Appendix P

Estimating Mode Effects

APPENDIX P

ESTIMATING MODE EFFECTS

Mode effects can encompass a wide variety of survey characteristics related to the mode of administration (Groves, 1989). This appendix discusses our approach to looking at differences between the mail and telephone results because of nonresponse, coverage, and other factors that would be manifested in the final distribution of respondents (e.g., different respondent selection methods). A second set of analyses will be conducted that attempt to assess the effects of the mode of communication across the mail and telephone surveys. *Mode of communication* refers to how the use of visual versus aural communication affects the interpretation and response to questionnaire items. As with maximizing response rates, the approach for the design of the surveys will be to minimize these effects as much as possible. However, as noted below, there are circumstances where it will be difficult to eliminate these effects for certain types of questions (e.g., Dillman, 2000, pp. 232ff). In these instances, we will estimate the effects cause by mode by taking advantage of the dual-frame design of the survey.

P.1 Minimizing Mode Effects

A number of studies have looked at differences between the visual mode of mail surveys and the aural mode of telephone surveys (de Leeuw, 2005). The differences that have been found are due to two primary sources: (1) social desirability bias and (2) the format, order, or context of items. For questions on behavior or facts, previous studies have not found large differences due to the mode of communication between telephone and mail surveys (e.g., Dillman and Tarnai, 1991; Brogger, et al, 2002; Link and Mokdad, 2005). For example, Link and Mokdad (2005) found no differences between mail and telephone questionnaires for reported levels of alcohol use. Similarly, Brogger and colleagues (2002) did not find many differences in the reporting of respiratory symptoms and risk factors. The major exception to this general rule is for questions that ask about socially desirable behaviors and/or opinions. McHorney (1994) observed higher, and presumably more accurate, levels of reporting in mail surveys than in interviewer-administered surveys for health conditions; while Link and colleagues (2006) found much higher rates of risk behaviors reported on the mail survey when compared to a telephone survey. Affirmative responses to questions of a personal nature, such as having certain types of cancer or receiving certain types of medical procedures (e.g., hysterectomy) might be reported at higher levels on the self-administered mail survey. Presenting a positive image may also relate to questions on less private

topics, such as eating habits, exercise, and knowledge of cancer and mental health. There is some evidence that measures of psychological well-being are affected by the presence of an interviewer (see HINTS 2005 item HC-02; Pruchano and Hayden, 2000; Springer and Hauser, 2006).

For certain type of questions, the use of a visual mode on the mail survey may lead to different response distributions when compared with the aural mode on the telephone. A mail survey is less likely to exhibit order and context effects because the respondent has the ability to read ahead. Mail survey respondents may adjust how they rate each item, based on how strongly they feel about all the items on the list. Telephone survey respondents will not be able to make these adjustments because their exposure is more segmented. When appropriate, we will change the order of the questions on the telephone, as other questions are similarly treated.

Several studies have found that telephone surveys exhibit a recency effect with respondents tending to choose the last category of a list of response options, while there is a primacy effect for mail surveys. It has also been hypothesized that telephone respondents are more likely to acquiesce and/or choose extreme responses on a scale. The empirical results related to these hypotheses are inconsistent across empirical analyses (de Leeuw, 2005). On HINTS, there are very few questions that are likely to be subject to recency and primacy effects. In those cases where it might be important, we will consider changing the order of the response options. Acquiescence can be addressed by reversing the wording of questions to require respondents to “agree” with one item in a battery and to “disagree” with adjacent items when indicating similar positions on a scale.

A difference that may be more of a concern are items that contain a relatively long list of response alternatives that are not read to respondents on the telephone. For a mail survey, listing all options and having respondents mark all that apply would be natural. However, this may lead to differences with the telephone survey because response categories will serve as cues on the mail survey that are different form the purely open-ended questions asked on the telephone survey. To address this, Westat will vary the order of the categories on the mail survey to minimize any order effects that might occur from the respondent reading all the categories.

P.2 Estimating Mode Effects

It will not be possible to design around all mode effects. To assess the extent and magnitude of any remaining effects, a two-stage analysis will be conducted. The first stage will examine whether there is differential unit- and item-level nonresponse across the two modes. These analyses would largely be done as part of the comparisons discussed above. Using information from these analyses, we will build a profile of those that are most likely to complete a mail or telephone survey. When building this profile, it will be important to understand how bias is affected for the measures included on HINTS 2007. For example, what are the implications for questions on cancer prevalence versus questions on methods used to find cancer-related information? The issue of bias could be different across these two measures. It will also be of interest to disaggregate the effects by coverage (i.e., those without landline telephones) and nonresponse (i.e., those that choose not to respond to the RDD survey).

