Reducing Survey Bias: A Few Old (Pretty Easy) Lessons are Especially Important for Web Surveys

Lisa A. Skumatz, Ph.D., Skumatz Economic Research Associates, Inc (SERA)
Jane S. Peters, Ph.D., Jane S. Peters Advising

ABSTRACT

Web surveys have been a significant boon to evaluation. However, the ease and relatively low expense of implementing these surveys has at the same time led to a drift away from taking the steps needed to ensure sufficient high-quality responses to support evaluation. This paper reviews the relatively few steps that turn a "responses we got back" survey into a valid, reliable set of data that minimizes bias, and is worthy of the time and effort put into evaluation analyses and recommendations. These steps are not unknown (Dillman's classic is the basis for many), but the authors review nearly a hundred surveys annually and the vast majority – from reputable consulting firms – seem to need reminders of these principles. The paper goes into detail on each of the steps, including:

- Early steps include optimal stratification/and sampling keyed to survey applications; quality instruments; pretesting through analysis steps (and additional pre-coding); up-front invitations and reminders; multiple attempts before deletion of sample points; and quota monitoring.
- A vital but commonly omitted step is a "non-response survey" of those not yet responding. Assigning budget to this effort is critical to reduce non-response bias.
- Finally, weighting against known demographics / firmographics reduces bias and improves representativeness, and defensible imputation of important missing data reduces bias and retains sample for analytical results. This, along with reporting of response rates and confidence intervals, enables the reader to justify the veracity of the findings.

These reminders for the web-based surveys will help reduce two key sources of survey bias, and improve the data supporting evaluation studies.

Introduction

Web surveys are easy and relatively inexpensive on the face of it, but the ease of use seems to be causing survey practices to become more casual and less rigorous. It is easy to say what not to do in conducting a web-based survey these days. Don't take the list of participating customers that provided emails and send out round after round of solicitations to take your survey until you run out of time or budget, or you get your desired response count or "N". We are seeing all too frequent devolution to this in the field as surveys have gotten easier to issue.

A few reminders can help reduce two important kinds of survey bias.

- Response bias. This is caused by inaccuracies in answers provided by respondents from a poorly designed survey. The information provided does not represent either the true respondent mean or the population's mean.
- Non-response bias. This derives from eligible survey sample that are unable to or unwilling to respond to a survey question or the entire survey. Bias occurs if the non-response is not random. Those who elected to respond are different, so the respondents do not represent the entire population.

The resulting data and subsequent analyses from surveys conducted using poorer practices suffer from inaccuracies in the estimated mean, larger variance for estimates, and the sample responses no longer being representative of the population as a whole. This paper provides reminders in a few key areas that will help improve the reliability and robustness of the survey. The three key areas addressed in this paper include:

- Sampling, survey development, and data collection / monitoring,
- Non-response survey,
- Missing data imputation, bias reduction and weighting.

The authors are addressing the suggestions to the application of energy evaluation surveys; for example, residential and commercial participant or non-participants surveys, baseline residential or commercial surveys and the like. However, these basic lessons are applicable to a wide range of projects that include survey data collection..

Generally, the amount spent on evaluation should relate to the cost of a wrong decision (or benefits from avoiding a wrong decision). That would imply that these best practices in surveys are especially important (but not only important) when the surveys are applied to quantitative analyses related to savings estimates used for planning or performance incentives (gross savings, net-to-gross, etc.), factors affecting financial decisions or benefit-cost, goals achievement, and other similar applications.

In this paper we are not providing a textbook of all steps in survey work. Rather, we are providing gentle reminders of key longstanding principles and rules; trying to be practical / real-world and correct, without being overly picky or pedantic; and with a few new tips from recent real-world experience added in.

