

Separating the Wheat from the Chaff: Ensuring Data Quality in Internet Samples

Reg Baker & Theo Downes-Le Guin

Abstract

The emergence of online access panels as a sample source has been a major factor in the rapid growth of online surveys. The increased use of non-probability samples in place of probability samples has generated considerable debate and discussion within the research industry about the levels of sampling error and bias inherent in these samples. At the same time, market research buyers, for the most part, have found the obvious advantages in cost and turnaround compelling enough to consciously overlook some of the unattractive aspects of online access panels. But this may begin to change as a new set of non-statistical concerns emerge. These concerns are focused on issues such as respondent qualifications, increased satisficing, and possible professionalization of survey respondents. In response to these concerns, both panel companies and research agencies have begun to look more closely at panel composition and online survey research results. Though this scrutiny happens with varying degrees of formality, the results across organizations have largely converged on three definitions of potentially problematic panel respondents: professionals (i.e., individuals belonging to several panels and completing a large volume of surveys); inattentives or satisficers (i.e., respondents who complete surveys but do so with little cognitive effort); and fraudulents (i.e., respondents who misrepresents themselves in order to qualify, either at the panel registration phase or in individual surveys). This paper reviews the emerging research record as both panel companies and research agencies begin to look more carefully at the behavior of online panelists. It describes the steps that researchers can take to ensure data quality when doing research that uses online access panels.

Keywords

Access panels, online research, data quality, panel quality

1. Introduction

To describe the growth of online research over the last ten years as “dramatic” is to understate the case. Virtually non-existent in 1996, *Inside Research* (2006) reports the global spend in online market research for 2006 to be \$2.6 billion. No surprise to anyone the bulk of this research, at least in the US, relies on online access panels. It is fair to say that online panels have made online research go.

The emergence and increasing acceptance of online panels as a legitimate sample source has not been without controversy. Probability sampling has been a key underpinning of survey

research since the 1930s. In the 1980s telephone surveys using random-digit dialing (RDD) emerged as the most popular methodology for U.S. market research studies, largely due to the availability of a sampling frame that included virtually the entire U.S. population. The most relevant exception to that generalization became mail panels, databases of volunteers, sometimes numbering in the hundreds of thousands, who had signed up through one means or another to do postal surveys. The general acceptance of this methodology throughout much of the industry likely paved the way for acceptance of additional non-probability sampling approaches.

Enter Gordon S. Black, who in the mid-1990s began to assemble a large panel of volunteers to do surveys on the Web. In 1998, he summarized the challenge as follows:

To do what we are doing we have had to set aside the staple of our industry -- the simple random sample... Random sampling is a very powerful tool in every avenue of science and industry for increasing the accuracy of estimates while decreasing the cost of the process... We are not challenging the validity of random sampling... We are instead investigating whether findings from huge samples of Internet respondents, coupled with sophisticated weighting processes, are as accurate as anything done on the telephone or door-to-door.

The extent to which Black's key requirement of sophisticated weighting processes has been met is outside the scope here. Suffice it to say that there are arguments both pro (Taylor (2007)) and con (Schonlau (2004)). The focus of this paper is the current challenge facing the panel industry, what we have described elsewhere (Downes-Le Guin et al. (2006)) as "a perfect storm." At its heart this challenge is about the evolving behavior of online panelists and the degree to which surveys done with them produce results that our clients can safely use to make business decisions.

2. The New Online Panelist Pathologies

One sure fire way to know when we have a problem is when it is given a name. So it is with online panelists. Panel companies and the researchers who use their services have begun to recognize three important panelist pathologies (See, for example, Smith and Brown, 2005; or Downes-Le Guin, 2005). Each of these pathologies is measurable to some degree, and each can affect individual survey results if present in sufficient magnitude.

Hyperactives

The first of these has been dubbed "hyperactives," sometimes also referred to as "professional respondents." These are panelist who do lots of surveys and who may also belong to multiple panels. There is some evidence that response patterns on certain types of surveys may differ based on the number of previous surveys the respondent has completed. (See, for example, Coen et al. (2005); and Cartwright and Nancarrow (2006)). The evidence on multiple panel membership is even more compelling. Table 1 combines findings from two studies of multiple panel membership, one by Smith and Brown (2006) using Harris Interactive in the UK and US and the other by Fine et al. (2006) using AMR Interactive in Australia. The data suggest that respondents who belong to more than one panel are more likely to be female and of lower socio-economic status. Fine et al. conclude that "there are major differences between people who are members of only one panel and those who are members of multiple panels... These differences are seen in demographics, attitudes, and behaviour." A

comprehensive study of 19 panels in the Netherlands (Willems et al. (2006)) found similar differences with hyperactives reporting, among other things, to be “... less satisfied with his or her income, less satisfied with how democracy is working, and less healthy.” The obvious implication is that the mix of hyperactives in a given sample may influence survey results.

