CHAPTER 13

Identifying and Addressing Response Errors in Self-Report Surveys

JAMES P. LYNCH AND LYNN A. ADDINGTON

INTRODUCTION

Much of the data used by criminologists is generated by self-report surveys of victims and offenders.[1] Although both sources share a common reliance on responses to questions, little overlap exists between the two traditions. They have largely talked past each other, especially with regard to sharing methodological studies and findings. At first glance, this miscommunication appears a bit strange since the basic cognitive task required for both sets of respondents is the same: report on criminal activity in which they have been involved. The division between the two traditions makes greater sense upon examination of their initial motivating goals and auspices.

Self-report offending surveys originated from a desire to identify predictors of offending in order to develop and test theories to explain the differences in criminal behavior across individuals. To achieve these goals, criminologists designed survey instruments that focused on individuals and the relative volume of crime across offenders. Researchers concentrated their methodological efforts on ways to maximize information collected about criminal offenders and their social context (see Thornberry and Krohn 2000; and Junger-Tas and Marshall 1999 for discussions of these issues). In contrast, the principal purpose of victimization surveys was to estimate the level and change in the level of crime across periods of time, which included generating annual crime rates. This goal required a focus on the criminal victimization incident rather than on an individual person. As a result, researchers emphasized methodological issues that maximized the accuracy of recalling and reporting crime events.

The distinct motivations for collecting data and the primary uses of the data have contributed to different interests in locating crime events in space and time. For example,

[1] We use the terms "self-report offending surveys" and "offender surveys" as well as "self-report victimization surveys" and "victimization surveys" to refer to data collection efforts that rely on respondents reporting their own experiences with crime (as either an offender or victim).

A.R. Piquero and D. Weisburd (eds.), *Handbook of Quantitative Criminology*,
DOI 10.1007/978-0-387-77650-7_13,

researchers use self-report offending surveys to identify offenders, distinguish high and low volume offenders, and understand why some people become offenders and others do not. Such concerns place less emphasis on the precise counting of crime events, whereas more on obtaining individual and contextual information to explain offending patterns. Self-report victimization surveys, on the other hand, are used to estimate crime rates. This task requires exhaustive reporting of target events and precise dating of their occurrence to ensure accurate rates. Victimization surveys also are frequently used to check crime data reported by police. This additional role requires measuring the precise location of events by jurisdiction as well as enough detail about events to reproduce common types of crime classifications used by the police.

These diverse interests and tasks across traditions produce differences in the design and emphasis among the major self-report offending and victimization surveys. Until recently these parallel worlds were able to operate separate and distinct from one another with little reason to commingle. Currently, changes in how these data are being used – especially with the self-report offending surveys – necessitate a re-examination of this division. Self-report offending data are now being used to study offending careers and career criminals using trajectory models.[2] The sophisticated statistical models used in this work assume a precision of the data and how the criminal event is reported. These new demands on self-report offending surveys are beginning to resemble the long-standing requirements of victimization surveys. For example, the use of trajectories and growth curves to distinguish the types of criminal careers now makes exhaustive reporting and the temporal placement of offenses more important (Thornberry 1989; Lauritsen 1999). In addition, the desire to separate serious from less serious crime events requires that the respondent provide more detail on the nature of offending events so that researchers can make such distinctions (Thornberry and Krohn 2000). With this increasing overlap in the need for both types of surveys to accurately locate events in time and space, the methodological work conducted in the context of victimization surveys becomes a relevant resource for self-report offending surveys. Research conducted in the victimization context can identify areas where offending surveys might not be providing enough precision and suggest ways to improve data accuracy.

In this chapter, we seek to highlight the methodological work conducted in the context of victimization surveys to suggest how this research can benefit self-report offending surveys. In this way, we can identify major response errors in both traditions. Before beginning this discussion, three caveats are important to mention.

First, our discussion is limited to response errors, that is, errors having to do with the performance of the cognitive task required of respondents and the recording of their responses. Errors resulting from sampling or nonresponse are not addressed. Our focus is restricted to response error because the cognitive task required in surveys of victimization and offending is similar. Both traditions require respondents to search their memory for crimes in which they were involved, to accurately locate these events in time, and to recount details of the event so that they can be classified. Other aspects of victim and offender surveys may vary more.

Second, our discussion reflects a reliance on a few main victimization and self-report offending surveys. One reason is practical as the number of self-report surveys of offending

[2] In particular, the use of sophisticated statistical tools such as latent growth curve analysis or semi-parametric group-based modeling (SPGM) of trajectories has increased dramatically in the past decade. Piquero (2008), identified over 80 studies that have employed these methods to analyze criminal activity over the life course, and many of these studies use data from self-report surveys of offending.

and victimization has exploded since the 1960s (Thornberry and Krohn 2000; Junger-Tas and Marshall 1999; Lynch 2006). Another reason is that each tradition has a handful of prominent surveys that serve as the benchmark for designing other surveys. These exemplars tend to be large-scale surveys where a great deal of time and money has been invested on design issues. As a result, the data are viewed as being of very high quality and are widely used among criminologists. For victimization surveys, our discussion relies most heavily upon the National Crime Victimization Survey (NCVS) and its predecessor the National Crime Survey (NCS) and to a lesser extent on the British Crime Survey (BCS).[3] In terms of surveys of offending, three leading surveys shape our paradigms. These include the National Youth Survey (NYS), the Cambridge Youth Study (CYS), and the Program of Research on the Causes and Correlates of Delinquency (PRCCD).

Finally, by taking this approach, we are not implying that no methodological work has been done in the self-report offending area or that victimization surveys cannot benefit from this literature. Rather we believe that communication across the two traditions is long overdue, and this chapter is a first attempt at cross-fertilization. We begin this dialog with what victimization surveys have to say to self-report offending surveys because we are more familiar with the victim survey tradition. We welcome a corresponding article that addresses the methodological work from the self-report offending tradition that can improve victimization surveys. Topics such as panel bias, use of calendrical devices, strategic response, and development of age-appropriate interviews have all received much more attention in surveys of offending than in those of victimization (Lauritsen 1998, 1999; Horney and Marshall 1992; Mennard and Elliot 1990; Thornberry 1989).

