unit response rates as defined above, conduct an analysis of nonresponse bias with an assessment of whether the data are missing completely at random.

Nonresponse Analysis

- For a multistage (or wave) survey, focus the nonresponse bias analysis on each stage, with particular attention to the "problem" stages.
- A variety of methods can be used to examine nonresponse bias, for example, make comparisons between respondents and nonrespondents across subgroups using available sample frame variables.
- Comparison of the respondents to known characteristics of the population from an external source can provide an indication of possible bias, especially if the characteristics in question are related to the survey's key variables.

Response Rate Issues with Panels

- Similar to a multistage survey
  - Several stages of agreement and participation by respondents over some period of time before they are panel "members"
  - Response rates need to take into account all stages
  - Attrition before their "term" is over may also introduce potential bias
- Agencies need to provide appropriate response rates and plans for nonresponse bias analysis

Implementation of Standards

- Application of standards requires judgment balancing use of the information and resources
- Agencies need to provide sufficient information in their Information Collection Requests to OMB to demonstrate whether they are meeting the standards.
  - uses of the information will be considered
  - agency should provide reasons why a standards could not be met and actions taken to address potential issues

MICHIGAN

JPSM The Joint Program in Survey Methodology

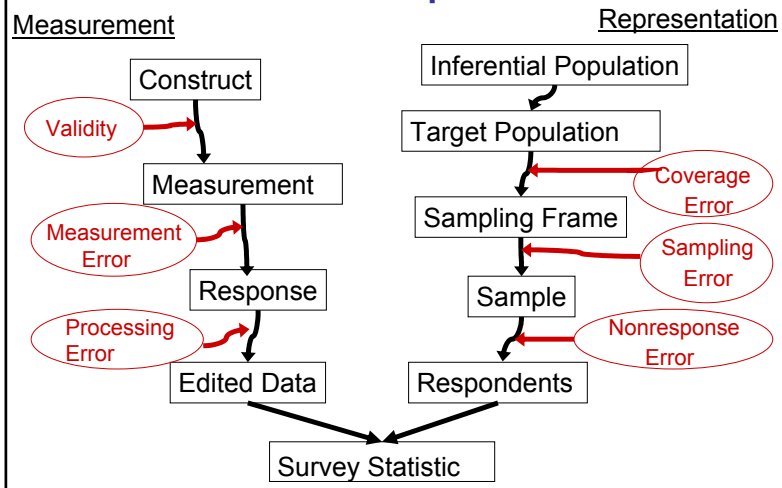
Response Rates and Nonresponse Bias - What We Know and What We Don't Know

Robert M. Groves
University of Michigan and the Joint Program in Survey Methodology

September, 2006

Funding from the Methodology, Measurement, and Statistics Program of the US National Science Foundation Grant 0297435.

Nonresponse Viewed in the Total Survey Error Perspective



Four Mutually-Problematic Observations

1. **With 100% response rates probability samples provide measurable uncertainties for unbiased estimates**
2. **Response rates are declining**
3. **Keeter *et al.* (2000), Curtin *et al.* (2000), Merkle and Edelman (2002) show no nonresponse bias associated with varying nonresponse rates**
4. **We continue to urge practitioners to achieve high response rates**

Result: Confusion among practitioners

Assembly of Prior Studies of Nonresponse Bias

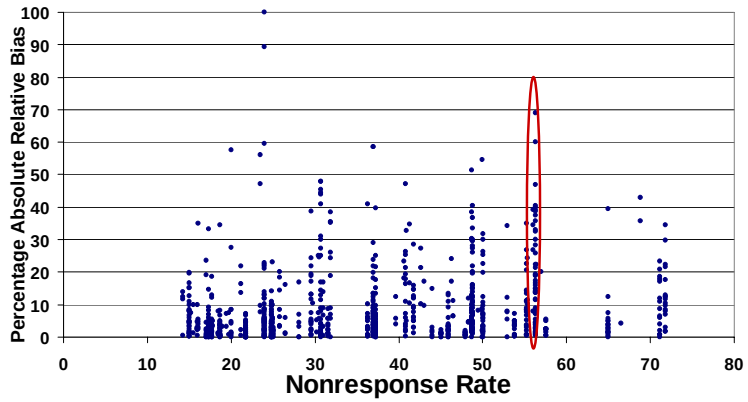
- **Search of peer-reviewed and other publications**
- **47 articles reporting 59 studies**
- **About 959 separate estimates (566 percentages)**
  - **mean nonresponse rate is 36%**
  - **mean bias is 8% of the full sample estimate**
- **We treat this as 959 observations, weighted by sample sizes, multiply-imputed for item missing data, standard errors reflecting clustering into 59 studies and imputation variance**

Meta-Analysis of Nonresponse Error Studies

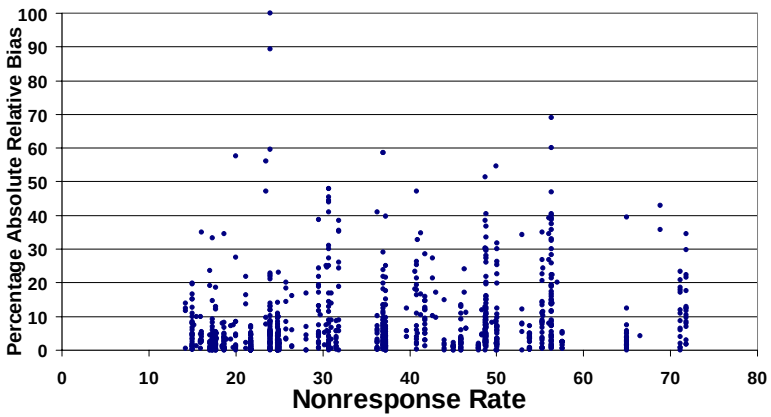
- **Each has data available to compute a relative bias due to nonresponse. The absolute value of the relative bias is**

$$\left|\frac{(\bar{y}_r - \bar{y}_n)}{\bar{y}_n}\right|$$

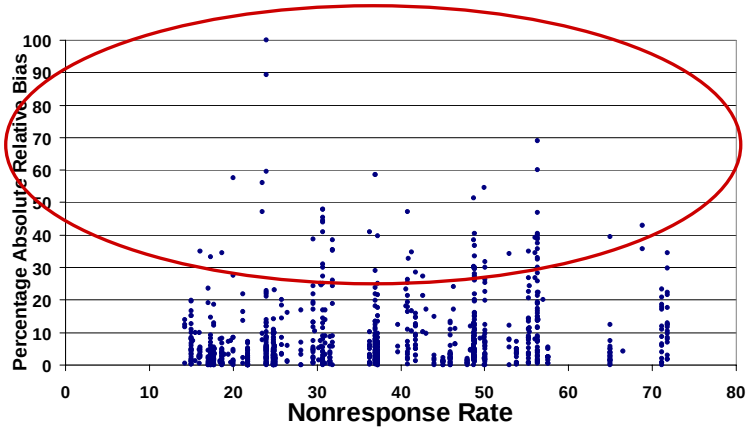
Percentage Absolute Relative Nonresponse Bias by Nonresponse Rate for 959 Estimates from 59 Studies



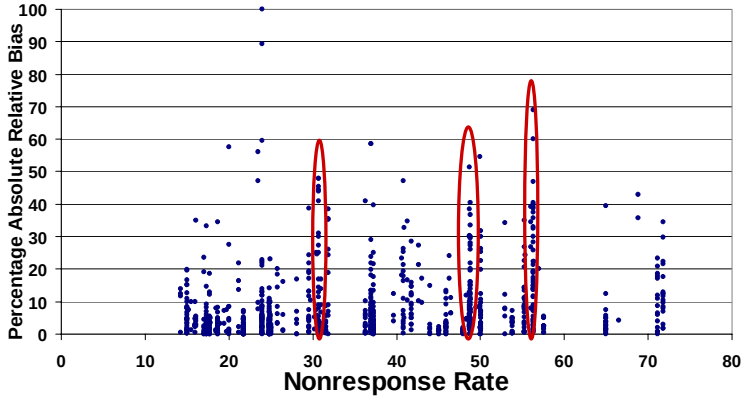
Percentage Absolute Relative Nonresponse Bias by Nonresponse Rate for 959 Estimates from 59 Studies



1. Nonresponse Bias Happens



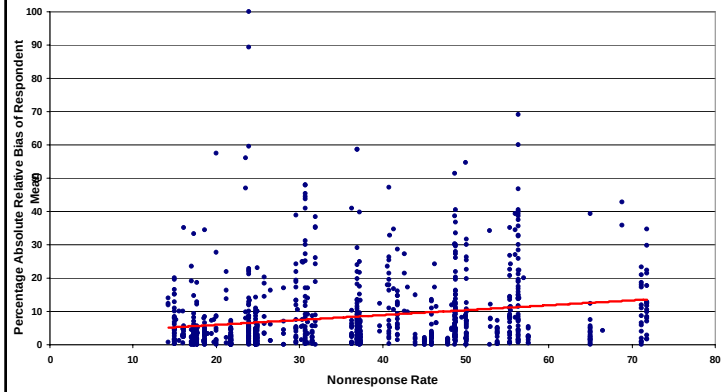
2. Large Variation in Nonresponse Bias Across Estimates Within the Same Survey, or



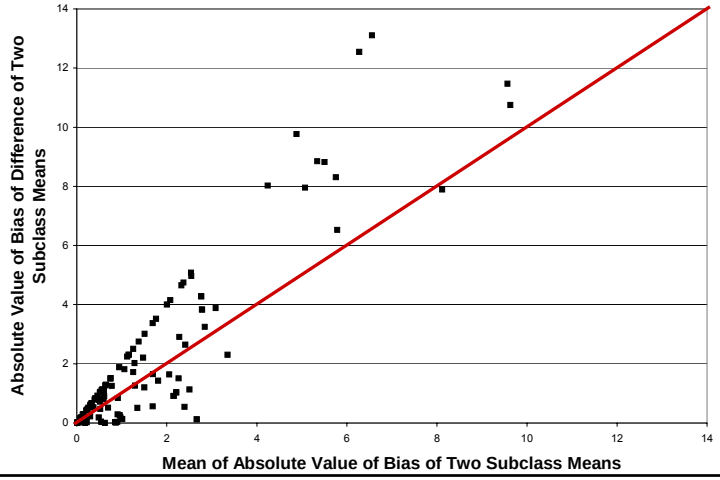
Do Differences of Subclass Means have Lower Nonresponse Bias?

