

The Honesty of Online Survey Respondents: Lessons Learned and Prescriptive Remedies

By Felicia Rogers and Michael Richarme

Overview

The front end of online data collection relies heavily on the honesty of respondents. This is particularly true with Internet-based surveying, when the contact between researcher and respondent is solely electronic.

Key areas of mitigating front-end data vulnerability include correct identification of respondents, constructing surveys that are both cognitively and affectively close to the actual consumer decision-making process, utilizing appropriate incentives that promote desired response behaviors, and identifying and eliminating the impact of survey cheaters.

The purpose of this paper is to present a series of preventative measures that researchers can and should take to reduce these vulnerabilities. The measures are based on consumer behavior, statistics, and psychology theory with empirical support. The measures have also been successfully utilized in practice by Decision Analyst and other professional research firms.

Respondent Identification

Various estimates of the Internet user population in the United States range from about 60% of the population to over 75% of the population. However, that doesn't mean that the Internet user population is proportionally spread across different subgroups and subcultures. Therefore, many firms that manage major online panels take care to ensure that potential respondents are recruited from a variety of disparate sources. This can be accomplished through multiple means, including

website links and banners, website sponsorships, email advertising and newsletters, press releases over traditional media, personal letters, search-engine listings, and other similar means.

The use of a double opt-in protocol for online research panels has become a de facto standard employed by reputable firms.

Detailed information is gathered about the respondent during the registration process, and respondents are carefully selected for specific survey opportunities based on their profiles. Potential survey participants are selected from these panels and are emailed invitations to participate in surveys.

To make sure that survey invitations reach respondents, Decision Analyst has gone through an extensive white-listing

procedure with Habeas, one of several firms that certifies safe lists and monitors the practices of legitimate email senders. The white-list certification code is embedded in email headers, which signals to over 800 large ISPs that the emails are legitimate and should not be caught in spam filters. This protects both the emailer and the recipient.

However, many large panel management firms are engaging in a research "arms race," in which the size of one's panel is purported to be directly related to the quality of one's panel. Some firms have, at times, purchased email address lists, which are then utilized to artificially boost the reported member counts of their panels. Clearly, this logic doesn't stand the scrutiny of reason and passes muster with only the most novice researcher. Panel providers should always disclose the need to supplement with additional sources if and when the

Various estimates of the Internet user population in the United States range from about 60% of the population to over 75% of the population. However, that doesn't mean that the Internet user population is proportionally spread across different subgroups and subcultures.

need occurs. Virtually every panel provider faces the need to do this at least occasionally.

In addition, some firms fail to purge their panels of records with bad email addresses (or at least partition them off), even after they have been bounced by the Internet Service Provider (ISP) as no longer active. There are multiple reasons an email might “bounce.” A soft bounce occurs when an email does not reach the intended recipient for a temporary reason, such as a full mailbox. In that case, the researcher should attempt to recontact the recipient up to three or four times over the course of some designated time period (typically a week or two). If the ISP generates a hard bounce code, that means the mailbox or the host server is no longer active, and the corresponding email address is invalid. In this case the record should be removed from the panel survey candidate list. The panel counts should be adjusted accordingly, so there is no risk of inflated expectations based on overstated panel counts.

A third condition also applies when an invitation email is successfully delivered, but the respondent does not respond to the invitation to take the survey. If a panelist is unresponsive to a certain number of consecutive invitations, that panelist should be removed from the database. Decision Analyst does this type of pruning on a continual basis, keeping the panel fresh and eliminating deadwood nonresponders.

With a clean database, the likelihood of higher response rates (and the corresponding reduction of concern about nonrespondent bias) is increased.

Survey Construction

Paper surveys are simple, low cost, and allow the communication of pictures and diagrams. But they are inflexible and do not easily facilitate list rotations, skip patterns, or other technical aspects of bias elimination. Computerized telephone surveys are often more costly, but they allow us to use some of the aspects of research methodology that paper surveys don't. Rotations and quota management are simplified, and project duration is typically shortened. However, telephone surveys also have some inherent drawbacks, including the potential for interviewer-induced bias and the inability to test visual media.