The remainder of this chapter is organized as follows. First, we identify the main sources of response errors studied in victimization surveys that are pertinent to surveys of offending. Each of these response errors is explored in turn. For each type of error, we discuss what is known about the error from the self-report victimization survey literature, we describe how surveys of offending have approached this issue, and we examine how research from self-report victimization surveys can benefit self-report offending surveys. We conclude by using the juxtaposition of the two survey traditions to suggest how to improve our understanding of response error in self-report surveys. We also recommend interim ways to address response errors based on currently available information.

SOURCES OF RESPONSE ERRORS IN VICTIMIZATION SURVEYS

The original development and subsequent redesign of the National Crime Survey (NCS) relied upon extensive methodological work concerning response and other types of survey error. In developing the NCS, two factors motivated the emphasis on and investment in methodological work. The primary reason was the federal government's reliance on victimization rates as a social indicator similar to the unemployment rate and its use of the Census Bureau to administer the survey (Biderman and Reiss 1967). As a result, survey statisticians at the Census Bureau, rather than criminologists or other social scientists, controlled the development of these surveys. As with other Census surveys, the Bureau devoted a significant amount

[3] Two smaller victimization surveys are of note with regard to their work on improving our understanding of recall and screening with regard to sexual assault and rape. These two surveys are the National Violence Against Women Survey (Tjaden and Thoennes 1998) and the Campus Survey of Sexual Violence (Fisher and Cullen 2000).

of attention to issues of sampling and measurement errors. Addressing these sources of error figured prominently in the NCS's design. A second reason for the emphasis on methodological work was more political. Leaders at the Census Bureau and the Justice Department had a keen interest in ensuring that the design of the NCS was defensible, since groups such as the powerful police lobbies viewed victimization surveys with suspicion.

Even after the NCS was developed, methodological work continued (Lehnen and Skogan 1981). This work combined with the National Academy of Sciences' influential publication *Surveying Crime* (Penick and Owens 1976) prompted the NCS Redesign. The Redesign constituted a multi-year, multi-million dollar effort to identify the error structure of the NCS and modify the survey's design to reduce errors (Biderman et al. 1986). The resulting changes were implemented in the National Crime Victimization Survey (NCVS). The initial development work and redesign coupled with a continuous stream of research from the Census Bureau as well as more modest efforts from the academic community generated a rich base of methodological work (Cantor and Lynch 2000). A comparable resource is not available with regard to self-report offending surveys, especially with regard to response errors. In addition to the different organizing principles mentioned earlier, the resources available to the self-report offending tradition have been more meager. In particular, the self-report offending tradition has lacked large and continuous government-sponsored national surveys.[4]

This extensive methodological research in the victimization context identified a number of response errors as well as determined the effect these errors have on estimates and the need to ameliorate the errors or adjust the results (if the errors cannot be eliminated). Three major sources of response error contribute to inaccuracies in victimization surveys: recall bias, high volume incidents, and temporal placement of events. Principal among these response errors is *recall bias* wherein respondents fail to report their involvement in a crime because either they did not understand the cognitive task or they forgot that the event occurred. Asking about and recording the number of *high volume incidents* involving the same victim or offender has also proved to be problematic in victim surveys because of the detail collected for each incident and the potential for overburdening the respondent. Problems with *temporal placement of events* constitute another well-documented issue in victimization surveys. In this case, respondents report their involvement in a crime but inaccurately date the incident, usually moving it closer to the interview than it actually was.[5] Each type of error is discussed more fully in the following section.

Recall Bias

Recalling and reporting criminal events are deceptively difficult tasks to perform. The victim survey tradition has devoted a great deal of time and attention to facilitating this task for

[4] In the United States, no federally-sponsored survey of offending that is comparable to the NCVS exists. The United Kingdom takes a different approach. There, the Home Office sponsors an *Offending, Crime and Justice Survey*, which was first fielded in 2003 (Budd et al. 2005).

[5] Martin et al. (1986) also identified strategic response as a problem. Here, respondents recall the event but refuse to report it because they are ashamed, embarrassed, or suspicious. This source of error is not included in our discussion here because the victim survey tradition has very little to contribute beyond that which the offender survey tradition has done already. Both types of surveys employ Computer-Assisted Self Interviewing (CASI) to reduce strategic response (Thornberry and Krohn 2000; Cantor and Lynch 2000).

respondents. Two techniques to improve the recall of particular interests are screening strategies and cues to prompt the recall of crime events and reference period length to limit memory decay and facilitate recall (Biderman et al. 1967; Reiss 1967; Dodge 1970; Turner 1972; Biderman et al. 1986; Koss 1996). Studying the effect of implementing these techniques provides estimates of the magnitude and distribution of recall bias in victimization surveys.

Screening Strategies and Cuing for Crime Events. Screening questions in victimization surveys are designed to cue respondents to report whether they have experienced a particular type of victimization or not. When the NCS was designed, some debate surrounded whether screening for eligible events should be conducted in one or two steps (Biderman et al. 1967; Cantor and Lynch 2000). One-step screening identifies and classifies eligible crime events based on the response to one screening question. For example, a "yes" to the screening question would both indicate that a respondent was a victim and count the respondent as the victim of a particular crime, such as robbery if the question were designed to elicit mention of a robbery. Two-step screening, on the other hand, separates identification of an eligible event from classification of the event. The screening interview is devoted solely to prompting recall of a potential victimization event. Here, a "yes" to a screening question results in the administration of a second set of questions designed to elicit very detailed information about the event. Gathering this additional information serves two purposes. First, the information provides greater detail about the victimization. Second, the details allow out-of-scope events to be excluded from victimization estimates. An out-of-scope incident could be one that took place outside the time period of interest or is not the type of crime included in the study. In addition to gathering more detailed and accurate information, this two-step procedure is believed to be superior to the one-step approach from a cognitive perspective because it allows the respondent to focus on the recall task exclusively. Early tests of victimization surveys found that separating the screening and description tasks resulted in more exhaustive reporting of victimization events (Biderman et al. 1967). As a result, the NCS and NCVS – as well as surveys modeled after such as like the BCS – utilize a two-step procedure.