Table 1

	Single Panel		Multiple Panels	
	HI	AMR	HI	AMR
Female	52.2%	61.9%	74.7%	71.3%
18-39	40.8%	41.7%	24.3%	49.0%
Less than HS	19.8%	19.4%	23.0%	22.9%
Employed FT	43.7%		35.6%	
Less than \$25K	20.3%		14.1%	

In our experience the levels of hyperactive respondents can vary substantially, both among panel companies and even within surveys using the same panel company. General population surveys of respondents with few qualifying questions often show the lowest levels of hyperactivity, while surveys targeting lower incidence or frequently surveyed respondents can be substantially higher. For example, in general population surveys we frequently see rates of multi-panel membership (based on self reports) of around 30 percent. By way of comparison, on surveys of physicians we often see rates of mulitpanel membership at 50 percent and higher, depending on specialty.

Inattentives

The second type of problematic panelist is the inattentive, that is, a respondent who gives less than full attention to reading the questions and providing well thought out answers. The phenomenon is not confined to online research and access panels. It was first identified by Krosnick (1991) who called it “satisficing.” Three main causes are typically cited: insufficient respondent motivation; reduced respondent cognitive ability; and poor questionnaire design. Satisficing in online research is most often characterized by behaviors such as short completion times, high levels of item nonresponse, non-differentiation or “straightlining” in matrix or grid questions, fewer selections in multiple response questions, and inconsistent or nonsensical responses. Recent studies by Lugtigheid and Rathod (2005) and Gilasic (2006) report that in online studies these behaviors become especially pronounced with surveys of 20 minutes or longer.

As discussed later in this paper, we frequently use a combination of “verification” items in grids and interview length to identify inattentives. Depending on the survey design, the combination of verification item failures and short completion times (e.g., 50 percent of the median interview length) will generally yield between two percent and 10 percent rates for inattentive respondents.

We are unaware of any studies that show whether levels of satisficing are worse with online than with other methodologies or whether online panelists are more likely to satisfice.

Nonetheless, there are a number of features of the online survey experience that would seem to create the necessary conditions for satisficing. These include a wealth of survey opportunities, universal use of incentives, self administration, increasing survey complexity, and, sadly, repetitive and boring survey design (the latter compounded by the fact that online surveys are frequently fielded on very short timelines and without adequate pretesting compared to other modes).

Fraudulents

Finally and perhaps most conspicuously in the minds of research buyers, are the “fraudulents,” sometimes called “gamers”. These are individuals who assume false identities or simply misrepresent their qualifications either at the time of panel registration or in the qualifying questions of individual surveys. Their primary motive is incentive maximization. They tend to be seasoned survey takers who can recognize filter questions and answer them in ways that they believe will increase their likelihood of qualifying for the survey; one classic behavior is selection of all options in multiple response qualifying questions; another is overstating purchase authority or span on control in a B2B survey.

In an earlier paper (Downes-Le Guin et al. (2006)) we described a number of first hand experiences with fraudulent panelists. For example, we described a study in which the target respondent was both home and business decision makers to represent potential purchasers of a new model of printer. The study was multinational with a mix of sample sources including a US customer list provided by the client, a US commercial panel, a European commercial panel, and an Asian phone sample that was recruited to do the survey online. One multiple response qualifying question asked about the ownership of ten home technology products. About 14 percent of the US panelists reporting owning all ten products, including the Segway Human Transporter, an expensive device known to have a very low incidence of ownership among consumers. This behavior was virtually nonexistent in the other sample sources.

A second example comes from a study of U.S. interventional cardiologists, a physician specialty that treats heart disease with non-invasive procedures such as catheterization. ICs are not trained nor licensed to perform open surgery procedures, yet in our survey when respondents were asked to indicate which of about 60 procedures they perform six percent claimed to be performing five or more open surgery procedures (including “extremity amputation”) and another six percent claimed to be performing two to four such procedures. Neither we nor the client believed these respondents were who they claimed to be, and so they were dropped from the study.

