

August 13, 2009

Literature Reviews: Examination of Data Collection Methods for the NCVS

Report

Prepared for

Bureau of Justice Statistics

810 7th Street, NW
Washington, DC 20531

Prepared by

RTI International

3040 Cornwallis Road
Research Triangle Park, NC 27709

RTI Project Number 0211889

RTI Project Number
0211889

Literature Reviews: Examination of Data Collection Methods for the NCVS

Report

August 13, 2009

Prepared for

Bureau of Justice Statistics
810 7th Street, NW
Washington, DC 20531

Prepared by

RTI International
3040 Cornwallis Road
Research Triangle Park, NC 27709

Contents

Section	Page
1. Address-Based Sampling	1-1
1.1 Summary	1-2
1.2 References	1-3
2. Mixed-Mode Surveys	2-1
2.1 Attractiveness of Mixed-Mode Data Collection	2-1
2.2 Mode Effects	2-2
2.3 Things to Consider When Mixing Modes	2-3
2.4 Summary	2-4
2.5 References	2-4
3. Self-Administered Modes of Data Collection	3-1
3.1 Mail Surveys	3-1
3.2 Self-Administered Modes and Sensitive Questions	3-2
3.3 Web Surveys	3-3
3.4 Summary	3-4
3.5 References	3-5
4. Use of Incentives	4-1
4.1 Theories on Incentive Effectiveness	4-1
4.2 Impact on Response Rate and Nonresponse Bias	4-2
4.3 Prepaid Versus Postpaid Incentives	4-3
4.4 Incentives and Survey Mode	4-3
4.5 Summary	4-4
4.6 References	4-4
5. Additional Issues in Measuring Crime Victimization in Surveys	5-1
5.1 Event Recall	5-1
5.2 Proxy Respondents	5-2
5.3 Crime Severity, Survey Context, Stigma, and Terminology	5-3

5.4 Summary5-4

5.5 References5-4

1. ADDRESS-BASED SAMPLING

The development and improvement of a database of addresses in the United States has provided a potential alternative to the costly creation of sampling frames for area probability surveys through field listing. Address-based sampling (ABS) is possible using the Delivery Sequence File (DSF), a computerized file that contains all delivery point addresses serviced by the U.S. Postal Service (with the exception of general delivery). So far, evaluations of DSF for replacing enumeration of household units have shown promise, with potential household coverage as high as 97% on average. All evaluations have shown higher household coverage in urban areas than in rural areas.

The survey literature so far has focused on various approaches to a sampling frame construction from an address list and evaluation of its coverage and usability properties. The different approaches yield a uniform finding: using mailing addresses to develop a sampling frame for metropolitan households is a good and less costly alternative to household enumeration. For example, Iannacchione, Staab, and Redden (2003) applied Kish's half-open interval (Kish, 1965) frame-linking procedure to evaluate the coverage of an ABS frame using DSF. It was estimated that half-open intervals could be constructed and located for 94% of the addresses in the newly constructed frame. In another study designed to compare the coverage of ABS to field enumeration, Iannacchione et al. (2007) used global positioning system (GPS) technology to match the housing units from each frame. Even though field enumeration yielded higher overall coverage (98% vs. 82% in ABS), there was no difference when the matching was restricted to occupied urban housing units. Morton et al. (2007) applied Geographic information system (GIS) and GPS technologies to match postal (mailing address lists by postal carrier routes) to census geography (tracts and blocks). Not surprisingly, housing units in urban areas were more likely to geocode to the correct census block than housing units in rural areas (73% vs. 38%). O'Muircheartaigh et al. (2006) compared the coverage and cost-benefit tradeoffs of traditional enumeration and ABS on a national scale, employing a process in which a benchmark frame was constructed and ABS and traditional enumeration were evaluated against it. Overall, ABS was found to be more effective than the traditional enumeration, with the exception of areas with irregular street patterns and high population growth rates.

A few studies present methods for improving the coverage of ABS. Dohrmann, Han, and Mohadjer (2006) proposed enhancing the existing "Waksberg approach" to select segments with high growth rates at higher probabilities and applying lower subsampling rates for inclusion of missed units in such segments. O'Muircheartaigh, English, and Eckman (2007) proposed a model-based approach to inform decisions prior to data collection on whether field enhancement to ABS would be needed in particular segments. ABS was found appropriate for small-scale, low-cost surveys but was seen as not yet ready to fully replace traditional enumeration for high-quality national surveys. McMichael, Ridenhour, and

Shook-Sa (2008) proposed an alternative to HOI—a three-component procedure called Check for Housing Units Missed (CHUM). In initial evaluation, the first component of CHUM picked 79% of the missing units, while the second component picked the remaining 21%.

The quality of the address lists and of their coverage varies by vendors. Various vendors maintain and provide current versions of DSF that could be purchased for surveys (USPS does not offer it to survey organizations). Dohrmann, Han, and Mohadjer (2006) compared list quality by vendors (Compact Information Systems [CIS], Donnelly Marketing, and ADVO) for an urban/suburban area and compared ABS to traditional enumeration in urban/suburban, very urban, and rural areas. CIS and ADVO were found to be comparable. Consistent with other findings, high match rates between ABS and traditional enumeration were reported mainly for urban areas.

Alternatives to DSF have also been considered in investigating alternative methods for sampling frame construction. For example, Kalsbeek, Kavanagh, and Wu (2004) examined the utility of using lists of property tax parcels in U.S. counties. A test of the proposed approach yielded high levels of validity and reliability, similar to the levels associated with the traditional housing unit enumeration.

Finally, the evaluation of ABS for sampling frame creation for the general population has been expanded by a comparison to random-digit dialing (RDD) sampling methods (Link et al., 2008). In addition to the lower cost of the ABS mail survey, ABS reported significantly higher response rates than RDD in five of the six studied states.

1.1 Summary

Overall, the existing research presents a promising future for ABS in survey design and suggests that its true potential may be in mixed-mode surveys. The attractiveness of ABS is that it is cost efficient and time efficient. Large-scale surveys often require several months to list all dwelling units in the selected segments (usually, census blocks). In contrast, ABS offers greater geographic diversity (selection of housing units is not restricted to small segments based on census blocks) and thus presents a potential for improving statistical efficiency. There are some drawbacks associated with the construction of an address-based sampling frame related to the overall completeness of the list, the current status of the addresses, and the adequacy of the list coverage in rural areas. The typical sources of undercoverage for ABS are post office boxes, when used as the only method for mail delivery (making up 1.3% of households in the United States, according to Staab and Iannacchione [2003]); rural routes (making up 3.9% of households nationwide); and noninstitutional group quarters (e.g., dormitories, assisted living facilities, shelters) that are not identified on the USPS lists because they operate their own post office or because mail is delivered to the business unit.

1.2 References

- Dohrmann, Sylvia, Daifeng Han, and Leyla Mohadjer. 2006. "Residential Address Lists vs. Traditional Listing: Enumerating Households and Group Quarters." In *Proceedings of the American Statistical Association, Section on Survey Research Methods*, pp.2959-64.
- Iannacchione, Vincent, Katherine Morton, Joseph McMichael, David Cunningham, James Cajka, and James Chromy. 2007. "Comparing the Coverage of a Household Sampling Frame Based on Mailing Addresses to a Frame Based on Field Enumeration." In *Proceedings of the American Statistical Association, Section on Survey Research Methods*, pp.3323-32.
- Iannacchione, Vincent, Jennifer Staab, and David Redden. 2003. "Evaluating the Use of Residential Mailing Lists in a Metropolitan Household Survey." *Public Opinion Quarterly* 67(2):202-10.
- Kalsbeek, William, Sarah Kavanagh, and Jingjing Wu. 2004. "Using GIS-Based Property Tax Records as an Alternative to Traditional Household Listing in Area Samples." In *Proceedings of the American Statistical Association, Section on Survey Research Methods*, pp.3750-57.
- Kish, Leslie. 1965. *Survey Sampling*. New York: Wiley
- Link, Michael, Michael Battaglia, Martin Frankel, Larry Osborn, and Ali Mokdad. 2008. "A Comparison of Address-Based Sampling (ABS) Versus Random-Digit Dialing (RDD) for General Population Surveys." *Public Opinion Quarterly* 72(1):6-27.
- McMichael, Joseph, Jamie Ridenhour, and Bonnie Shook-Sa. 2008. "A Robust Procedure to Supplement the Coverage of Address-Based Sampling Frames for Household Surveys." In *Proceedings of the American Statistical Association, Section on Survey Research Methods*, pp. 4329-35.
- Morton, Katherine, Vincent Iannacchione, Joseph McMichael, James Cajka, Ross Curry, and David Cunningham. 2007. "Linking Mailing Addresses to a Household Sampling Frame Based on Census Geography." In *Proceedings of the American Statistical Association, Section on Survey Research Methods*, pp. 3971-74.
- O'Muircheartaigh, Colm, Edward English, and Stephanie Eckman. 2007. "Predicting the Relative Quality of Alternative Sampling Frames." In *Proceedings of the American Statistical Association, Section on Survey Research Methods*, pp.3239-48.
- O'Muircheartaigh, Colm, Edward English, Stephanie Eckman, Heidi Upchurch, Erika Garcia, and James Lepkowski. 2006. "Validating a Sampling Revolution: Benchmarking Address Lists Against Traditional Listing." In *Proceedings of the American Statistical Association, Section on Survey Research Methods*, pp. 4189-96.
- Staab, Jennifer, and Vincent Iannacchione. 2003. "Evaluating the Use of Residential Mailing Addresses in a National Household Survey." In *Proceedings of the American Statistical Association, Section on Survey Research Methods*, pp.4028-33.