When the Internet was first utilized for research in the mid-1990s, many researchers quickly realized that this technology had the ability to eliminate or greatly reduce many of the concerns that prior methodologies introduced. Many of the limitations of computerized telephone surveying and paper surveying, particularly those conducted through the mail, were removed.

But there is still one huge step to take in this area. Many Internet-based surveys are simply paper surveys converted to electronic form. This is particularly true of the plethora of “do-it-yourself” (DIY) programs available to the research industry today. The big problem with this approach is that it isn't the way in which consumers view and evaluate the world. These simple surveys don't mimic the consumer decision-making processes. At best, they capture surface impressions rather than deep insights. So what researchers get is an analog of the consumer's opinions and beliefs rather than useable information.

Over the past few years, several research firms have been moving in the direction of developing more realistic surveys that are more engaging to respondents, elicit more thoughtful and meaningful responses, and are much better proxies for the actual decision process. Take the shopper insights arena, for example. Some newer techniques include the use of flash programs and 3D animation to provide simulation of a realistic purchasing environment, the use of multimedia stimuli, and the use of more interactive scales and response methods.

The Internet does facilitate this approach far better than any prior research methodology developed, perhaps with the exception of individual depth interviews. Some researchers will continue to utilize the DIY or outdated approaches and continue to wonder why their research isn't producing very reliable or meaningful results. Others, using the approach of integrating more realism into their research methodologies, will see more engaged respondents, more meaningful and insightful responses, and more actionable research results.

Respondent Incentives

One key to higher quality responses is establishing a mutually beneficial relationship with respondents. In a recent project

American Consumer Opinion® respondents were asked why they completed surveys for Decision Analyst. Responses included a desire to contribute direction and provide input, an interest in seeing what might be introduced in the future, and a general need to share opinions with others. Financial incentives (getting paid for completing surveys) were in the middle of the motivation rankings. This was consistent with prior research conducted by Franke and Shah (2003).

High quality responses are the most important consideration. Incentives are a means of providing some token recognition and appreciation to respondents for the time and attention they provide in answering surveys. There are a variety of means of recognition—some financial and some nonfinancial (e.g., customized websites with information, games, and other engaging items for specific groups of respondents). In addition to general rewards associated with being a member of a respondent panel, there are specific incentives associated with the completion of individual surveys. Different panel providers utilize different incentives, ranging from points accumulating toward specified prizes, to sweepstakes to cash.

Cash is king when it comes to respondent interest in incentives. A number of empirical studies (Cobanoglu, Eyerman, Kulka, and Zagorsky) have shown that cash incentives generate superior response rates and lower attrition rates. However, there are some arguments against cash. For one, it raises the cost of surveys. Although this is a valid argument that must be considered, it is inwardly-focused and should be weighed against the intangible value of a more engaged, more satisfied respondent.

Another argument against the use of cash incentives is that they will attract professional respondents, who are seeking to earn a supplemental income from completing surveys and whose qualifications and responses are suspect. One way to mitigate that concern is to limit the number of surveys an individual takes during a specific time period. The respondent should receive enough invitations to remain engaged with the panel, yet not enough invitations that survey fatigue or supplemental income motivations enter the equation.

Decision Analyst places strict controls on respondent participation at the sampling stage. Our policies do not allow

self-selection into surveys; respondents must be invited by us to participate. If the survey is a simple screener for qualifications, a respondent may re-enter the sample pool after four business days. If respondents complete a screener or survey, they are excluded from the sample pool for 15 days, and are also precluded from taking another survey on the same topic for six months. The typical respondent completes four or five surveys a year, which completely removes the supplemental income motivation concerns.

