

WITHOUT A NET

MAKING BUSINESS DECISIONS FROM ONLINE PANEL DATA

Reg Baker

INTRODUCTION

After a decade of extraordinary growth, online research has now entered a period in which there are significant concerns about panel data quality and even about the validity of online as a research method. Concerns about the emergence of professional respondents, unknown levels of fraudulent qualifying for surveys, unacceptable levels of survey satisficing, and a handful of well-publicized instances in which online surveys failed to replicate either offline or other online results have combined to create a renewed focus on data quality (Baker, 2008). Clients increasingly seek assurances that the results we produce with online studies are reliable enough to use as the basis for important business decisions. The research industry has responded in two ways.

First, individual companies – both panel companies and full service research agencies – have launched panel data quality initiatives designed to identify and root out the major threats. Panel vendors are focused in areas such as multiple and fraudulent registration, multipanel membership, and identification of repeat offenders who exhibit suspicious behaviors over a range of surveys. The full-service agencies that design, execute and report on the research now recognize the need to choose their panel partners more carefully, add additional quality assurance, and create a more positive and engaging online experience for survey respondents.

Second, the industry as a whole, acting through trade associations and professional organizations, is creating an expanding web of standards and guidelines crafted

to help the buyers of panel sample, whether research agencies or end clients, separate the wheat from the chaff. For example, the International Organization for Standardization (ISO) that in 2006 produced the first ever global standard for market, opinion and social research (ISO 20252), will soon release a standard for online access panels (ISO 26362). ESOMAR (2008) has just released an updated and expanded version of its “26 Questions to Help Research Buyers of Online Samples.” The Advertising Research Foundation (2007) has created the Online Research Quality Council with a “mission to develop a set of standards for research buyers and sellers.”

As welcome as these developments may be, sorting through them is no easy task, especially for clients who generally are more focused on how to use research to drive business results than the intricacies of online research methodology. And despite its current troubles, the online value proposition of speed, cost and design flexibility remains too compelling to ignore.

WORKING WITHOUT A NET

Throughout most of its history, market research, like the survey research discipline on which it has relied, has been grounded in the science of probability sampling. Simply stated, if we can create a sample frame of individuals that covers the entire population of interest, draw a random sample of people from that frame, administer a well-designed questionnaire and achieve a reasonably good response rate we can safely assume that any estimates we make from that sample will have a calculable range of possible error. The larger the sample,

the smaller the likely error. Put another way, probability sampling maximizes the likelihood that we will achieve a fully representative sample of the population of interest, although it cannot guarantee it.

Critics have been quick to point out the dramatic fall in survey response rates over the last 20 years has eroded the representativeness of modern probability samples and the accuracy of its estimates (see, for example, Kellner, 2007). Despite these challenges, research continues to show that even with low response rates the robustness of the methodology can still produce reliable estimates (Holbrook et al, 2008; Baker, 2007; Groves, 2007; Keeter et al, 2007).

Most online panels are not built using probability sampling. Rather, the panel paradigm violates two of the most basic tenets of probability theory. First, the sample frame does not include the entire population. At best it includes only the online population, which depending on the country, can vary dramatically in its representativeness. Second, panel members are not selected randomly but rather volunteer to join the panel. Relatively few panels employ probability-based methods in panel recruitment either from the entire population or the online population. These twin biases of undercoverage and self selection mean that the traditional safety net of probability sampling is lost online. To put it bluntly, there is no truly scientific or theoretical foundation for research with access panels (Bethlehem, 2007).

Despite this apparent shortcoming, online now accounts for over four billion dollars (US) of research worldwide (*Inside Research*, 2008). The online advantages of speed and lower cost have been major drivers of this growth, but it simply would not have been possible to establish the methodology so quickly and so broadly were it not for the very simple fact that in a multitude of side-by-side comparisons with other methods it has been shown to work (Miller, 2000). To quote Humphrey Taylor (2007), chairman of the Harris Poll, "Newton had no theory that explained gravity or that justified his 'laws' of gravity, dynamics or optics. But they came to be widely accepted because they worked."

