Mixed Mode Data Collection Using Paper and Web Questionnaires

Peter Werner, Statistics Sweden
Gösta Forsman, Swedish Road Administration

1. Introduction

Web surveys are becoming increasingly popular, perhaps mostly because of advantageous cost factors. The reverse of the medal is that Web access is far from universal, and hence Web surveys only permit inference to populations of Web users. However, the Web can be used in combination with traditional response modes in a mixed-mode approach, in which the respondents are given the opportunity to choose between a Web questionnaire and, for example, an ordinary paper questionnaire. Such an approach enables data collection from a *probability sample of a general population*, while still aiming at reducing costs by utilizing the Web.

There are many conceivable ways of designing a mixed-mode survey offering a choice between a Web questionnaire and a paper questionnaire, but the number of possible designs is limited by the amount of auxiliary information contained within the sampling frame. For general target populations, the frame will neither contain e-mail addresses nor information on Web access. Hence, under normal circumstances, the primary mode of contact would be regular mail, and *all* elements of the sample would have to be offered a paper questionnaire in addition to the Web response option. Could a mixed-mode survey designed this way be more cost-efficient than a standard mail survey? In order to find some kind of answer, we conducted a survey experiment in a population of university students.[*]

In this paper, we give a brief description of the experiment and present some of the results. More about the experiment and the results thereof can be found in Werner (2005).

2. The survey experiment

The survey experiment was conducted in collaboration with the City of Linköping in Sweden in spring 2004, and the target population was students at Linköping University − a population in which the Web access rate is supposedly 100 percent. The official survey objective was to investigate the students housing conditions, so as to obtain a basis for future housing planning. The sampling frame was constructed from a register containing various data concerning each registered student, such as the postal address, phone number, and e-mail address. A simple random sample of $3,000$ students was drawn from the frame population. The students of the sample were then randomly assigned to five experimental groups, each of which corresponded to a specific data collection approach. The five data collection approaches were the following (shorthand notation and group size in brackets):

- *Mixed-mode approach I (MI, 750).* The first mailout included a paper questionnaire and a cover letter offering a Web response option. The reminder mailout only included the Web response option.

- *Mixed-mode approach IIa (MIIa, 375).* The first mailout only included a cover letter with the Web response option, but informed the recipients that a paper questionnaire was to be sent out in a week or so to those who could not or would not respond using the Web. The reminder mailout included a paper questionnaire as well as the Web response option.

- *Mixed-mode approach IIb (MIIb, 375).* Identical to *MIIa* with one small but potentially important exception: the cover letter contained *no information* about the forthcoming paper questionnaire.

- *Reference approach (R, 750).* A standard mail approach not offering a Web response option. Both of the two mailouts included a paper questionnaire.

- *E-mail approach (E, 750).* A standard e-mail approach not offering a paper questionnaire. Each of three e-mails sent included a unique link to the Web questionnaire.

Note that two contacts were made in each approach except E, and that e-mail contacts were made only in approach E. The cover and reminder letters used in the three mixed-mode

approaches contained the survey Web site URL and a unique PIN code. In each approach, non-respondents were followed-up by telephone interview.

For reasons mentioned above, an approach such as E, where e-mail contacts are made, is usually not an alternative to a standard mail approach in practice. Therefore, approach E is actually not of primary interest in this particular study. However, the results of using this approach might be of interest in other contexts, and hence we decided to publish some of them here.

3. Response rates

The field period started off with the first mailout in mid-March, and the last telephone interview was made in the end of May. The field period can be subdivided into different phases defined by non-overlapping time intervals according to the following illustration.

Note that for approach E, three phases precede the follow-up phase. The table below shows observed response rates *at the end of* each specific phase (1, 2, 3, and F), by approach.

	MI	$MIIa$	$MIIb$	R	E
1	0.41	0.33	0.23	0.36	0.40
2	0.62	0.64	0.63	0.66	0.62
3	-	-	-	-	0.68
F	0.84	0.85	0.85	0.84	0.85

As the first-phase response rates show, $MIIb$ did not start out too well. The relatively large difference between the two variants of MII is interesting. It could, although unlikely, be mere chance, or it could actually be an effect of informing/not informing recipients of the forthcoming questionnaire mailout. After two phases of data collection, the response rates are quite similar, and $MIIb$ has almost caught up with $MIIa$. The response rates after the follow-up are even more similar, and also quite high. Hence, in terms of final response rates, all five approaches worked out well in practice.

The next table shows, for each mixed-mode approach, the proportion of Web responses among all responses at the end of the second phase.

MI	$MIIa$	$MIIb$
0.35	0.67	0.58

Interestingly, the difference in Web response rates between MI and each of the two variants of MII is very large, in particular the difference between MI and $MIIa$. In our study, offering the Web option prior to sending out the paper questionnaire proved to be successful, generating an increased Web response rate. The difference between $MIIa$ and $MIIb$ indicates that the recipient should be given information about the forthcoming paper questionnaire, although the difference is somewhat to small to allow for such a conclusion.