Fraudulence has arguably garnered disproportionate focus in the market research industry because, at first blush, it seems more obviously wrong and severe a behavior than panelist hyperactivity and inattentiveness. Left unmonitored, it can have a profound effect on both response error and credibility of the online panel as sample source. At the same time, gross examples of fraudulence are increasingly rare as public panel providers have converged on a set of best practices around identity verification and deduplication. While these practices are far from foolproof, it’s useful to remember that “curbstoning” and other forms of interviewer falsification have haunted both in-person and call center research from their beginnings. Hyperactive and inattentive panelists, on the other hand, are far harder to consistently define, detect and treat in a consistent fashion from survey to survey.

3. Separating the Wheat from the Chaff

We hope the foregoing discussion makes clear that conducting online research with access panels has multiple pitfalls. Avoiding them requires that researchers develop a comprehensive strategy that includes criteria for panel vendor selection, thoughtful questionnaire design, and careful cleaning of survey response data.

Panel Vendor Selection

Deming's (1986) Point 4 seems instructive here: "End the practice of awarding business on the basis of price tag alone." The online panel industry, at least in the U.S., has evolved into something approaching a commodity business. It is difficult to see how that leads to quality research. We favor an approach to vendor selection that starts with careful evaluation of panel recruitment, panelist registration, sample development, and panel management practices.

As in most evaluative exercises knowing the questions to ask is at least half the battle. ESOMAR has been especially helpful in this regard, having developed a set of questions known as "The ESOMAR 25" (2005) that focus on key elements of panel recruitment, panel management, and sample development. The International Standards Organization (ISO) has convened a working group from the same technical committee that developed the ISO 20252 Standard for Market, Opinion, and Social Research with the goal of providing a comprehensive framework for establishment and auditing of access panel standards. Other trade and professional organizations (e.g. the Market Research Association in the US) have developed or are in the process of developing similar guidelines. Of special interest should be questions about:

- The recruitment process and especially steps taken to ferret out multiple registrations.
- Panelist verification, that is, steps taken to verify that a panelist is who s/he says s/he is. This is especially important on B2B panels or panels of professionals such as physicians.
- Limits placed on participation.
- Policies for replacing "bad" respondents on individual surveys.

Panel companies are mostly aware of the ESOMAR 25 and many can easily provide their response to them. Some have even posted those responses on their Web sites. They also are aware of the panelist pathologies described above. They recognize the threat this poses to their business, and many have instituted special programs and procedures aimed at reducing multiple and/or fraudulent registrations. These procedures include scrubbing of email addresses against list of known hyperactives, monitoring of IP address frequency, verification against external sources, and registration and profiling surveys that contain some of the same techniques described below. Panel companies that have embraced these kinds of procedures enthusiastically are the only companies one should seriously consider.

Questionnaire Design

While panel vendors can do a great deal more than they currently are doing to ensure the legitimacy and suitability of the sample members they send us the responsibility for valid survey results ultimately rests with the researcher. This is the minimum that our clients expect from us.

As researchers, our interaction with panelists is pretty much confined to the survey experience, or more specifically, the survey questionnaire. The main focus of questionnaire design has rightfully been the substantive issues needed to answer the basic research questions the client has posed. But there are other important foci that are especially important for online research with access panel respondents.

The first is providing the respondent with as engaging an experience as we can. Not all inattentive behavior can be blamed on the panelist; bad questionnaire design can quickly wear down even the most motivated respondent. Classic questionnaire features that encourage inattentive behaviors include screen after screen of grid questions, continuing iterations of conjoint screens on which the changes from screen to screen are almost imperceptible, and questions that are so poorly worded as to not be taken seriously. Many online questionnaires seem to have evolved little beyond the webbification of CATI screens or mail surveys.

Luetters et al. (2007) have proposed the use of innovative answering devices they call “Sniperscales” that bring some video game techniques into surveys. They willingly admit that they are walking what they call “the narrow path between science and entertainment” but claim that the use of such devices reduces inattentive behavior. At the same time, the widespread assumption that respondents must be entertained and even visually stimulated seems to fly in the face of the best practices that have emerged around virtually every form of web interface design and human/computer interaction. If these disciplines serve as an effective analog for good online survey design, our goal should be to create user interfaces that are clean and uncluttered – focusing on succinct and unambiguous wording rather than gratuitous visual elements.