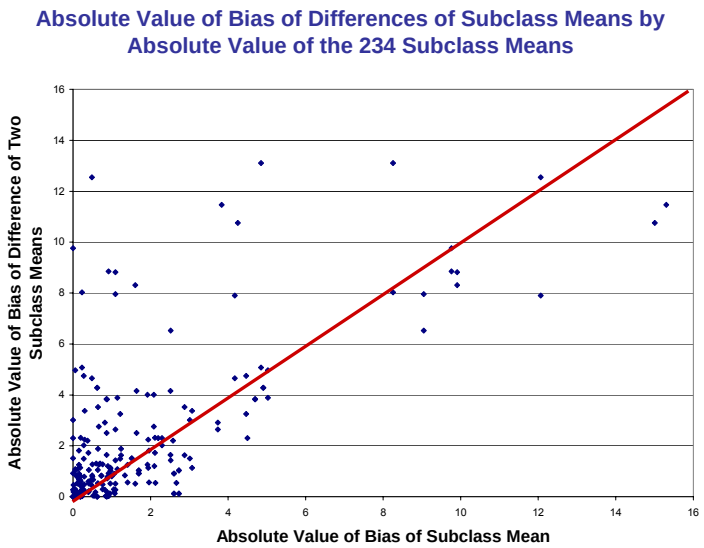
- **When estimating subclass differences, we hope that nonresponse biases of the two estimates cancel**
- **234 reported estimates of subclass means and their 117 differences**
- **Only 41 of the 117 differences have bias of the differences of subclass means lower than average bias of the two subclass means**

3. The Nonresponse Rate of a Survey is a Poor Predictor of the Bias of its Various Estimates (Naïve OLS, R^2=.04)



Absolute Value of Bias of 117 Differences of Subclass Means by the Average Absolute Value of the 234 Subclass Means





Nonresponse Error for Sample Mean

In simplest terms $\bar{Y}_r = \bar{Y}_n + \left(\frac{m}{n}\right)\left[\bar{Y}_r - \bar{Y}_m\right]$

OR

Respondent Mean = Full Sample Mean + (Nonresponse Rate)*(Respondent Mean – Nonrespondent Mean)

OR

Survey Results = Desired Results + Error

OR

Nonresponse Error = f(Rate, Difference between Respondents and Nonrespondents)

Does the Bias Level of Demographic Variables Predict the Bias of Substantive Estimates?

- Common to adjust for socio-demographic variables in hopes of eliminating nonresponse bias
- 23 studies measure biases on both
- Bias levels of demographic variables only weakly related to bias levels of substantive variables (r = .28, "explained variance"=8%)
- Naïve selection bias models or propensity models probably have little effect on bias reduction

A Stochastic View of Response Propensities

$$Bias(\bar{y}_r) = \frac{\sigma_{yp}}{\bar{p}} = \left(\frac{\rho_{yp}}{\bar{p}}\right)\sigma_y \sigma_p$$

where σ_{yp} = covariance between y and response propensity, p

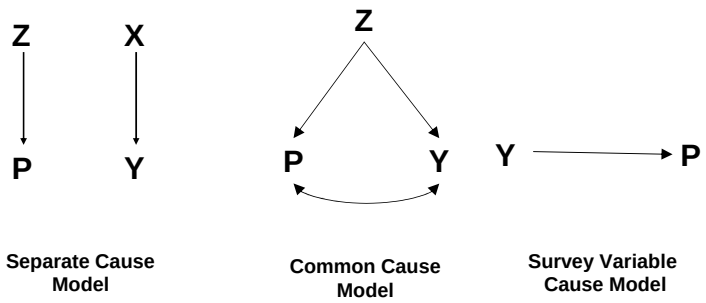
$\bar{p}$ = mean propensity over the sample

ρ_{yp} = correlation between y and p

σ_y = standard deviation of y

σ_p = standard deviation of p

Viewing Nonresponse Error from a Causal Modeling Perspective



Three Key Candidate Design Features to Produce Nonignorable Nonresponse

- Topic interest
- Sponsorship
- Incentives

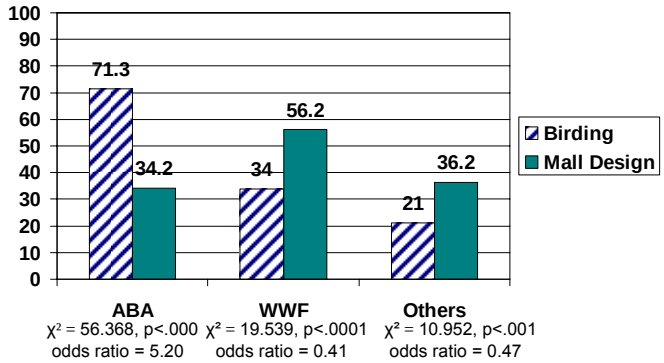
So What to Do?

- Reduce importance given to response rates
- Increase attention to identifying estimates most affected by mechanisms that produce nonresponse
- At Michigan we are mounting experiments attempting to systematically produce nonresponse bias

Experiments Producing Nonresponse Bias

- Survey topics
  - Birds, Bird-watching, and Birding
  - Design of Indoor Shopping Malls
- Target populations
  - members of American Birding Association (ABA)
  - donors to World Wildlife Fund (WWF)
  - other adults (Others)

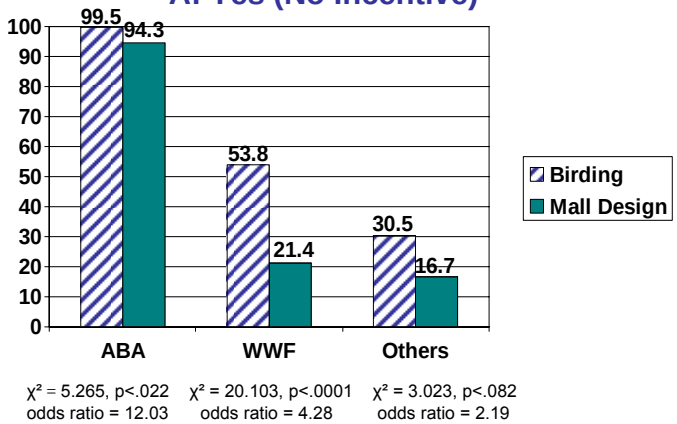
Response Rate by Topic and Population (No Incentive)



Five Summary Statements of What We Know

1. Large nonresponse biases can happen in surveys
2. Most variation in nonresponse biases lie among estimates in the same survey
3. The search for mechanisms that link nonresponse rates to error should be focused on the item-level
4. Differences of subclass means do not tend to have lower biases that the individual subclass means
5. Demographic variables are poor predictors of nonresponse bias in the meta-analysis
6. Topic interest, incentives, and sponsors appear to be influences affecting bias

Q1: Did you participate in bird watching/ birding during the past 12 months? A: Yes (No Incentive)



So, Where's the Research Field Going Now?

- Reducing emphasis on response rate; increasing emphasis on nonresponse error
  - OMB statistical standards
  - papers at International Workshop on Nonresponse
- Building in nonresponse bias studies into surveys
  - rich frames
  - supplementary records
  - seeded samples
  - interviewer observations of respondents and nonrespondents
  - nonresponse followup studies

So, Where's the Research Field Going Now?

- **Rethinking right-side variables in propensity and selection models**
  - **Collect auxiliary data helpful to understanding linkage between response propensities and survey variables**
  - **Proxy indicators of the *y* variable**

A Meta-analysis of Response Rates to Contingent-Valuation Studies Conducted by Mail

Sonia Aziz,
Kevin J. Boyle,
V. Kerry Smith and
Daniel McCollum

Heberlein and Baumgartner (1978)
meta analysis of 98 surveys

- market research background (-),
- government sponsor (+),
- general population (-)
- employee population (+),
- school/army population (+),
- salient topic (+),
- length (pages) (-),
- follow-up mailings (+),
- $$ incentive 1st contact (+), and
- special 3rd contact (+).