To ensure enough contact with respondents, Decision Analyst conducts its own research using panel members. Examples of this include things like the monthly Economic Tracker we conduct in the U.S. and several other countries around the globe, split-sample studies where different methodologies and approaches are being examined, and our quarterly panel screenings to identify low-incidence category users. These activities allow us to connect with respondents who may not qualify for client-sponsored projects, or respondents who are overrepresented in the panel population, and keep them involved with the survey process. When a client-sponsored survey comes around, that engagement translates into a higher response rate and more thoughtful responses.

Finally, an often overlooked benefit of using cash incentives for survey completion is the final quality-assurance step it can provide. Decision Analyst physically mails incentive checks to panelists, so we need to know the legal name and mailing address of each panel member to complete the transaction. This adds an additional element to our fraud-prevention efforts in managing our panels. Such a quality-assurance step would not be possible through other incentive methods.

Mitigating Survey Cheaters

Even with all the controls and measures taken in recruiting and motivating respondents, cheaters can occasionally sneak through the various traps and provide responses that could mislead the researchers. In addition to the recruitment and panel management techniques discussed above, researchers should incorporate survey-specific practices to flag potential “cheater” data to determine whether or not the panelist’s data should be removed. Identification of respondents who exhibit cheating behaviors is also important for removal of those re-

spondents from the panel to prevent them from participating in subsequent surveys.

There are a number of a priori and post hoc techniques that help keep cheaters out of surveys, and also help clean the data set for those who have managed to sneak into the survey. The a priori techniques generally take place prior to the invitation emails being released and, at Decision Analyst, include intensive efforts to identify bad registration data, such as invalid addresses, illogical education-occupation or age-education combinations, too many occupations, duplicate registrations, and patently false information provided during registration.

In addition trap questions are included within surveys to identify respondents who are not reading the questions before selecting responses or who are using automated response methods. Examples of trap questions include a simple direction to choose a specific response from a short list of items, such as asking the respondent to choose “Cat” from a short list of animals that includes the response “Cat.” A variation widely used in many industries is to have the respondent type a specific word or two into a text box that replicates the stylized version of that word in the survey.

A more complex process, but one equally valuable and routinely carried out by Decision Analyst, is to examine open-ended responses for haphazard or illogical answers.

Research firms should also keep track of the survey completion length of time, and those respondents who “speed” through a survey should be flagged for data examination. If, for example, a survey should have a typical completion time of 15 minutes, any respondents who complete the survey in an extraordinarily fast manner should be scrutinized more closely. Firms who watch for speeders utilize a variety of methods to determine the boundaries for each project. A simple but effective technique is to look at all respondents who exceed four or five standard deviations from the mean completion time.

Another technique, more applicable with surveys that contain grids of attitudinal Likert-type or semantic differential scales, is to examine each response for patterns of answers. Simple straightlining or regular geometric patterns of responses should be flagged for further scrutiny. Often a well-constructed survey will have logic check questions in different

sections to help identify whether the person was responding honestly or just speeding through the survey.

When a survey response is flagged for a possible cheater, the panel provider and data analysts must decide whether to keep or delete the data. Key things to consider include:

- Length of time to complete (speeders).
- Answer patterns (e.g., straightlining).
- Completeness/Appropriateness of open-ended responses.
- Responses to cheater trap questions.

If the conclusions from these examinations do not immediately lead the researchers to eliminate the record from the data set or the panel, an important step can be to remove the record(s) from the data set and look at a topline tabulation. This will help determine whether or not the record(s) make(s) a material difference to the overall survey results. If so, the survey should be removed from the data set. The panelist should be marked in the database as a cheater and removed from the panel for future surveys. If not, the respondent should be tagged as suspicious and any future surveys from that person should be examined for possible cheating behavior. Repeated offenses generally confirm suspicions of cheating behavior and the respondent would then be removed permanently from the database. Having conducted Internet-based surveys since the mid-1990s, Decision Analyst has found that less than 1% of all respondents are considered cheaters.