In the end probably the biggest design challenge we face is survey length. With sufficient incentive we usually can cajole the needed number of respondents to complete even the longest online survey, but the quality of the data may not be what it needs to be.

A second focus is detection of the bad respondent behaviour. In our organization we have come to refer to this as “setting traps” to identify respondents who are candidates for deletion from the survey dataset because they exhibit one or more of the panelist pathologies. Techniques include:

- Use of multiple response qualifying questions that include very low incidence behaviors (e.g., the earlier Segway example) or inconsistencies (i.e., you can’t be both a and b)
- Inclusion of consistency checks by asking essentially the same question but in two different ways in separate parts of the questionnaire.
- Inclusion of “verification items” in grid questions that span multiple screens. These might be questions that reverse direction in scales or simply give an instruction such as, “For verification purposes please click the item on the far right.”
- Collection of time stamps to measure how long a respondent took to answer individual questions or complete the entire survey.

Editing/Cleaning Response Data

The final element in a comprehensive panel quality assurance strategy is post data collection editing designed to identify problematic respondents who are candidates for deletion and possible replacement. This starts with checking the just-described traps as well as evidence of other problematic behaviors. In this context, the researcher should keep in mind that every

survey regardless of sample quality or quality of questionnaire design will have some level of error. The challenge is identify patterns that reflect bad respondent behavior rather than survey error. And so we advocate a scoring system across multiple items and then only delete those respondents who fail multiple items. Elsewhere (Downes-Le Guin (2005)) we have referred to this as “three strikes and you’re out.” Smith and Brown (2006) characterize this more eloquently and accurately as a “quality index” on which the researcher sets cut-off scores. High scoring respondents are deleted while low scoring respondents are retained.

Among the items that might make up such an index are these:

- Unusually short completion times as compared to the median interview length
- Selection of all items in a multiple response or other obvious cheating behavior in qualifying questions
- Selection of bogus or low probability answers
- Internal inconsistencies
- Low differentiation or “straightlining” in grids
- High levels of item nonresponse
- Failure of verification items in grids
- Gibberish or duplicated responses in open ends

The checks that might be included in such an index are limited only by the researcher’s imagination and the effort required to implement them. Mostly we have found that a modest amount of effort (e.g., four hours or less) in post data collection edit can identify the most serious problems. In many instances we perform our initial checks during data collection so that additional sample can be arranged and the survey timeline more or less preserved. We also have experimented with building these QA checks into the Web questionnaire program so that they are performed on a respondent-by-respondent basis in real-time as individual surveys are completed. A key to success of this type of index, however, is that it be specific to each study. The categories of indexed items may be essentially the same from study to study, but how they are treated and the weight given in the index should be sensitive to the specific panel being used, the subject matter of the study, practical limitations on how many panelists we can afford to eject from the final dataset, and so forth.

4. Conclusion

The online panel industry and the methodology it makes possible are now about ten years old. Both have been enormously successful in a relatively brief span of time, but both now face a major challenge. Whether one calls it growing pains or a major crisis, it is clear that our approach to building and operating panels must change, and so must the ways in which researchers use them. Panel companies for the most part recognize that the quality issues are real, and most have begun to compete on quality as aggressively as they have in the past competed on size, coverage, and price. At the same time, definitions of “quality” vary widely and are inordinately focused on the most easily identified issues around fraudulence that can be managed by the panel providers. Those who specify the sample design and deal with the data – that is, full service research firms and their clients – must take increased responsibility for defining and addressing other aspects of quality. Industry organizations such as ESOMAR must consider whether their obligation should extend beyond offering question and guidelines to some form of auditing. The previously-mentioned ISO Working Group would seem to be moving the industry in that direction. That being said, the industry has awakened to the

problems, and hopefully have begun to recognize the key role they must play. Our clients ultimately rely on us as researchers to deliver valid results and recommendations. It may be gratifying for us to know that our online sample suppliers are taking quality seriously, but they will only meet that responsibility if we continue to insist on it.

