Schonlau M. Will web surveys ever become part of mainstream research? Journal of Medical Internet Research. 2004; 6(3):e31. Available at http://www.jmir.org/2004/3/e31/.

Final version differs slightly from the pre-publication version in this document.

Guest Editorial

Will web surveys ever become part of mainstream research?

Matthias Schonlau , Ph.D.

Corresponding Author:
Matthias Schonlau
RAND
1700 Main Street
Santa Monica, CA 90407
email: matt@rand.org

This journal issue contains two interesting papers on web survey methodology that reach different conclusions about the potential use of web surveys, in particular which of two modes achieves the higher response rate. High response rates are commonly seen as an indicator for the validity of surveys. Leece et al. [1] used systematic sampling to assign half of a list of orthopedic surgeons to a web survey and the other half to a mail survey. They concluded that the web survey had a significantly lower response rate than the mail survey, and warn "Researchers should not assume that the widespread availability and potential ease of Internet-based surveys will translate into higher response rates". In contrast, Ritter et al [2] recruited participants from the Internet and randomly assigned them to a mail or a web survey and came to a different verdict. They compared the responses on 16 health related questions/instruments and find that none of 16 instruments were significantly different among the two study arms. Ritter et al. [2] found that among those assigned to the web survey participation was at least as good if not better than participation among those assigned mailed questionnaires.

The different findings can be explained by the respective recruiting strategies. Ritter et al recruited participants over the Internet. Clearly, respondents recruited on the web are more likely to respond to a web survey than the general population. The finding is nonetheless interesting because it is not obvious that the response rate to a web survey would be higher than the one to a mail survey even among Internet savvy respondents. A web survey typically achieves a higher response rate when respondents are contacted by email rather than by mail [3]. Analogously, a mail survey typically achieves a higher response rate when respondents are contacted by mail rather than by email. It is possible that recruiting respondents on the web also reduces the response rate of a mail survey because the recruiting mode is different from the response mode.

Both Ritter et al. [2] and Leece et al [1] survey special rather than general populations. Ritter et al [2] recruits respondents from the Internet. Leece et al [1] have a master list of orthopedic surgeons and they have email addresses for 79% (all but 45 respondents) of the respondents in the web survey arm. A much greater challenge would be to conduct a web survey of a general population for which no master list of email address is readily available. One approach, contacting respondents by mail and encouraging response by web with a mail fallback option, is discussed in Schonlau et al. [4]. This approach is not very practical because the second response mode requires additional resources and slows the survey down.

Ritter et al.'s [2] and most web surveys are conducted with convenience samples rather than with random samples. A convenience sample is a sample where the participants are selected, in part or in whole, at the convenience of the researcher. In contrast, a random sample is one where the researcher insures that each member of that population has a known probability (for example, equal probability) of being selected. For example, a sample of respondents recruited from newsgroup postings is a convenience sample for most populations of interest. Eysenbach and Wyatt note "In `open' web-based surveys, selection bias occurs […] through self-selection of participants." [5]. Such selection bias implies a convenience sample because the probability of selection is unknown.

Whether or not web surveys will some day be part of the mainstream in survey research hinges on the question whether it is possible to draw inferences from convenience samples. Conventional survey sampling wisdom states that one cannot draw inferences from convenience samples and that convenience samples are therefore useless, except perhaps for pilot studies. Still, convenience samples can be used to conduct experiments within that sample, and Ritter et al [2] gave a nice example that shows how convenience samples can be effectively used for this purpose. They conducted a properly randomized experiment; whether or not the underlying sample is representative is a secondary consideration. Of course, one limitation is that the findings in Ritter et al [2] do not hold for people who do not have access to the Internet. One can think of other experiments that can be conducted with a single convenience sample, including experiments to test response order effects (in visual response modes the first answer choice tends to be chosen more often) and anchoring effects (the answer choice may be affected by the context, including what was asked in previous questions). It may also be possible to use vignettes and factorial experiments in web surveys based on convenience samples – however more research is needed to support this claim.

The mere suggestion that one day it may be possible to draw inferences based on a convenience sample is provocative to most survey researchers. Skepticism is often useful in curbing overly enthusiastic claims.