SAMPLING, SURVEY DEVELOPMENT, AND DATA COLLECTION PROCEDURES

Populations, Samples, and Stratification

Certainly, all are aware that a sample pulled from a population can at best represent that population as a whole; and not a population you do not have. However, even getting it to well-represent that population requires some thinking. What is the core question for the study: for example, are you primarily interested in answers that are representative of the participant experience or representative of energy use? The population for the first one would tend to be participants and you'd sample randomly; the second would require knowledge of the energy use of each participant and then stratifying the sample based on the energy use. Generally, include a stratified sampling design if any of the following match your condition:

- you want to be able to provide representation for specific sub-groups, or
- you have reason to believe that some respondents will tend to avoid responding disproportionately.
- Some of the population are especially important in explaining your value of interest (e.g. explaining energy in C&I may require well-oversampling the largest energy users.)

There are optimal rules for choosing strata (Schaeffer, Mendenhall, and Ott 1986, Section 5.9). Random selection is the norm once you have prepared your population and developed your strata. However, up-front work to develop a population without important gaps is critical. If you're planning a web survey, but only have emails for 60% of your population, that will be a very poor survey for representing your population, and your client should not consider it acceptable quality. Some options include[1]:

- Using address (or phone) information that is often available for nearly every participant, and solicit surveys via postcard or letter (or phone). This option is discussed below. This is a vital step to reduce bias, even though more expensive, more complex (logos and more utility review), and slower. Quality is an important objective when taking peoples' time and data, and as underpinnings to decisions being made from the research.
- Samples for some types of surveys may be drawn from alternate sources, like tax records, purchased samples or other sources.
- Conducting work in advance to drive email engagement (like incentivizing customers to move to paperless billing, or requiring emails before program incentive checks are provided), or specific outreach to gather emails with other known contact information.

It is not acceptable to use a population lacking a substantial share of your key contact variable as your sampling frame. If the contact variable is missing in any systematic ways it cannot reliably represent the full population.

Recall that sample sizes are computed differently depending on whether they are sampling for responses for opinions and characteristic, vs. load-based applications. Assure you are using the proper formula and assumptions. Remember, also, that one piece of information needed for setting sample sizes is the coefficient of variation (C.V.). Information on actual coefficients of variation from the previous study, or from similar studies of the same types of respondents, may give you information that allows a refined (maybe even smaller) sample size than the default you may be using.

---

[1] Purchasing emails (or phone numbers) matched by address from third-party firms can be an option. However, it has several drawbacks important in energy applications. 1) There is no guarantee of email quality, chances for duplication for poorly matched email names/addresses, and customer (renter) turnover. This leads to a net negative customer experience in the name of data collection. 2) Mass emails with purchased emails typically gets the purchaser spam-trapped, which can take months to fix. 3) Utilities already need to send out mass communications for things like storms - getting blacklisted hurts the ability of utilities to reach customers at relevant times. Thanks to a reviewer for identifying these important utility considerations.

Issue of Survey Length

It is usually assumed that in survey length, short is good – that survey length affects non-response, likely because of the burden on respondents. As it turns out, the research on the link between survey length and completion / non-response is surprisingly weak and inconsistent. One study looked at the issue of non-response being caused by burdens on respondents (Bogen 1996). It found respondents' professed concerns about survey length did not correspond to their actual completion behavior. They also found basic follow-up procedures usually eliminated length-related response rate differences. One study used regression analysis to examine the most important factors affecting response rate, and found the odds were approximately 50% higher for the short version of a questionnaire than the long (Sahlqvist 2011) and that the cost-per-returned questionnaire was 82% higher for the long survey compared to the short. Several others were unable to find strong relationships; we found the literature was often vague in the conclusions. However, one other meta-type study came to a very interesting conclusion worth considering. After a review of scores of studies, they stated, "Given the inherently problematic nature of comparing questionnaires of various lengths, it is preferable to base decisions on the use of instruments on the content rather than the length per se" (Rolstad et. al. 2011).

A conclusion may be that the relationship seems to be more often positive than not, but length is not necessarily the primary driver; interesting content and relevance to the respondent may be very important. Our experience also argues that questions that are worded very formally or passively or with too many caveats, can be very off-putting and harder to answer (quickly) than ones that get to the point! Also, the cost information indicates, in some cases, it may be quite cost-effective to issue two short surveys to get the information you need. Another solution may be to double your sample (or split your sample) and craft a survey with 3 modules: one that is issued to both (e.g. demographics, participation satisfaction) and then use an in-survey randomizing question that selects whether any single respondent gets module A or B (example – Net-to-Gross vs. Non-energy impacts). An easy in-survey randomizer is to ask people the first letter of their last name, which in the US splits at A-L, and M-Z for 50%.