One cannot help but wonder whether the very success of online panel research is not now the cause of its potential undoing. By one estimate, after deduplication, there are about 18 million people in the US who have signed up to do online research (Rivers (2006)). That amounts to about nine percent of the adult population. We know little about what portion of that 18 million participates in surveys on a regular basis, but it likely is a still smaller group. (By way of contrast, studies show that, on average, respondents to telephone studies in the US participate at a rate of no more than one or two studies per year.) At the same time the demand for online panelists continues to soar, creating the very conditions that lead to the pathologies we have described. In this context the QA techniques put forth in this paper might be little more than a holding action while we refine or perhaps even redefine the paradigm. The most obvious next step is finding ways to encourage a greater share of the population to join panels and to participate regularly. Less obvious is some alternate means for getting the respondents we need for our online surveys that not only increase the size of the respondent pool but also allow us to reduce the bias that has always been inherent in the volunteer panel design

References

- Black, G. and Terhanian, G. (1998) "Using the Internet for Election Forecasting." *The Polling Report*. October 28.
- Cartwright, T. and Nancarrow, C. (2006) "The Effect of Conditioning When Re-interviewing a Pan-European Panel." Presented at ESOMAR World Research Conference: Panel Research 2006, Barcelona.
- Coen, T., Lorch, J. and Piekarski, L. (2005) "The Effects of Survey Frequency on Panelists' Responses." Presented at ESOMAR World Research Conference: Panel Research 2005, Budapest.
- Deming, W.E. (1986) *Out of the Crisis*. MIT Press.
- Downes-Le Guin, T. (2005) "Satisficing Behavior in Online Panelists: What's a Research Buyer to Do?" Presented at the Annual MRA Conference, Chicago, IL.
- Downes-Le Guin, T., Mechling, J., and Baker, R. (2006), "Great Results from Ambiguous Sources: Cleaning Internet Panel Data." Presented at ESOMAR World Research Conference: Panel Research 2006, Barcelona.
- ESOMAR (2005) *Conducting Market and Opinion Research Using the Internet*.
http://www.esomar.org/uploads/pdf/ESOMAR_Codes&Guideline-Conducting_research_using_Internet.pdf.
- Fine, B., Menictas, C., and Casdas, D. (2006) "Attitudinal Differences: Comparing People Who Belong to Multiple Versus Single Panels." Presented at ESOMAR World Research Conference: Panel Research 2006, Barcelona.

- Gilasic, M. (2006) "The Effects of Questionnaire Length on Quality of Data in a Web Survey." Joint Program in Survey Methodology, University of Maryland.
- Inside Research* (2007) "Special Report: Online MR Continues Strong." January.
- Krosnick, J. (1991) "Response Strategies for Coping with the Cognitive Demands of Attitude Measures in Surveys." *Applied Cognitive Psychology*, 5, pp. 213-236.
- Lugtigheid, A. and Rathod, S. (2005) *Questionnaire Length and Response Quality: Myth or Reality*. Survey Sampling International.
- Luetters, H., Westphal, D., and Heublein, F. (2007) "*SniperScale: Graphical Scaling in Sata Sollection and its effect on the response behaviour of participants in online studies.*" Presented at the General Online Conference, Leipzig.
- Rivers, D. (2007) "Are We Running Out of Respondents? Understanding Survey Participation." Presented at AMA Executive Insights Conference, Phoenix, AZ.
- Schonlau, M., Zapert, K., Simon, L., Sanstad, K., Marcus, S., Adams, J., Spranca, M., Kan, H., Turner, R., and Berry S. (2004) "A Comparison Between Responses from a Propensity-Weighted Web Survey and an Identical RDD Survey." *Social Science Computer Review*. 22: 128-138.
- Smith, R. and Brown, H. (2005) "Assessing the Quality of Data from Online Panels: Moving Forward with Confidence." Harris Interactive.
- Smith, R. and Brown, H. (2006) "Panel and Data Quality: Comparing Metrics and Assessing Claims." Presented at ESOMAR World Research Conference: Panel Research 2006, Barcelona.
- Taylor, H. (2007) "The Case for Publishing (Some) Online Polls." *The Polling Report*. January 15.
- Willems, P., Ossenbruggen, R., and Vonk, T. (2006) "The Effects of Panel Recruitment and Management on Research Results: A Study Across 19 Online Panels." Presented at ESOMAR World Research Conference: Panel Research 2006, Barcelona.

About the Authors

Reg Baker is Chief Operating Office of Market Strategies International.

Theo Downes-Le Guin is a Senior Vice President of Market Strategies International.