The Issue

- NOAA Panel (1993) implicitly contested mail implementation → "… *CV study should be conducted with personal interviews*" (p. 4607).
- OMB (2006) asserts that (ICRs) "… *with expected response rates lower than 80 percent need …" (p. 61)*
  - *description of expected response rate*
  - *steps to be taken to maximize response*
  - *plans to evaluate nonresponse bias"*

Bruvold and Comer (1988)
20 categories, 50 features

- expanded survey sponsorship categories (-)
- subject matter relationship to self, family (+)
- data requested (attitudinal (-) own behav. (+))
- nature of data (embarrassing/confidential (-))
- year survey was conducted (+)

SALIENCE

- TAILORING – (Groves and Couper, 1998)
- LEVERAGE-SALINECY (Groves, Singer and Corning, 2000)
- Saliency made relevant to different respondents by the interviewer
- Mail surveys require careful pretesting to tailor (Dillman, 2000), but one-size-fits-all
- Decomposed saliency into elements of the CV exercise

SURVEY VARIABLES	STATISTICS
RESPONSE RATE	μ = 0.62, 0.14 – 1.0
GOVERNMENT	14%
UNIVERSITY	69%
GENERAL PUBLIC	42%
USER	45%
LENGTH	μ = 13, 2 - 30
CONTACTS	μ = 4, 1 - 6
CERTIFIED 3RD MAILING	14%
MONETARY INCENTIVE	24%
NON-MONETARY	78%
PRETEST	94%
PRENOTIFICATION	43%
PERSONALIZATION	33%
YEAR CONDUCTED	μ = 1988, 1975 - 1995
BULK POSTAGE	7%

Current Study

Sample Frame – Carson et. al. 1994

- 1672 citations
- 124 unique, original studies
- Mail survey of PI's – 74% response rate
- 146 studies, 188 unique observations
- Logit model corrected for heter.

CV VARIABLES	STATISTICS
WATER QUAL.	12%
AIR QUALITY	2%
HUNTING	16%
FISHING	15%
TOTAL	36%
NONUSE	2%
INCREASED PROV.	59%
CURRENT PROV.	30%
QUANTITY Δ	74%
WTP	94%
DICHOTOMOUS	41%
OPEN ENDED	20%

CV VARIABLES *CONT.*	STATISTICS
ACCESS FEE	4%
TAX	9%
INCOME	11%
COST OF LIVING	5%
TRIP EXPENSES	21%
INDIVIDUAL	69%
CERTAINTY	80%
SUBSTITUTES	16%
CV QUESTIONS	μ = 3.5, 1-36
INVEST. #1	28%
INVEST. #2	3%
INVEST. #3	5%
INVEST. #4	6%

DISCUSSION

- Investigator Control/Influence.

LIMITED CONTROL	CONTROL
GENERAL PUB.	CONTACTS
USER	MONETARY INC
WATER QUAL.	BULK POSTAGE
HUNTING	OPEN ENDED
FISHING	
INVESTIGATOR	

VARIABLES	MHB	SR	CV	SW
GENERAL PUB.	-0.78	-0.40	-0.36	-0.54
USER		0.52	0.28	
CONTACTS	0.37	0.39	0.29	0.24
MONETARY INC.	0.32	0.38	0.28	
BULK POSTAGE		-0.67	-0.92	-0.85
WATER QUAL.			0.75	0.64
HUNTING			0.72	0.79
FISHING			0.18	0.30
OPEN ENDED			-0.18	-0.31
INVEST. #1			0.71	0.56
INVEST. #2			1.44	
INVEST. #3			0.42	0.48

DISCUSSION – cont.

- Very few survey features influence response rates
- Investigator experience
- Falling response rates make OMB 80% threshold more of a challenge
- Insights apply to other modes of implementing CV (stated preference) surveys

DISCUSSION – cont.

Daubert guidelines for rules of evidence:

- **response rates to CV surveys conducted by mail have been *tested*;**
- **addresses issue of the *error rate* through identifying design features that enhanced response rates;**
- **development and refinement of *standards* for design and implementation; and**
- **CV surveys conducted by mail do *enjoy widespread acceptance (or use)*.**

Practitioners' Perspectives

Sample Representativeness Workshop

Resources for the Future

October 2, 2006

RESOURCES FOR THE FUTURE

PI	Sample population	Mode	Frame	Cases fielded; completion %
Dickie, Gerking (DG)	Non-AA parents w/ kids, Hattiesburg, MS; U.S. Households	Central facility on computer; Internet	RDD; RDD Panel (KN)	RRD: 6723 post-screen; 9.4% Panel: 63%
Viscusi et al. (VHB)	Adults (>18), U.S.	Internet	RDD Panel (KN)	5655; 75%
Cameron et al. (CGL)	Households, U.S.	Mail; Internet	Direct mail database; Convenience sample	8576; 19%
M. Evans, Smith, Poulos (ESP)	Adults	Internet	RDD Panel (KN)	2670; 79%
Krupnick et al. (KBEB)	Adults, NYS	Internet; Internet; Mail	RDD Panel (KN); former panel (KN); RDD (KN)	Panel: 1284; 74% Former: 2120; 14% Mail: 2372; 24%
Johnston et al. (JSSB)	Households with DVD player, RI (most likely)	Phone/Mail	RDD	1500+; NA
Besedin, Johnston (BJ)	Households, U.S.	Internet; Phone; Phone (non-response)	RDD Panel (KN); former panel (KN); RDD	3900; NA 600 for non-response study
Lew et al. (LRL)	Adults, U.S.	Phone/Mail	Purchase sample (Sampling Int'l)	4000 expected, 34-70% in pre-test (424)
Boyle, Paterson (BP)	Households, U.S.	Phone/Mail	RDD	NA, NA
Chapman, Bishop (CBcr)	Households, U.S.	Internet	RDD Panel (KN);	Pre-test: 312; 69% Final: 1500 completes expected
Chapman et al. (CBrw)	Households, U.S.	TBD (Internet or Mail)	RDD (KN or Survey Sampling Inc.)	1200-1500

RESOURCES FOR THE FUTURE

PI	Survey Study (type)	Sponsor	Status (expected completion date)
Dickie, Gerking (DG)	Environmental Risk to Children's Health (CV)	EPA, ORD/NCER	Complete
Viscusi, Huber, Bell (VHB)	Assessing the Value of Improvements to Lakes and Rivers (CE, Ref)	EPA, OW	Complete
Cameron, Gerdes, Lee (CGL)	Popular Support for Climate Change Policy (CE, CV)	NSF	Complete
M. Evans, Smith, Poulos (ESP)	Environmental Vulnerability (Pilot) (CV)	EPA, ORD/NCER	All but write-up complete (11/06)
Krupnick, Banzhaf, D. Evans, Burtraw (KBEB)	Adirondacks (CV)	EPA CAMD	Complete
Johnston, Besedin (JB)	Ecological Benefits of Aquatic Living Resources (CE)	EPA, ORD/NCER	Survey Design (9/08)
Besedin, Johnston (BJ)	§316(B) Entrainment (CE)	EPA OW/EAD	Awaiting OMB Review (NA)
Lew, Rowe, Layton (LRL)	Steller Sea Lion (CE)	NOAA Fisheries	Pretest Complete (11/07)
Boyle, Paterson (BP)	Visibility in National Parks (CE)	NPS	Survey Design (NA)
Chapman, Bishop et al. (CBcr)	HI Coral Reef Valuation Study (CE)	NOAA, NOS	Awaiting OMB Review (2007)
Chapman, Bishop, Tourangeau, Bisack (CBrw)	North Atlantic Right Whale Study (CE)	NOAA, NFSC	Awaiting OMB Review (2008)

RESOURCES FOR THE FUTURE

Ex-Ante Strategies

- Initial and follow-up contacts (Total Design Method)
- Over-sampling under-represented groups
- Incentives
- Signaling of importance
  - Priority mailing
  - Individualizing
- Identification of survey sponsor
- Non-response conversion (reminder calls)
- Data collection on non-responders

RESOURCES FOR THE FUTURE

Ex-Post Strategies

- Comparison of sample stats with population stats; unweighted and weighted estimates of WTP
- Comparison of characteristics and WTP across modes
  - Boyle calibration equation approach
- Comparison of attitudinal and demographic characteristics of responders:
  - and non-responders
  - or responders to a follow-up survey of non-responders or different survey
- Heckman selection/correction (for non-observables)
  - Cameron use of geo-data
- Comparison of WTP and other responses based on various measures of "engagement"
- Test for conditioning using time on panel/experience
- Survey former panel members

RESOURCES FOR THE FUTURE

Research Needs

- Analysis of effort and by mode (DG)
- Is KN panelist more ideal? (VHB)
  - Distortion Indicator
- Disentangle mode and frame effects (KBEB)
- Selection effects of panel stages and of other modes (various)
- Heckman correction for CE (various)
- Incentives and response rates

RESOURCES FOR THE FUTURE

Ex Post Strategy Issues

- Are demographics, knowledge or performance better/worse from KN than from other modes? Conflicting evidence from DG vs.VHB
- Does engagement matter? Inconsistent evidence from DG, VHB, KBEB
- Are non-responders different from responders? Yes; consistent evidence from VHB, KBEB, CGL. But small effect and Heckman correction not important.
- Are frame/mode effects important?
- Does weighting matter? Not much

RESOURCES FOR THE FUTURE

Should we move to panels?

Joel Huber
Duke University

Why go to panels?