Survey Instrument Reminders

We assume the reader is preparing a well-written survey. This implies: clear questions, questions linked to analytical needs, no compound questions, language and jargon matched to the audience, and other steps in developing a quality instrument. As noted in the previous section, make the questions as direct and engaging as possible. Check that the survey responses will actually support the analytical uses or computations you plan, and run a test on those desired calculations prior to survey launch.

One common type of question bears a short discussion, Likert Scales. These 5, 7, or 9-point questions ask respondents to indicate a selection along a linear scale marked with labels on the extremes (very dissatisfied and very satisfied), or with labels at each numeric point. Analysis of Likert questions from energy efficiency (and other) surveys usually assign equal differences in scores for each numeric increment, even though the originators of Likert scales warned about the inappropriateness of this practice. There are also issues with end point avoidance. This is a considerable weakness, and research (Gibbs and Skumatz 2022) suggests considering the use of VAS (Visual Analog Scale), allowing answers along a 100-length slider, or better yet, consider using labeled scaling (LS) options, rather than Likert. LS uses research that shows that respondents are very consistent in their "understanding" of labels (like much more, somewhat

more, slightly less, and other "labels"), and in a way that lets researchers assign actual multipliers to those terms, and that these multipliers are decidedly not linear. This is a significant improvement over a linear Likert Scale, and can be readily used in analytical applications.

Do not forget to randomize response category lists, and other good practices. Also consider programming the survey with internal logic or range checks and choose which key variables will have required responses. Drop-downs and multiple choice will help make responding easier than open-ended questions, will help you code responses at the end, and will increase response rates. It is important to recognize that any missed or don't know responses reduce the valid observations count available for analysis.

Careful programming, including skip patterns, is vital. After programming, you and colleagues (and perhaps friends unfamiliar with the topic) should pre-test the survey rigorously for clarity, skip patterns, length, etc.

Steps for Soliciting Responses and Collection the Data – Address-based Web Survey Data Collection.

The following recommended steps are adapted from Dillman (2014).

After programming, careful review of the survey instrument, and selecting the sample, conduct a pretest with a soft launch to enough of the sample to achieve 10 responses. Check the data carefully and prepare a memo that describes the issues with the survey, and the corrections you'll make to address problems identified. Identify places where certain "other / specify" responses are frequent enough to allow additional pre-coded response categories. Then make the corrections and revisions prior to the full launch.

The recommended steps for survey administration include:

- Prepare an introductory postcard or letter with an invitation to the survey link and/or phone number for self-administration.
    - If emails are nearly universally available, prepare the introductory email invitation with the web survey link.
- Review the data and check the responses for problems and code additional options as useful.
- Two weeks after the first postcard / letter is sent, provide a reminder postcard with invitation to survey link or phone number (this will ensure all sample points have access to the survey).
    - If the outreach is via email, provide the second outreach one week later.
- For the primary sample, conduct four attempts before deleting the sample point.
- Check quotas for stratified samples and see if additional targeted outreach is needed (using fresh sample if available).
- Complete 4 outreach attempts to added sample points prior to discarding the sample point.

Following up repeatedly with the original sample, rather than "throwing" more sample at it, is the preferred option, for statistical reasons. Tellingly, even "Statistics for Dummies," (Rumsey 2003) notes, "selecting a smaller initial sample and following up aggressively is better than selecting a bigger sample that ends up with a low response rate. Aggressive follow-up reduces bias."

Getting more Responses in your Primary Sample: Addressing problems of Declining Response Rates

We recognize response rates are falling dramatically. A few tips for increasing residential and commercial response rates follow. Other considerations are discussed below.