In short, online's departure from our industry's long-term reliance on probability-based methods should not disqualify it as a useful survey methodology. However, it is the proverbial "horse of a different color" and as such should be thought about and evaluated differently from traditional probability-based methods.

ELUSIVE BEST PRACTICES

The case for and against online is further muddled by its novelty. Despite considerable ongoing research, we still lack an industry-wide consensus around best practices for panel recruitment, incentive structures, panel management, and sample selection. We feel in our bones that these things make a difference, and from time to time we see examples in our work. We are extending our understanding of these dynamics, but have yet to come close to being definitive.

The same may be said about online questionnaire design. The collective experience of our industry is with surveys by interviewers, mostly telephone in North American and in-person in the rest of the world. Writing questions and designing questionnaires for self-administration by Web has posed a new set of challenges for us. Simply migrating over postal survey designs has proved problematic. Recent attempts to create a more engaging and effective survey experience for respondents are still in the early stages and their ultimate effects on participation and response quality are yet to be settled. While we have learned that design matters, the elements of what constitutes good design remain elusive.

REDUCING UNCERTAINTY

This paper looks at the panel data quality issue from the client's perspective. It starts from the premise that in almost all online research the safety net of probability theory no longer exists, that all online research is inherently biased, and that the challenge for agencies and clients alike is to understand that bias and determine whether it is likely to lead to an incorrect business decision. It lays out five key issues for clients to consider when choosing online.

1. Be sure that it is safe to ignore the offline population

A fundamental issue in almost any research project is the identification of a sample that is representative of the target population. Our chances of doing so online clearly are greater in a country like The Netherlands where Internet penetration hovers close to 90% than in, say, Turkey, where less than 25% of the population is online (Miniwatts Marketing Group, 2008). According to ESOMAR (2007) "50 - 60 percent seems to be the tipping point for acceptance that representative samples can be achieved." Just over half the countries in the European Union pass this test. Even then, depending on the research topic, this may be too low a threshold.

Much depends on how the target population is defined. In France, for example, where household Internet penetration is around 60% only about a third of persons aged 55 years or older and one in five of those who terminated their education at 15 years or younger are online (Blyth, 2008). In the Czech Republic where overall Internet penetration is 46% the comparable numbers are 16% and 6% respectively. Though less often catalogued, Internet penetration rates also can vary substantially by ethnicity and urban versus rural residence.

It should be obvious that it is difficult to achieve a representative online sample when research is targeted at these underrepresented groups. While a panel may have sufficient numbers to fill quotas, that does not guarantee representativeness.

Equally important is an understanding of the underlying attitudes and behaviors that cause some people to go online while others do not, and how those attitudes and behaviors might correlate with the research topic. The roughly four in ten households in France who have yet to go online may be different in important ways from those who have, and those differences may be reflected in their receptivity to certain products or features. Put another way, might the survey topic be correlated, positively or negatively, with the likelihood of being online? Surveys focused on widely used household products such as toothpaste or cereals may

be unaffected, but other products such as consumer electronics or behaviors such as newspaper readership may be studied best with other methods. People who are online may also have early adopter tendencies that would make them more receptive to new products than those who have remained offline. Based on his comparison of telephone and online Chakrapani (2007) suggests that online may be the wrong choice if one of the research goals is estimation of market share.

2. Require that the research agency address directly the potential for bias and propose specific plans for correcting it

The pending ISO standard for access panels (2008) requires that reports to clients include "evidence that leads to limitations in the projectability to the defined target population / universe." To once again quote Humphrey Taylor (2007), "The issue we address with both our online and our telephone polls is not whether the raw data are a reliable cross-section (we know they are not) but whether we understand the biases well enough to be able to correct them and make the weighted data representative."

The failure of probability samples (such as telephone surveys) to be representative is generally due to response rate. The resulting bias is typically expressed in terms of demographic imbalances that we correct with weighting. The underlying assumption of this demographic model is that values, attitudes, and to a certain extent behaviors are strongly correlated with life stage, gender, and socioeconomic status. In theory, the sample is drawn in a way that is unbiased so any bias in the final survey data is due primarily to differential response rates among demographic groups.