Lessons from Victimization Surveys. The need to improve respondents' recall was a motivating factor that prompted redesigning the NCS. During the NCS Redesign, Albert Biderman was the first to bring survey researchers and cognitive psychologists together to discuss how best to stimulate the recall and reporting of crime events (Biderman and Moore 1982). The hope was that theories of cognition would serve as a framework for screening strategies to promote more complete recall of crimes. This Cognition Workshop in 1982 served as the stimulus for more extensive collaboration between cognitive psychologists and survey researchers as well as more elaborate theories of how respondents stored and retrieved memories of crime events (Jabine et al. 1984). These theories, in turn, were the basis of survey procedures that fostered more complete reporting of crimes.

On the basis of the discussions generated by the Cognition Workshop, the failure to report crime events was attributed to conceptual failure, memory failure, response inhibition, or a combination of these (Biderman et al. 1986). *Conceptual failure* occurs when the survey instrument does not correctly convey the cognitive task concerning the detail expected, the activity or time period of interest, or the concept at issue. Respondents in victimization surveys may, for example, believe that general and approximate reporting of crime events is sufficient, when the survey actually is demanding greater specificity in reporting attributes of

crime event. Alternatively, the respondent may report on crime-like events rather than actual crimes or events outside the reference period because he misunderstood his task. Another issue with conceptual failure is that the respondent's definition of crimes may differ from that used in the survey such that he omits "gray area" events such as assaults in which his brother was the offender. *Memory failure* is the inability of respondents to locate in their memory the target event when they clearly understand the nature of that event. *Response inhibition* occurs when respondents are ashamed or embarrassed to report events accurately and completely, so they fail to report or they report inaccurately.

This simple typology of why respondents do not report eligible crimes was used as the blueprint for redesigning the screening procedures in the NCVS, with the greatest emphasis given to reducing memory and conceptual failures. The principal method for reducing memory failure was to substantially increase the number of prompts given to the respondent to help them search their memory for crime events. This goal was accomplished by using a "short cue" screening interview.[6] Traditional victimization surveys (including the NCS) simply asked respondents "If they had been hit, kicked or punched?". The NCVS still asks the initial question ("Were you ever hit, kicked, or punched?"), but follows up with a number of other possible attributes of events in rapid succession. These questions are designed to cue incident attributes like weapons ("with a stick?"), known offenders, ("By a friend? By your brother?"), and locations ("At work or at school?"). Using these incident attributes as prompts sought to help the respondents locate the event in their memory. By asking the cues in rapid succession, the goal was to deliver as many as possible in a short period of time.

To take into account conceptual failures in which the respondent did not define eligible events as crimes, alternative frames of reference were added to the screener. Here as well, additional cues were added to reduce memory failure. Alternative frames of reference are important because victims may not treat the criminal act as the most salient aspect of the crime event. If the event happened at work, for example, respondents may not regard the incident as a victimization, especially for respondents who work in professions such as police officers or bouncers where crimes such as assaults may frequently occur. If respondents store memory of the assault at all, they may keep it in the "at work" memory file rather than the "victimization" memory file. By explicitly including "work" as a frame of reference, this question both signals to the respondent that these incidents are eligible events to report to the survey and it prompts him to search this part of his memory. Failures of concept were also addressed by asking explicitly about some forms of crime such as rape[7] as well as firming up "gray area" events by indicating that these events are within the scope of the surveys. Cues such as, "Sometimes people do not report attacks by people they know…", were used to prompt respondents to include events that the respondent might not classify as a victimization incident, but that the NCVS includes in its estimates.

The major contribution of the NCS Redesign was the improvement of screening procedures to reduce underreporting. Much of the increase in reporting of victimization events as a result of the redesign was due to improved screening procedures (Hubble and Wilder 1988; Persley 1995; Kindermann et al. 1997). Specifically, these strategies contributed to a 40%

[6] In fact, the screening interview more than doubled in the amount of time it took to administer. The screener developed for the NCVS requires about 18 minutes to complete as compared to the screener previously used in the NCS, which only took about 8 minutes.

[7] The NCS did not specifically ask about rape because of concerns that such questions were too sensitive to be asked, especially in a government-sponsored survey. During the redesign, it was determined that societal norms had changed enough to permit directly asking questions about rape and sexual assault (Rennison and Rand 2007).

increase in the reporting of victimization events (Martin et al. 1986; Hubble and Wilder 1988; Rand et al. 1997; Cantor and Lynch 2005). The research and development work on screening as well as the results of the redesign indicate that a lot of crime remains unreported both to the police and to casual crime surveys and one must work hard at screening to get the most complete reporting possible.

Approach in Self-Report Offending Surveys. In comparison, self-report surveys of offending have devoted less attention to screening. The approach taken to screening in offender surveys actually tends to be odds with the best practices identified in the self-report victimization literature. For example, self-report offending surveys generally use the one-step approach to inquiring about offending. This reliance on the one-step approach generates other screening problems including the minimal use of cues, multiple cognitive tasking, mix of screening and detail recall, and inability to exclude ineligible events. The discussion in the following section examines these problems using specific examples.

One primary problem in screening questions used in offending surveys concerns the minimal use of cues presented to stimulate recall of criminal events. The instrument used in one of the PRCCD instruments (Rochester Youth Development Study), for example, asks respondents:

> Since your last interview have you damaged, destroyed or marked up somebody else's property on purpose?