- Lower cost/completed survey
- Much faster/more current
- Projectable to US adult population
- Can target subsets by region, behavior demographics
- Can be used in longitudinal studies
- Willingness to answer difficult surveys

Selection Tests on Water Quality Values

- Many tract-level demographic characteristics predict a person's likelihood to reach each each level of the Knowledge Networks panel.
- Censored regression analysis predicting water quality value based upon tract-level data show minimal differences between surveyed and non-surveyed people.
- Heckman Selection model of water quality value does not produce a significant difference in the selection equation.

Conclusion: For our variables, selection is minor, but it could be greater for other outcome variables

Problems with panels

- Respondents become professional survey takers—worn out or trivially sharpened
- Panelists are not ordinary people
  - Like to answer questions
  - Have time to answer them
  - Value the rewards from answering them
  - Are at ease with computer interface
- Need more panelist, competitors

Adjusting for selection bias

- Characteristics within panel: include terms for time of tenure, time lag to take survey, survey volume and likelihood to leave panel later
- Characteristics of panelists compared with nation
  - Demographics—age, education, income, life cycle
  - Psychographics—Activities, interests and opinions

Factors likely to alter propensity to be in the sample

- Greater propensity to be in the sample
  - Computer/internet use
  - Open to different views
  - Free time index
- Less propensity to be in the sample
  - Geographic mobility
  - Get help with taxes
  - Typical of local community

Developing distortion indicator questions

- These would be part of the panel's information about respondents
- If they were independent of the critical decision variable then there is no problem
- If correlated with the decision variable then these could be the basis for an adjustment

Scavenging for covariates to use in selectivity assessment and correction strategies

Trudy Ann Cameron
University of Oregon

Sketch of projects undertaken in collaboration with J.R. DeShazo (UCLA), Mike Dennis (KN), Rick Li (KN), and Jason Lee (ICFI)

prepared for workshop on

Sample Representativeness: Implications for Administering and Testing Stated Preference Surveys

Resources for the Future
October 2, 2006

SUMMARY: Concerns and prospects

1. Exogenous weights based on relative frequencies in estimating sample and the population—"look busy";

2. There are several phases of selection between the initial random contact with potential panelists and final membership in estimation sample—evidence of different processes at each stage
   - need proprietary variables to expand set of regressors at each stage;

3. "Marginal" versus "conditional" selection models for each phase (can be explored even with identical regressors);

4. Comprehensive selectivity (overall selection from RDD contacts to estimating sample) – for "govt" variable;

5. Propensities and probabilities as ad hoc shifters for preferences;

6. Climate study – final-step selection (invited participants to estimating sample): capturing "salience" of survey topic

7. Advocate:
   - make maximum use of opportunities to geocode respondents' locations and take advantage of indicators of neighborhood attitudes (related to salience) that can also be geocoded;
   - use salience variables to model selection that occurs AFTER potential respondents become aware of the survey topic;
   - appropriate variables will be idiosyncratic to each study;
   - requires considerable ingenuity and persistence.

8. Near-term research agenda?

About those weights….

Common strategy:
1. Define "bins"
2. Calculate relative freq.. in each bin for population
3. Calculate relative freq. in each bin for estimating sample
4. Calculate "exogenous weights":
   rel. freq. (pop) / rel. freq. (sample)

Problem:
1. With sample, can create any cross-tabulation that seems to be relevant or important; e.g. age by gender by income by … by political ideology by avidity concerning resource….?
2. Some cross-tabs are provided by the Census, some are not
3. Cross-tabs involving specialized variables are not available in the Census at all? Harris Interactive (HI)-type benchmarking RDD attitude surveys? Still not enough to capture salience for all possible survey topics.

Work-around?

1. Assume distributions of different population variables are independent
2. Construct *"joint"* distribution by using *products of marginals*
3. Skip desired variables that are not available in Census

A few misgivings…

Usual weights are not going to be "perfect"

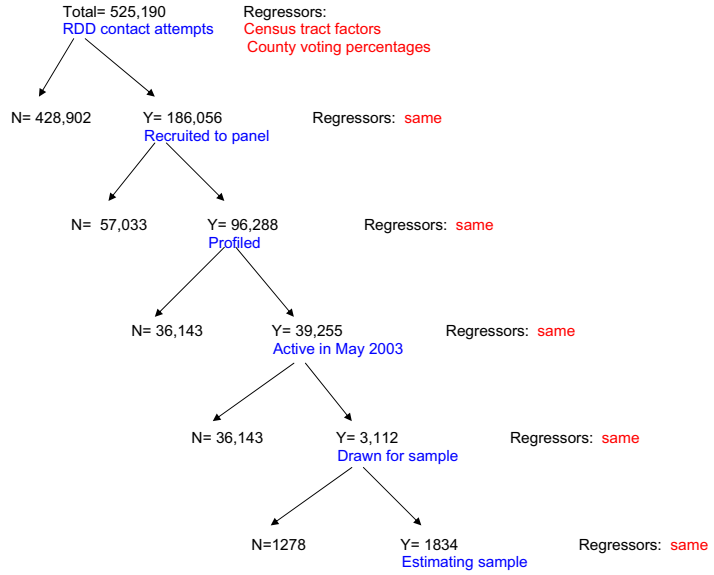
1. not sufficient resolution; some variables unavailable
2. needed Census cross-tabs not necessarily available (may require independence assumption and products of marginals)

Well known problem:

Exogenous weights based on *observables* cannot capture systematic differences in distributions of *unobservables* between population and sample.

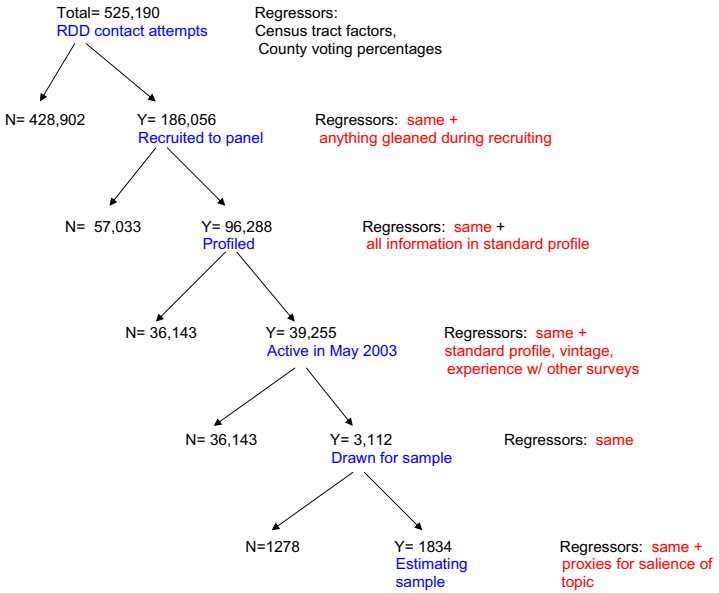
Exploring the "phases" of selection (identical regressors)

Statistical Lives Project (Knowledge Networks); Cameron and DeShazo (2006)
"Attrition Patterns in a Major Internet Consumer Panel"



Exploring the phases of selection – expanding the set of regressors

Statistical Lives Project (Knowledge Networks) – other variables exist



Marginal versus Conditional Selection Models

"Marginal" models:

1. RDD contacts → Recruited to panel
2. RDD contacts → Profiled
3. RDD contacts → Still active in panel in May 2003
4. RDD contacts → Drawn for survey
5. RDD contacts → Present in estimating sample for model

"Conditional" models:

6. RDD contacts → Recruited to panel
7. Recruited to panel → Profiled
8. Profiled → Still active in panel in May 2003
9. Still active → Drawn for survey
10. Drawn for survey → Present in estimating sample for model

Problem: Most response rate calculations, and most existing selection-correction models address only the transition in 10. We have estimated all ten models in Cameron and DeShazo "Attrition Patterns…"

Table 2 – **Marginal** participation models (n=525,190) from initial RDD contacts down to "private preferences" estimating sample)