Access panels offer a greater challenge since the samples themselves are biased at the outset. Some people choose to go online while others do not, and the differences between these groups may be important. People who join access panels to do surveys may be more different still. We know little about why one person may choose to join Panel X while another person may choose to join Panel Y. And, as noted above, these biases

likely are more severe in countries with lower Internet penetration than in those with near universal use. Simple demographic weighting is insufficient to correct for these kinds of differences.

Harris Interactive (Terhanian et al, 2001) pioneered the use of logistic regression based weighting methods that essentially add attitudinal and behavioral measures into post stratification adjustments to online samples. Their method involves using a parallel probability-based telephone survey with a small set of shared questions to estimate the propensity of an online respondent being a telephone respondent. The research record suggests that sometimes these adjustments work and other times they don't (Duffy et al, 2005; Lee, 2006).

Psychology has often been criticized by other social sciences as a discipline built on poor sampling practices, yet that discipline's habit of testing hypotheses via a series of test and retests with multiple samples from different populations may have applicability for online research. De Gaudemar (2007) describes the use of multiple panels to mediate the bias of a single panel, whether done as a single survey or as a series of individual surveys. "Measure twice, cut once," as the carpenters like to say. Of course, with significant rates of multipanel membership there is the strong possibility that some number of individuals will participate in the same survey via multiple panels. Techniques for deduping respondents across panels are only the early stages of development. The likelihood of multiple completions from a high incidence general population survey is probably quite low and perhaps negligible. However, on surveys of low incidence populations that may require multiple panels just to fill quotas the potential duplication could be high.

In sum, there are a number of different techniques that might be employed, although none provide the same level of reassurance we have been accustomed to with demographic weighting of probability samples. Nonetheless, recognizing the possible bias and taking steps to correct it is essential.

3. Understand that there are mode effects

The earlier discussion of probability and nonprobability sampling is one reason why online results may not match those from telephone or in-person interviews. Differences due to survey mode are another. Two types of mode effects are commonly reported: social desirability and differing use of scales.

Social desirability is the reluctance of respondents to reveal socially undesirable attitudes or behaviors to others (Groves, 1989). The tendency for survey respondents to present themselves in more positive terms to interviewers than in more anonymous, self-administered surveys was recognized long before there were online surveys (Hochstim, 1967). An example of this tendency is almost always seen in a standard question used in many US healthcare surveys. The question asks the respondent to report his or her current health across a five point scale from "excellent" to "poor." Online results typically show significant fewer respondents choosing "excellent" than do telephone surveys (Schonlau, 2004). Generally speaking, the more sensitive the question the more likely we will see significant social desirability bias.

While much of the research on social desirability bias has focused on especially sensitive topics such as drug use and abortion, there also is evidence that it can play a role in more mundane self-presentation issues. For example, in a telephone/Web mode comparison study Harris Interactive found significant differences across mode on questions about church attendance, charitable giving, personal hygiene, and frequency of exercising. In all instances, telephone respondents gave more social desirable responses than online respondents (Gamerman, 2008).

There also is considerable evidence that respondents use scales differently online than in interviewer-administered surveys. A frequent finding when transitioning customer satisfaction from telephone to online is a fall-off in satisfaction scores. Christian, Dillman, and Smyth (2008) analyzed results from a large number of mode comparison studies and concluded that "telephone respondents provide more positive ratings and are more

likely to use the extreme positive end point than are Web respondents.” These results seem to be relatively indifferent to the number of scale points or labeling used.

The point here is not to make a judgment about which methodology comes closer to the truth, or to invalidate specific methodologies for some kinds of research. Rather, it is to caution that online may produce results that are significantly different from interviewer-based methods and that those differences are due to the method of interview rather than sample bias or real differences in respondent attitudes. One may see higher reports of sensitive behaviors and fewer top-box endorsements on questions using scales. These results should be interpreted at least partially as mode effects, especially when comparing online results to those of previous studies done in other modes. If direct comparability is important, such as maintaining a time series, and mode effects are likely, online may not be the best choice.

4. Insist that online results be interpreted in the broader context of other sources of industry, business, and marketing information

Online results should be at least partially consistent with other sources of information on the same topic. These sources might include other surveys, focus groups, syndicated data, government databases, sales data or other kinds of industry information. Where there are differences, it is the job of the researcher to understand why and what that may tell us about the research questions.