This question provides only two cues that prompt recall of the target crime: the action (destruction of property) and the victim (somebody else's property). The question fails to clarify gray areas, such as instances wherein ownership or intentionality is vague. For example, a teenage respondent's parents buy a car for his use, but he damages the car to antagonize his parents. In the respondent's mind, the car is his property and no delinquent act would be reported in response to this question. As a result, the survey would undercount these types of activities. The question also fails to use multiple frames of reference, such as school or home or certain types of leisure behavior during which the destruction took place.[8] The vandalizing of playground equipment, for example, may be so common place that it would not be singled out unless specific reference was made to this activity or location. Property damage also is common place in sibling disputes, but the incident most likely would not be recalled and reported unless some specific reference was made about the sibling. The failure to recall such an event may be due to the event being stored in a "sibling" memory file as opposed to a "property destruction" file.

A second screening problem in self-report offending surveys is that respondents often are asked to do a number of complex cognitive tasks rather than focusing on the single task of searching their memory for potentially eligible events. This "multi-tasking" requires respondents not only to find the target event in their memory, but also to identify how many times it has occurred in the past year. This problem can be illustrated by the following question on vandalism from the National Youth Survey (NYS):

> How many times in the last year have you purposely damaged or destroyed OTHER PROPERTY that did not belong to you, not counting family or school property?

[8] Other self-report offending surveys such as the NYS do screen with some reference to school, family, and other as frames of reference. However, the density of cues is more limited than in the NCVS.

Here, the question requires respondents to search for instances involving damage to property, then exclude those that did not happen in the last year as well as those that did not involve property owned by family members or the school. A significant effort is demanded from the respondent in a single operation, especially since the respondent is also expected to answer carefully and accurately.[9]

A third common problem with screening questions regarding offending is the interspersing of screening with short descriptions of events. This structure adds unnecessary complexity to the respondent's cognitive task. For example, the NYS asks the screening question:

> How many times in the Last Year have you:
>
> 226. Stolen or tried to steal a motor vehicle such as a car or motorcycle?
>
> If the respondent answers something other than "none" they are asked:
>
> 226. Thinking of the last time:
>
> (a) What type of vehicle was it?
> (b) Did you actually steal it?
> (c) Who did the vehicle belong to?
> (d) How did you get the vehicle started?
> (e) What were you going to do with the vehicle?
> (f) Were you alone or did others take part in this event?
> (g) Had you been drinking or taking drugs before the incident?
> (h) Did you report this same event for any other question in this set?

Here, the question requires the respondent to move between the cognitive task of recalling any events of motor vehicle theft to recalling details of the last theft. If the entire survey instrument is examined, the cognitive task is even more complicated. The next question requires the respondent to return to searching his memory to count another type of crime, such as robbery, and then provide specific details. This switching between recalling types and counts of events to recalling details of a specific event can make it more difficult for a respondent to perform either task well for that particular question. Describing the specifics of one event can make it more difficult to search one's memory for other types of crime referred to in subsequent screening questions in the survey instrument.

A fourth and final problem frequently observed in the screening questions for offending surveys concerns the failure of the follow-up questions to gather information that can be used to exclude ineligible events and that can serve as a check on the performance of the cognitive tasks required of respondents in the screening questions. This problem includes both the failure to ask any follow-up questions at all as well as the failure to ask questions that would provide these details. For example, the vandalism screening question described earlier, did not include any follow-up questions that asked respondents for the date of the event, which would ensure that the event happened during the reference period. Respondents also are not asked about ownership of the property or the intentionality of damage in the vandalism events. Asking respondents about these aspects of the event as a follow up to the screen questions would permit the exclusion of ineligible events.

[9] In the victimization survey context, studies have compared early surveys that asked respondents to do these complex exclusions within one question with surveys that isolated identifying potential events from exclusion of ineligible events. Respondents were more productive when asked separate questions that allowed them to identify potential events and exclude ineligible ones (Biderman et al. 1967; Reiss 1967; Cantor and Lynch 2000).

IMPLICATIONS. The evidence from victimization surveys suggests that the current screening strategies employed in self-report surveys of offending likely suffer from a mix of both underreporting due to limited cuing and complex cognitive tasks and over reporting due to the inclusion of ineligible events. While these errors may be equal and off-setting for aggregate estimates of the volume of offending, they may have serious repercussions for correlational analyses due to the selectivity of offenders and events captured by the surveys. In particular, underreporting could be very consequential especially for estimates of the offending rates of individuals in a given unit of time (also referred to as *lambda*) which is used in creating offender typologies. The effect on other analytical uses of the data is less clear since underreporting could be distributed differently across values of predictor variables. In the victimization survey context, initial assessments of the change in the NCVS design suggested that increases in reporting were greater for low risk groups than high risk groups, such as older respondents as opposed to younger and wealthy as opposed to poor respondents (Kindermann et al. 1997). This finding suggests that underreporting is not uniformly distributed across values of independent variables. Cantor and Lynch (2005) estimated models with data collected before and after the NCVS redesign and found substantial stability in the models predicting victimization. The changes in cuing had the expected effects on the level of victimization reported, and the coefficients for predictor variables had the same sign and relative magnitude before and after the redesign. The constancy of the effects of changes in the cuing strategy across predictor variables was especially the case for clearly defined crimes like robbery. This tendency was much less true for less well defined crimes like aggravated assault and simple assault because these incidents are more susceptible to interactions between design characteristics of the survey and attributes of respondents and crime events. The interactions between design changes and attributes of victims and crime events suggest that the inclusion of more extensive screening for crime events may result in quite different correlates of offending than are currently the case, at least for ill-defined types of crime like simple assault.

REFERENCE PERIOD. As discussed earlier, work done as part of the NCS Redesign identified three sources for respondents' failure to report crime: conceptual failure, memory failure, and response inhibition or distortion (Biderman et al. 1986). Of relevance to reference periods is *memory failure*, which is the inability of respondents to locate in their memory the target event when they clearly understand the nature of that event. This problem increases with the passage of time such that the longer the period between the interview and the occurrence of the crime event, the less complete the reporting. Memory decay occurs more slowly for events that have been rehearsed or that have consequences that anchor the event in memory. Cues and other memory aids can facilitate more complete recall, but the decay of memory over time is a universal in retrospective surveys.

