

Choosing Between Telephone and Online for Survey Data Collection

By Reg Baker, Ph.D, President and Chief Operating Officer, Market Strategies International

August, 2010

Introduction

The growth of online survey research over about the last 15 years has been truly remarkable. While estimates vary, most agree that online now accounts for somewhere between one third and one half of the total volume of survey data collection being done by US market research firms (See, for example, ESOMAR, 2009). Key to this growth has been the positioning of online as a faster and less expensive method capable of delivering the same or better survey results as traditional methods (telephone, mail, and in-person). Indeed, online often is the least expensive and fastest survey method for studying a wide range of business problems, although this speed advantage is clearest when samples sizes are large and/or expected incidence is low. It is less certain that the survey results achieved are comparable to other methods, especially in those cases where one of the objectives is to accurately measure some characteristic, attitude, or behavior in a target population, whether it be a customer base, individuals with certain characteristics, or the general population.

This issue of accuracy is the main focus in this paper. In it we offer a framework for evaluating an online design versus a telephone design when a central goal of the research is to produce an accurate estimate of some personal characteristic, attitude or behavior in a target population. In this context, accuracy means a survey result that is as close as possible to the “true value” in that population.

We recognize that not all research has as its goal the development of accurate estimates of population values.

For example, in some instances directional measures may be sufficient. In others we might only want to test hypotheses about the interrelationships among characteristics, attitudes, and buying behavior. Under these and similar circumstances we might well use different criteria than described below to select a data collection method.

The Case Against Telephone

The most frequently-heard argument against telephone research has been that the long-term decline in response rates has undermined the accuracy of it as a research method. How can we expect to have accurate and unbiased results when the vast majority of people in our sample are never reached or, if reached, refuse to participate? How can we trust our results when we have response rates of 10 percent or less and achieved samples that have too many older people and nowhere near enough younger people? In the language of surveys, the difference between a survey result or estimate and the true value in the target population caused by the failure to interview everyone in the sample is called *nonresponse error*.¹ Our results may be biased because those we did not interview hold different opinions than those we did interview.

More recently, we have seen significant concerns emerge about the breakdown of the telephone sampling frame as more and more Americans give up their landline telephones in favor of a cell phone.² One of the historic advantages of telephone has been the existence of a sampling frame that included the entire US

¹ For a thorough discussion of the types of errors that can affect survey results see Groves (1989).

² Some also have expressed concern about the steady migration to VOIP but those numbers generally are part of the standard landline frame, although they often are not designated as being VOIP.

household population, except for the roughly four to five percent without any telephone service. The existence of this frame meant that we could draw fully representative samples of the general population. But now, according to the most recent data reported by the National Health Interview Survey, the government-sponsored in-person survey that has become the gold standard for measuring cord cutting, more than one in five US homes (22.7 percent) has only a cell phone (Blumberg and Luke, 2010).³ In the language of surveys, when our survey results do not match the true values in the population because the sampling frame used to draw the sample did not include the entire population we have *coverage error*. In such instances, results may be biased because the people not included in our sampling frame hold differing opinions that those who are included.

The Telephone Defense

Advocates for telephone research argue that while high response rates generally are preferred over low response rates the actual amount of error in surveys with low response rates is quite small. Often it can be effectively addressed by simple demographic weighting. For example, a study conducted by the Pew Research Center investigated the impact of nonresponse on survey results for a broad set of demographic, attitudinal, and behavioral items, many of which could be verified against external benchmarks such as the Census, party registrations, or election results (Keeter, Kennedy, Dimock, Best, and Craighill, 2007). The researchers fielded the same survey to two samples. They worked one sample in their standard way and the other more aggressively, for example, by placing significantly more calls. The standard survey achieved a response rate of 25 percent and the more rigorous survey produced a 50 percent response rate. They then compared the results from the two surveys for 84 different items, finding just seven that were significantly different. The mean difference in these seven items was five percentage points. In other

words, the lower response rate survey produced essentially the same results for the vast majority of items as the survey with the higher response rate.

We conducted a similar study at Market Strategies with a large telephone healthcare survey. In our case we achieved a response rate of just six percent on the standard survey and 40 percent with the more rigorous approach (Baker, 2007). We found virtually no differences in demographics, self-reports of health status, frequency of exercise, or absence of health insurance.