2. MIXED-MODE SURVEYS

Researchers are continually trying to find the optimal mix of methods to minimize total survey error in survey estimates. Declining response rates, increasing costs, coverage issues, and data collection deadlines have all led to the increasing use of mixed-mode survey designs. With the popularity of telephone surveys in the 1970s, the mix of face-to-face and telephone data collection modes soon became attractive for large national surveys (e.g., the Current Population Survey). The development of computer technology marked the next change in data collection—computer-assisted equivalents were implemented in all major modes of data collection (de Leeuw and Collins, 1997; Couper and Nicholls, 1998). The development of web surveys gave rise to a combination of mail and web surveys.

When discussing mixed-mode surveys, it is important to investigate the reasons for mixing modes, mode effects, and issues to consider when mixing modes. These items are discussed in more detail below.

2.1 Attractiveness of Mixed-Mode Data Collection

Groves et al. (2004) identified three main reasons for using mixed-mode data collection: cost reduction, response rate maximization, and money saving in longitudinal surveys. The use of a combination of data collection methods reduces cost, as it typically involves an attempt to collect data in a cheaper mode (e.g., mail), followed by a more expensive mode (e.g., telephone), and possibly moving to an even more costly mode (e.g., face-to-face interviewing) for the nonrespondent sample persons. The American Community Survey is an example of this approach: it starts in a mail mode; this is followed by telephone follow-up of nonrespondents; and then there are face-to-face follow-ups with a subsample of the remaining nonrespondents (Alexander and Wetrogan, 2000). Maximization of response rates is often achieved through mixed-mode data collection. For example, the Current Employment Statistics program offers multiple modes of data collection, such as web, fax, inbound interactive voice response (IVR), telephone, and mail. While the Current Employment Statistics survey, which includes 390,000 business establishments, employs six methods of data collection, the use of two or three modes is more common in increasing response rates and decreasing costs.

Longitudinal surveys also employ mixed-mode data collection to reduce cost in later waves, when rapport between the interviewer and the respondent has already been established in the first wave, usually administered in face-to-face mode. An example of this approach is the Current Population Survey, where interviewers obtain telephone numbers in the first wave of data collection that are to be used in subsequent rounds.

Biemer and Lyberg (2003) note that mixed-mode designs have now become the norm of data collection in the United States and Western Europe. The attractiveness of mixed-mode

designs is in their ability to compensate for the weaknesses of individual modes. For example, to reduce coverage bias in the early days of telephone data collection, mixed-mode dual frame designs were often employed, benefiting from the cost savings of telephone interviewing and the complete coverage of face-to-face data collection (e.g., Massey, Marquis, and Tortora, 1982; Marquis and Blass, 1985; for a detailed discussion, see Groves and Lepkowski, 1985). Another feature that makes mixed-mode designs attractive is their application in reducing nonresponse bias. Since nonresponse includes both noncontacted respondents and those who refuse to cooperate under the initial protocol, implementing a different mode of data collection can be addressed both by changing the method of contact and using different persuasive techniques, particularly through the use of interviewers. It is not necessarily that some modes are better than others for a particular population; to the extent that individuals vary in their likelihood to participate across modes and that respondents to different modes are somewhat different, the threat of nonresponse bias is minimized through the use of multiple modes.

The possibility that some respondents prefer one mode over another has been recognized. Often, however, the mode in which respondents are asked about their mode of preference is selected as the mode of choice. For example, Groves and Kahn (1979) reported that among respondents in a national telephone survey, 39% expressed a preference to be interviewed by phone, 23% in a face-to-face setting, and 28% by mail. The preferred mode of interview in a face-to-face survey was overwhelmingly face-to-face (78%), followed by mail (only 17%). Some studies suggest that giving the choice of mode to the respondent does not necessarily improve response rates. For example, Dillman, Clark, and West (1995) showed that offering the respondent the choice of returning a questionnaire by mail or calling in to be interviewed did not improve response rates. On the other hand, sequential change of modes has been reported to significantly improve response rates. For example, Shettle and Mooney (1999) reported a response rate of 68% after four mailings and an incentive, which increased to 81% with telephone follow-up and to 88% with a final switch to face-to-face interviewing.

2.2 Mode Effects

Different data collection modes possess different strengths and weaknesses. In searching for reasonable alternatives, studies have contrasted pairs of modes. Compared with face-to-face surveys, telephone surveys have been found to yield lower response rates (Groves and Kahn, 1979; Cannell et al., 1987; Sykes and Collins, 1988), shorter responses to open-ended questions (Groves and Kahn, 1979; Sykes and Collins, 1988; Kormendi and Noordhoek, 1989), and higher rates of satisficing and socially desirable responding (Holbrook, Green, and Krosnick, 2003; Kirsch, McCormack, and Saxon-Harrold, 2001). In addition, sensitive questions have been found to increase mode differences (Aquilino and LoSciuto, 1990). Similarly, comparisons between mail and telephone modes show higher

social desirability effects (Dillman and Tarnai, 1991; Walker and Restuccia, 1984) and increased response order and question order effects (Bishop et al., 1988) for telephone surveys. A meta-analysis of face-to-face versus mail response rates did not find significant differences between modes (Goyder, 1986). Research so far has produced mixed results on the effect of these modes on reports of sensitive behaviors. For example, Bongers and van Oers (1998) found no difference between mail and face-to-face interviewing on responses to alcohol-related questions, but Hochstim (1967) and Tourangeau and Smith (1996) found greater reporting of sensitive behaviors in self-administered surveys.

2.3 Things to Consider When Mixing Modes

There are potential drawbacks to using mixed-mode survey designs, affecting different sources of survey error: coverage, nonresponse, measurement, and processing. Coverage error can be affected in mixed-mode designs when multiple sampling frames are needed. Although the use of multiple frames can reduce undercoverage, it involves the use of statistical adjustments to sample weights to merge data from each mode—a procedure that can induce varying, and often unknown, amounts of error, depending on the particular frames and study design.

As noted earlier, mixed-mode designs are often used to increase response rates, but when they are used to reduce costs, they can lead to lower response rates—likely respondents to face-to-face survey requests may be less likely to participate if first asked in a different mode, such as by mail. Apart from the choice of modes to be implemented in a study, the order of modes can also have an impact on cost and response rates—and may likely result in a different mix of survey errors. An equally important decision is whether to implement modes simultaneously, giving the choice of mode to the respondent, or sequentially, often offering the lower-cost modes first. While this is an important design decision, one that could affect response rates, nonresponse bias, and the measurement properties of the data, it is still in need of empirical research.

Perhaps the greatest source of error from implementing a mixed-mode design is from measurement. Differences across modes have been identified in the research literature, which for the most part can be attributed to three factors: interviewer versus self, visual versus auditory, and computer versus paper-and-pencil administration. In a seminal paper covering two of these three dimensions, Tourangeau and Smith (1996) found greater reporting of sensitive behaviors in computer-assisted self-interviewing than in computer-assisted personal interviewing, and even greater reporting of sensitive behaviors in audio computer-assisted self-interviewing.

A large body of literature reports that interviewer-administered modes evoke socially desirable reporting to a greater extent than do self-administered modes (Aquilino, 1994; de Leeuw, 1992; De Maio, 1984; Hochstim, 1967). It has also been suggested that

respondents are more likely to acquiesce in the presence of an interviewer (Schuman and Presser, 1981). Additionally, the presentation of the survey questions (visual vs. auditory) in each mode contributes to primacy or recency effects, as described by Krosnick and Alwin (1987).

Finally, the mix of modes in a survey can result in different processing errors. Often overlooked, the errors made by interviewers (e.g., coding of occupation) are different from the errors made in the processing of paper questionnaires, which in turn are different from those in computerized self-administered modes. Like measurement error, this is particularly threatening when these mode-specific errors are not randomly distributed across different sample members—and the interview mode is seldom, if ever, a random choice or assignment.

2.4 Summary

Overall, mixed-mode designs will continue to gain popularity mainly because of their ability to reduce costs and maximize response rates. However, careful consideration should be given to the potential impact of such designs on the coverage, nonresponse, and measurement properties of the data.

2.5 References

- Alexander, C.H., Jr., and S. Wetrogan. (2000). "Integrating the American Community Survey and the Intercensal Demographic Estimates Program." In *Proceedings of the American Statistical Association*.
<www.amstat.org/sections/srms/proceedings/papers/2000_046.pdf>
- Aquilino, W.S. 1994. "Interview Mode Effects in Surveys of Drug and Alcohol Use." *Public Opinion Quarterly* 58:210-40.
- Aquilino, W.S., and L.A. LoSciuto. 1990. "Interview Mode Effects in Drug Use Surveys." *Public Opinion Quarterly* 54(3):362-95.
- Biemer, P.P., and L.E. Lyberg. 2003. *Introduction to Survey Quality*. New York: John Wiley.
- Bishop, G., H. Hippler, N. Schwarz, and F. Strack. 1988. "A Comparison of Response Effects in Self-Administered and Telephone Surveys." In *Telephone Survey Methodology*, R. Groves, P. Biemer, L. Lyberg, J. Massey, W. Nicholls, II, and J. Waksberg, eds., pp. 321-40. New York: Wiley.
- Bongers, I.M.B., and J.A.M. van Oers. 1998. "Mode Effects on Self-Reported Alcohol Use and Problem Drinking: Mail Questionnaires and Personal Interviewing Compared." *Journal of Studies on Alcohol* 59:280-85.
- Cannell, C.F., R.M. Groves, L. Magilavy, N. Mathiewetz, P. Miller, and O. Thornberry. 1987. *An Experimental Comparison of Telephone and Personal Health Interview Surveys*. Vital and Health Statistics, series 2, no. 106. DHHS Pub. No. (PHS)87-1380. Washington, DC: U.S. Government Printing Office.