4. Costs

The table below presents the realized *variable* cost (in Swedish currency, SEK) of each approach except E. The variable cost includes costs of envelopes, address labels, cover letters, questionnaires, and postage. Moreover, it includes the labor costs of assembling mailout packages and handling incoming paper questionnaires, as well as interview costs. The costs are calculated in accordance with the cost situation at the City of Linköping as of March 2004. The costs of $MIIa$ and $MIIb$ have been doubled to compensate for smaller sample sizes.

MI	$MIIa$	$MIIb$	R
40,200	34,900	37,900	46,300

As the table shows, each of the three mixed-mode approaches resulted in a variable cost lower than that of the standard mail approach, R. In fact, the difference between R and $MIIa$ is as large as SEK $11,400 -$ nearly 25 percent of the cost of R. Taking fixed costs into account, however, the mixed-mode approaches will appear less successful. First, the Web server used for surveys at the City of Linköping is leased from another company on a monthly basis (SEK $3,000$/month). Second, the Web survey software used implies an annual license renewal cost (SEK $40,000$/year). Assuming that ten Web-based surveys of similar proportions are conducted over a year, a fixed cost of SEK $7,600$ could be added to the variable costs of MI, $MIIa$, and $MIIb$, respectively. This fixed cost does not affect the standard mail approach.

5. Cost-efficiency

Suppose that the objective of a survey is to estimate a population total t from data collected

using a specific approach A_g. Let $\hat{t}_g$ be the estimator of t, and let C_g be the cost of the data collection. The cost-efficiency, CEF, of the approach A_g with respect to the estimation of t can then be defined as $CEF(A_g) = [MSE(\hat{t}_g)]^{-1}/E(C_g)$, where $MSE(\hat{t}_g)$ is the mean square error of $\hat{t}_g$, and $E(C_g)$ is the expected cost of the data collection. In words, the cost-efficiency is defined as the amount of information (the inverse of the MSE) per unit of expected cost, which is an intuitively reasonable measure. The data collection approaches in our study could be compared with respect to this cost-efficiency measure, by constructing confidence intervals for cost-efficiency differences. We are currently working on this, aiming at publishing the results elsewhere.

6.   Previous research

Previous research in this particular field (mail-based mixed-mode surveys with a Web response option) has resulted in diverse results, from which it is hard to come to any general conclusions. The studies differ with respect to type of target population, overall response rates, the number of and contents of the mailouts, the response modes offered in addition to the Web questionnaire, and the way of offering these response modes. Designs corresponding to MI seem to be quite frequently tested (e.g., Griffin et al., 2001, Tarnai and Paxson, 2004), whereas sequential designs corresponding to MII are not as oftenly seen among published results (e.g., Schonlau et al., 2003). In fact, we do not know of any study in which the recipients have been *informed* of a forthcoming paper questionnaire mailout, as in $MIIa$. However, in some cases the opportunity of *requesting* a paper questionnaire has been offered (e.g., Lozar Manfreda et al., 2001). Cost comparisons in this type of studies are extremely rare, in spite of the fact that reducing costs is often said to be the primary objective of offering a Web response option. Lozar Manfreda et al. (2001) constitutes an exception to this rule.

7.   Concluding remarks

The results of the empirical study in this paper show that a probability sampling based mail survey with a Web response option can lead to lower costs than a traditional mail survey. Moreover, the results show that a sequential mixed-mode approach, in which the Web option is offered prior to the mail questionnaire (MII), has the *potential* of being less costly than a mixed-mode approach in which the mail questionnaire and the Web option are offered simultaneously (MI). Concerning response mode preferences, the Web response rate of the $MIIa$ group was nearly twice as high as that of the MI group − 67% versus only 35%. This result suggests that with little effort, it is possible to have people responding by a mode they normally do not prefer.

The target population in this study consisted entirely of university students − an ideal population in this context, since it consists of people who should be quite familiar with using the Web. Although the results of the study do not generalize to populations of more general character at present, they might have something to say about what to expect from such populations in a near future as the Web becomes more integrated into our daily lives.

REFERENCES

Griffin, D. H., Fischer, D. P., and Morgan, M. T. (2001). Testing an Internet Response Option for the American Community Survey. *Paper presented at the annual conference of the American Association for Public Opinion Research, Montreal, Quebec, Canada, May 17-20, 2001.*

Lozar Manfreda, K., Vehovar, V., and Batagelj, Z. (2001). Web versus Mail Questionnaire for an Institutional Survey. In: Westlake et al. (eds.), *The Challenge of the Internet: Proceedings of the ASC International Conference on Survey Research Methods*, pp. 79-87. Association for Survey Computing.

Schonlau, M., Asch, B. J., and Du, C. (2003). Web Surveys as Part of a Mixed-Mode Strategy for Populations That Cannot Be Contacted by E-Mail. *Social Science Computer Review* **21**, 218-222.

Tarnai, J., and Paxson, M. C. (2004). Survey Mode Preferences of Business Respondents. *Paper presented at the annual conference of the American Association for Public Opinion Research, Phoenix, Arizona, May 13-16, 2004.*

Werner, P. (2005). On the Cost-Efficiency of Probability Sampling Based Mail Surveys with a Web Response Option. *Linköping Studies in Statistics No. 6. Doctoral Thesis, Linköping University.*