

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

David Cantor  
Westat and  
Joint Program for Survey Methodology

Prepared for the 2010 Consumer Expenditure Surveys Methods Workshop,  
December 8-9, 2010.

# 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**

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

<sup>‡</sup> 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

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

\*\* p<.05

Source: Hubble, D. and B.E. Wilder (1988) "Preliminary results from the National Crime Survey CATI Experiment." Proceedings of the American Statistical Association: Survey Methods Section.

# Centralized vs Decentralized Interviewing; 3 vs 6 Month Reference periods

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

\*\* p<.05

Source: Hubble, D. and B.E. Wilder (1988); Kobilarcik, et al., (1983);

# 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.

	Recounting Period Length				
	3	4	6	9	12
Rape	60.2	54.1	47.8	39.8	35.0
Robbery	23.2	20.9	18.4	15.4	13.5
Aggravated Assault	18.7	16.8	14.9	12.4	10.9
Simple Assault	14.8	13.3	11.8	9.8	8.6
Personal Theft	5.9	5.3	4.6	4.0	3.6
Burglary	9.1	8.1	6.9	5.9	5.3
Household Larceny	7.5	6.7	5.8	5.0	4.5
Motor Vehicle Theft	21.2	18.8	15.9	13.7	12.4

# 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 assessments 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