Variable	recruited	profiled	active	drawn	estimating
census factor avail.	0.8850	0.6580	0.5685	-0.5344	-0.6734
	(21.11)***	(13.11)***	(8.17)***	(8.07)***	(9.65)***
well-to-do prime	-0.0543	-0.0521	-0.0953	-0.1165	-0.0888
	(29.15)***	(24.13)***	(32.93)***	(15.02)***	(9.42)***
well-to-do seniors	-0.0538	-0.0179	0.0263	0.0106	0.0352
	(27.88)***	(8.12)***	(9.34)***	(1.54)	(4.02)***
single renter twenties	-0.0662	-0.0474	-0.0545	-0.0253	-0.0318
	(35.91)***	(22.50)***	(19.43)***	(3.56)***	(3.61)***
unemployed	-0.0082	-0.0398	-0.0246	-0.0050	-0.0200
	(4.25)***	(17.84)***	(8.54)***	(0.70)	(2.21)**
minority single moms	0.0124	-0.0521	-0.0446	-0.0101	-0.0273
	(5.63)***	(20.41)***	(13.73)***	(1.32)	(2.74)***
thirty-somethings	-0.0502	-0.0368	-0.0195	-0.0035	-0.0169
	(22.74)***	(14.79)***	(5.78)***	(0.39)	(1.57)
working-age disabled	0.0193	0.0018	0.0099	0.0211	0.0069
	(8.79)***	(0.73)	(3.14)***	(2.81)***	(0.73)
some college, no graduation	0.0439	0.0520	0.0376	-0.0108	-0.0067
	(22.16)***	(23.06)***	(12.89)***	(1.49)	(0.74)
elderly disabled	-0.0407	-0.0337	-0.0102	0.0180	0.0241
	(20.96)***	(14.96)***	(3.53)***	(2.62)***	(2.84)***
rural farm self-employed	0.0177	0.0050	0.0153	0.0439	0.0419
	(5.64)***	(1.38)	(3.38)***	(4.38)***	(3.31)***
low mob. stable neighborhd	0.0352	0.0171	0.0158	-0.0029	-0.0016
	(16.32)***	(6.95)***	(5.01)***	(0.37)	(0.16)
Native American	0.0409	0.0546	0.0528	0.0343	0.0285
	(16.92)***	(19.48)***	(14.59)***	(4.04)***	(2.62)***
Female	0.0038	0.0015	0.0099	0.0066	0.0009
	(1.64)	(0.57)	(2.88)***	(0.79)	(0.09)
health-care workers	0.0168	0.0185	0.0202	0.0372	0.0374
	(7.78)***	(7.50)***	(6.29)***	(4.78)***	(3.82)***
Asian-Hisp-lang isolated	-0.0310	-0.0562	-0.0651	-0.0454	-0.0604
	(15.07)***	(23.41)***	(20.49)***	(5.75)***	(5.78)***
Vote percentage avail.	0.1142	0.2026	0.2399	-0.9356	-1.0114
	(2.47)**	(3.81)***	(3.31)***	(12.12)***	(11.94)***
Gore percent in 2000	-0.0840	-0.0430	-0.0355	-0.1049	-0.2096
	(4.34)***	(1.95)*	(1.25)	(1.51)	(2.41)**
Nader percent in 2000	0.0132	0.0914	-0.0277	0.6544	0.9018
	(0.11)	(0.70)	(0.16)	(1.60)	(1.81)*
Constant	-1.3075	-1.7341	-2.2139	-0.9995	-0.9354
	(22.27)***	(24.97)***	(23.02)***	(13.23)***	(12.08)***
Observations	525,190	525,190	525,190	525,190	525,190

Table 3 - **Conditional** participation models (declining sample sizes); from initial RDD contacts down to "private preferences" estimating sample

Variables	recruited	profiled	active	drawn	estimating
census factor avail.	0.8850	-0.0834	0.0998	-0.1366	-1.1354
	(21.11)***	(0.87)	(0.76)	(0.49)	(2.51)**
well-to-do prime	-0.0543	-0.0229	-0.1053	-0.0885	0.1019
	(29.15)***	(7.42)***	(23.33)***	(7.72)***	(3.42)***
well-to-do seniors	-0.0538	0.0377	0.0787	-0.0273	0.0898
	(27.88)***	(11.85)***	(17.44)***	(2.63)***	(3.62)***
single renter twenties	-0.0662	0.0032	-0.0356	0.0140	-0.0354
	(35.91)***	(1.03)	(8.01)***	(1.28)	(1.33)
unemployed	-0.0082	-0.0587	0.0124	0.0195	-0.0476
	(4.25)***	(18.40)***	(2.71)***	(1.83)*	(1.89)*
minority single moms	0.0124	-0.1057	-0.0069	0.0292	-0.0694
	(5.63)***	(30.02)***	(1.32)	(2.49)**	(2.55)**
thirty-somethings	-0.0502	0.0001	0.0149	0.0103	-0.0555
	(22.74)***	(0.04)	(2.90)***	(0.75)	(1.58)
working-age disabled	0.0193	-0.0218	0.0150	0.0253	-0.0550
	(8.79)***	(6.05)***	(2.97)***	(2.19)**	(2.03)**
some coll, no graduation	0.0439	0.0372	-0.0076	-0.0516	0.0141
	(22.16)***	(11.45)***	(1.63)	(4.69)***	(0.53)
elderly disabled	-0.0407	-0.0069	0.0339	0.0324	0.0230
	(20.96)***	(2.07)**	(7.15)***	(3.02)***	(0.92)
rural farm self-employed	0.0177	-0.0165	0.0243	0.0680	-0.0031
	(5.64)***	(3.22)***	(3.30)***	(4.35)***	(0.09)
low mob stable neighborhd	0.0352	-0.0148	0.0027	-0.0172	0.0023
	(16.32)***	(4.13)***	(0.53)	(1.46)	(0.08)
Native American	0.0409	0.0444	0.0196	0.0096	0.0219
	(16.92)***	(11.19)***	(3.34)***	(0.70)	(0.67)
female	0.0038	-0.0017	0.0188	0.0104	-0.0177
	(1.64)	(0.44)	(3.42)***	(0.80)	(0.59)
health-care workers	0.0168	0.0138	0.0117	0.0408	-0.0017
	(7.78)***	(3.88)***	(2.30)**	(3.42)***	(0.06)
Asian-Hisp-lang isolated	-0.0310	-0.0637	-0.0439	0.0001	-0.0695
	(15.07)***	(18.78)***	(8.70)***	(0.01)	(2.46)**
Vote percentage avail.	0.1142	0.2172	0.1336	-0.5793	-0.5950
	(2.47)**	(2.93)***	(1.20)	(2.51)**	(1.39)
Gore percent in 2000	-0.0840	0.0318	-0.0093	-0.0189	-0.4453
	(4.34)***	(1.01)	(0.21)	(0.18)	(1.75)*
Nader percent in 2000	0.0132	0.2183	-0.1174	0.4108	0.7355
	(0.11)	(1.16)	(0.45)	(0.66)	(0.50)
Constant	-1.3075	-0.1171	-0.4341	-0.6385	2.1218
	(22.27)***	(0.97)	(2.54)**	(1.79)*	(5.95)***
Observations	525,190	186,056	96,288	35,463	3,112

Comprehensive selection: "RDD contacts" to "Estimating sample"

From Cameron and DeShazo (2006) "A Comprehensive Assessment of Selection in a Major Internet Panel for the Case of Attitudes toward Government Regulation"

"Outcome" variable is "govt":

"People have different ideas about what their government should be doing. How involved do you feel the government should be in regulating environmental, health and safety hazards?"

1=minimally involved ……. 7=heavily involved.

The "govt" question was asked in both the "prevention" and the "treatment" version of the "public health policy" surveys

Selectivity equation from jointly estimated model		
Census tract factors available [a]	-0.7427	(12.50)***
"well-to-do prime aged"	-0.1185	(15.00)***
"well-to-do seniors"	0.0324	(4.62)***
"single renter twenties"	-0.0324	(4.34)***
"unemployed"	0.0062	(0.85)
"minority single moms"	-0.0194	(2.37)**
"thirty-somethings"	-0.0181	(2.13)**
"working-age disabled"	-0.0013	(0.17)
"some college, no graduation"	-0.0183	(2.44)**
"elderly disabled"	0.0055	(0.76)
"rural farm self-employed"	0.0425	(4.09)***
"low mobility stable neighborhood"	-0.0094	(1.18)
"Native American"	0.0407	(4.59)***
"female"	0.0112	(1.29)
"health-care workers"	0.0192	(2.33)**
"Asian-Hisp language isolated"	-0.0666	(7.75)***
2000 vote percentage available	-1.1264	(15.32)***
Gore percent (county)	-0.1769	(2.45)**
Nader percent (county)	1.6150	(3.93)***
Constant	-0.6224	(9.02)***
ρ (implied Heckman error correlation)	0.08462[b]	(1.52)

Why not use Heckman correction models for every study?

Cameron and DeShazo (2006) uses "govt" variable—allows us to use packaged selective correction algorithms (in Limdep: ordered probit with selection correction; or Stata, if we treat the govt variable as continuous)

More typical stated preference estimation context:

Conjoint choice experiments—multiple-alternative conditional logit (or analogous)

Problem:

No easy FIML selectivity correction models for these models

Potential solutions:

Jointly estimated mixed logit model and selection equation. At a minimum, allow "intercept" in selection equation sub-model to be correlated with parameter that gives "intercept" in valuation sub-model.

My one attempt at such a model was "uncooperative."

Dan Hellerstein's work?

What might we do instead? (…or, meanwhile?)

Estimated selection equation provides:

1. Fitted participation "propensities":
$$P_i = z_i'\gamma$$
2. Fitted participation "probabilities":
$$\Pi_i = \Phi(z_i\gamma)$$

- Point estimates for P_i and Π_i can be produced for every initial RDD recruiting contact (e.g. N = 525,000+ for KN).
- Calculate central tendency across set of RDD contacts.
- Calculate *deviations* from this "average" across all RDD recruiting contacts; call these p_i and π_i.
- If every RDD recruiting contact was equally likely to show up in the estimating sample, P_i and Π_i would be same for all, and p_i= 0 and π_i = 0 for all i.
- Want to be able to simulate the estimation model under these more-desirable conditions

One strategy for assessing the potential effects of selection

Q: Do the estimated preference parameters—within the estimating sample—vary systematically with the participation propensities (participation probabilities)?