Based on issues of memory failure, the reference period has implications for recall and the reporting of crime events. *Reference period* refers to the length of time that respondents are asked to search their memory to find eligible crimes. Longer reference periods mean more underreporting of crime events in general. The exception arises for serious criminal events, which are remembered longer and more uniformly than less serious crimes. While the effects of reference period length are reasonably consistent across both traditions, the research conducted in the victimization survey tradition employs more rigorous research designs and less highly selective populations. As a result, this research strongly suggests that shorter reference periods may change the results of analyses of self-report offending data.

Lessons from Victimization Surveys. A consistent finding from both work done as part of the NCVS redesign as well as methodological studies in other retrospective surveys is that shorter reference periods produce more complete reporting (Lepkowski 1981). In theory, no point occurs at which shortening the reference period does not improve data quality (Biemer et al. 2004). One reason is that the work on victimization surveys suggests that memory decay happens very quickly so that reference periods of a few days would be optimal. In practice, however, reducing recall error is not the only concern of researchers, and a trade-off exists between reference period length and sampling error. Short reference periods require large samples to ensure including enough respondents who experienced victimization or offending in that very short span of time to provide reliable estimates.[10] In other words, reporting of crimes is more complete in a short reference period, but there will be less of it.

The issue with regard to reference periods and recall becomes one of finding the optimal balance between measurement and sampling error as well as ensuring the ability to report on an appropriate unit of time, such as a particular year or the past 12 months. The NCVS uses a 6-month reference period. Other surveys use longer periods such as the British Crime Survey, which employs a 12-month reference period. The decision to use a 6-month reference period for the NCVS was the basis of findings from randomized experiments that compared various reference period lengths and supported after the fact by analyses of recency slopes. When Census researchers compared 3-, 6-, and 12-month reference periods in the NCS, they found shorter periods resulted in significantly higher reporting rates than did longer reference periods (Bushery 1981). Biderman and Lynch (1981) obtained similar results when examining recency slopes in the NCS. Here 26% of all victimizations reported in the survey were reported in the month of the reference period most proximate to the interview and only 12% of the incidents were reported in the most distant month of the 6-month reference period. While many factors contribute to the negative relationship between recency and reporting, the most plausible explanation is memory decay.

This underreporting due to memory decay and long reference periods may vary by types of events and respondents; however, the evidence for such a conclusion is mixed. NCS reference period experiments (Kobelarcik et al. 1983) indicated that reporting is more complete for younger persons across all types of crime when shorter reference periods are employed. Minority respondents reported more violent crimes than non-minority respondents with the 3-month reference period but no differences were observed between the 6- and 12-month reference periods. On the other hand, Biderman and Lynch (1981) did not find many differences between respondents or events reported in the most proximate as opposed to the most distant months of the reference period. Younger respondents had a slight tendency to be more affected by recency bias than older respondents. The largest effect observed was in police-reported events. Here a greater proportion of the crimes reported to the survey in the most distant month of the reference were reported to the police as compared to the most proximate month of the reference period. This finding is consistent with findings that events that are rehearsed or have consequences are less subject to memory decay.

In sum, the methodological work on victimization surveys indicates the longer the reference period, the less complete the reporting of crime events. This underreporting does not seem to be correlated with attributes of respondents when reference periods of 6 months or more are compared, but there are differences in reporting when reference periods of 3 months or less are compared to longer reference periods. Even when longer reference periods are used,

[10] In addition, sample size, in turns, brings issues of cost into play.

crime events that result in some additional repercussion such as reporting to the police, reporting to insurance, obtaining treatment for injury are reported more completely than events without these consequences.

Approaches in Self-Report Offending Surveys. Self-report offending surveys often use much longer reference periods than the NCVS's 6-month reference period. The original self-report surveys of offending were cross-sectional surveys that asked respondents to recount their offending experience over reasonably long periods of time such as 3 years (Gold 1972). No consistent reference period is used. The NYS uses a 12-month reference period. One of the interviews in the Cambridge Youth Study, however, asks its 32-year-old respondents to report on their offending since 16 years of age. Using reference periods of several years or one's lifetime will result in substantial underreporting of offending behavior. More recently, most of surveys of offending have been longitudinal panel designs that interview respondents repeatedly at intervals that allow for a reference period of a year or two at the most.

A number of studies have compared offending rates produced by prospective (longitudinal surveys) and retrospective (cross-sectional surveys) designs. These studies are essentially comparisons of different reference periods. In their comparison of different reference periods, Mennard and Elliot (1990) found longer reference periods produced much lower estimates of the incidence and prevalence of juvenile offending. They also found differences in models predicting the frequency and patterns of offending using these different reference periods. Unlike the methodological studies conducted with victimization surveys, most of the work concerning reference periods in the self-report offending tradition compares already lengthy recall periods such as 1 year, with even longer periods such as 2 or more years. This comparison almost certainly underestimates the magnitude of the reference period effect and distorts the distribution of these events across attributes of respondents and events.

In the offender survey tradition, only one comparison has included an "ultra-short" reference period. Roberts et al. (2005) found large discrepancies between a retrospective survey that employed a 1-week reference period as compared to one utilizing a 5-month reference period. The ultra-short, 1-week reference period produced much higher estimates of offending and even affected the classification of individual as high or low volume offender. The Roberts study is not very generalizable, however, for three reasons. It utilized a very small sample ($n = 75$). The sample was from a very selective population (young people admitted to an emergency room who were identified as having histories of violence). In addition, the surveys used different formats. The group with the 5-month reference period was interviewed using life events calendar while the group with the 1-week reference period was interviewed with more traditional formats with more highly structure and uniform questioning.[11] While research in the self-report offending tradition established the superiority of reference periods of a year relative to longer reference periods, no work has been done on the effects of ultra-short reference periods relative to more commonly used reference periods of a year or two.

[11] Life event calendars use the recounting of life events on a calendar as a means of improving recall of target events in retrospective surveys and increasing the accuracy of dating these events. For an application in offending surveys, see Horney and Marshall (1992).