- Tips for residential surveys:
- Clear statement of purpose, simple wording / get to the point
- Consider providing monetary incentives sent in the advance letters
- Consider offering a two-part incentive – a small monetary incentive to all respondents and an entry into a lottery for a higher incentive for all respondents.
- Consider using combinations or phases of web and phone approaches,
- Knowledgeable interviewees for phone (not call centers),
- Assure the survey is mobile-friendly

- Tips for C&I surveys:
- Consider that sometimes gatekeepers know a lot of key information (e.g., very useful in some elements for builder surveys and others).
- Monetary incentives need to respect the value of their time.
- Offer information sharing (or the report) may also have value to potential respondents.
- Consider working with professional associations
- Consider snowball samples where random isn't vital (e.g., vendors, stakeholders, etc.).

There are multiple ways to increase response rates. As mentioned, "engaging", and if possible, short and concise surveys, are key. Incentives (entering all respondents into a lottery for a big prize, or smaller incentives for all respondents) certainly helps, especially with trade allies and non-participant surveys (Na'im, Wilson-Wright, and Rathbun 2015). Note that trade allies sometimes are unable to accept incentives due to company policy. Incentives tend to increase response rates, but one issue is whether or not it changes who responds. It likely does, and so self-report bias occurs, and it may be more of an issue with incentives; however, having a large enough response rate to do statistics is also important.

Response rates are also increased by a subject line that is inviting, the words you use in the text, time of day / day of week that you send the survey, who sends the survey, and other strategies. Adding incentives that the respondent can tailor to their preferred stores, adjusting the length of the survey and many others that have been noted by other authors (Buhr and Greco 2015; Andryauskas 2015; Loomis et al. 2019). The Total Design Method from Dillman (2014), mentioned above, is a strong approach to achieving strong samples and response rates.

Panels

Panels are another option for increasing or achieving response rates. You may be in a position to create your own panel. You can develop a well-designed, representative set of respondents to a survey, designed to be re-surveyed periodically, with replacements for attrition or sample drift. The respondents expect to be surveyed, and are willing to do so, based on your (monetary or other) agreement. This repeated panel helps assure you get responses you need. These are not inexpensive, and need maintenance to keep strong properties and representativeness. Subjects can need replacement as they become more knowledgeable and other reasons.

If you are on a tight budget, or cannot bear budget risk, or need responses in a hurry, another option is to use a purchased panel. SurveyMonkey™, Qualtrics™, Dynata™, and others have such panels available that can be used for your web survey. You work with the vendor to identify the panel characteristics of interest and the number of responses needed. They can usually meet requests for largest cities, as well as counties, states, and national panels. Given the incidence expected for your survey, they'll quote you a cost per completed response, usually a very reasonable dollar amount, and far less hassle and faster than running your own (or starting a survey from scratch). You'll work with them to estimate incidence, which varies if you want any respondent vs., for instance, only people older than 21 who know about their energy bill in a medium-sized city. The negatives are that you have less control, you need to investigate the quality of their panel, and you very much need to investigate the responses for "click throughs," to watch for respondents that are becoming near professionals in answering questionnaires. It is also important to specify the demographic/firmographic characteristics for the panel recognizing that the panel provided will only provide what you request and if you miss a characteristic in your specifications, you may be surprised. However, you only pay for "good" responses, and you get them quickly and affordably. This may be something worth considering in some situations.

Factors to Consider in Web Survey Platforms?

If you are considering using a web survey platform or purchased panel, a list of a few features of interest for an effective web survey platform follow. Consideration of these factors will help automate reminders and facilitate other elements consistent with best practices mentioned above.

- Survey cloud-based software solutions that include effective survey design and delivery tools that allow the survey to be delivered at specific times of day and week (time and date and different batches),
- options that track responses and allow reminders to be set to go only to non-respondents,
- have systems that get through blocking software that big companies use and that google uses on Gmail,
- options that include the ability to customize the look and design of the survey,
- options that that allow the author to set the subject line,
- options that include the ability to offer incentives, and offer gift cards easily,
- options that have services to do A&B testing of delivery methods as part of pre-testing, and options that offer a variety of formats for the survey from mobile devices to desktop --- especially since there is an increasing number of surveys conducted on phones and tablets these days.
- Handles multiple language panels in large states.
- If panels are considered :
  - have very strong panel support; you can design the response rate to match multiple demographic characteristics (and important feature because any key variable you don't match to will be off).
  - All panels have risks, there are lots of panel providers. Qualtrics has their own and uses other specialized panels.