It often is difficult to estimate just how much the results of any given survey are likely to be affected by nonresponse. Knowing something about how nonresponders differ from responders can be very helpful because it can tell us something about how our results might change if the former participated in the survey. In the case of RDD surveys it generally is difficult to know much more than demographics but in other kinds of surveys, such as those that use samples from a company's list of customers, the company may know a great deal about nonresponders that can help us understand how representative our results are, how nonresponse may be affecting results, and how we might correct any bias.

The second major concern about the accuracy of telephone surveys is the coverage error that might result due to the increasing number of households that no longer have a landline and therefore are not included in the standard telephone sampling frame. Fortunately, it is possible to acquire representative samples of cell phones that include cell phone users with and without landlines.⁴ By putting the landline and cell phone samples together coverage error is significantly reduced to the point where it no longer is a serious threat to representativeness or survey accuracy. Further, the inclusion of cell phones in the survey sample generally results in a higher proportion of younger adults completing the survey, a much-needed correction given the demographic bias of landline-only samples.

³ This same survey also reports that an additional 14.7 percent of homes have a landline but people living there claim that they take most of their calls on a cell phone. This also could contribute to lower response rates and potentially significant nonresponse error. However, recent research has shown that a significant proportion of these "cell mostly" households still respond to surveys on their landlines (Boyle, Lewis, and Tefft, 2009).

⁴ In the case of studies that use lists of customers, cell phones are often included as the main contact number and those numbers can be included in the sample to be called. This calling can be somewhat less efficient due to regulatory restrictions against the use of automated dialers to call cell phones, but the resulting costs generally are not significant.

To conclude, the main source of concern about the accuracy of telephone surveys has been nonresponse error, that is, the failure to interview everyone in the sample. While the recent and significant migration of US households from landlines to cell phone has caused considerable concern about coverage error, there are remedies, albeit at some additional cost.

The Case Against Online

Whereas the case against telephone has focused mainly on nonresponse error the case against online has been mostly about coverage error. Estimates of Internet use and penetration in US households tend to vary widely. Arguably the most accurate are those collected by the Current Population Survey (CPS), a government in-person survey noted for its accuracy. The CPS reports that 68.7 percent of US households had an Internet connection while 76.7 percent had household members who connected to the Internet from home or some other location such as their workplace (Current Population Survey, 2009). This compares favorably with the most recent data from the Pew Research Center showing that, as of December of 2009, 74 percent of US adults use the Internet either at home or some other location (Rainie, 2010). Thus, in round numbers, one in four US adults cannot be reached by Internet. Further, there continues to be a well-documented demographic bias by age, education, income, race, and ethnicity when comparing those who use the Internet with those who do not.

This roughly 25 percent coverage error is made worse by the lack of a sample frame composed of all Internet users from which we might draw samples. Without such a frame a number of companies have built and now maintain large online panels of Internet users who have volunteered to do surveys. These companies use a variety of means to enlist volunteers with one of the most popular being the posting of banner ads on Web sites across the Internet.⁵ The resulting panels, often comprised of several million members,

are then used like sampling frames to draw samples for specific surveys. Unfortunately, even the largest panels include two percent or less of US adults and we know very little about the kinds of people who volunteer, especially how they may differ behaviorally and attitudinally from those who do not volunteer, and how those differences might affect survey results. Put another way, we might say that the coverage error in a sample drawn from an online panel may be 98 percent or greater and likely has substantial bias due to the way in which the panel was built.

The response rate perspective is no better. To compute a response rate for a survey using an online panel that is comparable to the response rate commonly used for a telephone survey with a probability-based sample we must factor in the response rate in the original panel recruitment stage as well as at the specific survey stage (Callegaro and Disogra, 2008). While panel companies seldom disclose the response rates to their recruitment campaigns most observers agree that they hover around 1 percent. If we assume that the response rate to a specific survey using that panel is likely to be 10 percent or less than the true response rate, comparable to what we would recruit for a telephone surveys, is 10 percent times 1 percent or about .1 percent.