IMPLICATIONS. Exploring reference periods in self-report offending surveys is particularly important due to the new use of these data in trajectory studies. These statistical methods are sensitive to data accuracy and even modest errors in recall may affect trajectory groupings (Eggleston et al. 2004; Sampson et al. 2004). In their call for more investigation of the effects of data quality, Eggleston et al. (2004) explicitly mention the length of the reference period in self-report offending surveys as a potential source of distortion. Departure from complete reporting or the inaccurate dating of crime events that can result from the use of long reference periods likely has consequences for identifying offending trajectories and for distinguishing types of trajectories.

Methodological testing in both victimization and offending survey traditions support the idea that shorter reference periods promote more complete recall and reporting than longer reference periods do. The work in victimization surveys, however, informs with regard to how best to balance the need for accuracy with the practical needs of sample size and implementation. The reference periods of 1 and 2 years that are customary in offender surveys, however, may result in substantial underreporting and perhaps even bias. The comparisons of reporting using different reference periods across offender surveys may be understating the extent and nature of under-reporting because they do not capture the steepest part of the recency slope which occurs in the days and weeks immediately after the event. Work from the victimization survey tradition suggests that self-report offending surveys could benefit from methodological studies that explored using very short (such as a month) reference periods and comparing these findings with the more commonly used annual reference period to better estimate the magnitude of underreporting and any potential bias in reporting of offenses.

High Volume Repeat Events

A second type of response error concerns high volume repeat events. These events are crimes that occur frequently in a given period of time and share many of the same characteristics, such as the type of criminal act, the same offender or the same victim. A cab driver, for example, who gets robbed every Friday night, would be experiencing high volume repeat victimization. Intimate partner violence is another example commonly associated with high volume repeat events (Dodge and Balog 1987). Here the difficulty – and potential for error – arises from how to accurately count and capture these incidents. Different techniques have arisen from the victimization and offending survey traditions.

LESSONS FROM THE VICTIMIZATION SURVEYS. Asking respondents about high volume repeat events has proven to be difficult in victimization surveys. The incident focus of the surveys required that an incident form be completed for each positive response to a screener question. In some cases, respondents, such as those experiencing domestic violence, indicate that a large number of similar events have occurred and that they cannot date them or provide distinct information on each crime. Under these conditions, the interviewer has the option of following a "series incident" procedure in which one incident form is completed on the most recent incident and the number of times this type of crime occurred in the reference period is recorded. The series incident procedure was developed to reduce the reporting burden for respondents.

While this strategy helps respondents, it introduces error into the estimates since it amounts to asking the more imprecise question of "how many times" the respondent has been victimized with no incident report requiring greater specificity. Such concerns are significant

enough that the Justice Department does not include series events in the annual crime counts because it considers these counts to be unreliable. Studies examining the effects of including series counts results in annual rates that fluctuate wildly (Biderman and Lynch 1991; Planty 2007). Planty (2007), for example, estimated that in 1993, rape and sexual assault estimates would have increased 225% if series victimizations were included. This change is driven by 21 unweighted incidents with very high estimates of the number of crimes they have experienced. The maximum number of aggravated assaults reported in a series event over the period 1993–2002 ranged from 75 in 2000 to 750 in 1996. Planty (2007) also observed that estimates of the number of events in a series tended to bunch around figures that were multiples of months or weeks in the reference period. This pattern suggested that respondents were estimating the frequency of these events rather than recalling events separately. These findings indicate that very imprecise crime estimates will result from asking respondents with a lot of crime experience to report on how many crimes occurred unless these estimates are subjected to the rigor of an incident form.

The inclusion of high volume repeat events clearly has implications for estimates of crime overall, but the treatment of these experiences also affects estimates of crime across population groups. Planty (2007) found that when series victimizations were excluded from the aggravated assault rates, blacks (35) had higher rates per hundred thousand than whites (27) or Hispanics (29). When the victims' estimates of the number of events in the series were included in the rates, Hispanics became the highest risk group (77) compared to blacks (47) and whites (39).

Approaches in the Self-Report Offending Tradition. Since offending surveys do not have an incident focus, they have relied almost exclusively on the respondent's estimate of "how many times" an offense has occurred. For example in the PRCCD's Rochester Youth Development Study, a screening question asks about a certain type of offending. If the answer is positive, then the respondent is asked "How many times?" The NYS employs a more elaborate procedure that asks respondents to report on how many crimes of a certain type occurred and, if over ten crimes are reported, questions are asked about the rate at which these crimes occurred such as two or three per month.

Implications. Asking respondents to indicate whether they were involved in a specific number of crimes of certain type yields imprecise information on the number of crimes that occurred and the nature of those crimes. When respondents are asked to report in detail about individual crimes, rather than simply reporting counts of crime, estimates of aggregate crime rates are more stable over time. The victimization survey technique of forcing respondents to recount details of specific crime events in the form of an incident report appears to encourage more reliable reporting. The instability in aggregate victimization rates when reported counts of events are included was observed when they were used for persons who reported six or more victimizations. It may be that reporting counts of crimes is accurate when a respondent only has a few crimes to report and it becomes progressively less accurate when more crimes are reported. The instability in the aggregate crime rates was similar in the NCVS trends when the series victimization procedure was used for three or more events and when it was used for six or more events (Biderman and Lynch 1991; Planty 2007). This finding suggests that after a relatively small number of events, respondents start estimating rather than recounting the number of crimes they were involved in. At minimum, the experience from victimization surveys suggests that responses to the "how many" question in screening for offenses that involve large numbers should be regarded with caution.

Temporal Placement of Crime Events

A third type of response error concerns the temporal placement of crime events. Here, the respondent is required to accurately date when the criminal incident occurred. Errors in temporal placement more often move the target events forward in time so that they are reported as closer to the interview than they actually are. This phenomenon is often referred to as "telescoping." As a result, a concern for victimization surveys is the effect this error has on annual rates and change estimates.