If significant variables are found at the national or exchange level, Westat will use this information in the nonresponse adjustment for the weights. This will allow for combining across modes while compensating for differential nonresponse observed for the variables in these data sets.

The second stage of the analysis will assess whether there are differences because of the mode of communication for HINTS variables. This will be completed by comparing response distributions on HINTS measures across modes. For a continuous HINTS variable of interest, the basic test would be a linear regression of the form:

$$y_{ir}=a_i+b_{im}M+\sum c_{ij}x_{jr}+e_{i\,r}$$

Where:

y_{ir} = Value of i^{th} variable of interest for the r^{th} respondent
a_i = Average value for i^{th} variable
b_{im} = Regression coefficient for the mode effect on the of i^{th} variable of interest
M = 0 if the mode is telephone
1 if the mode is mail
c_{ij} = Regression coefficient for the effect of the j^{th} control variable on the i^{th} variable of interest
x_{jr} = Value for the r^{th} respondent of the j^{th} variable needed to control for differences in coverage and non-response between the two frames.
e_{ir} = Stochastic error term for i^{th} variable of interest and r^{th} respondent

For categorical and ordinal HINTS variables, logistic regression will be used. The x variables are those that are found to be significant in the unit or item nonresponse and coverage analysis. To the extent that the final survey weights adjust for some of these differences, inclusion of these variables will be minimized. The significance of b_{im} would be used to judge whether there is an significant mode effect for this particular variable.

If there were no significant differences found with respect to nonresponse and coverage from the first stage of the analysis, then the above regression equation would be reduced to a series of standard difference tests (e.g., t-tests, chi-square).

P.3 Further Exploring Mode Effects

One of the differences hypothesized to contribute to mode differences is related to the context and pace of the interview. Mail surveys allow respondents time for more thoughtful answers. Similarly, with no interviewer involved, a mail survey is not subject to differences related to between-interviewer variance. On the other hand, mail surveys remove the benefit of being able to have the interviewer explain unclear questions if the respondent is confused, as is done in the CATI. Other issues related to filling out a mail survey are the processes used when responding to the survey. One supposed advantage of a CATI is the ability to answer questions and clarify the meaning of particular questions. There is no real equivalent process for the mail survey, that may result in the respondent leaving the question blank. The design will build in measures that track the occurrence of each of these behaviors (i.e., asking questions of interviewers or leaving an item blank) to assess their relative frequency.

To assist in the evaluation of the different methodologies, we propose collecting additional information for both modes. One will be to conduct debriefing interviews with respondents. The mail respondents will be asked some additional information on their survey. The telephone respondents will be asked an additional set of questions after the main survey is completed.

The mail respondents will be asked questions related to the circumstances and process of filling out the survey (see Appendix J, page J-68). Examples of the types of questions to be included are listed here:

- Did they do it in more than one sitting?
- Did they fill out the survey with other members of the household?
- What did they think of the length of the questionnaire?
- How long did it take them to complete the questionnaire?
- (If receiving an incentive:) Would they have filled out the questionnaire if there was no incentive?

For those who were originally offered the option of completing by mail but completed by telephone, we will ask questions about the role of the incentive, as well as their reasons for not completing the mail (see Appendix J, pages J-68 and J-69). Examples of possible questions include the following:

- Would they have filled out the questionnaire if there was no incentive?
- Did they receive the mailing?
- What did they think of the length of the questionnaire?

For both the telephone and the telephone followup to the mail survey, we will collect observations of respondent performance. For the telephone survey, we will record the respondent-interviewer interactions for selected questions on the survey. The goal is to measure these interactions for questions that we anticipate to be different between the mail and telephone modes. For example, open-ended questions will likely be different between the two modes, at least with respect to the number of responses. Some of these questions will be targeted on the telephone to observe the extent that interviewers probe and the types of responses they get as a result of those probes.

To record these interviews, respondents will be told during the consent statement that some of the questions may be recorded for quality control reasons. A similar approach has been used on other Westat RDD surveys with no noticeable effect on the response rate and with a high degree of agreement by respondents.

