

Establishing Confidence About Data and Results: A Comment on Sample Size and Response Rates

June 20, 2014

Note: This editorial was written by Howard J. Shaffer, PhD, CAS, Editor-in-Chief for The BASIS.

For many years, I have had the privilege of being a gatekeeper of science. This responsibility is stimulating, challenging, and demanding. During my tenure as a journal editor, reviewer, and editorial board member, I repeatedly observed and evaluated three important features of social science research: sample size, response rate, and retention rate. These study characteristics have established or diminished the value of many research projects. The hopes of many scientists have been dashed by poor samples and low response or retention rates.

Gatekeepers of science (e.g., journal editors) have the responsibility of determining whether they have sufficient confidence in the data, results, and reporting of research to allow the publication of research projects - which holds the potential to extend the body of scientific knowledge. Interpreting the value of scientific research requires some judgment about the integrity of data. The process that leads to the establishment of such value is just one reason why peer-reviewed research is held in high esteem compared to grey literature (i.e., not peer reviewed). Grey literature might make a contribution to science, but it fails to bring the imprimatur of scientific standards. Many a report released to the public would not survive peer review. This circumstance can challenge the scientific literacy of the public and lead to confusion about what we know. The public needs standards to help them distinguish important science from pseudoscience. Peer review certainly isn't a perfect process, but it does represent the best standard currently available. Because the issues of scientific literacy and research design extend beyond the scope of this comment, in this article, I will limit my remarks regarding ongoing concerns about scientific research to a brief discussion about sample size, response rates, and retention rates.

Statistical power for identifying differences between groups is a function of sample size and effect size (e.g., robustness of a treatment). In general, to assure statistical power, during the planning phase of research, investigators determine

their sample size need by identifying the smallest groups that they might want to compare (e.g., comparing two groups from two different jurisdictions). Researchers need to be sure that there is sufficient statistical power (e.g., 0.8) to make this comparison. Assuming a consistent effect size, if there is a sufficient sample size for this small group analysis, then all of the other comparisons of larger groups will have the same or more statistical power. To illustrate, imagine a study with 1000 people and a disorder with 0.5-1.5% prevalence. We can expect to identify 5-15 new cases. This group is too small to yield sufficient statistical power for investigators to distinguish it from other groups of equal or smaller size. In other words, this sample is too small to permit comparisons across regions or events. Consequently, researchers need to consider the problems with sampling low base rate disorders and choose a research design that can yield higher numbers either through sample size or other methodological features.

The next issue reflects response rate. The question is whether a sample, regardless of size, is representative of the population from which it was drawn and which it is meant to represent. Response rate (Frankel, 1982) is one measure of confidence that the sample represents both the population (e.g., community) from which it was drawn and the target group (e.g., disordered gamblers) that segments of this sample should represent. Non-response rates have been increasing for several decades and hold the potential to bias survey results (Groves, 2006; Groves & Couper, 1998; Johnson & Owens, 2003). In fact, some would argue that increasing non-response rates represent a serious crisis for the field of survey research (Johnson & Owens, 2003). Complicating this matter, "A non-response rate of 25%, although a good achievement in many settings, can seriously distort the observed prevalence of a disease when the disease itself is a cause of non-response" (Hulley, Gove, Browner, & Cummings, 1988, p. 27). To illustrate, because they commonly experience financial difficulties, community surveys based on telephone service might not identify a representative sample of disordered gamblers. Consequently, address based sampling (ABS) has emerged as a better methodology to random number dialing.

Further complicating the issue of response rates, a survey of 18 scholarly scientific publications revealed, "None of the journals reported having an established minimal response rate standard. One editor, however, did report that despite the absence of a formal policy, the journal did expect 'at least a 60% response rate with rare exceptions'" (Johnson & Owens, 2003, p. 129). "The editor of another journal agreed, adding that 'in most instances, 20% is too low,

and 80% is a de facto standard, but there is a considerable gray area” (Johnson & Owens, 2003, p. 130). Citing Babbie (2007, p. 262), Goves (2006) notes, “A review of the published social research literature suggests that a response rate of at least 50 percent is considered adequate for analysis and reporting. A response of 60 percent is good; a response rate of 70 percent is very good.” Finally, Singleton and Straits (2005, p. 145) note, “... it is very important to pay attention to response rates. For interview surveys, a response rate of 85 percent is minimally adequate; below 70 percent there is a serious chance of bias.”

On a slightly different but related topic, there is a tendency for repeated observation or longitudinal studies to evidence reducing retention rates over time. However, there is evidence that, with the proper attention to detail and comprehensive tracking strategies, very high rates of retention (e.g., 96.6%) can be obtained with very difficult target populations (Cottler, Compton, Ben-Abdallah, Horne, & Claverie, 1996).

As you can see, target response rates reflect a variety of values and result from many methodological influences. As a guide, researchers typically seek response rates of at least 70% to feel confident that their sample is representative of the community. Alternatively, response rates less than 50% are intuitively unacceptable because with a rate this low more people opted out of the research than opted in. There simply is no way to feel confident that a study with such a low response rate reflects an unbiased look at the community. In addition to raising important questions about the general population, we also cannot be confident that the respondents accurately represent the target segment of the population (e.g., disordered gamblers). Trying to solve the sampling problems associated with representativeness, researchers can statistically “correct” the data by weighting it after it has been collected. Although scientists have used this strategy, ultimately it is not as suitable a solution as using better sampling strategies. Post hoc data weighting simply emphasizes the data collected from outliers, thereby permitting the possibility that anomalous data is magnified.