Lessons from Victimization Surveys. Both studying the effect of and ameliorating the error generated by telescoping has been given much more emphasis in the victimization surveys than in the self-report offending tradition. One source of data on temporal placement in the victim survey tradition is reverse record checks. In these studies, a sample of crimes reported to police is taken from police records. The victims are contacted and interviewed to see if they will report the events to the survey. The date of the event as reported by the victim is compared to the date of the police report to determine the respondent's accuracy in dating the event. Several reverse-record check studies conducted prior to designing the original NCS (Dodge 1970; Turner 1972) and additional record check studies performed as part of the NCS Redesign (Miller and Groves 1986). These studies indicated a substantial amount of error in the temporal placement of crime events in retrospective surveys. Moreover, the errors in temporal placement increased as the reference period lengthened (Dodge 1970).

Another source of data on the temporal placement of crime events comes from the testing of techniques designed to reduce temporal displacement. Reducing telescoping has been a major influence on the design of victimization surveys. The NCVS, for example, employs a rotating panel design in large part to control telescoping. Respondents are interviewed seven times at 6 month intervals. The first interview is a "bounding interview" that is not used for estimation purposes but simply to bound the reference period for subsequent interviews.[12] Respondents are asked to report crime that occurred since the last interview, so the previous interview serves as a cognitive bound on the recall task. When respondents report a victimization, an incident form is administered, and as one of the first questions, respondents are asked the date of the event. This question provides another opportunity for the respondent to not report an event out of the reference period. The events reported in the incident form are checked against crimes reported in the previous interview and respondents are asked to affirm that similar events were not actually reported in the previous interviews. Respondents in unbounded interviews report about 40% more victimizations than respondents in unbounded interviews, so telescoping arguably has a considerable effect (but see Addington 2005, questioning the effect of telescoping across types of crimes). Studies of bounding also show that bounding effects vary across types of crime with property crimes being affected more than violent crimes, attempts more than completed crimes and crimes not reported to the police more than those reported (Murphy and Cowan 1984; Addington 2005).

Rotating panel designs are not the only technique to limit telescoping errors. Temporal anchors are useful, especially for cross-sectional victim surveys. Loftus and Marburger

[12] Three reasons have been posited for how bounding in the crime survey reduces telescoping (see, Biderman and Cantor 1984; Skogan 1981). First, the information from the bounding interview generates a reference list for the interviewer. Second, the bounding interview provides a cognitive reference point for the respondent. Third, the bounding interview educates respondents about precision expected in their responses.

(1983), as part of the NCS Redesign, demonstrated the power of temporal anchors when they interviewed persons at the Seattle Airport and asked them about their victimization experience in the past 12 months. After they responded, the interviewers asked if the incident occurred before or after the eruption of Mount St. Helens which happened exactly a year before. With this temporal anchor about 40% of the respondents allowed that the event they reported did not occur within the reference period. This finding indicates a massive amount of foreword telescoping and the potential of temporal anchors for more precise temporal placement. Because of the use of rotating panel designs in the major victimization surveys, the use of temporal anchors and other calendrical devices, such as life event calendars have not made their way into the major victimization surveys discussed here.

Limitations in Self-Report Offending Surveys. A number of studies in the self-report offending literature have examined the accuracy of respondents' dating of crime events. Like the comparable victimization studies, these analyses use comparisons both with police records and of alternative survey procedures meant to reduce temporal misplacement. The record check studies find reasonably close correspondence between the temporal location of crime events in self-report surveys and in police records. Comparisons of dating across different types of surveys find larger differences.

Record check studies conducted with offending surveys, however, are not the same as the reverse-record checks performed in the victimization survey context. The former compare counts of events reported as occurring in a specific time period while the latter compare specific incidents occurring in a particular point in time.[13] Matching overall counts of events is considerably easier than matching specific incidents. Matching counts of crimes is easier in part because one can have the same count of crimes per month in the survey and police records with each source including very different events. The police counts can be low because not all events come to the attention of the police and the police fail to record many events, while the respondents' counts can be low because of underreporting. Assessed on an event or incident basis, these two data sources would have very low correspondence, but assessed in the aggregate the correspondence is high. It is not surprising, therefore, that comparisons of survey responses and police records in the offender survey tradition find reasonably high correspondence.

Comparisons of the temporal placement of events across different survey procedures are rare, and they have usually examined the effect of reference period length on the temporal placement of crimes. Two studies are noteworthy. One is from the work conducted by Roberts et al. (2005) discussed earlier. In this study, the rate of reporting an event and placing that event in the correct month is 13.6 per hundred, while the probability of reporting an event is 43.6 per hundred. This finding suggests that a substantial number of errors occur in dating offenses in retrospective surveys and that these errors increase with the length of the reference period. The applicability of the results of the Roberts study to self-report surveys more generally is complicated by the selectivity factors in the Roberts study that were discussed in the foregoing section. A second study is that of Kazemian and Farrington (2005). Kazemian and Farrington (2005) compared a prospective survey with a 2-year reference period to a retrospective survey

[13] Huizinga and Elliot (1985) come the closest in the self-report offending survey tradition to an incident-based reverse record check where arrests incidents found in police records are matched on a one to one basis with reports from the self-report survey. Even in this study, the amount of information used to match is quite limited, and dates are not included.

with a reference period of approximately 16 years. In examining the time of onset of offending in prospective and retrospective surveys of offending, they found little agreement between the two procedures. Respondents in retrospective surveys typically reported the age of onset as being later, that is, they reported it closer to the interview than was the case in prospective surveys. Temporal placement errors occurred more frequently with more minor as opposed to more serious events.

Self-report offending surveys differ considerably in the procedures used to promote accurate temporal placement of events. Early cross-sectional surveys used a 3-year reference period that was not bounded in any way, although respondents were asked to date the events separately (Gold 1972). The National Youth Survey (Elliot 2008) used Christmas as a bounding event. The PRCCD's Rochester Youth Development Study (RYDS) survey asks respondents to report on offending that occurred since the last interview, usually a year or two ago.[14] The prior interview serves as the far bound of the reference period and should reduce telescoping into the reference period. While this procedure appears to be similar to the NCVS bounding method, the RYDS instrument does not give as much emphasis to minimizing telescoping as does the NCVS. For example, both the RYDS and the NCVS interviews caution the respondent about the length of the interview period and makes reference to "since the last interview." Unlike the RYDS interview, the NCVS incident form asks the respondent a second time about temporal placement when requiring the month of the criminal incident. In addition, the NCVS interviewers keep a narrative description of the crimes reported in the prior interview on the control card. When crimes are reported in the current interview, the interviewers compare the events reported in the current interview with those reported in prior interviews. If an event is similar, then respondent are asked if the current event is the same as one they reported in the previous interview. If they respond affirmatively, the event is not counted. This "unfounding" task is not done in RYDS.

