Discussion of Plans for Designing the Recall Period for the Consumer Expenditure Interview Survey

David Cantor
Westat and
Joint Program for Survey Methodology

Outline

• General Issues
• Key aspects of Recall Period
• Key tradeoffs to balance
• Best practices for designing recall period
General Issue - Motivation

• Recall task is among the hardest for a survey respondent
  – Requires time to search memory. Need to prevent premature denial of eligible events.
  – May need to try different search strategies

• Motivation to complete these tasks has dramatic effects on data quality
  – Success of the Event History Calendar (Belli and Callegaro, 2008)
  – Cognitive interview debriefing
Table 1. Comparison of estimated NCS and NCVS victimization rates, 1992

<table>
<thead>
<tr>
<th>Number of victimizations per 1,000 persons or households</th>
<th>Post-redesign NCVS</th>
<th>Pre-redesign NCS</th>
<th>NCVS/NCS ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Personal crimes</td>
<td>49.6</td>
<td>34.4</td>
<td>1.44</td>
</tr>
<tr>
<td>Crimes of violence</td>
<td>47.8</td>
<td>32.1</td>
<td>1.49</td>
</tr>
<tr>
<td>Rape</td>
<td>1.8</td>
<td>.7</td>
<td>2.57</td>
</tr>
<tr>
<td>Robbery</td>
<td>6.1</td>
<td>5.9</td>
<td>1.03</td>
</tr>
<tr>
<td>Assault</td>
<td>40.0</td>
<td>25.5</td>
<td>1.57</td>
</tr>
<tr>
<td>Aggravated</td>
<td>11.1</td>
<td>9.0</td>
<td>1.23</td>
</tr>
<tr>
<td>Simple</td>
<td>28.9</td>
<td>16.5</td>
<td>1.75</td>
</tr>
<tr>
<td>Personal theft</td>
<td>1.8</td>
<td>2.4</td>
<td>.75</td>
</tr>
<tr>
<td>Property crimes</td>
<td>325.3</td>
<td>264.5</td>
<td>1.23</td>
</tr>
<tr>
<td>Household burglary</td>
<td>58.6</td>
<td>48.9</td>
<td>1.20</td>
</tr>
<tr>
<td>Household theft</td>
<td>248.2</td>
<td>195.5</td>
<td>1.27</td>
</tr>
<tr>
<td>Motor vehicle theft</td>
<td>18.5</td>
<td>20.1</td>
<td>.92</td>
</tr>
</tbody>
</table>

*The ratio of the NCVS to the NCS estimates was statistically significant at the 90-percent level of confidence.*

Burden Affects Motivation

• Current CE = 1 hour?
  – Mock interview took 2.5 hours
  – Significant time is needed to review general and specific questions

• Important to set realistic limit
  – Depends on methods to promote recall (e.g., use of records), use of proxy interviewing
  – 1 hour is upper limit (probably too long)
    • Cognitive interview and focus group experience
General Issue - Interviewers

• Interviewers play a key role in promoting recall
• Natural tension between gaining cooperation and effective probing
  – Current timings: 1 hour CE interview involves significant shortcuts on probing.
  – Panel context – respondent has been exposed to questionnaire. This may encourage shortcuts
• Interviewer variance studies for CE (Cho, et al, 2006; others?)
NCS Redesign: Monitoring Matters

• Biderman et al., 1985. CATI interviews conducted by SRC (Michigan) produced victimization rates twice as high as equivalent Census field interviews.

• Hubble and Wilder, 1988. Census CATI interviews produced significantly higher rate than Census Field interviews (see next slides).

• Effects seemed to have carried over to production NCVS (Cantor and Lynch, 2005).
## Comparison of Field and CATI Telephone Interviews

<table>
<thead>
<tr>
<th>Type of Crime</th>
<th>De-centralized</th>
<th>CATI</th>
<th>Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>All Personal</td>
<td>73.9</td>
<td>115.5</td>
<td>1.56**</td>
</tr>
<tr>
<td>Robbery</td>
<td>3.0</td>
<td>9.3</td>
<td>3.10**</td>
</tr>
<tr>
<td>Assault</td>
<td>17.5</td>
<td>24.2</td>
<td>1.38**</td>
</tr>
<tr>
<td>Theft</td>
<td>53.4</td>
<td>81.9</td>
<td>1.53**</td>
</tr>
<tr>
<td>All Household</td>
<td>161.1</td>
<td>198.9</td>
<td>1.23**</td>
</tr>
<tr>
<td>Burglary</td>
<td>54.1</td>
<td>73.4</td>
<td>1.36**</td>
</tr>
<tr>
<td>Theft</td>
<td>78.4</td>
<td>99.6</td>
<td>1.27**</td>
</tr>
<tr>
<td>Motor Vehicle Theft</td>
<td>28.6</td>
<td>25.9</td>
<td>ns</td>
</tr>
</tbody>
</table>

** p<.05

Centralized vs Decentralized Interviewing; 3 vs 6 Month Reference periods

<table>
<thead>
<tr>
<th>Type of Crime</th>
<th>Ratio for CATI and Decentralized Telephone</th>
<th>Ratio of 3 and 6 month recall periods</th>
</tr>
</thead>
<tbody>
<tr>
<td>All Personal</td>
<td>1.56**</td>
<td>1.22**</td>
</tr>
<tr>
<td>Robbery</td>
<td>3.10**</td>
<td>ns</td>
</tr>
<tr>
<td>Assault</td>
<td>1.38**</td>
<td>1.28**</td>
</tr>
<tr>
<td>Theft</td>
<td>1.53**</td>
<td>1.19**</td>
</tr>
<tr>
<td>All Household</td>
<td>1.23**</td>
<td>1.16**</td>
</tr>
<tr>
<td>Burglary</td>
<td>1.36**</td>
<td>ns</td>
</tr>
<tr>
<td>Theft</td>
<td>1.27**</td>
<td>1.18**</td>
</tr>
<tr>
<td>Motor Vehicle Theft</td>
<td>ns</td>
<td>ns</td>
</tr>
</tbody>
</table>