The advice here is to practice what Smith and Fletcher (2004) have called the “holistic” method. The essence of this approach is to recognize that we “can no longer restrict (ourselves) to working with single, reasonably robust, sources of customer opinion.”

On the one hand, we live in a period when a variety of social and technological changes have made it increasingly difficult to do “good” research in the traditional sense. People have become increasingly more difficult to reach, their willingness to participate in research has declined dramatically, and good usable sample frames

are harder and harder to come by.

But we also live in an incredibly rich information environment with powerful tools to filter content down to that which is most pertinent to the business problem we are trying to solve. We now have available to us a significantly broader set of research tools capable of delivering all kinds of interesting insights, but each of these has their strengths and their weaknesses. The real value in research lies in understanding those strengths and weaknesses so that inconsistencies can be understood and findings synthesized into actionable insights. Online is an extremely valuable research tool, but it should never be the sole basis for an important business decision. Reconciliation of online results with other available research and market information is a must.

5. Work with research agencies that understand the relevant industry and business issues as well as online methodologies

As the foregoing discussion suggests, it is essential that the online results be interpreted in the broader context of the business problem in all its complexity. For an agency to meet that challenge it must understand the business dynamics the client company faces, the competitive landscape, its business strategy, its brand position, and the long-term trends within the industry.

With the online revolution has come a whole new breed of research supplier whose principle point of differentiation has been the ability to execute online studies quickly and cheaply. The Internet has made it possible for almost anyone to do market research without the traditional investment in infrastructure such as field staffs and telephone centers. These new entrants might be researchers, but too often they are technologists or simply entrepreneurs in the dot com mold.

Execution is important and analysis even more so, but *interpretation* is what gives research its value. The ultimate goal of every market research study is to help a client make a business decision by reducing the uncertainty around likely outcomes. Achieving this goal requires not just the technical skills of data

PART 1 / BUSINESS NEEDS

collection and analysis, but strong business acumen focused on the problems clients face in the context of their specific industry and markets. Of all of the considerations described above, this may be the most important.

CONCLUSION

It is unfortunate that in too many instances the arguments for online have been advanced by claims that traditional, probability-based methods are no longer viable. In the spirit of turnabout being fair play, the new breed of online community and Web 2.0 evangelists often sell their wares with full throated indictments of online panels. It's hard to see how any of this is helpful. In too many instances it only serves to confuse clients and devalue what we do.

The bad news is that all of our methods are imperfect. If there was once a time when a single well-designed study could tell us exactly what we needed to know and do so with certainty, that time has passed. The good news is that we have an increasingly broad number of methods to choose from. Each of these methods has its own strengths and weaknesses. Our challenge as researchers is to understand those strengths and weaknesses and then choose those from which we can generate the most useful insights into our client's business problems.

The online method and the access panels that are its foundation have made it possible for us to execute complex research designs more quickly and at a lower cost than we may have ever thought possible. But they are not the answer to every research question, and even when online may be the right choice, strong execution and interpretation informed by the broader context of available information and industry issues are essential.

References

Advertising Research Foundation (2007), "Foundations of Quality Initiative Launched by ARF Online Quality Research Council," www.thearf.org/assets/pr-2008-04-24a.

Baker, R. (2008). "A Web of Worries," *Research World*. June, pp. 8-11.

Baker, R. (2007). "Worse than Web? Another Look at Telephone Data Quality." Research Industry Summit on Respondent Cooperation. Chicago.

Bethlehem, J. and Stoop, I. (2007). "Online Panels – A Paradigm Theft?" in Trotman, M., Burrell, T., Gerrard, L., Anderton, K., Basi, G., Couper, M., Morris, K., Birks, D., Johnson, A. J., Baker, R., Rigg, M., Taylor, S., and Westlake, A. (Eds.), *The Challenges of a Changing World: Proceedings of the Fifth International Conference of the Association of Survey Computing*. Berkeley, UK: Association for Survey Computing.

Blyth, B. (2008). "Getting the right mix for quality," *Research World*. June, pp. 16-18.