The final disadvantage of online is that standard measures of sampling error (also known as margin of error) cannot be calculated. Because the panel itself was not selected randomly but by people volunteering to join there are unknown and unmeasured biases that do not exist when there is random selection from a sample frame that contains the full population of interest. In the case of panels, the connection back to the target population is lost and it therefore is impossible to make mathematically sound statements about the relationship of estimates from an online survey to the true values in the target population.⁶

To conclude, the primary source of concern about online is the coverage error created by the dual forces of lack

⁵ A handful of panels have been recruited using traditional probability-based methods (generally telephone) but they tend to be significantly smaller and somewhat more expensive than the large volunteer panels. Their smaller size makes it difficult to use them for studies focused on lower incidence populations or smaller geographic areas.

⁶ For further discussion see, "Opt-in Surveys and Margin of Error" at <http://www.aapor.org/Content/NavigationMenu/PollampSurveyFAQs/OptInSurveysandMarginofError/default.htm>

of access to the Internet by the full US population and the volunteer or self-selection method by which online panels are built. For those who worry about the decline of response rates in telephone surveys we also note that the true response rates for most Internet panel surveys are a fraction of those for telephone surveys.

The Online Defense

Advocates for online frequently argue that while online panels are not themselves representative their inherent bias can be corrected either at the time a sample is drawn or during post-survey stratification. One common approach is *purposive sampling* in which knowledge about the panel members is used to draw samples with characteristics that match those in the target population (Kellner, 2008). The underlying assumption is that if these characteristics (typically key demographics such as gender, age, race, etc.) match the target population then all other survey variables will fall in line. However, this approach generally fails to acknowledge that the biases in online panels are much more complex than demographics and are rooted in the behavioral and attitudinal differences that cause some people to adopt the Internet while others do not and some to join panels to do surveys online while others do not.

A second common approach is post survey adjustment by standard demographic weighting. While this can sometimes produce results that mirror the target population in demographic distributions it does not address the likely attitudinal bias rooted in how the panel was initially recruited. There is a more complex class of adjustments, generally referred to as *propensity weighting*, that attempts to correct for behavioral and attitudinal bias by including measures of these in the weighting procedure. Rigorous tests of propensity weighting procedures have generally found that sometimes they work and other times they do not (see, for example, Lee and Valliant, 2009).

To conclude, the primary concern with the accuracy of online surveys using commercial Web panels is coverage error and the resulting bias due to less than universal Internet access and non-systematic way in which volunteers are

solicited. While there are methods proposed for correcting these biases those methods generally do not effectively address the fundamental source of that bias, i.e., the behavioral and attitudinal differences between Internet users and non users on the one hand and those who volunteer to do surveys versus those who do not on the other.

Telephone and Online Results Compared

There have been literally dozens of studies over the last several years comparing results for telephone and online and virtually all of them have shown different results across the two methods. Some of these studies have included comparisons to benchmark data established by non-survey means such as the Census, industry data, or publicly available administrative data. Some, have looked at not just one but several online panels as points of comparison. In addition to differences across methods, these studies also have found significant variation among online panels. For example, Krosnick, Nie, and Rivers (2005) found that while a single telephone RDD sample was off an average of 4.5 percent from benchmarks, six different online panels were off an average of five percent to 12 percent, depending on the sample supplier. In an extension of this same research, Yeager et al. (2009) found that an RDD telephone survey was consistently more accurate than a series of online panel surveys even after post-stratification by demographics. Results from a much larger study by the Advertising Research Foundation (Walker, Pettit, and Rubinson, 2009) using 17 panels has also documented even greater divergence from benchmarks as well as “a wide variance particularly on attitudinal and/or opinion questions (purchase intent, concept reaction, and the like).” Finally, a recently-completed review of the empirical literature on online panels undertaken by a task force assembled by the American Association for Public Opinion Research, the premier professional association for US survey researchers, found that “empirical research to date comparing the accuracy of surveys using non-probability online panels with that of probability-based methods finds that the former are generally less accurate when compared to benchmark data from the Census or administrative records” (AAPOR, 2010).