- Couper, M.P., and W.L. Nicholls, II. 1998. "The History and Development of Computer-Assisted Survey Information Collection Methods." In *Computer-Assisted Survey Information Collection*, M.P. Couper, R.P. Baker, J. Bethlehem, C.Z.F. Clark, J. Martin, W.L. Nicholls, II, and J.M. O'Reilly, eds., pp. 1-22. New York: John Wiley.
- de Leeuw, E.D. 1992. *Data Quality in Mail, Telephone, and Face-to-Face Surveys*. Amsterdam: TT-Publicaties.
- de Leeuw, E., and M. Collins. 1997. "Data Collection Methods and Survey Quality: An Overview." In *Survey Measurement and Process Quality*, L. Lyberg, P. Biemer, M. Collins, E. de Leeuw, C. Dippo, N. Schwarz, and D. Trewin, eds., pp. 199-220. New York: John Wiley.
- De Maio, T.J. 1984. "Social Desirability and Survey Measurement: A Review." In *Surveying Subjective Phenomena, Vol. 2*, C. Turner and E. Martin, eds., pp. 257-82. New York: Russell Sage Foundation.
- Dillman, D.A., J.R. Clark, and K.K. West. 1995. "Influence of an Invitation to Answer by Telephone on Response to Census Questionnaires." *Public Opinion Quarterly* 51:201-19.
- Dillman, D., and J. Tarnai. 1991. "Mode Effects of Cognitively Designed Recall Questions: A Comparison of Answers to Telephone and Mail Surveys." In *Measurement Errors in Surveys*, P. Biemer, R. Groves, L. Lyberg, M. Mathiowetz, and S. Sudman., eds., pp. 73-93. New York: Wiley.
- Goyder, J. 1986. "Surveys on Surveys: Limitations and Potentialities." *Public Opinion Quarterly* 50:27-41.
- Groves, R.M., F. Fowler, M. Couper, J. Lepkowski, E. Singer, and R. Tourangeau. 2004. *Survey Methodology*. New York: Wiley.
- Groves, R.M., and R. Kahn. 1979. *Surveys by Telephone: A National Comparison With Personal Interviews*. New York: Academic Press.
- Groves, R.M., and J.M. Lepkowski. 1985. "Dual Frame, Mixed-Mode Survey Designs." *Journal of Official Statistics* 1:263-86.
- Hochstim, J.R. 1967. "A Critical Comparison of Three Strategies of Collecting Data From Households." *Journal of the American Statistical Association* 62:976-89.
- Holbrook, A.L., M.C. Green, and J.A. Krosnick. 2003. "Telephone vs. Face-to-Face Interviewing of National Probability Samples With Long Questionnaires: Comparisons of Respondent Satisficing and Social Desirability Response Bias." *Public Opinion Quarterly* 67:79-125.
- Kirsch, A.D., M.T. McCormack, and S.K.E. Saxon-Harrold. 2001. "Evaluation of Differences in Giving and Volunteering Data Collected by In-Home and Telephone Interviewing." *Nonprofit and Voluntary Sector Quarterly* 30:495-504.
- Kormendi, E., and J. Noordhoek. 1989. *Data Quality in Telephone Surveys*. Copenhagen: Danmarks Statistik.

- Krosnick, Jon A., and D.F. Alwin. 1987. "An Evaluation of a Cognitive Theory of Response-Order Effects in Survey Measurement." *Public Opinion Quarterly* 51:201-19.
- Marquis, K.H., and R. Blass. 1985. "Nonsampling Error Considerations in the Design and Operation of Telephone Surveys." In *Proceedings of the First Annual Research Conference of the U.S. Bureau of the Census*, pp. 301-29.
- Massey, J.T., K. Marquis, and R. Tortora. 1982. "Methodological Issues Related to Telephone Surveys by Federal Agencies." In *Proceedings of the Social Statistics Section, American Statistical Association*, pp. 63-72.
- Schuman, H., and S. Presser. 1981. *Questions and Answers in Attitude Surveys: Experiments in Question Form, Wording, and Context*. New York: Academic Press.
- Shettle, C., and G. Mooney. 1999. "Monetary Incentives in U.S. Government Surveys." *Journal of Official Statistics* 15:231-50.
- Sykes, W., and M. Collins. 1988. "Effects of Mode of Interview: Experiments in the UK." In *Telephone Survey Methodology*, R. Groves, P. Biemer, L. Lyberg, J. Massey, W. Nicholls, II, and J. Waksberg, eds., pp. 301-20. New York: John Wiley and Sons.
- Tourangeau, R., and T. Smith. 1996. "Asking Sensitive Questions: The Impact of Data Collection, Question Format, and Question Context." *Public Opinion Quarterly* 60:275-304.
- Walker, A.H., and J.D. Restuccia. 1984. "Obtaining Information on Patient Satisfaction With Hospital Care: Mail Versus Telephone." *Health Services Research* 19:291-306.

3. SELF-ADMINISTERED MODES OF DATA COLLECTION

Self-administered surveys involve indirect contact with the respondent, may utilize both visual (e.g., mail) and aural (e.g., audio computer-assisted self-interviewing [ACASI]) channels of communication, and usually do not allow for complex instruments (unless computer administered). Self-administered modes can be used as stand-alone modes, in mixed-mode designs, or in portions of face-to-face surveys where sensitive questions are asked. A common feature in self-administered modes (when used as stand-alone modes or in mixed-mode designs) is the sequence of the distribution of materials—such as advance letters; the cover letter and questionnaire; and the reminder message and follow-up questionnaire—used to maximize response rates (see Dillman, 1978; Dillman, 2000).

There are various types of self-administered methods of data collection that differ largely in the extent to which they employ technology and utilize aural and visual presentation. Mail surveys remain one of the most popular modes, in part due to the ability to use address sampling frames. Other self-administered modes include e-mail, web, fax, optical character recognition (OCR), disk-by-mail (DBM), touchtone data entry (TDE), voice recognition entry (VRE), automatic speech recognition (ASR), and inbound interactive voice response (IVR). Several self-administered modes are used as part of an interviewer-administered survey, where the interviewer sets up the equipment, instructs the respondent in how to use it, and is available during the interview to assist, if necessary: computer-assisted self-interviewing (text-CASI), audio-CASI (ACASI), video-CASI (V-CASI), and audio-visual CASI (AV-CASI).

From a cost-and-error perspective, self-administered modes are often characterized by relatively low costs when used as the primary mode but are associated with lower response rates than interviewer-administered surveys. This leaves a substantial potential for nonresponse bias in self-administered surveys. Often, it is not possible to disentangle refusals from noncontacts—for example, Mathiowetz, Couper, and Singer (1994) reported that in 63% of households in the United States, one person is responsible for opening mail and that 63% of households throw some mail away without opening. In addition, even in interviewer-administered surveys, breakoff rates are very high: for example, in their review of IVR studies, Tourangeau, Steiger, and Wilson (2002) reported breakoff rates as high as 31%; similarly, Couper, Singer, and Tourangeau (2004) reported a 24% overall breakoff rate in outbound IVR, and Gribble et al. (2000) reported a 24% breakoff rate for telephone-CASI, compared with 2% for computer-assisted telephone interviewing (CATI).

3.1 Mail Surveys

Mail surveys continue to be one of the most popular methods for data collection. The research on mail surveys is voluminous: a bibliography compiled in the 1990s on research to improve mail survey procedures published since 1970 included more than 400 entries (Dillman and Sangster, 1990). Nonresponse has been the biggest challenge to mail surveys

so far; thus, various studies have focused on procedures and techniques for maximizing response rates. Such techniques include incentives; personalization of correspondence; content of cover letter; questionnaire layout, length and color; follow-up reminders; and so forth (see Dillman, 1978; Dillman, 2000). The Total Design Method (TDM) proposed by Dillman (Dillman, 1978) and later renamed the Tailored Design Method (Dillman, 2000) utilizes social exchange theory to guide the integration of specific procedures. The theory posits that sample members are more likely to return the questionnaire if the perceived benefit of doing so outweighs the perceived cost of responding. This has led to practical recommendations on how to design a mail survey that appears interesting, trustworthy, easy, and less time consuming to complete.

In terms of coverage, mail surveys so far have not enjoyed the degree of coverage accomplished by face-to-face surveys. However, with the development and improvement of a database of addresses and the promising future of the Delivery Sequence File¹ for address-based sampling, mail surveys may become a mode that offers almost complete coverage of households in the United States at a relatively low price.