Chakrapani, C. (2007). "On the validity of online panels," *Canadian Journal of Marketing Research*, 24, 1, pp. 7-15.

Christian, L. M., Dillman, D. A., and Smyth, J. D. (2008). "The Effects of Mode and Format on Answers to Scalar Questions in Telephone and Web Surveys," in Lepkowski, J.M., Tucker, C., Brick, M. J., de Leeuw, E. D., Japec, L., Lavrakas, P.I. J., Link, M. W., and Sangster, R. L. (Eds.), *Advances in Telephone Survey Methodology*, New York: Wiley.

de Gaudemar, O. (2006). "Benefits and challenges of multi-sourcing: Understanding differences between sample sources," Panel Research 2006: ESOMAR World Research Conference, Barcelona.

Duffy, B., Smith K., Terhanian, G., and Bremer, J. (2005). "Comparing data from online and face-to-face surveys," *International Journal of Market Research*, 47, 6, pp. 615-639.

ESOMAR (2008). "26 Questions to Help Research Buyers of Online Samples." <http://194.38.169.84/uploads/pdf/professional-standards/26questions.pdf>

ESOMAR (2007), *Global Market Research 2007*: ESOMAR Industry Report, Amsterdam: ESOMAR.

Gamerman, E. (2008). "When Voters Lie," *Wall Street Journal*, August 2.

Groves, R. M. (2007). "Nonresponse Rates and Nonresponse Bias in Household Surveys," *Public Opinion Quarterly*, 70, 5, pp. 646-675.

Groves, R. M. (1989). *Survey Errors and Survey Costs*, New York: Wiley.

Holbrook, A. L., Krosnick, J.A., and Pfent, A. (2008). "The Causes and Consequences of Response Rates in Surveys by the News Media and Government Contractor Survey Research Firms," in Lepkowski, J. M., Tucker, C., Brick, M. J., de Leeuw, E. D., Japec, L., Lavrakas, P. J., Link, M. W., and Sangster, R. L. (Eds.), *Advances in Telephone Survey Methodology*, New York: Wiley.

Inside Research (2008), "WW Online Spend Growth: Strong & Steady." March, p. 3.

ISO (2008). Access panels in market, opinion and social research -- Vocabulary and service requirements. ISO/FDIS 2632, Geneva: International Organization for Standardization.

Keeter, S., Kennedy, C., Dimock, M., Best, J., and Craighill, P., "Gauging the Impact of Growing Nonresponse on Estimates from a National RDD Telephone Survey," *Public Opinion Quarterly*, 70, 5, pp. 759-779.

Kellner, Peter (2007). "Down with random samples," *Research World*, May, p. 31.

Lee, S. (2006). "Propensity Score Adjustment as a Weighting Scheme for Volunteer Panel Web Surveys," *Journal of Official Statistics*, 22,2, pp. 329-350.

Miller, J. (2000). "Going Online with Consumer Research: the General Mills Experience." 2000 EXPLOR Forum: Continuing Progress in Online Research, Chicago.

Miniwatts Marketing Group (2008). <http://www.internetworldstats.com/stats4.htm#europe>.

Schonlau, M., Zapert, K., Payne Simon, L., Sanstad, K., Marcus, S., Adams, J., Spranca, M., Kan, H-J., Turner, R., and Berry, S. (2004), "A Comparison Between a Propensity Weighted Web Survey And An Identical RDD Survey," *Social Science Computer Review*, 22, pp. 128-138.

Smith, D.V.L. and Fletcher, J.H. (2004). *The Art & Science of Interpreting Market Research Evidence*. Chichester, England: Wiley.

Taylor, Humphrey (2007). "The Case for Publishing (Some) Online Polls," *The Polling Report*, January. Washington, DC: *The Polling Report*, Inc.

Terhanian, G., Smith, R., Bremer, J., and Thomas, R.K. (2001). "Exploiting Analytical Advances: Minimizing the Biases Associated with Non-Random Surveys," in *Proceedings from ESOMAR/ARF Worldwide Audience Measurement Conference*, pp. 247-272.

The Author

Reg Baker is Chief Operating Officer, Market Strategies International, United States.