As a general rule, Decision Analyst accepts a small number of completes beyond each project’s overall quota. This allows us to eliminate respondents who are speeding through surveys or otherwise cheating, without compromising the desired completed response sample size.

Decision Analyst maintains a database of cheaters, and any new registrants to the American Consumer Opinion® Online panel, or any of the B2B specialty panels owned and operated by Decision Analyst, are screened against this database to ensure that the identified cheaters are not allowed to rejoin.

Conclusions and Recommendations

The issue of data quality is an ongoing one for professional market researchers, whether they are on the supplier side or client side of the business. Researchers are continually striving

to improve methods and techniques to better understand the conscious and unconscious motivations and behaviors of consumers. The depth of our understanding is confounded by respondents who honestly don't know the answers to probing questions, leading us to the development of more sophisticated techniques to dig deeper into the mind of the consumer. Dishonest respondents, who either utilize surveys as an outlet for destructive behaviors or who simply want "free money," can cause havoc with these sophisticated techniques if the dishonest respondents are not identified and removed from the process. Ultimately, they can cost marketing research buyers millions of dollars in the form of bad (misinformed) business decisions.

This paper presents some thoughts on identification and removal of dishonest respondents, both prior to and during the survey process. Each technique, taken separately, provides some marginal value to the process. A comprehensive program of ensuring data quality utilizes each technique in combination with the others, giving researchers a higher level of confidence that their insights into the motivations and behaviors of consumers are more meaningful and actionable.

As with any industry, there will be those who take shortcuts and make assumptions about data quality in the interest of time- and/or cost-savings. One can only hope that the decisions made with that type of research are not costly or damaging to either individuals or to society.

References

- Cobanoglu, Cihan, and Nesrin Cobanoglu (2003), "The Effect of Incentives in Web Surveys: Application and Ethical Considerations," *International Journal of Market Research*, 45.
- eTForecasts & Computer Industry Almanac, February 2008.
- Eyerman, J., K. Bowman, and D. Wright (2005), "The Differential Impact of Incentives on Refusals: Results from the 2001 National Household Survey on Drug Abuse Incentive Experiment," *Journal of Economics and Social Measurement*, pp. 157-169.
- Franke, N., and S. Shah (2003), "How Communities Support Innovative Activities: An Exploration of Assistance and Sharing Among End-Users," *Research Policy*, 32, pp. 1199-1215.
- Kulka, R., J. Eyerman, and M. McNeeley (2005), "The Use of Monetary Incentives in Federal Surveys on Substance Use and Abuse," *Journal Of Economics And Social Measurement*, 30, pp. 233-249.
- Richarme, M., and J. Colias (2008), "Realism in Research: Innovative Uses of 3D Animation Qualitative and Quantitative Research Methodologies," *ESOMAR Congress 2008 Proceedings*, September, pp. 346-361.
- Zagorsky, J., and P. Rhoton (2008), "The Effects Of Promised Monetary Incentives on Attrition in a Long Term Panel Survey," *Public Opinions Quarterly*, 72, No. 3, pp. 502-513.
- Zijlstra, W., L.A. van der Ark, and K. Sijtsma (2007), "Outlier Detection in Test and Questionnaire Data," *Multivariate Behavioral Research*, 42, pp. 531-555.

About the Authors

Felicia Rogers (frogers@decisionanalyst.com) is an Executive Vice President of Decision Analyst. Michael Richarme (mrichar@decisionanalyst.com) is a Senior Vice President of Decision Analyst. The authors may be reached by phone at **1-800-262-5974** or **1-817-640-6166**.

Decision Analyst is a leading international marketing research and analytical consulting firm. The company specializes in advertising testing, strategy research, new product ideation, new product research, and advanced modeling for marketing-decision optimization.



604 Avenue H East • Arlington, TX 76011-3100, USA
1.817.640.6166 or 1.800. ANALYSIS • www.decisionanalyst.com