** p<.05

Monitor Interviewers

• Use of CARI
  – Census Bureau is developing this capability
  – Provide systematic and timely feedback

• Other paradata (e.g., timings)
Key Aspects of the Recall Period

• Method of Recall
  – Methods for cueing
  – Role of the interviewer
  – Use of visual aids (calendar – Belli and Callegaro, 2008; attention to format - Redline, et al, 2009)

• Bounded vs Unbounded recall period
  – Effects of bounding by recall period
  – Dependent vs independent bounding
  – Internal telescoping
Sample size, Low Incidence Items and Shorter Recall Period

- Explore which items can be collected using a longer reference period
  - Large variation in data quality by type of item and length of period (Neter and Waksberg, 1963)
  - What is optimum length for main items?
- Increase collection using self-administered methods to reduce the cost of collection
  - Shift more items to Diary Survey
  - Shorten reference period for interview survey for particular types of items.
Alternative Designs

• Two independent samples:
  – Large, salient purchases use a long, unbounded, reference period
  – Separate sample with shorter, bounded, reference periods for less salient purchases.

• Single survey with one long reference period.
  – Mix reference periods, depending on purchase

• Single Panel Design: Monthly interviews followed by a longer reference period.
What is the Tradeoff Between Variance and Bias?

• For fixed costs, it will be difficult to shorten the reference period without sacrifice of precision.

• What are the minimum precision requirements?
  – Use prior research to assess tradeoffs
  – Evaluate from experiments conducted as part of the redesign
Table 3.1. Estimated yearly percent change in the crime rate required to be 90% sure that the NCS rate will show a significant (95%) difference by recounting period length and type of crime.

<table>
<thead>
<tr>
<th></th>
<th>Recounting Period Length</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>3</td>
</tr>
<tr>
<td>Rape</td>
<td>60.2</td>
</tr>
<tr>
<td>Robbery</td>
<td>23.2</td>
</tr>
<tr>
<td>Aggravated Assault</td>
<td>18.7</td>
</tr>
<tr>
<td>Simple Assault</td>
<td>14.8</td>
</tr>
<tr>
<td>Personal Theft</td>
<td>5.9</td>
</tr>
<tr>
<td>Burglary</td>
<td>9.1</td>
</tr>
<tr>
<td>Household Larceny</td>
<td>7.5</td>
</tr>
<tr>
<td>Motor Vehicle Theft</td>
<td>21.2</td>
</tr>
</tbody>
</table>
Other Tradeoffs to Consider?

• Respondent burden increases with length of recall period
  – May need to ask about fewer items
  – Enhancing recall may increase burden

• Panel conditioning – does it increase with shorter reference periods?
  – Prior research (Silberstein and Jacobs, 1989; Cho, et al., 2004)) did not find large effects of time-in-sample
  – Will this change if shorter periods are used?
Best Practices: Basics

• Develop appropriate interviewing protocol:
  – Interview structure; interviewer procedures and methods to monitor; Visual and other aids (Conrad; Peytchev; Stafford)
  – Use of proxy interviewing (Mathiowetz; Schaeffer)

• Analyze existing data. Examine recency curves by type of item (Silberstein and Jacobs, 1989; Steinberg, et al., 2006; Biderman and Lynch, 1981; Fay and Li, 2010).
  – Some guide to recall effects by type of item

• Conduct scale tests comparing different recall periods
  – Small scale could be done in the lab (easier to get validation data)
  – Large scale test, under field conditions (Neter and Waksberg, 1964; Sudman and Ferber, 1971; Chu, et al., 1992)
Best Practices: Assess Tradeoffs

• Non-response bias vs. measurement error
  – Non-response is not directly related to bias (Groves, 2006; Keeter, et al., 2006)
  – Efforts to reduce measurement error may lead to higher non-response (e.g., use of proxy interviewing; more complete interviewer protocols)

• Differential effects of recall error by population group (Kobilarcik, et al., 1983; Cantor, 1986)

• Do you need to test effects on level vs change?
  – Effects of recall period may be less on measures of change
Validation is Needed to Assess Tradeoffs

• Small scale experiments:
  – Collect records
  – Intensive debriefing

• Larger scale
  – Reverse record checks. Sample from retail records (especially large purchases). Perhaps combine this with selective forward record check methods
  – Have respondent keep a diary and/or conduct short interviews on a regular basis (Rips, et al., 2003; Millen, et al., 2005))

• Randomization should reduce the impact of these errors on assessment of comparative validity
Methods of Validation: Is More Better?

• General belief there is under-reporting
  – Comparison to National Accounts
  – Prior research has made this assumption (Neter and Waksberg, 1964; Sudman and Ferber, 1971)

• Validity of assumption is likely to vary by type of purchase
  – Dangerous for small purchases that are re-occurring using rule-based recall
  – Enhanced protocols may lead to more over-reporting
Summary

• Initial efforts should be put into improving the interviewing protocol, including interviewer monitoring and training

• When investigating optimum recall period
  – Compare tradeoffs between variance, non-response bias and measurement error
  – Assess possible differential bias across population groups
  – Is it important to test for bias in change estimates?

• Include external validation criterion when evaluating optimum periods
Thank-you
DavidCantor@Westat.com