These extra options add to the cost of a survey tool but ensure that the evaluator can better manage the survey consistently with best practices and get the most out of the survey process.

THE FORGOTTEN, DELIBERATE NON-RESPONSE SURVEY

One of the most important omissions being made since the shift to using web surveys is the near-exclusive use of one-stage surveys, rather than a primary survey with a deliberate follow-up of a sample of the non-respondents to your survey. As mentioned above, non-response bias results when eligible respondents in the survey sample are unable to, or unwilling to, respond to a survey question or the entire survey. Bias occurs if the non-response is not random. The end result of this problem is that the respondents do not represent the entire population. The resulting data suffers inaccuracies in the estimated mean from the true mean, and larger variance for estimates. The crux of the problem is that the sample responses are not representative of the population *as a whole*. Good evaluation work requires good data. Reserve a share of the survey budget for a second stage non-response survey, and allocate time for this effort. This dedicated non-response survey effort is crucial to provide significant improvements in these statistical properties of your survey. A common no-response survey approach follows.

After the completion of the primary survey:

- Select a Non-respondent sample –After the initial wave is complete, take about a 10% sample[2] of the entire population of non-respondents and conduct 3- 4 additional attempts. You may wish to increase the incentive for this subsample. You should consider a different outreach look or method, and potentially different completion method (e.g. phone, etc.) where possible.
    - Use a similar (or more aggressive) reminder process as above, for example 2 weeks after, send reminder postcard / notice. (1 week after for Email).

Why is this a substantially better approach? Because, with a one-stage survey, you are assuming that everyone in your population is just like your respondents. With a two-stage survey, you are recognizing that those who respond relatively easily (the primary stage respondents) may differ from those who only responded after considerably more effort. These two populations will tend to differ. The weights for the follow-up will be larger than the main sample because they are representing the non-respondents, and the representativeness of the survey is greatly improved. Fewer total respondents in a two-stage survey have better representativeness than a one-stage survey with more respondents. With web-based approaches, this is an easily implemented modification with big returns and low incremental cost.[3] For those interested, key formulae for the mean and variance are reproduced below which the reader may find useful. The source also provides formulae for selecting the optimal size of the non-response survey based on survey cost and variances.

---

[2] Ten percent is the traditional follow-up percent, but other values can be used.

[3] Formulae copied from https://online.stat.psu.edu/stat506/lesson/11/11.1. Deriving or explaining formula are not the purpose of this paper. They are reproduced here merely for reader convenience in assembling information relevant to the paper's subject.

1. Mean:

$$\bar{y}_d = w_1 \bar{y}_1 + w_2 \bar{y}_2$$

2. Variance:

$$\hat{V}ar(\bar{y}_d) = \frac{N - n'}{N(n' - 1)} \sum_{h=1}^{L} w_h (\bar{y}_h - \bar{y}_d)^2 + \frac{N - 1}{N} \sum_{h=1}^{L} \left(\frac{n'_h - 1}{n' - 1} - \frac{n_h - 1}{N - 1} \right) \frac{w_h s_h^2}{n_h}$$

3. Selecting the number of call-backs (non-response sample completes) that minimize cost for a desired fixed value of Variance of the mean (denoted as V_0)

Where Total cost=$(n' * c_0) + (n'_1 * c_1) + (n_2 * c_2)$
Finding optimal value of k (k>1), where: $n_2 = n'_2 / k$, when N is large

$$k = \sqrt{\frac{c_2(\sigma^2 - w_2 \sigma_2^2)}{\sigma_2^2(c_0 + c_1 w_1)}}$$

$$n' = \frac{N(\sigma^2 + (k - 1)w_2 \sigma_2^2)}{NV_0 + \sigma^2}$$

Where:

n' is the initial simple random sample selected from a population of N units
n'$_1$ of these respond (stratum 1)
n'$_2$ of these do not respond
Nonresponse survey: call back n$_2$ samples (completes) by simple random sampling from the n'$_2$ non-respondents