From a measurement error perspective, mail surveys have been reported to be less susceptible to response order effects (mainly recency effects, i.e., choosing the last response category) relative to telephone surveys (Bishop, Hippler, and Schwarz, 1988; Ayidiya and McClendon, 1990). Another difference between mail and interviewer-administered modes that is frequently observed in research is the tendency for mail respondents to use the entire scale when vague quantifiers are used as scale categories rather than selecting the extremes. Such an effect was first reported by Hochstim (1967) and was later supported by studies on mode comparisons by Dillman and Mason (1984), Mangione, Hingson, and Barrett (1982), Talley et al. (1983), Walker and Restuccia (1984), and Zapka, Stoddard, and Lubin (1988). One possible explanation for the observed differences is that respondents do not interact with an interviewer and thus are less concerned about self-representation and less likely to provide socially desirable responses (extremes on scales). In fact, self-administered surveys in general have been reported to yield higher rates of sensitive and socially undesirable behaviors and attitudes, possibly due to the increased social distance between respondent and researcher and the private environment in which the survey can be filled out.

3.2 Self-Administered Modes and Sensitive Questions

In response to the need for a private data collection environment, various (usually CASI) techniques in which the respondents interact directly with a laptop computer for a portion of the face-to-face interview have been utilized. A seminal article by Tourangeau and Smith

¹ A computerized file that contains all delivery point addresses serviced by the U.S. Postal Service with the exception of general delivery.

(1996) examined responses to computer-assisted personal interviewing (CAPI) and interviews conducted using text-CASI and ACASI. Topics ranged from illicit drug use to sexual behavior. The findings supported the notion that the privacy of the CASI setting encouraged respondents' honesty in reporting such sensitive behaviors. It was also demonstrated that the audio component of the interview (ACASI) enhanced the feeling of privacy, thereby increasing the level of reporting. Similar findings were reported by Aitken et al. (2000), Hewett et al. (2004), Fu et al. (1998), Kissinger et al. (1999), and Moskowitz (2004). A recent study by Couper, Tourangeau, and Marvin (2009) demonstrates that the gains from using ACASI are modest relative to text-CASI and that most respondents make limited use of the audio component.

Many national surveys that gather data about sensitive topics employ self-administration for part of the interview. For example, the National Survey of Family Growth administers items about pregnancies and abortions in ACASI and also in the main CAPI module. A difference of 17% in reports of abortions has been reported between ACASI and CAPI (Fu et al., 1998). Similar findings have been reported on illicit drug use in the National Longitudinal Study of Youth (Schoeber et al., 1992) and in a randomized experiment embedded in the 1990 National Household Survey on Drug Abuse (NHSDA) field test.

Recently, the effects of self-administered modes on socially desirable and sensitive reporting were reexamined by Kreuter, Presser, and Tourangeau (2008). The authors used survey and university record data to look at mode effects on the reporting of potentially sensitive information by a sample of recent university graduates. Conventional CATI, IVR, and web modes were compared. Web administration was found to increase the level of reporting of sensitive information and reporting accuracy relative to conventional CATI, followed by IVR.

No significant differences in reports to sensitive and socially desirable questions have been reported *across* self-administered modes (e.g., Dillman and Tarnai, 1991; Knapp and Kirk, 2003; Lensvelt-Mulders et al., 2006). Generally, computerization does not add an additional advantage (e.g., Dillman and Tarnai, 1991), even though the use of ACASI can be invaluable for low-literate populations.

3.3 Web Surveys

With the mass use of the Internet, web surveys became popular very fast. Web surveys offer access to millions of potential respondents, at low cost and with rapid turnaround. Coverage remains the biggest threat to inference from web surveys (unless the target population is made up entirely of web users). Sampling frames for web surveys are hard to construct because the "internet population" is different in many aspects from the general population in the United States (Couper, 2000). Thus, web surveys often use non-probability-based sample designs. Many survey organizations create panels of web respondents that are recruited via a probability mode, such as phone, face-to-face, or mail.

However, this strategy adds another layer of concern—panel conditioning that occurs with continuous experience with a survey over time (Kalton and Citro, 1993; Kalton, Kasprzyk, and McMillen, 1989).

When frames are available and probability methods employed (e.g., lists of e-mail addresses of university students), web surveys generally produce lower response rates than mail surveys (e.g., Guterbock et al., 2000; Kwak and Radler, 2002; Lesser and Newton, 2001; Lesser and Newton, 2002). The reasons for this may be many—the fact that techniques that have proven successful in increasing response rates in mail surveys may not work for web surveys, technical difficulties, and so forth. Concerns of privacy and confidentiality may be a crucial factor affecting not only web survey response rates but also the ability to collect sensitive information with less social desirability bias (Couper, 2000). We are not familiar with research that examines the extent to which the use of web surveys negates the ability of self-administered surveys to collect sensitive information.

From a measurement error perspective, web surveys possess unique features, such as the ability to deliver multimedia content to respondents; however, there may be variation in how a survey appears on a respondent's screen (dependent on browser settings, screen size, etc.). Various aspects of visual design features have been tested, including the use of progress indicators (e.g., Crawford, Couper, and Lamias, 2001; Conrad et al., 2005; Heerwegh and Loosveldt, 2006); paging versus scrolling web survey design (e.g., Peytchev et al., 2006); definitions (e.g., Conrad et al., 2006); visual analog scales (e.g., Couper et al., 2006); response formats (e.g., Heerwegh and Loosveldt, 2002); and interviewer pictures, scale colors, and other visual features (e.g., Couper, Conrad, and Tourangeau, 2007; Tourangeau, Couper, and Conrad, 2007).

Web surveys are increasingly becoming a popular option in mixed-mode designs using the choice of completion method, where the focus is on minimizing respondent burden and cost (rather than concern about possible mode effects). Many government agencies have introduced a web option (usually in panel surveys of establishments): for example, the Current Employment Statistics program at the Bureau of Labor Statistics (Clayton and Werking, 1998); and the U.S. Census Bureau's Library Media Center survey (see Tedesco, Zuckerberg, and Nichols, 1999; Zuckerberg, Nichols, and Tedesco, 1999).

3.4 Summary

Self-administration is a preferred mode of data collection for survey questions related to sensitive or socially undesirable events and behaviors. This is usually achieved through the use of various CASI techniques, even though research suggests that it is the use of self-administration rather than computerization of the survey interview and audio components that is believed to enhance a respondent's privacy. Mail and web modes are the dominating self-administered options that are used as stand-alone modes or in mixed-mode designs.

3.5 References

- Aitken, Sherrie S., James DeSantis, Thomas C. Harford, M. Fe Caces. 2000. "Marijuana Use Among Adults: A Longitudinal Study of Current and Former Users." *Journal of Substance Abuse* 12(3):213-26.
- Ayidiya, S.A., and M.J. McClendon. 1990. "Response Effects in Mail Surveys." *Public Opinion Quarterly* 54:229-47.
- Bishop, G.G., H. Hippler, and F. Schwarz. 1988. "A Comparison of Response Effects in Self-Administered and Telephone Surveys." In *Telephone Survey Methodology*, R.M. Groves, P. Biemer, L. Lyberg, J. Massey, W. Nicholls, II, and J. Waksberg, eds., pp. 321-40. New York: Wiley and Sons.
- Clayton, R.L., and G.S. Werking. 1998. "Business Surveys of the Future: The World Wide Web as a Data Collection Methodology." In *Computer Assisted Survey Information Collection*, M.P. Couper, R.P. Baker, J. Bethlehem, C.Z.F. Clark, J. Martin, W.L. Nicholls, II, and J. O'Reilly, eds., pp. 543-562. New York: Wiley.
- Conrad, F.G., M.P. Couper, R. Tourangeau, and A. Peytchev. 2005. "Impact of Progress Feedback on Task Completion: First Impressions Matter." In *SIGCHI 2005: Human Factors in Computing Systems*, pp. 1921-24. Portland, OR: ACM Press.
- Conrad, F.G., M.P. Couper, R. Tourangeau, and A. Peytchev. 2006. "Use and Non-Use of Clarification Features in Web Surveys." *Journal of Official Statistics* 22(2):245-69.
- Couper, M.P. 2000. "Web Surveys: A Review of Issues and Approaches." *Public Opinion Quarterly* 64:464-94.
- Couper, M.P., F.G. Conrad, and R. Tourangeau. 2007. "Visual Context Effects in Web Surveys." *Public Opinion Quarterly* 71(4):623-34.
- Couper, M.P., E. Singer, and R. Tourangeau. 2004. "Does Voice Matter? An Interactive Voice Response (IVR) Experiment." *Journal of Official Statistics* 20(3):551-70.
- Couper, M.P., R. Tourangeau, F.G. Conrad, and E. Singer. 2006. "Evaluating the Effectiveness of Visual Analog Scales: A Web Experiment." *Social Science Computer Review* 24(2):227-45.
- Couper, M. P., R. Tourangeau, and T. Marvin. 2009. "Taking the Audio Out of Audio-CASI." *Public Opinion Quarterly* 73(2):281-303.
- Crawford, S.D., M.P. Couper, and M.J. Lamias. 2001. "Web Surveys—Perceptions of Burden." *Social Science Computer Review* 19(2):146-62.
- Dillman, D.A. 1978. *Mail and Telephone Surveys: The Total Design Method*. New York: Wiley and Sons.
- Dillman, D.A. 2000. *Mail and Internet Surveys: The Tailored Design Method*. New York: Wiley and Sons.