Generalize each preference parameter

$$\beta_j = \beta_{j0} + \beta_{j1} p_i, \ \forall i \quad \text{or}$$
$$\beta_j = \beta_{j0} + \beta_{j1} \pi_i, \ \forall i$$

Model predicts that "true" preference parameters—i.e. if every observation in the estimating sample was equally likely to be present—will be give by the baseline parameter β_{j0}.

- If shift parameters β_{j1} *are not* significantly different from zero, this may suggest that the lack of any relationship may extend to the non-participants among the RDD contact (but this is not proof).
- If shift parameters β_{j1} *are* statistically significant, researchers should be concerned that preferences also differ with response propensities between the participant and non-participant subsets.

Selection models for "Value of a Statistical Illness Profile" study

Classes of explanatory variables:

- Some are general response-rate predictors
- Some are specific to the salience of health issues

1. 15 Census tract factors (including "health care workers" factor)
2. County voting percentages in 2000 Presidential election
3. Number of hospitals in county
4. Actual mortality from selected causes (in same county over previous 12 years)

Fitted comprehensive participation probability (π_i) significantly affects only the marginal utility of the log of the present value of future sick-years. Shift is about 3 on a base parameter of about -50. Mean value of probability difference is also *far* less than one.

Model for Conditional Selection Process #10: transition from the "drawn" set of target households to the "estimating" sample

Dredging up proxies for the salience of the survey topic
Example: climate policy survey (http://globalpolicysurvey.ucla.edu)

Q: What variables might capture the *salience* of climate change impacts AND can be *linked* to each potential respondent in some way

Current versus historical temperatures

Avg temp, mailday+5 in zipcode
Positive temp dev. from normal
Negative temp dev. from normal

Perception of vulnerability to climate change

Insurance payments in state, 78-03
Natural disaster (same state)
Natural disaster (neighboring state)
=1 if county flooding during 1998-2000
County with flooding in 2000
County with flooding in 1999

Geographic risk data available (indicator)
Tornados recorded >6/1000 sq. mi.
Tornados recorded >11/1000 sq. mi.
Tornados recorded >15/1000 sq. mi.
<100 mi from coast, South hurricane zone
<100 mi from coast, North hurricane zone
1/(Distance from ocean)

Season when survey was completed

January	July
February	August
March	September
April	October
May	November
June	December

Major events

Survey mailed 9/11/01-10/9/01 (World Trade Center)
Mailed to anthrax localities
Major sporting event

Features of survey instrument

Design value for program cost

Replacement packet sent
Included a post-it note
=1 if shortened survey version

County political mix

County Nader vote percentage
County Bush vote percentage
County voter turnout percentage

Attributes of address/addressee

Midpoint of addressee age bracket
Midpoint addressee family inc. bracket/1000
Addressee household size
Census tract median house value
Addressee length of residence
1=rural, 0=urban
Addressee female
Address is an apartment
Address is a condominium
Address is a house

Neighborhood characteristics Census factor scores

Well-to-do prime
Well-to-do seniors
Single renter twenties
Unemployed
Minority single moms
Thirty-somethings
Working-age disabled
Some college, no grad
Elderly disabled
Rural, farm, self-empl
Low mobility older neighb
Native American
Asian-Hisp. language isol.

Selection portion of climate model (persistent variables)		
Features of survey instrument	Coef.	t-stat
=1 if shortened survey version	0.2084	(2.74)***
Vulnerability		
Natural disaster (same state)	-0.1006	(1.06)
Geographic risk data available	0.1622	(1.82)*
Tornados recorded >11/1000 sq. mi.	-0.1586	(1.89)*
<100 mi from coast, North hurricane zone	-0.1675	(3.50)***
Seasonality		
September	-0.1264	(2.19)**
October	-0.0333	(0.65)
County political mix		
County Nader vote	1.7355	(1.71)*
Attributes of address/addressee		
Midpoint of addressee age bracket	0.0106	(2.07)**
(Midpoint of addressee age bracket)2/100	-0.0100	(2.03)**
Midpoint of addressee family inc. bracket/1000	0.9575	(1.96)**

Continued….

Continued….

Census tract factor analysis scores		
Well-to-do prime	0.0716	(3.87)***
Well-to-do seniors	0.0690	(3.56)***
Unemployed	-0.0750	(3.92)***
Minority single moms	-0.1113	(4.82)***
Working-age disabled	-0.0513	(2.48)**
Some college, no grad	0.0599	(3.15)***
Native American	0.0610	(2.38)**
Asian-Hisp. language isol.	-0.0664	(3.16)***
Constant	-1.3036	(11.69)***
Estimated error correlation (rho)[a]	-0.089	-0.36
Observations	7527	
Log L	-4806.24	

Research Agenda

1. Figure out some easily implemented models that allow us to mimic Heckman selectivity correction models in the context of multiple conditional logit estimation (e.g. Stata clogit).
2. Find everything you can that might account for differences in the salience of your specific survey topic across individuals
3. There is “deeper” systematic selection in going from the original RDD contact attempts to the set of active panelists in the KN sample at any point in time.
   - Whether these common phases of selection matter for a specific survey is an empirical question to be asked for every study.
   - A finding of minimal evidence of selection bias for one study using KN is encouraging, but does not yet guarantee that no other study will be affected by selectivity biases.
   - Need enough studies to produce a “weight of the evidence.”

Correcting for Non-response Bias in Discrete Choice Models: A Two-Stage Mixed Logit Model

Daniel Hellerstein
USDA/Economic Research Service

For the Oct 2, 2006 RFF Workshop…
Sample Representativeness: Implications for Administering and Testing Stated Preference Surveys

ERS
ECONOMIC RESEARCH SERVICE
United States Department of Agriculture

The Problem:

Respondents to a survey may **systematically differ** from the general population.

This difference may yield misleading inferences – conclusions based on information provided by survey respondents may not hold true for the underlying population

Systematic differences may be observable, or may be unobservable

- **Observable**: measurable factors can be identified for respondents and non-respondents
- **Unobservable**: factors that can not be measured, in either respondents or non-respondents

Observable differences

Example:
the average income of the survey is not the same as the general population

In most cases, I contend that …
differences in measurable factors should not yield biased conclusions.

Example:
Although OLS models are most efficient when the X (independent) variables are broadly distributed, a clumpy distribution of X will not cause inconsistent estimates.

Example:
Within-sample data, when combined with observation-specific population weights, can be used to make overall population predictions.

Unobservable differences

Example:
due to idiosyncratic personality factors, people who return a survey may be the most avid users of a resource

For obvious reasons, direct use of measurable characteristics (of the sample & population) won't help control for unobservable differences.

- However, indirect measures can be used.
- In particular, sample selection models.

Sample Selection Models:

Basic Idea:

1. Respondents are "selected into" the sample
2. A *1st-stage* model predict who will be selected
3. A *2nd-stage* model predicts the variable of interest
4. The random components (RV) in both stages are potentially correlated

Which means …

1. The first-stage RV of **participants** may *not* have an expected value of zero.
2. Hence, if the two RVs are correlated, the 2nd-stage RV may **not** have an expected value of zero.
3. Hence, the 2nd stage model must control for such inconveniences.

A digression: moments of the incidentally truncated bivariate normal (BVN) distribution:

$$E[y \mid z > 0] = \mu_y + \rho\sigma_y \lambda(\alpha_z)$$

$$Var[y \mid z > 0] = \sigma_y^2 \left(1 - \rho^2 \delta(\alpha_z)\right)$$

Where....

$$\alpha_z = \gamma w / \sigma_z,$$

$$\lambda(\alpha_z) = \phi(\alpha_z) / \Phi(\alpha_z)$$

$$\delta(\alpha_z) = \lambda(\alpha_z)\left(\lambda(\alpha_z) + \alpha_z\right)$$

The $\rho\sigma\lambda$ component of the Heckit is an observation-specific estimate of the (non-zero) expected value of the 2nd stage random error.

Example: the Heckit Model

Assumptions:

(1) $Z_i^* = \gamma W_i + \tau_i$

$Z_i = 1$ if $Z_i^* > 0$ $\qquad$ $Z_i = 0$ if $Z_i^* < 0$

$\text{Prob}(Z_i = 1) = \Phi(\gamma W)$ $\qquad$ $\text{Prob}(Z_i = 0) = 1 - \Phi(\gamma W)$

(2) $Y_i = \beta X_i + \varepsilon_i$

Y_i is observed only if $Z_i = 1$

τ_i and $\varepsilon_i \sim \text{bivariate normal}(0,0,1,\sigma_\varepsilon,\rho)$

Model:

1) *Using data from those in the sample and those not in the sample,* use a Probit to estimate γ

2) For each observation in the sample, compute: $\lambda = \phi(\gamma w_i) / \Phi(\gamma w_i)$

3) Estimate β and β_λ using a linear regression of Y on X and λ

Notes: $\beta_\lambda = \rho\sigma_\varepsilon$

λ is some times called an *inverse-mills ratio*

Discrete Choice Models and non-response bias:

Discrete choice models, such at the Multi Nomial Logit (MNL) are often used in environmental valuation.

?? Can sample selection models be used to control for non-response bias in MNL models ??

Naïve answer:

a) estimate a 1st probit,
b) compute appropriate values of λ,
c) include λ them in the MNL 2nd stage.

Problem… what is the first stage RV correlated with?