$w_1 = n'_1 / n'$
$w_2 = n'_2 / n'$
y_d=mean (y bar d)
c_0=set-up costs for survey per respondent
c_1=cost of standard response (incremental)
c_2=cost of call back response (non-response survey)
σ^2 is the variance of the entire population
σ^2_2 is the variance of the non-response group
σ^2 is estimated as s^2, the sample variance

MISSING DATA, BIAS REDUCTION AND WEIGHTING

Defensible Imputation Methods for Missing Data

Logic and range checks and drop-downs embedded in the survey programming will simplify the effort involved in cleaning the data. Cleaning and inspecting the data (really *looking* at it) is a key step. After cleaning, review the database for patterns of missing data in priority variables. You will almost certainly have patterns of missing data in your responses. Some questions are harder to answer (e.g. amount of wall insulation in a household).

If these data are important, there are bad ways to fill in the data, and good ways. The poor method of filling in the missing data is to use the mean of other responses – even using the mean for a drilled-down group of most-closely-related respondents. This provides poor statistical properties – particularly through understating the variance. The better method (described simply here, and in more detail in sources like Ong et.al (1989)), is to use a "hot-deck method". The principle is to assemble the most closely matched groups / cells / strata of respondents for the

missing data of interest. Then assemble all the responses from households that did answer the question. Randomly sort them and insert these values into the respondents with the missing data. This assures the mean is still very close to that for the group but allows for variation in the values for improved computations of standard error, for modeling, and other purposes.

A related approach is to classify the closely grouped respondents using propensity scoring (PS) criteria, match the responses to the most closely related respondents, and use that data to fill in the missing data directly – or potentially assigned via a sorted hot-deck method.

Bias Reduction and Weighting

Weight the data based on your pre-stratification plan, taking into account the number of responses you received per cell. Then conduct several checks:

1. The weights will be affected by how many responses were actually received in some cells. Examine the range of the weights and consider how much weight is reasonable to put on some of the responses. If the data quality is known to be excellent, you may be comfortable having high weights for some observations. If the data quality is normal, most researchers are uncomfortable having a few observations have very much higher weights that cause them to be counted very much more heavily than other observations in the analytical work. All data, including these highly weighted observations, are likely to be imperfect. Adjustments may be desirable, necessitating merging "adjacent" cells if responses are zero or so small, they lead to huge weights. This is an important step.

2. Compare the results to your initial population, or to census, if that is a reasonable comparison (like a RASS or C&I baseline, or programs without targeted participants). Compare the demographic / firmographic results to census (or known characteristics of your initial population) and note the groups that may have responded more-than proportionally and less than proportionally. One common example is higher-than-census responses from heads of households over 65 years of age. That self-selection bias can be reduced by the research by adjusting weights for some identifiable subgroups downward (65+ in this case) and others upward. It is highly recommended that multiple variables be used as the basis of these adjustments (e.g., age, own / rent, income, race / language, geography). The adjustments will not lead to a perfect fit; there will be tradeoffs between the variables to better match census or the characteristics of your initial population. However, the researcher can make tradeoffs and reach an optimum that reduces self-selection bias in the analytical data through weighting.

This is the initial weighting of the data for the primary sample. If you conduct a dedicated non-response survey, your weighting will have a secondary round to merge in and properly weight the joint set of responses.

Finally, good survey reporting is needed to provide confidence in your data collection, a key element to providing confidence in your analytical results. Provide tables of response rates in all cases. Discuss survey disposition in the appendix and document the coefficients of variation found in the work to advise future surveys. Provide means with confidence intervals, significant differences, and other information in your results tables.

CONCLUSIONS

These reminders for the web-based world will benefit survey budgets and evaluation quality. There are some variations in these procedures for phone, mail, web, and other options, but the key reminders remain the same.