Conclusion

Selecting an appropriate survey method for understanding markets and customers has become a challenging and complex task. The traditional method of telephone research has come under substantial fire, first out of concern for the potential bias created by alarmingly high levels of nonresponse and more recently by concerns about coverage error due to the increasing abandonment of landline telephones. The latter problem is largely manageable, but the former continues to be cause for concern. Online data collection using samples from panels of volunteers has been widely proposed as an alternate approach. While these surveys are almost always cheaper and often faster than telephone, the combination of extensive coverage error and high levels of nonresponse make them subject to substantial bias. Research has yet to demonstrate the consistent effectiveness of various methods that might correct this bias. Head-to-head comparisons of telephone and online have generally found telephone to be more accurate, and comparisons of results across panels have shown wide variability.

The evidence to date clearly suggests that if the goal of the research is to accurately estimate the true values of survey variables in the target population then a well designed telephone survey fielded to a sample from a frame with full or near full coverage of the target population (e.g., landlines plus cell phones or a list of customers) is the method of choice. However, not all surveys have the goal of accurately measuring the prevalence of a behavior or attitude in the population. Sometimes we are interested in how personal characteristics might drive product preference or how various attitudes might interact to create openness to different advertising messages. In these and similar instances, online may be a good choice, especially given its generally lower cost, faster turnaround, and the flexibility offered by an interactive survey platform like the Web.

There is no one-size-fits-all approach to data collection method. Making good choices requires careful consideration of the bias involved in competing methods and

some assessment of the degree to which we might expect that bias to interact with the survey topic (For additional discussion, see Groves, 2006; Chakrapani, 2007). Like the design of the survey itself, knowing the right questions to ask is key.

References

- Baker, R. (2007). "Worse than Web? Another Look at Telephone Data Quality." Presented at the Research Industry Summit on Data Quality. Chicago.
- Boyle, J., Lewis, F., and Tefft, B. (2009). "Cell Phone Mainly Households: Coverage and Reach for Telephone Surveys Using RDD Landline Samples." *Survey Practice*. December.
- AAPOR (2009). AAPOR Report on Online Panels. Downloaded from <http://www.aapor.org/AM/Template.cfm?Section=Left&Template=/CM/ContentDisplay.cfm&ContentID=2553>
- Blumberg, S. and Luke, J.V. (2010). "Wireless Substitution: Early Release of Estimates from the National Health Interview Survey, January - June 2009." <http://www.cdc.gov/nchs/data/nhis/earlyrelease/wireless200912.htm>
- Callegaro, M., & Disogra, C., (2008). "Computing Response Metrics for Online Panels." *Public Opinion Quarterly*, 72(5): 1008-32.
- Chakrapani, C. (2007). "On the Validity of Online Panels," *Canadian Journal of Marketing Research*. 24, 1, 7-15.
- Current Population Survey (2009). Downloaded from <http://www.census.gov/population/www/socdemo/computer/2009.html>
- ESOMAR (2009). *Global Market Research 2009*. Amsterdam.
- Groves, R.M. (2006). "Nonresponse Rates and Nonresponse Bias in Household Surveys," *Public Opinion Quarterly*. 70, 5, 646-675.
- Groves, R. M. (1989). *Survey Errors and Survey Costs*. New York: John Wiley & Sons.

- Keeter, S., Kennedy, C., Dimock, M., Best, J., and Craighill, P. (2007). "Gauging the Impact of Growing Nonresponse on Estimates from a National RDD Telephone Survey," *Public Opinion Quarterly*, 70,5, 759-779.
- Kellner, P. (2008). "Down with random samples." *Research World*, June, 31.
- Krosnick, J. A., Nie, N., and Rivers, D. (2005). "Web Survey Methodologies: A Comparison of Survey." Paper presented at the 60th Annual Conference of the American Association for Public Opinion Research in Miami Beach, Florida.
- Miller, J. (2008). "Burke Panel Quality R and D." Cincinnati: Burke, Inc.
- Lee, S. and Valliant, R. (2009). "Estimation for Volunteer Panel Web Surveys Using Propensity Score Adjustment and Calibration Adjustment." *Sociological Methods and Research*, 37: 319-343.
- Rainie, L. (2010). "Internet, Broadband, and Cell Phone Statistics," Pew Internet and American Life Project. Pew Research Center.
- Walker, R., Pettit, R., and Rubinson, J. (2009). "The Foundations of Online Research Quality - Executive Summary 3: Inter Study Comparability and Benchmark Analysis." New York: Advertising Research Foundation.