Interestingly, response and retention rates reflect the popular adage that “you can pay us now or you can pay us later.” Research shows that scientists can increase follow-up retention rates by repeatedly contacting potential respondents (e.g., Kleschinsky, Bosworth, Nelson, Walsh, & Shaffer, 2009), or converting those who have declined to participate by providing sufficient incentives. Of course, participant incentives also hold the potential to bias the sample because the

incentive is more meaningful to some types of potential participants than others. Kessler and his colleagues have provided us with an excellent review of the array of procedural strategies and tactics that they employed to produce a response rate greater than 70% with a complex and difficult national sample (i.e., the National Comorbidity Study Replication; Kessler et al., 2004); in addition to establishing acceptable response rates, many of these methods also can improve retention rates within longitudinal studies. For example, Kessler et al. (2004) observed that survey construction features, interviewer supervision, quality control, training and retraining were study procedures that helped to avoid a poor response rate. Similarly, these strategies can help to improve longitudinal retention rates.

Recognizing the trend toward lower response rates and higher non-response rates, some researchers have capitulated to the problem by suggesting that low response rates are just the state of contemporary research - and they are willing to accept the status quo. There is little doubt that response rates gradually have been declining for social science research. However, there is no acceptable justification for this state of affairs. Rather than accepting the declines as inevitable, researchers need to recognize this circumstance as a crisis that demands the development of new methods and strategies. Insufficient sample sizes are insufficient. Low response rates and the consequent small sample sizes are unacceptable because these do not permit investigators to be confident about their samples. Insufficient response rates lead to small groups, leaving researchers unable to compare group differences. Many researchers simply do not build the features necessary to obtain adequate response rates into their studies because of insufficient funding, impatience, lack of methodological skill, or a combination of these factors. Absent sufficient response rates, interpreting results will remain uncertain and research dollars potentially will be squandered. I don't want to suggest that we have all the answers. Although researchers have some tools to obtain sufficient response rates, it is time for scientists to recognize and reconsider the problems associated with response and retention rates. It is time to avoid lowering our scientific standards simply because it is easier than developing new methods. Now is the time to develop innovative methods and raise our standards to a level that inspires confidence in our data and the interpretation of our results.

In closing, it is not unusual to find that some researchers recognize the value of response rates, but instead of applying innovative and creative solutions to

improve their response rates, they rate-hack. That is, they report awkwardly calculated and inflated response rates. Rate-hacking occurs when researchers engage in mathematical calisthenics to reduce the denominator (i.e., participants / eligible participants) of those who were eligible to participate in their study - thus inflating and obscuring the actual response rate. As a result of this kind of methodological sleight of hand, CASRO has published ethical guidelines for survey research and for calculating response rates ((CASRO), 2008; Frankel, 1982). Ultimately, the responsibility for designing, implementing, collecting, analyzing, and reporting the data and results associated with scientific research remains a scientific challenge; however, it also remains a challenge for the integrity of scientists.

- Howard J. Shaffer, Ph.D., C.A.S.

What do you think? Please use the comment link below to provide feedback on this article.

References

(CASRO), C. o. A. S. R. O. (2008). Code of Standards and Ethics for Survey Research: Council of American Survey Research Organizations.

Babbie, E. R. (2007). The practice of social research (11th ed.). Belmont, CA: Thomson Wadsworth.

Cottler, L. B., Compton, W. M., Ben-Abdallah, A., Horne, M., & Claverie, D. (1996). Achieving a 96.6 percent follow-up rate in a longitudinal study of drug abusers. *Drug & Alcohol Dependence*, 41(3), 209-217.

Frankel, J. R. (1982). On the definition of response rates: A special report of the CASRO task force on completion rates. Port Jefferson, NY: The Council of American Survey Research Organizations.

Groves, R. M. (2006). Nonresponse rates and nonresponse bias in household surveys. *Public Opinion Quarterly*, 70(5), 646-675.

Groves, R. M., & Couper, M. P. (1998). *Nonresponse in Household Interview Surveys*. New York: John Wiley & Sons, Inc.

Hulley, S. B., Gove, S., Browner, W. S., & Cummings, S. R. (1988). Choosing the

study subjects: Specification and sampling. In S. B. Hulley & S. R. Cummings (Eds.), *Designing clinical research: an epidemiologic approach* (pp. 247). Baltimore: Williams & Wilkins.

Johnson, T., & Owens, L. (2003). Survey response rate reporting in the professional literature. *American Association for Public Opinion Research*, 127-133.

Kessler, R. C., Berglund, P., Chiu, W. T., Demler, O., Heeringa, S., Hiripi, E., Jin, R., Pennell, B., Walters, E. E., & Zaslavsky, A. (2004). The US National Comorbidity Survey Replication (NCS-R): Design and field procedures. *International Journal of Methods in Psychiatric Research*, 13(2), 69-92.

Kleschinsky, J. H., Bosworth, L. B., Nelson, S. E., Walsh, E. K., & Shaffer, H. J. (2009). Persistence pays off: Follow-up methods for difficult-to-track longitudinal samples. *Journal of Studies on Alcohol and Drugs*, 70(5), 751-761.

Singleton, R., & Straits, B. C. (2005). *Approaches to social research* (4th ed.). New York: Oxford University Press.