- Dillman, D.A., and R.G. Mason. 1984. "The Influence of Survey Methods on Question Response." Paper presented at the Annual Meeting of the American Association for Public Opinion and Research, Delavan, WI, May 17-20.
- Dillman, D.A., and R.L. Sangster. 1990. *Mail Surveys: A Comprehensive Bibliography 1974-1989*. Technical Report. Pullman, WA: Social and Economic Sciences Research Center, Washington State University.
- Dillman, D., and J. Tarnai. 1991. "Mode Effects of Cognitively Designed Recall Questions: A Comparison of Answers to Telephone and Mail Surveys." In *Measurement Errors in Surveys*, P. Biemer, R. Groves, L. Lyberg, N. Mathiowetz, and S. Sudman, eds., pp. 73-93. New York, NY: John Wiley and Sons.
- Fu, Haishan, Jacqueline E. Darroch, Stanley K. Henshaw, Elizabeth Kolb. 1998. "Measuring the Extent of Abortion Underreporting in the 1995 National Survey of Family Growth." *Family Planning Perspectives* 30(3):128-38.
- Gribble, James N., Heather G. Miller, Joseph A. Catania, Lance Pollack, and Charles F. Turner. 2000. "The Impact of T-ACASI Interviewing on Reported Drug Use among Men Who Have Sex with Men." *Substance Use and Misuse* 35:869-90.
- Guterbock, T.M., B.J. Meekins, A.C. Weaver, and J.C. Fries. 2000. "Web Versus Paper: A Mode Experiment in a Survey of University Computing." Paper presented at the Annual Meeting of the American Association for Public Opinion Research, Portland, OR, May 18-21.
- Heerwegh, D., and G. Loosveldt. 2002. "An Evaluation of the Effect of Response Formats on Data Quality in Web Surveys." *Social Science Computer Review* 20(4):471-84.
- Heerwegh, D., and G. Loosveldt. 2006. "An Experimental Study on the Effects of Personalization, Survey Length Statements, Progress Indicators, and Survey Sponsor Logos in Web Surveys." *Journal of Official Statistics* 22:191-210.
- Hewett, P.C., B. S. Mensch, and A. Erulkar. 2004. "Consistency in the Reporting of Sexual Behaviour by Adolescent Girls in Kenya: A Comparison of Interviewing Methods." *Sexually Transmitted Infections* 80(Suppl.2):43-8.
- Hochstim, J. R. 1967. "A Critical Comparison of Three Strategies of Collecting Data From Households." *Journal of the American Statistical Association* 62:976-89.
- Kalton, G., and C.F. Citro. 1993. "Panel Surveys: Adding the Fourth Dimension." *Survey Methodology* 19(2):205-15.
- Kalton, G., D. Kasprzyk, and D.B. McMillen. 1989. "Nonsampling Errors in Panel Surveys." In *Panel Surveys*, D. Kasprzyk, G. Duncan, G. Kalton, and M.P. Singh, eds., pp. 249-70. New York: Wiley.
- Kissinger, Patricia, Janet Rice, Thomas Farley, Shelly Trim, Kayla Jewitt, Victor Margavio, and David H. Martin. 1999. "Application of Computer-Assisted Interviews to Sexual Behavior Research." *American Journal of Epidemiology* 149(10):950-54.

- Knapp, H., and S.A. Kirk. 2003. "Using Pencil and Paper, Internet and Touch-Tone Phones for Self-Administered Surveys: Does Methodology Matter?" *Computers in Human Behavior* 19(1):117-34.
- Kreuter, F, S. Presser, and R. Tourangeau. 2008. "Social Desirability Bias in CATI, IVR, and WEB Surveys: The Effects of Mode and Question Sensitivity." *Public Opinion Quarterly* 72(5):847-65.
- Kwak, N., and B. Radler. 2002. "A Comparison Between Mail and Web Surveys: Response Pattern, Respondent Profile, and Data Quality." *Journal of Official Statistics* 18(2):257-74.
- Lensvelt-Mulders, G.J.L.M., P.G.M. van der Heijden, O. Laudy, and G. van Gils. 2006. "A Validation of a Computer-Assisted Randomized Response Survey to Estimate the Prevalence of Fraud in Social Security." *Journal of the Royal Statistical Society Series A-Statistics in Society* 169:305-18.
- Lesser, V.M., and L. Newton. 2001. "Mail, Email, and Web Surveys. A Cost and Response Rate Comparison in a Study of Undergraduate Research Activity." Paper presented at the Annual Meeting of the American Association for Public Opinion Research, Montreal, Canada, May 17-20.
- Lesser, V.M., and L. Newton, L. 2002. "Comparison of Response Rates and Quality of Response in a Survey Conducted by Mail, E-mail, and Web." Paper presented at the Annual Meeting of the American Association for Public Opinion Research, St. Pete Beach, FL, May 16-19.
- Mangione, T.W., R. Hingson, and J. Barrett. 1982. "Collecting Sensitive Data: A Comparison to Three Survey Strategies." *Sociological Methods Research* 10:337-46.
- Mathiowetz, Nancy, Mick P. Couper, and Eleanor Singer. 1994. "Where Does All the Mail Go? Mail Receipt and Handling in U.S. Households." Survey Methodology Program Working Paper No. 25. Ann Arbor: University of Michigan.
- Moskowitz, J.M. 2004. "Assessment of Cigarette Smoking and Smoking Susceptibility Among Youth—Telephone Computer-Assisted Self-Interviews Versus Computer-Assisted Telephone Interviews." *Public Opinion Quarterly* 68(4):565-87.
- Peytchev, A., M.P. Couper, S.E. McCabe, and S. Crawford. 2006. "Web Survey Design: Paging vs. Scrolling." *Public Opinion Quarterly* 70(4):596-607.
- Schoeber, Susan E., M. Fe Caces, Michael R. Pergamit, and Laura Branden. 1992. "Effect of Mode of Administration on Reporting of Drug Use in the National Longitudinal Survey." In *Survey Measurement of Drug Use: Methodological Studies*, Charles F. Turner, Judith T. Lessler, and Joseph C. Gfroerer, eds., pp. 267-76. DHHS Publication no. (ADM) 92-1929. Washington, DC: U.S. Department of Health and Human Services, Public Health Service, Alcohol, Drug Abuse, and Mental Health Administration.
- Talley, J.E., J.C. Barrow, K.F. Fulkerson, and C.A. Moore. 1983. "Conducting a Needs Assessment of University Psychological Services: A Campaign of Telephone and Mail Strategies." *Journal of American College Health* 32:101-03.

- Tedesco, H., R.L. Zuckerberg, and E. Nichols. 1999. "Designing Surveys for the Next Millennium: Web-Based Questionnaire Design Issues." In *Proceedings of the Third Association for Survey Computing International Conference*, pp. 103-12. Edinburgh, September 22-4.
- Tourangeau, R., M.P. Couper, and F. Conrad. 2007. "Color, Labels, and Interpretive Heuristics for Response Scales." *Public Opinion Quarterly* 71(1):91-112.
- Tourangeau, R., and T.W. Smith. 1996. "Asking Sensitive Questions: The Impact of Data Collection Mode, Question Format, and Question Context." *Public Opinion Quarterly* 60:275-304.
- Tourangeau, R., D. Steiger, and D. Wilson. 2002. "Self-Administered Questions by Telephone: Evaluating Interactive Voice." *Public Opinion Quarterly* 66:265-78.
- Walker, A.H., and J.D. Restuccia. 1984. "Obtaining Information on Patient Satisfaction With Hospital Care: Mail Versus Telephone." *Health Service Research* 19:291-306.
- Zapka, J.G., A.M. Stoddard, and H. Lubin. 1988. "A Comparison of Data From Dental Charts, Client Interview, and Client Mail Survey." *Medical Care* 26:27-33.
- Zuckerberg, A., E. Nichols, and H. Tedesco. 1999. "Designing Surveys for the Next Millennium: Internet Questionnaire Design Issues." Paper presented at the annual conference of the American Association for Public Opinion Research, St. Petersburg Beach, FL, May 16-19.

4. USE OF INCENTIVES

The use of incentives in surveys has been studied for decades. The literature surrounding the use of incentives details multiple dimensions that impact the effectiveness of incentives. These include theories on why incentives work, impact on response rate and nonresponse bias, prepaid versus postpaid incentives, and mode differences.

4.1 Theories on Incentive Effectiveness

Different reasons for the effectiveness of incentives have been provided in the literature. The theory of social exchange in the field of social psychology suggests a mechanism of social indebtedness, in response to which the individual cooperates with a survey request (Dillman, 1978). While social exchange would require that the sample member does not link the incentive to the survey request, a feature of the use of social exchange is that the incentive is rather small, so it is construed as a token of appreciation rather than a form of compensation for time and effort. This would suggest an incentive amount that has a small value. Kulka (1994) conducted an extensive overview of the existing literature and concluded that there was support for the belief that small monetary incentives increased response rates—a phenomenon largely attributed to social exchange.