These will be tabulated as part of the mode analysis to describe the situational and contextual circumstances surrounding the response process. They will also be correlated with variables that exhibit a mode effect. For example, are people more likely to report a sensitive behavior when they are alone?

P.4 Consequences and Actions Related to Significant Mode Effects

Any mode effect has to be analyzed on a variable-by-variable basis. As the analysis described above proceeds, Westat staff will discuss the various options with NCI. When a significant mode effect is found, there are three different strategies that can be taken: (1) concentrate on one specific mode, (2) enter mode as a control variable in analyses, or (3) ignore the mode altogether. The situations when each of these strategies might be used are discussed below.

The first strategy will be used when conducting trend analyses by comparing the RDD sample for HINTS 2007 with the comparable samples from HINTS 2003 and 2005. This is an acceptable way to control for significant changes in the design of ongoing surveys while still preserving the survey trends (e.g., see Kinderman, et al., 1997). To fully exploit this strategy, it will be important to keep the telephone questionnaire as similar as possible to the previous HINTS designs to preserve comparability for trends.

When restricting attention to HINTS 2007 data, the analyst can control for mode in analyses when significant mode effects have been identified in the analyses discussed above. These analyses will use the final weights that accounted for differential coverage and nonresponse across the two modes (and found in the analyses described above). It may also be reasonable to ignore mode effects, as long as those effects do not interact with variables that are targeted in the analyses. The random assignment of mode and the coverage-nonresponse adjustment may be enough to negate any effects. For example, one might be interested in testing whether age and gender are significant predictors of finding cancer information on the Internet. If mode is related to finding cancer information, but if this difference does not differ by age or gender (e.g., older subjects do not report Internet use differently by mode), then estimates will be unbiased even if mode is not included in the equation.

P.5 Cost Estimates

Estimates of costs for completing interviews, by mode, will be generated. The costs will be broken into three categories: (1) preparation, (2) data collection and (3) data capture and editing. From these components, we will estimate the cost of completing an interview from each mode.

From this cost information, a unit-cost for each mode will be calculated. The unit cost will provide an accounting of where the major expenses are for each mode. These costs will also be used to calculate the optimal mix of telephone and mail surveys.

REFERENCES

Brogger, J., Bakke, P., Eide, G.E., and Gulsvik, A. (2002). Comparison of telephone and postal survey modes on respiratory symptoms and risk factors. *American Journal of Epidemiology,* 155(6), 572-576.

Dillman, D.A. (2000). *Mail and Internet surveys: The tailored design method,* (2nd ed.). New York: John Wiley & Sons.

Dillman, D. and Tarnai, J. (1991). Mode effects of cognitively designed recall questions: A comparison of answers to telephone and mail surveys. In P. Biemer, R. Groves, L. Lyberg, N. Mathiowetz, & S. Sudman (Eds.), *Measurement errors in surveys* (pp. 73-93). New York: John Wiley & Sons.

de Leeuw, E.D. (2005). To mix or not to mix data collection modes in surveys. *Journal of Official Statistics*, 21(2), 233–255.

Groves, R. (1989). *Survey errors and survey costs*. New York: John Wiley and Sons.

Kindermann, C., Lynch, J.P., and Cantor, D. (1997). Effects of the redesign on victimization estimates. National Crime Survey Victimization Survey, Bureau of Justice Statistics, NCJ-164381.

Link, M. and Mokdad, A. (2005). Are Web and mail modes feasible options for the Behavioral Risk Factor Surveillance System? Paper presented at the Health Survey Research Methods Conference.

Link, M., Battaglia, M., Frankel, M.R., Osborn, L., and Mokdad, A.H. (2006, November). Address-based versus random-digit-dial surveys: Comparison of key health and risk indicators. *American Journal of Epidemiology,* 164, 1019-1025.

McHorney, C.A., Kosinski, M., and Ware, J.E., Jr. (1994). Comparisons of the costs and quality of norms for the SF-36 health survey collected by mail versus telephone interview: Results from a national survey. *Med Care,* 32(6), 551-567.

Pruchno, R.A. and Hayden, J.M. (2000). Interview modality: Effects on costs and data quality in a sample of older women. *Journal of Health and Aging,* 12, 3-24.

Springer, K.W. and Hauser, R.M. (2006). Survey measurement of psychological well-being [CDE working paper No. 2002-9]. Wisconsin Longitudinal Study, Center of Demography and Ecology, University of Wisconsin-Madison.