A digression: the MNL model:

Individual i can choose from 1...J possible alternatives.

$$m_j = x_j\beta + v_j$$

Respondent chooses j* such that $m_{j^*} > m_{j'}$ $\left[j' = 1,..,j^* - 1, j^* + 1,..,J\right]$

If v_j has an extreme-value distribution, than the probability of choosing alternative j is:

$$\exp(x_j\beta) \Big/ \sum_j \exp(x_j\beta)$$

BIG PROBLEM:

For many models, the "order of alternatives" is arbitrary.

Example:

In a freshwater recreation model, the choice set of accessible waterbodies is different for individuals in different parts of the nation (though the same set of measured variables exist for all waterbodies).

Hence, across all observations, for any "alternative j"…

there is no obvious reason for the same correlation to exist between u and v_j

A possible workaround is to carefully order alternatives, so that each individuals j'th alternative is somehow similar.

Instead of this, I consider a 2-stage mixed logit model

Simple case:

If *τ and* v_j *are distributed as bivariate normal …*

Then

In the MNL estimator, include

$$ß_{\lambda,j} * \lambda$$

term(s) in one (or more) of the m_j, and estimate $ß_{\lambda,j}$

Small problem:

What does inclusion of separate $ß_{\lambda,j}$ *terms (for more than one alternative) imply about the independence of* v_j *?*

A digression: the Mixed MNL model:

Starting with the standard MNL: $prob_j = \exp(x_j\beta) \Big/ \sum_j \exp(x_j\beta)$

However, each individual (n) has unique coefficient values (β_n)

The kth coefficient in β_n is modeled as: $\beta_{k,n} = \tilde{\beta}_k + \eta_{n,k}$

where...

$\tilde{\beta}_k$ is the systematic portion of $\beta_{n,k}$

$\eta_{n,k}$ is an observation & coefficient specific random variable drawn from a MVN(0,Ω) random variable

The mixed logit: $$\mathbf{prob_j} = \sum_r \left(\exp(x_j\beta_n^r) \Big/ \sum_j \exp(x_j\beta_n^r) \right) \Big/ R$$

where

$$\beta_n^r = \hat{\beta} + \hat{\Omega}\xi_{nr}$$

$\hat{\beta}, \hat{\Omega}$ = estimated values of $\tilde{\beta}$, estimated value of the "square root" of Ω

ξ_{nr} = randomly generated Kx1 vector of standard normal errors.

R = R different replications (different draw of ξ_{nr} for each r = 1...R replications)

The 2-stage mixed logit.

Basic idea:

Instead of assuming that

τ and v_j are BVN distributed …

Assume that

τ and η_k are BVN distributed

In other words, the values of one (or several) of the varying parameters will be correlated with the first stage decision.

This assumption removes the need to have special ordering for alternatives!

BUT DOES IT WORK?

that is still under investigation.

Some results from artificial data …

- 1600 observations
- 8 alternatives
- A first stage model determines whether a response is observed.
- The rv in the first stage model is correlated with the rv used to generate the varying parameter.

An operational version of the 2-stage mixed logit.

Assuming a diagonal Ω covariance matrix

then, where the mixed uses...

$$\beta_{n,k}^{r} = \hat{\beta}_k + \sigma_k \xi_{nr,k}$$

the 2-stage mixed logit uses...

$$\beta_{n,k}^{r} = \hat{\beta}_k + \boxed{\left(\tilde{\sigma}_k \tilde{\rho}_k \lambda\right)} + \boxed{\left(\left(\tilde{\sigma}_k \sqrt{1-\tilde{\rho}_k^2}\,\delta\right)\xi_{nr,k}\right)}$$

where...

$\tilde{\sigma}_k, \tilde{\rho}_k$: estimates of the sd of η_k, and the correlation between η_k and τ

As described before ... λ and δ are observation specific values derived from a first-stage PROBIT :

$$\alpha_z = \gamma w$$
$$\lambda(\alpha_z) = \phi(\alpha_z)/\Phi(\alpha_z)$$
$$\delta(\alpha_z) = \lambda(\alpha_z)(\lambda(\alpha_z) + \alpha_z)$$

Artificial data 1: correlation=0.5 | 45% rate of response | average WTP=108
MCI, CRsq and WTP computed using all respondents

Model	dβ1	dβ3	dβ3	MCI	CRsq	WTP
Mixed Logit all obs	11%	5%	%3	0.86	0.82	105
MNL, respondents	12%	8%	8%	0.86	0.83	110
Mixed Logit, respondents	50%	3%	2%	0.86	0.83	110
2-stage mixed logit	40%	1%	2%	0.87	0.96	109

dβk : % difference between actual beta (for k'th coefficient)

MCI : McFadden index (1=perfect fit)

CRsq : Cramer R-square (higher values indicate better fit)

WTP : average Estimated willingness to pay for choice opportunity

Actual beta	3.0	-0.2	2.0
(sd)	(2.6)		

Artificial data 2: correlation=0.95 | 31% rate of response | average WTP=108
MCI, CRsq and WTP computed using all respondents

Model	dβ1	dβ3	dβ3	MCI	CRsq	WTP
Mixed Logit all obs	7%	5%	6%	0.83	0.78	104
MNL, respondents	53%	70%	70%	0.74	0.77	120
Mixed Logit, respondents	160%	5%	5%	0.78	0.78	125
2-stage mixed logit	3%	5%	10%	0.83	0.78	106

dβk : % difference between actual beta (for k'th coefficient)

MCI : McFadden index (1=perfect fit)

CRsq : Cramer R-square (higher values indicate better fit)

WTP : average Estimated willingness to pay for choice opportunity

Actual beta	3.0	-0.2	2.0
(sd)	(5.6)		

Artificial data 4: corr =0.57 & 0.13 | 46% rate of response | average WTP=33
MCI, CRsq and WTP computed using all respondents

Model	dβ1	dβ3	dβ3	dβ4	MCI	CRsq	WTP
Mixed Logit all obs	5%	12%	16%	20%	0.77	0.70	31
MNL, respondents	22%	39%	37%	35%	0.73	0.69	35
Mixed Logit, respondents	95%	32%	36%	48%	0.76	0.70	37
2-stage mixed logit	20%	32%	37%	77%	0.76	0.69	35

dβk : % difference between actual beta (for k'th coefficient)

MCI : McFadden index (1=perfect fit)

CRsq : Cramer R-square (higher values indicate better fit)

WTP : average Estimated willingness to pay for choice opportunity

Actual beta	4.0	-0.4	1.0	1.2
(sd)	(3.6)			(0.8)

Artificial data 3: correlation=0.58 | 43% rate of response | average WTP=23
MCI, CRsq and WTP computed using all respondents

Model	dβ1	dβ3	dβ3	MCI	CRsq	WTP
Mixed Logit all obs	12%	13%	12%	0.79	0.72	21
MNL, respondents	43%	3%	37%	0.74	0.71	24
Mixed Logit, respondents	90%	17%	20%	0.78	0.74	27
2-stage mixed logit	45%	17%	20%	0.79	0.74	24

dβk : % difference between actual beta (for k'th coefficient)

MCI : McFadden index (1=perfect fit)

CRsq : Cramer R-square (higher values indicate better fit)

WTP : average Estimated willingness to pay for choice opportunity

Actual beta	4.0	-0.4	1.0
(sd)	(5.6)		

Conclusions

•The 2-stage Mixed Logit offers a somewhat theoretically appealing method of controlling for non-response biase in discrete choice models.

•Under very limited testing, it has shown capacity to improve estimates.

•More work is needed. In particular, the current "heckit like" 2-stage estimator uses limited information when linking correlations to simulated draws . A more ambitious estimator could use a Gibbs sampler to estimate both stages nearly simultaneously

Internet vs. Phone Survey: An Application to Alcohol Research

Timothy Heeren, PhD; Erika Edwards, MPH;
Michael Winter, MPH; Ralph Hingson, ScD, MPH;
David Rosenbloom, PhD

Youth Alcohol Prevention Center,
Boston University School of Public Health

Supported by the National Institute on Alcohol Abuse and Alcoholism
Grant # P60-AA013759

BOSTON UNIVERSITY SCHOOL of Public Health

Background

- National Epidemiologic Survey on Alcohol and Related Conditions (NESARC)
- Sponsored by NIAAA
- In-home interviews with 43,093 adults ages 18+ completed by Census Bureau in 2001-2002
- Weighted data represent civilian noninstitutionalized population in 2000

Draft – do not cite without permission

Background

- A goal of the Youth Alcohol Prevention Center is to identify associations between age of drinking onset and adverse alcohol-related outcomes
- Focus on adults ages 18-39 who ever consumed ≥ 12 drinks in a 12 month period

Draft – do not cite without permission

Background

- Pilot RDD survey
- Sample of 1000 confirmed residential numbers
  - Unable to make contact: 31%
  - Refused participation: 31%
  - Response rate: 38%
  - 7 hours per completed interview
- *Conclusion:* RDD telephone survey of targeted population yields low response and is inefficient

Draft – do not cite without permission

Goals of This Analysis

- Test online opportunity
- Evaluate mode effects of online vs. telephone
- Assess differences between panel members and non-members
- Compare online responses to NESARC
- Ask questions not in NESARC

Draft – do not cite without permission

Demographics – Ever Drinkers

	Internet (n=3406)	On-panel telephone (n=330)	Off-panel telephone (n=603)	NESARC (n=11,549)
Age	%	%	%	%
18-25	22	22	27	35
26-35	52	48	49	45
36-39	26	30	24	20