Another reason for the effectiveness of incentives is more direct and can be described by theories such as economic exchange: an incentive is a form of compensation for participating in the survey. For some respondents, a particular compensation amount may be below a threshold level at the time of a survey request, but the higher the incentive, the more respondents decide to participate in the survey. Indeed, multiple studies have demonstrated that, for incentives, more is usually better. Trussel and Lavrakas (2004) examined the effect of incremental incentive increase in an experiment launched in a large-scale, mixed-mode survey. The levels of tested incentives ranged from \$0 to \$10. Consistent with previous findings, sending \$1 versus not sending an incentive at all resulted in higher response rates. The incremental increase in the incentive amount had a differential effect, depending on the outcome of the prior contact with the household. For households with positive outcome, it was not until the amount of \$5 was reached when the response and cooperation rates became significantly higher, relative to \$1. More interestingly, the \$7 to \$10 condition did not differ significantly from \$6. In contrast, in households that were never initially contacted or had negative outcome, each incremental dollar had a larger impact on response and cooperation than the previous dollar amount. The result suggests that when there is negative previous contact with the sample person, researchers should spend the maximum allowed in the budget on incentives.

Brick et al. (2005) compared the effectiveness of prepaid \$0, \$2, and \$5 incentives at various stages of a random-digit dialing (RDD) survey on educational topics. Brick et al. (2005) found that \$5 was more effective than \$2 in achieving initial cooperation, but the

relative effectiveness of the incentive (defined as the percentage point increase in the initial cooperation rate per dollar when compared to no incentive) was higher for the \$2 condition. Furthermore, incentives provided at the refusal conversion stage (a letter was mailed before calling) were more effective than incentives provided at the recruitment stage of the survey.

It is yet unknown whether respondents really construe a small incentive as a token of appreciation as opposed to a small amount of compensation, but in addition to the cognitive mechanism at play, a small token of appreciation can have a very different impact on survey costs compared to a larger compensation.

4.2 Impact on Response Rate and Nonresponse Bias

Offering respondent incentives is a demonstrated method to increase cooperation and response rates, but more importantly, it is also a method to decrease nonresponse bias. Sample members participate in surveys for various reasons. The leverage-salience theory (Groves, Singer, and Corning, 2000) posits that different people place different levels of importance on features of the survey request, such as the survey topic, survey sponsor, interview length, and so forth. Depending on what is made salient when the sample person is approached, the outcome of the survey request can be a refusal or an acceptance. For example, those less interested or involved in the survey topic can cooperate at a lower rate, leading to nonresponse bias in estimates based on the respondents. Incentives have been shown to increase cooperation *particularly* among sample persons with lower topic involvement. In a study that tests the theoretical framework based on the leverage-salience hypothesis, Groves, Singer, and Corning (2000) compared incentive and no-incentive treatments in a survey about political and community involvement. As expected, incentives significantly increased response rates. More interestingly, however, the effect of incentives was diminished for sample persons with high community involvement. Similar results were reported earlier by Baumgartner and Rathbun (1996), who found that monetary incentives increased cooperation more among those less interested in the survey topic. Such findings suggest that by attracting respondents who normally would not take part in the survey, incentives also changed the mix of sample persons who are measured, thus presenting a potential for reducing nonresponse error. However, in another test of the leverage-salience theory, Groves, Presser, and Dipko (2004) failed to find significant effects of monetary incentives in reducing the effect of topic interest on survey participation.

The link between response rates and nonresponse bias arises when there is a clear connection between response propensity and a survey variable of interest. The use of incentives may influence both the participation decision and survey variables. In a series of experiments launched to test whether those interested in the survey topic participate at higher rates and whether nonresponse bias on estimates of variables reflecting the survey topic was affected by this, Groves et al. (2006) also examined whether the use of incentives affected the link between topic interest and nonresponse bias. Incentives did not reliably

dampen the effect of topic interest, even though the results were in the hypothesized direction.

4.3 Prepaid Versus Postpaid Incentives

Another important factor when considering incentives is whether to offer them in advance, regardless of the sample person's decision to participate in the survey (prepaid), or after the respondent has agreed and completed the survey (promised). Some studies have found only prepaid incentives to be effective in reducing nonresponse in interviewer-administered surveys (Berk et al., 1987; Cantor et al., 1998; Singer, Van Hoewyk, and Maher, 2000), while Bosnjak and Tuten (2003) have found no difference between prepaid and promised incentives in web surveys.

Various studies have demonstrated the stronger effect of prepaid versus promised monetary incentives in mail surveys (for an overview, see Linsky, 1975; Armstrong, 1975). A meta-analysis of the experimental work on incentives in mail surveys by Church (1993) concluded that prepaid incentives yielded higher response rates than promised incentives or gifts sent with the initial mailing (65% average increase). Furthermore, it was concluded that an increase in the amount of money sent translated to an increase in response rates (but as Armstrong [1975] and Fox, Crask, and Kim [1998] suggest, at a decreasing rate).

Certain designs do not allow for prepaid incentives (e.g., most RDD surveys, or surveys of the whole household when the number of household members is unknown). In such cases, the amount offered may determine to a large degree the effectiveness of the incentive. For example, Cantor, Wang, and Abi-Habibm (2003) found an almost 10% increase in the response rate when promising \$20 (vs. no incentive) in an RDD survey of caregivers to children aged newborn to 17. Strouse and Hall (1997) recommend that for a survey to be successful, promised incentives have to be quite large (in the \$15 to \$35 range).

Promised incentives are fairly common at the refusal conversion stage. A number of studies have reported gains in response rates through offering relatively large amounts of money (\$25 or greater) at the end of the data collection period (e.g., Olson et al., 2004; Curtin, Presser, and Singer, 2005).

4.4 Incentives and Survey Mode

Comparison of the respondent conditions in self-administered versus interviewer-administered surveys suggests that the need for incentives will be greater in self-administered modes, where the persuasive presence of an interviewer is missing. In a meta-analysis that included face-to-face, telephone, and mixed-mode surveys, Singer et al. (1999) found that the effect of incentives was largely the same across modes. The results suggested that prepaid incentives yielded significant improvement in response rates, and

gifts were found to be significantly less effective than monetary incentives, even controlling for the value of the incentive.

It remains unknown whether nonmonetary incentives that appeal only to some respondents produce the same expected reduction in bias that is usually associated with monetary incentives. To an extreme, it is unclear whether such incentives may even induce bias in survey estimates—similarly to the bias induced through topic interest.

4.5 Summary

Despite these arguments and empirical findings, incentives may not be included in a study design due to their cost. Yet incentives can reduce the cost per case through the need for fewer interviewer call attempts to sample members and for the more costly refusal conversion attempts, as evidenced by the incentive experiments conducted for the National Survey on Drug Use and Health. The cost per interview in the \$20 group was 5% lower than the control; in the \$40 group, costs were 4% lower than the control. The cost savings were gained by interviewers spending less time trying to obtain cooperation from respondents (Kennet et al., 2005).

4.6 References

- Armstrong, J. Scott. 1975. "Monetary Incentives in Mail Surveys." *Public Opinion Quarterly* 39:111-16.
- Baumgartner, Robert, and Pamela Rathbun. 1996. "Prepaid Monetary Incentives and Mail Survey Response Rates." Paper presented at the Joint Statistical Meetings, Chicago, IL, Aug 4-8.
- Berk, M.L., N.A. Mathiowetz, E.P. Ward, and A.A. White. 1987. "The Effect of Prepaid and Promised Incentives: Results of a Controlled Experiment." *Journal of Official Statistics* 3(4):449-57.
- Bosnjak, Michael, and Tracy Tuten. 2003. "Prepaid and Promised Incentives in Web Surveys: An Experiment." *Social Science Computer Review* 21(2):208-17.
- Brick, Michael, Jill Montaquila, Mary Collins Hagedorn, Shelley Brock Roth, and Christopher Chapman. 2005. "Implications for RDD Design From an Incentive Experiment." *Journal of Official Statistics* 21(4):571-89.
- Cantor, D., Allen B., P. Cunningham, J.M. Brick, R. Slobasky, P. Giambo, and G. Kenny. 1998. "Promised Incentives on a Random Digit Dial Survey, Appearing in Non-Response in Survey Research." In *Proceedings of the Eighth International Workshop on Household Survey Non-Response*, A. Koch and R. Porst, eds., pp. 219-28. Mannheim, Germany: ZUMA.
- Cantor, D., K. Wang, and N. Abi-Habibm. 2003. "Comparing Promised and Prepaid Incentives for an Extended Interview on a Random Digit Dial Survey." Paper presented at the Annual Conference at the American Association for Public Opinion, Nashville, TN, May 15-18.