On the other hand, in health services research and biostatistics researchers have long drawn conclusions from observational studies. The purpose of the notorious "Table 1" in epidemiological cohort studies, in which demographical and other information on both cases and controls is commonly displayed, is to argue that (hopefully) exposed and non-exposed groups are not really different with respect to important potentially confounding

variables (such as for example age, education, etc), and that therefore any observed risk differences between the groups are indeed due to the exposure (to a risk factor or intervention) and not other observed confounding factors such as demographic differences. If a study is randomized the covariates should be "automatically" balanced by design. For example, it is unlikely that participants in the exposed (intervention or treatment) group are significantly older than in the non-exposed (control) group. If a study is not randomized such systematic differences are likely to occur due to selection bias. If in a non-randomized study one can show that the covariates are balanced, the argument goes, then there is little reason to distrust regression results or other inference based on observational data.

Rubin's framework for causal inference [6] has gone a step further: rather than hoping that the covariates in "Table 1" are balanced Rubin's approach ensures that they are balanced. This is accomplished by constructing propensity scores from logistic regression on baseline variables that are thought to capture the difference between web respondents and the general population. The propensity scores can then be used, for example, to construct subclasses in which covariates are approximately balanced. There is one very important assumption: there are no important unobserved variables that affect treatment assignment. Rubin's approach is universally accepted and frequently used in biostatistics.

Harris Interactive, a company that commercially conducts web surveys, has adapted Rubin's approach for drawing inferences using web surveys. Assignment to treatment or control corresponds to "assignment" of a respondent to a random or a convenience sample. If it is possible to capture the selection mechanism that distinguishes a random sample from the convenience sample then one can adjust for it. The Harris Interactive approach is described in more detail in Schonlau et al. [7]. The approach of Harris Interactive is theoretically sound. In practice, the challenge is to ask the right questions to capture the difference between the online and offline populations. I am involved in a study that explores whether it is feasible to move a portion of the Health and Retirement Survey (HRS), a large-scale U.S. panel survey, on the Internet in future survey waves. I have recently applied the propensity scoring approach to the HRS. This is work in progress, but early results are encouraging [8].

Will mainstream research ever draw inferences based on convenience samples? This would be a breakthrough, and I believe it will eventually be possible. There was a time when researchers did not believe that it was possible to draw inferences from mail surveys because they were self-administered. These days mail surveys are certainly considered "mainstream". I consider the possibility of inference based on convenience samples to be a very important goal in web survey research. But it is by no means the only one: Leece et al [1] and Ritter et al [2] successfully demonstrate that this research field is much larger.

References

[1] Leece P , Bhandari M , Sprague S, Swiontkowski M, Schemitsch E, Tornetta P,. Devereaux PJ, Guyatt G. Getting Surgeons to Respond to Surveys: J Med Internet Res 2004. 6(3): e###

[2] Ritter P , Lorig K, Laurent D, Matthews K A Comparison of Internet Versus Mailed Questionnaires: J Med Internet Res 2004. 6(3): e###

[3] Schonlau, M., Fricker R., Elliott, M. *Conducting Research Surveys via Email and the Web*, Santa Monica, CA: RAND. 2002.

[4] Schonlau M, Asch BJ, Du C. Web surveys as part of a mixed mode strategy for populations that cannot be contacted by e-mail. *Social Science Computer Review.* 2003; 21(2): 218-222.

[5] Eysenbach G, Wyatt J. Using the Internet for Surveys and Health Research. Journal of Medical Internet Research 2002;4(2):e13 <URL:http://www.jmir.org/2002/2/e13/>

[6] Rosenbaum, PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. *Biometrika.* 1983;70:41-55.

[7] Schonlau M, Zapert K, Payne Simon L, Sanstad K, Marcus S, Adams J, Spranca M, Kan H-J, Turner R, Berry S. A comparison between a propensity weighted web survey and an identical RDD survey. *Social Science Computer Review.* 2004; 22(1):128-138.

[8] Schonlau M., Van Soest A, Couper M, Winter J, Kapteyn . Attempting to adjust for selection bias in Web surveys with propensity scores: the case of the Health and Retirement Survey (HRS). In *Proceedings of the Section on Survey Statistics*, American Statistical Association; 2004, CD-ROM.(also available at www.schonlau.net)