Draft – do not cite without permission

Three Samples from KN

	On-panel Internet	On-panel telephone	Off-panel telephone
Enrolled	3,406	330	603
Cooperation rate	93%	55%	30%

Overall Web-enabled panel cooperation rate: 89%

Draft – do not cite without permission

Demographics – Ever Drinkers

	Internet	On-panel telephone	Off-panel telephone	NESARC
Gender	%	%	%	%
Male	51	44	34	57
Female	49	56	66	43

Draft – do not cite without permission

Demographics – Ever Drinkers

Race / Ethnicity	Internet	On-panel telephone	Off-panel telephone	NESARC
	%	%	%	%
White	71	75	70	70
Black	8	6	12	10
Hispanic	13	12	12	14
Other	8	8	5	6

Draft – do not cite without permission

Demographics – Ever Drinkers

Marital Status	Internet	On-panel telephone	Off-panel telephone	NESARC
	%	%	%	%
Married	51	55		44
Never married	41	35		42
Other	7	10		15

Draft – do not cite without permission

Demographics – Ever Drinkers

Education	Internet	On-panel telephone	Off-panel telephone	NESARC
	%	%	%	%
Not HS grad	9	6	6	11
HS grad	26	30	24	26
Some college	33	31	34	35
Bachelor's or higher	33	32	36	28

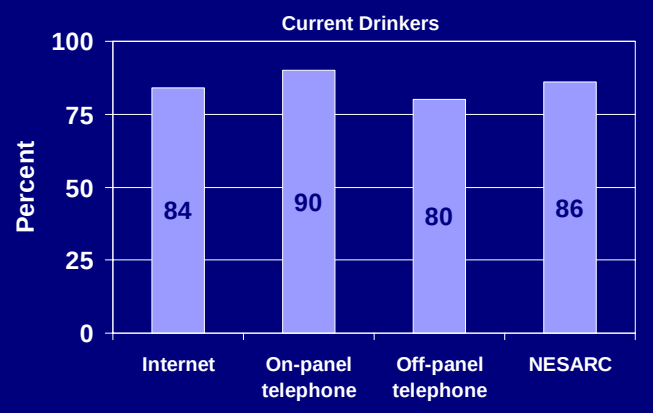
Draft – do not cite without permission

Demographics – Ever Drinkers

Occupation	Internet	On-panel telephone	Off-panel telephone	NESARC
	%	%	%	%
Full-time	56	52		65
Part-time	10	12		8
Unemployed	5	5		6
Homemaker	9	12		4
In school and working	13	13		8
In school and not working	6	6		8

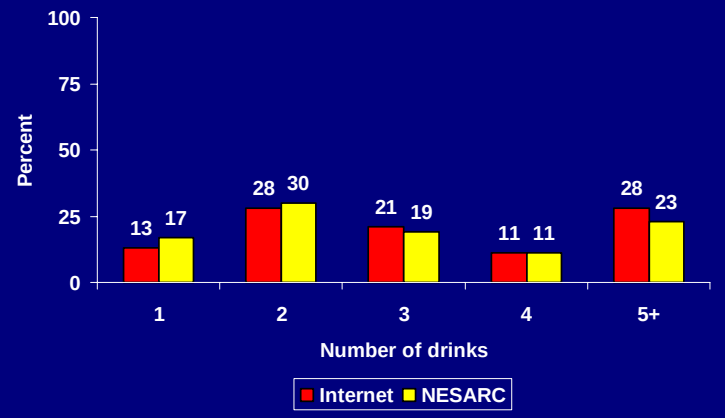
Draft – do not cite without permission

Drinking Status of Ever Drinkers



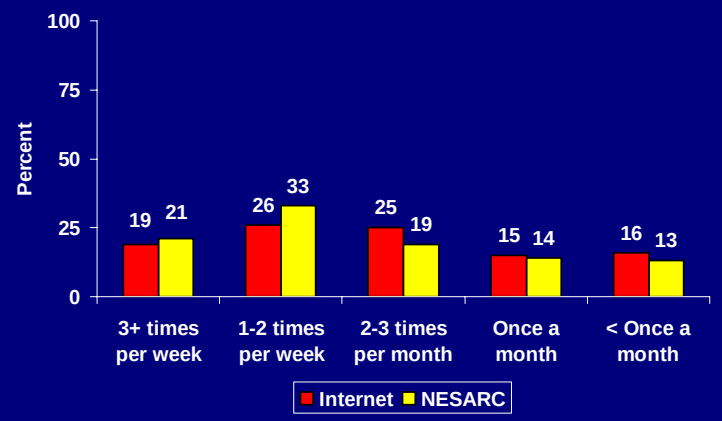
Draft – do not cite without permission

Usual Quantity, Current Year



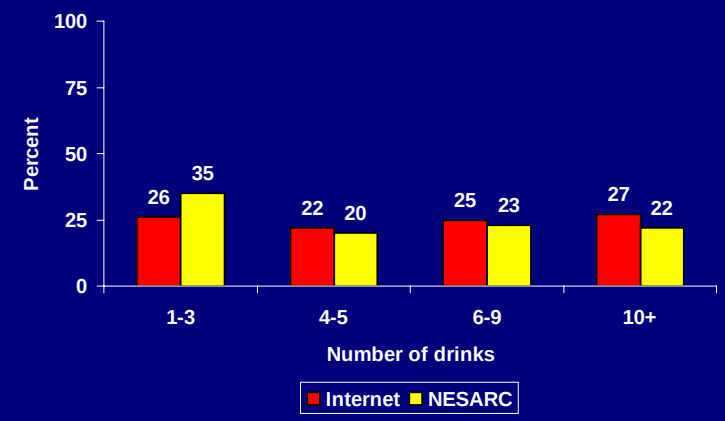
Draft – do not cite without permission

Frequency of Drinking, Current Year



Draft – do not cite without permission

Largest Quantity, Current Year



Draft – do not cite without permission

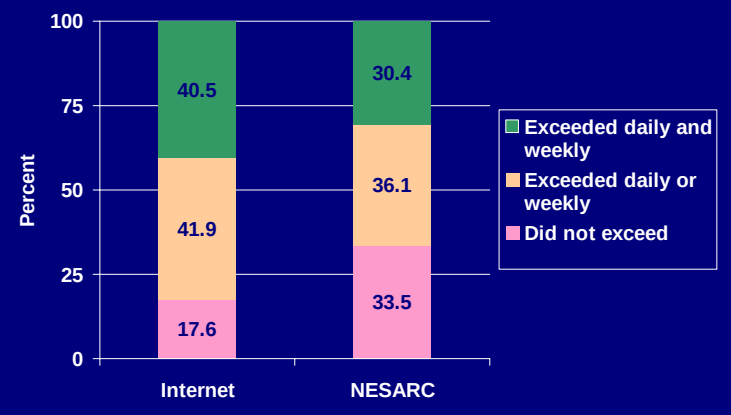
NIAAA Guidelines

- Men (≤ age 65)
  - No more than 4 drinks in a day AND
  - On average, no more than 14 drinks per week
- Women (and men > age 65)
  - No more than 3 drinks in a day AND
  - On average, no more than 7 drinks per week

Source: NIAAA. Helping Patients Who Drink Too Much: A Clinician's Guide. 2005.

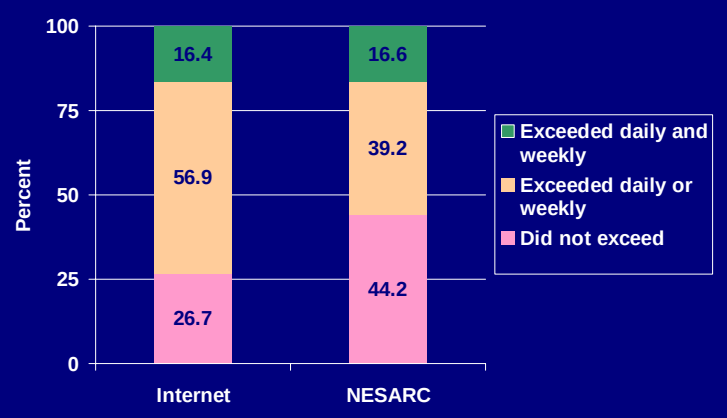
Draft – do not cite without permission

Heaviest Drinking Period



Draft – do not cite without permission

Current Year



Draft – do not cite without permission

Risky drinking and dependence

Heaviest Drinking Period	Internet Survey % dependent	NESARC % dependent
Exceeded both	58.5	51.0
Exceeded either	24.2	18.1
Did not exceed	4.1	3.3

Draft – do not cite without permission

Association between risky drinking and lifetime risk of dependence

Heaviest Drinking Period	Internet Survey	NESARC
	Odds Ratio (95% CI)	Odds Ratio (95% CI)
Exceeded both	33.0 (21.9, 49.7)	30.5 (24.3, 38.2)
Exceeded either	7.5 (4.9, 11.3)	6.5 (5.1, 8.2)
Did not exceed	1.0	1.0

Draft – do not cite without permission

Conclusions

- Internet panel is a better choice for targeted surveys than telephone
- Minor mode differences
- On-panel and off-panel discrepancies
- Good approximation to national survey

Draft – do not cite without permission