- Church, Allan H. 1993. "Estimating the Effect of Incentives on Mail Survey Response Rates: A Meta-Analysis." *Public Opinion Quarterly* 57:62-79.
- Curtin, R., S. Presser, and E. Singer. 2005. "Changes in Telephone Survey Nonresponse Over the Past Quarter Century." *Public Opinion Quarterly* 69:87-98.
- Dillman, Don A. 1978. *Mail and Telephone Surveys*. New York: Wiley.
- Fox, R.J., M.R. Crask, and J. Kim. 1988. "Mail Survey Response Rate: A Meta-Analysis of Selected Techniques for Inducing Response." *Public Opinion Quarterly* 52:467-91.
- Groves, Robert M., Mick P. Couper, Stanley Presser, Eleanor Singer, Roger Tourangeau, Giorgina Piani Acosta, and Lindsay Nelson. 2006. "Experiments in Producing Nonresponse Bias." *Public Opinion Quarterly* 70(5):720-36.
- Groves, Robert M., Stanley Presser, and Sarah Dipko. 2004. "The Role of Topic Interest in Survey Participation Decisions." *Public Opinion Quarterly* 68(1):2-31.
- Groves, Robert M., Eleanor Singer, and Amy Corning. 2000. "Leverage-Saliency Theory of Survey Participation—Description and an Illustration." *Public Opinion Quarterly* 64(3):299-308.
- Kennet, Joel, Joseph Gfroerer, Katherine R. Bowman, Peilan C. Martin, and David Cunningham. 2005. "Introduction of an Incentive and Its Effects on Response Rates and Costs in NSDUH." In *Evaluating And Improving Methods Used in the National Survey on Drug Use and Health*, J. Kennet and J. Gfroerer, eds., pp. 7-17. DHHS Publication No. SMA 05-4044, Methodology Series M-5. Rockville, MD: Substance Abuse and Mental Health Services Administration, Office of Applied Studies.
- Kulka, Richard. 1994. "The Use of Incentives to Survey 'Hard-to-Reach' Respondents: A Brief Review of Empirical Research and Current Research Practices." Paper presented at the Seminar on New Directions in Statistical Methodology, Council of Professional Agencies on Federal Statistics, Washington, DC, May 25-26.
- Linsky, Arnold. 1975. "Stimulating Responses to Mailed Questionnaires: A Review." *Public Opinion Quarterly* 39:82-101.
- Olson, L., M. Frankel, K.S. O'Connor, S.J. Blumberg, M. Kogan, and S. Rodkin. 2004. "A Promise or a Partial Payment: The Successful Use of Incentives in an RDD Survey." Paper presented at the Annual Conference of the American Association for Public Opinion Research, Phoenix, AZ, May 13-16.
- Singer, Eleanor, John Van Hoewyk, Nancy Gebler, Trivellore Raghunathan, and Katherine McGonagle. 1999. "The Effect of Incentives on Response Rates in Interviewer-Mediated Surveys." *Journal of Official Statistics* 15:217-30.
- Singer, E., J. Van Hoewyk, and M.P. Maher. 2000. "Experiments With Incentives in Telephone Surveys." *Public Opinion Quarterly* 64(2):171-88.
- Strouse, R.C., and J.W. Hall. 1997. "Incentives in Population Based Health Surveys." In *Proceedings of the American Statistical Association, Survey Research Section*, pp.952-57.

Trussel, Norm, and Paul Lavrakas. 2004. "The Influence of Incremental Increase in Token Cash Incentives on Mail Survey Response: Is There an Optimal Amount?" *Public Opinion Quarterly* 68(3):349-67.

5. ADDITIONAL ISSUES IN MEASURING CRIME VICTIMIZATION IN SURVEYS

Several methodological issues are of particular relevance to surveys collecting data on crime victimization, including problems with respondents' recalling and dating victimization incidents correctly, the use of proxy respondents, perceptions of crime severity, survey context, stigma, and terminology used in survey questions. The purpose of this document is to provide the Bureau of Justice Statistics with an overview of additional issues in measuring crime victimization surveys. This information will be used to inform the data collection methods project being conducted as part of the overall redesign of the National Crime Victimization Survey (NCVS).

5.1 Event Recall

Survey designers rely on respondents' recall when collecting reports of past behaviors and events. Accuracy of self-reports of past behaviors and autobiographical events is challenged by the failure to encode the event initially, telescoping (reporting of events outside the reference frame), or other sources of recall loss (Tourangeau, Rips, and Rasinski, 2000).

Not all encoded events are easily retrieved. Various studies have demonstrated that accuracy of responses to autobiographical questions depends on passage of time (Cannell, Miller, and Oksenberg, 1981; Loftus et al., 1992; Means et al., 1989; Smith and Jobe, 1994), length of reference period (for a meta-analysis, see Sudman and Bradburn, 1973), event salience characteristics (e.g., Thompson et al., 1996; Wagenaar, 1986), and question aids used to improve recall (e.g., Brewer, 1988; Wagenaar, 1986). Commonly used question aids are situational cues (e.g., physical context, date) and retrieval cues (e.g., examples of similar events). To improve crime report accuracy, the NCVS 1992 redesign introduced the *short-cue* screener strategy. The short-cue screener model attributed the failure to report crime incidents to a lack of conceptual question understanding, memory failure, or intentional misreporting; the redesign attempted to address the first two sources by using person and location reference frames and by increasing the number and variety of cues presented to the respondent. Preliminary tests of the short-cue screener yielded crime report rates 19% greater than the rates produced when the original screening questions were used (Martin et al., 1986). Several field tests conducted by the U.S. Census Bureau reported similar findings—significantly higher rates of violence and crime reporting for the short-cue screener group relative to the original screener group (Hubble, 1990). As expected, the introduction of the short-cue screener in 1992 yielded more reports of victimizations and captured types of crimes that were previously undetected (Rand, Lynch, and Cantor, 1997). It improved the measurement of traditionally underreported crimes (e.g., rape and aggravated assault) and crimes committed by family members and acquaintances (Kindermann, Lynch, and Cantor, 1997). The differences were largely

attributed to explicit cueing of certain crime types (e.g., rape and sexual assaults) and the addition of two reference frames to aid recall: the first, related to crimes committed by someone the respondent knew; the second, related to the location of the crime (U.S. Bureau of the Census, 1994).

In the search for strategies that improve recall, a number of studies have examined the issues of *forward telescoping*, the reporting of events that occurred prior to the reference period (e.g., Neter and Waksberg, 1964; Loftus and Marburger, 1983; Brown, Rips, and Shevell, 1985; Loftus et al., 1990), and *backward telescoping*, the reporting of events that occurred after the reference period (e.g., Sudman and Bradburn, 1973; Means et al., 1989). One of the design strategies used to reduce telescoping in panel surveys and employed by NCVS is *bounded recall* (Neter and Waksberg, 1964), a technique where the responses from the first interview are used to anchor responses from following interviews.

Another approach to assist event recall is known as *anchoring*, which uses events such as holidays, major public events, personal landmarks, and so forth (Linton, 1975; Loftus and Marburger, 1983; Brown, Shevell and Rips, 1986; Means et al., 1989). Yet another approach is to vary the length of the reference period depending on how salient and rare the event is judged by the researcher, the premise being that longer reference periods can be used for rare and salient events (Sudman and Bradburn, 1974; Mathiowetz, 1988; Warner et al., 2005). An examination of recall biases in NCVS revealed that rates of victimization decreased significantly as the length of the reference period increased (Bushery, 1981). In a reverse record-check study of victims of robbery, burglary, and assault, Czaja et al. (1994) found that the length of reference period and anchoring did not affect victimization rates; however, both factors influenced reports of victimization dates.

Furthermore, Event History Calendars (EHCs) have been employed to facilitate recall. EHCs facilitate the use of all memory retrieval mechanisms (top-down, sequential, and parallel). Such calendars rely on inherent cueing mechanisms: noteworthy events can be dated precisely and used as landmarks for other events; events remembered in one life domain can cue events that happened in another; and inconsistencies can be spotted easily and addressed. Freedman et al. (1988) found almost 90% agreement between monthly reports in EHC for events that occurred 5 years prior and validation data. Similar rates were reported by Caspi et al. (1996) when retrospective reports were matched to concurrent reports 3 years prior. Further, Belli, Shay, and Stafford (2001) found that EHC led to better-quality retrospective reports on key social and economic events measured by the Panel Survey of Income Dynamics.

5.2 Proxy Respondents

Many surveys use proxy respondents when the sample member is not available for an interview. Proxy reports offer time and cost savings, but they often do so at the price of

data quality. The validity of a proxy report depends largely on the relationship of the proxy to the respondent, the saliency of the event being reported, and the proxy's knowledge of the event. Cantor and Lynch (2000) discussed the results of a pilot test of NCVS that found far greater reporting of victimization with self-reports compared with proxy reports. Such findings are consistent with other studies comparing self-reports and proxy reports in other surveys (e.g., Hyland et al., 1997; Perruccio and Badley, 2004; Rajmil et al., 1999).

5.3 Crime Severity, Survey Context, Stigma, and Terminology

Respondents may not report smaller, less severe crimes, such as simple assault (attack without a weapon resulting in minor or no injury), because they may not believe the incident was serious enough to be considered a crime. Respondents may fail to recall the incident or may choose not to report it due to the perceived ambiguity of the crime. Crimes committed by nonstrangers (e.g., family members, intimates, acquaintances) may also be underreported for this reason (Kinderman, Lynch, and Cantor, 1997).

Surveys that measure victimization outside the context of criminal behavior, such as the National Violence Against Women Survey, have produced higher estimates of rape, domestic violence, and assault than the crime-focused NCVS. The contextual differences in the surveys may contribute to the different estimates of victimization because respondents are less likely to report incidents to NCVS that they do not consider to be criminal (Rand and Rennison, 2004). Additionally, the social and cultural stigmas attached to rape and domestic violence may result in underreporting.

Much attention has been given to the measurement of rape, including wording in survey questions (Fisher and Cullen, 2000). Research has demonstrated that the terms used and the specificity of questions can influence victimization reports. Different terms used to ask about sexual victimization may have different meanings to different respondents and, as a result, may influence respondents' understanding of the question and, ultimately, their reporting (Hamby and Koss, 2003). The use of legal terms may also impede comprehension. Behaviorally specific questions and specific descriptions of sexual acts produce higher rates of sexual victimization than the use of legalistic terms such as "rape" and "sexual assault" (Fisher, 2004; Hamby and Koss, 2003).