- Good quality sample with reliable / full (unbiased) contact information
- Suitable stratification matching survey use
- Possible consideration of purchased panels
- Good practices for survey recruitment (a la Dillman 2014)
- Well-designed survey including a concentration on engaging, clear questions (which may help trade off against length)
- Avoid Likert for surveys needing key types of analytical results and comparisons

- Tips for addressing declining response rates
- Multiple contacts
- Non-response survey
- Careful weighting (including against known characteristics) and checking for bias
- Defensible imputation methods for missing data
- Good reporting of survey statistics

Take the extra steps to provide high quality data. Certainly, there are budget implications from this, but remember, a survey with a planned follow-up of non-respondents provides better quality data than a survey with fewer total responses completed with one phase. Evaluations are conducted to inform very expensive programmatic decisions; spending wisely to provide good quality data to underpin these analyses is essential. You can spend a lot to analyze poor quality data, and it will result in poor quality results. We need quality data as the basis of evaluation results that will guide programs well. Remind yourself of principles you already know. Take the time to do it right. It does not have to be more expensive, and it'll be well worth it.

REFERENCES

Andryauskas, E. K. 2015. "Want Better Response Rates? Go Mobile! Best Practices for the Mobile." In *Proceedings of the IEPEC Conference*. Seattle, WA.

Bogen, K. 1996. "The Effect of Questionnaire Length on Response Rates – A Review of the Literature", US Bureau of the Census Working Paper. www.census.gov/content/dam/Census/library/working-papers/1996/adrm/kb9601.pdf

Buhr, T., and V. Greco. 2015. "Conducting Evaluation in an Era of Declining Telephone Survey Response Rates." In *Proceedings of the IEPEC Conference*. Seattle, WA.

Dillman, D. A., J. D. Smyth, and L. M. Christian. 2014. *Internet, Phone, Mail, and Mixed-Mode Surveys: The Tailored Design Method Fourth Edition.* John Wiley and Sons, Inc.. Hoboken New Jersey.

Gibbs, A., and L. A. Skumatz. 2022. "Taking the Bias out of Likert Scales: Four Examples using Better Alternatives." In *Proceedings of the EEDAL Conference*. France.

Jenssen, R. J. 1978. *Statistical Survey Techniques*. John Wiley & Sons. New York.

Loomis, J., E. Focella, A. Weaver, and R. Bliss. 2019. "Increasing Responses to Web Surveys: No Tote Bag Required." In *Proceedings of the IEPEC Conference*. Portland, OR.

Na'im, A., L. Wilson-Wright, and P. Rathbun. 2015. "How Abraham Lincoln Can Improve Your Survey: The Costs and Implications of Offering an Incentive and Mixed-Mode Survey." In *Proceedings of the IEPEC Conference.* Seattle, WA.

Ong, P., S. Holt, R. Barnes, and L. Skumatz. 1989. "Hot Deck Imputation Method." *The Energy Journal*.

Penn State University, Eberly College of Science. 2024. "Stat 506 / Sampling Theory and Methods." Web citation: online.stat.psu.edu/stat506/lesson/11/11.1

Rolstad, S., J. Adler, and A. Ryden. 2011. "Response Burden and Questionnaire Length: Is Shorter Better? A Review and Meta-analysis." *Value in Health* Volume 14, Issue 8, December 2011, Pages 1101-08. doi.org/10.1016/j.jval.2011.06.003

Rumsey. 2003. *Statistics for Dummies.* Wiley Publishing. Hoboken, NJ.

Sahlqvist, S., Y. Song, F. Bull, E. Adams, J. Preston, D. Ogilvie, and the iConnect Consortium. 2011. "Effect of questionnaire length, personalisation, and reminder type on response rate to a complex postal survey: randomized controlled trial." *BMC Medical Research Methodology* Article 62 (2011). May 6. doi.org/10.1186/1471-2288-11-62

Schaeffer, R., W. Mendenhall, and L. Ott. 1986. *Elementary Survey Sampling Third Edition*. PWS-KENT Publishing Company. Boston.

Skumatz, L. 2002. "Comparing Participant Valuation Results Using Three Advanced Survey Measurement Techniques: New Non-Energy Benefits (NEB) Computations of Participant Value." In *Proceedings of ACEEE 2002 Summer Study*. Monterey, CA.