In addition, the way a crime is enumerated affects the accuracy of the survey estimates. For example, repeated victimizations are common in cases of domestic violence. Concerns have been raised on how to accurately count repeated victimizations or series of victimizations. NCVS counts six or more similar victimizations that happened within a 6-month reference period as one incident (based on the most recent incident). Other surveys that calculate each incident separately produce higher estimates (e.g., Rand and Rennison, 2005).

5.4 Summary

A wide body of research addresses issues relevant to collecting valid survey data on crime victimization. The length of the reference period, questionnaire design aids used to improve recall, proxy respondents, perceptions of crime, stigma associated with the crime, and the choice of words in survey questions are among the factors that can affect the accuracy of crime reports. Careful consideration of such features at the survey design stage and the selection of the most appropriate mode of data collection may drastically improve report accuracy.

5.5 References

- Belli, Robert F., William L. Shay, and Frank P. Stafford. 2001. "Event History Calendars and Question List Surveys: A Direct Comparison of Interviewing Methods." *Public Opinion Quarterly* 65:45-74.
- Brewer, W.F. 1988. "Memory for Randomly Sampled Autobiographical Events." In *Remembering Reconsidered: Ecological and Traditional Approaches to the Study of Memory*, U. Neisser and E. Winograd, eds., pp. 21-90. Cambridge: Cambridge University Press.
- Brown, N.R., L.J. Rips, and S.M. Shevell. 1985. "Subjective Dates of Natural Events in Very-Long-Term Memory." *Cognitive Psychology* 17:139-77.
- Brown, N.R., Shevell, S.K, and Rips, L.J.(1986). Public Memories and Their Personal Context, In *Autobiographical Memory*, D.G. Rubin (ed.). Cambridge: Cambridge University Press.
- Bushery, J.M. 1981. "Recall Biases for Different Reference Periods in the National Crime Survey." In *Proceedings of the Section on Survey Research Methods, American Statistical Association*, pp. 238-43.
- Cannell, C., P. Miller, and L. Oksenberg. 1981. "Research on Interviewing Techniques." In *Sociological Methodology 1981*, S. Leinhardt, ed., pp. 389-437. San Francisco: Jossey-Bass.
- Cantor, David, and James P. Lynch. 2000. "Self-Report Surveys as Measures of Crime and Criminal Victimization." *Measurement and Analysis of Crime and Justice* 4:85-138.
- Caspi, A., T.E. Moffitt, A. Thornton, D. Freedman, J.W. Amel, H. Harrington, J. Smeijers, and P.A. Silva. 1996. "The Life History Calendar: A Research and Clinical Assessment Method for Collecting Retrospective Event-History Data." *International Journal of Methods in Psychiatric Research* 6:101-14.
- Czaja, Ronald, Johnny Blair, Barbara Bickart, and Elizabeth Eastman. 1994. "Respondent Strategies for Recall of Crime Victimization Incidents." *Journal of Official Statistics* 10(3):257-76.
- Fisher, B.S. 2004. *Measuring Rape Against Women: The Significance of Survey Questions*. NCJ 199705. Washington, DC: U.S. Department of Justice, National Institute of Justice.

- Fisher, B.S., and F.T. Cullen. 2000. "Measuring the Sexual Victimization of Women: Evolution, Current Controversies and Future Research." In *Criminal Justice 2000, Vol. 4: Measurement and Analysis of Crime*, D. Duffee, ed., pp. 317-90. Washington, DC: U.S. Department of Justice, National Institute of Justice.
- Freedman, D., A. Thornton, D. Camburn, D. Alwin, and L. Young-DeMarco. 1988. "The Life History Calendar: A Technique for Collecting Retrospective Data." In *Sociological Methodology*, C. Clogg, ed., pp. 37-68. New York: Academic Press.
- Hamby, S.L., and M.P. Koss. 2003. "Shades of Gray: A Qualitative Study of Terms Used in the Measurement of Sexual Victimization." *Psychology of Women Quarterly* 27(3):243-55.
- Hubble, D.L. 1990. "National Crime Survey New Questionnaire Phase-in Research: Preliminary Results." Paper presented at the International Conference on Measurement Errors in Surveys, Tucson, AZ, November 11-14.
- Hyland A., K.M. Cummings, W.R. Lynn, D. Corle, and C.A. Giffen. 1997. "Effect of Proxy-Reported Smoking Status on Population Estimates of Smoking Prevalence." *American Journal of Epidemiology* 145(8):746-51.
- Kinderman, J.L., J. Lynch, and D. Cantor. 1997. *Effects of the Redesign on Victimization Estimates*. NCJ 164381. Washington, DC: U.S. Department of Justice, Bureau of Justice Statistics.
- Linton, M. 1975. "Memory for Real World Events." In *Explorations in Cognition*, D.A. Norman and D.E. Rumelhart, eds., pp. 376-404. San Francisco: Freeman.
- Loftus, E.F., M.R. Klinger, K.D. Smith, and J. Fiedler. 1990. "Tale of Two Questions." *Public Opinion Quarterly* 54:330-35.
- Loftus, E.F., and W. Marburger. 1983. "Since the Eruption of Mt. St. Helens, Has Anyone Beaten You Up? Improving the Accuracy of Retrospective Reports With Landmark Events." *Memory and Cognition* 11:114-20.
- Loftus, E.F., K.D. Smith, M.R. Klinger, and J. Fiedler. 1992. "Memory and Mismemory for Health Events." In *Questions About Questions: Inquiries Into the Cognitive Basis of Surveys*, J.M. Tanur, ed., pp. 102-37. New York: Russell Sage Foundation.
- Martin, E., R.M. Groves, V.J. Matlin, and C. Miller. 1986. *Report on the Development of Alternative Screening Procedures for the National Crime Survey*. Washington, DC: Bureau of Social Science Research.
- Mathiowetz, N.A. 1988. "Forgetting Events in Autobiographical Memory: Findings From a Health Care Survey." In *Proceedings of the Section on Survey Research Methods, American Statistical Association*, pp. 167-72.
- Means, B., A. Nigam, M. Zarrow, E.F. Loftus, and M. Donaldson. 1989. *Autobiographical Memory for Health-Related Events Series 6, No. 2*. Department of Health and Human Services (DHHS) Publication No. (PHS) 89-1077. Washington, DC: U.S. Government Printing Office.

- Neter, J., and J. Waksberg. 1964. "A Study of Response Errors in Expenditures Data From Household Interview." *Journal of the American Statistical Association* 59:18-55.
- Perruccio, A.V., and E.M. Badley. 2004. "Proxy Reporting and the Increasing Prevalence of Arthritis in Canada." *Canadian Journal of Public Health* 95(3):169-73.
- Rajmil, L., E. Fernandez, R. Gispert, M. Rue, J.P. Glutting, A. Plasencia, and A. Segura. 1999. "Influence of Proxy Respondents in Children's Health Interview Surveys." *Journal of Epidemiology and Community Health* 53(1):38-42.
- Rand, M., J. Lynch, and D. Cantor. 1997. *Criminal Victimization, 1973-95*. NCJ-163069. Washington, DC: Bureau of Justice Statistics.
- Rand, M.R., and C.M. Rennison. 2004. *How Much Violence Against Women Is There?* NCJ 199702. Washington, DC: U.S. Department of Justice, National Institute of Justice.
- Rand, M.R., and C.M. Rennison. 2005. "Bigger Is Not Necessarily Better: An Analysis of Violence Against Women Estimates From the National Crime Victimization Survey and the National Violence Against Women Survey." *Journal of Quantitative Criminology* 21(3):267-91.
- Smith, A.F., and J.B. Jobe. 1994. "Validity of Reports of Long-Term Dietary Memories: Data and a Model." In *Autobiographical Memory and the Validity of Retrospective Reports*, N. Schwarz and S. Sudman, eds., pp. 121-40. Berlin: Springer.
- Sudman, S., and N. Bradburn. 1973. "Effects of Time and Memory Factors on Response in Surveys." *Journal of the American Statistical Association* 68:805-15.
- Sudman, S., and N.M. Bradburn. 1974. *Response Effects in Surveys: A Review and Synthesis*. Chicago: Aldine.
- Thompson, C.P., J.J. Skowronski, S.F. Larsen, and A.L. Betz. 1996. *Autobiographical Memory*. Mahwah, NJ: Erlbaum.
- Tourangeau, Roger, Lance J. Rips, and Kenneth Rasinski. 2000. *The Psychology of Survey Response*. Cambridge: Cambridge University Press.
- U.S. Bureau of the Census. 1994. *The National Crime Victimization Survey(NCVS) Redesign: Technical Background*. NCJ 151172, Washington, D. C.: Bureau of Justice Statistics.
- Wagenaar, W.A. 1986. "My Memory: A Study of Autobiographical Memory Over Six Years." *Cognitive Psychology* 18:225-52.
- Warner, M., N. Schenker, M.A. Heinen, and L.A. Fingerhut. 2005. "The Effects of Recall on Reporting Injury and Poisoning Episodes in the National Health Interview Survey." *Injury Prevention* 11:282-87.