IMPLICATIONS. Research in victimization and offending survey traditions indicates that the temporal placement of events is difficult for respondents, who tend to move events forward in time both into the reference period from outside and within the reference period. The evidence for this forward telescoping is much stronger in the victimization survey tradition largely because of the need for victimization surveys to isolate and attempt to date events. The methodological testing done with victimization surveys used samples more similar to the general population than many of the analyses conducted in the self-report offending tradition. The methodological work based upon police record checks in the victimization area also compared individual events reported in the survey to events record by the police, while studies in the offending tradition compared counts of events reported in the surveys to counts of events in official records. The former provides a more stringent test of temporal placement. Evidence also suggests that bounding procedures such as those used in the NCVS reduce telescoping. Most of the major offending surveys employ substantially less extensive bounding procedures. Consequently, offending surveys likely suffer from a substantial amount of forward telescoping and this telescoping probably involves more minor crimes that are not reported to the police.

[14] The authors are grateful to Terry Thornberry for providing a copy of the Rochester Youth Development Study interview schedule.

DISCUSSION

As with much of the measurement error in criminological data, response error and the problems it generates do not lend themselves to quick or easy solutions. These problems are compounded by the fact that little attention has been devoted to studying measurement error let alone investigating the magnitude and direction of measurement error. In this context, our knowledge of measurement error in the self-report survey tradition is much greater than for data based on official records. Within the self-report area, more resources have been invested in understanding response error for victimization surveys than those for offending. Even for victimization surveys, the amount of existing research is much too small for us to speak definitively and specifically about the effects of response errors on various uses of these data. It is not surprising, then, that our first suggestion is for more research on this topic. Specifically, we recommend a two-prong approach of long- and short-term research, which are discussed in more detail in the following section. In the long-run, the research community needs to invest in the creation of a "gold-standard" survey. Such a project is well beyond the means of individual researchers and will require concerted efforts of teams of criminologists and other social scientists dependent upon self-report methodologies. In the short term, interim strategies can be adopted by individual researchers. In addition to strategies to reduce error, we recommend specific approaches that individual researchers can take to avoid the effects of response errors or to test for the magnitude of their effects.

The Need for a "Gold Standard" Survey

Much of the foregoing discussion is predicated on our ability to know how much crime there is and when it occurs. Talking intelligently about response errors is difficult unless some standard is available against which responses can be compared. Tests of content validity are easy to do, but are subject to substantial disagreement about what constitutes the content of the concept. Construct validity tests require robust theory and are open to debate as to the strength of the relationships among the relevant concepts. Criterion validity is more persuasive than content and construct validity as it is less subject to debate so long as an agreed upon standard exists against which ones measure of crime can be compared. Unfortunately, no gold standard exists in the area of either self-report victimization or offending. As we have seen in this chapter, the traditional criteria for establishing the accuracy in self-report surveys are official records or surveys with different (and arguably superior) methodology. Both of these criteria have their problems. Police records are highly selective and have their own error structure that lessens their ability to serve as a standard for valid reporting (Biderman and Lynch 1981; Huizinga and Elliot 1985, 1986; Miller and Groves 1986). Moreover, Biderman (1966) argued that the social organization of surveys is so different from the social organization of official record systems that they could not and should not arrive at similar counts and characterizations of crime events. Using alternative surveys as standards for judging the validity of self-report surveys would not be subject to the distortion introduced by administrative records systems. Moreover, surveys are under the control of researchers, whereas administrative record systems are not, so it is, in principle, simpler to eliminate selectivity and other sources of non-comparability across the surveys than it is when comparing survey results to administrative records (Biderman and Lynch 1991). Alternative surveys would provide a much clearer assessment of the effects of specific survey procedures.

To investigate an alternative survey strategy, the initial question is whether theories of response in self-report surveys (for either victimization or offending) are sufficiently developed to identify clearly superior methodologies for self-report surveys that can be used as "the standard" against which lesser methods can be compared. When the self-report method was first developed, this question could not be answered. In the interim, a considerable amount of progress has been made in the development of response theories, particularly in the last 25 years. Today, we are better able to identify a package of instrumentation and procedures that would provide the most complete recounting of criminal involvement possible in a self-report survey. For these reasons, we believe that the greatest advances in outlining the error structure of self-report surveys will come from studies using "best practices" surveys as their standard for accurate reporting.

Developing a "gold standard" self-reporting survey is not a far-fetched idea. From the research on victimization and offending surveys reviewed in this paper, we know that a number of substantial response errors affect estimates of the level of crime, the distribution of both victimization and offending across population groups and the temporal placement of crime events. We also know that self-report surveys employing extensive screening interviews and very short reference periods can reduce many of these response errors. In addition, requiring detailed questioning about each event identified in the screening interview will provide better estimates of the number of crimes committed, especially in the case of high volume offenders. As such, we could develop an offender survey, for example, with these design characteristics. This survey would be done purely for methodological purposes, to be used as a "gold" standard against which on-going surveys of offenders can be compared. Such a survey would be devoted entirely to getting the most complete reporting of offenses rather than gathering information on the offenders or the social context of offending. A split sample could be employed with one half receiving this "state of the art" design and the other the design currently employed in major prospective surveys of offenders. Conducting such a survey would help address whether the response errors identified in the foregoing sections affect the characterization of offending trajectories and the attributes of offenders that predict whose career will follow a specific trajectory. Research done to date in the victimization survey strongly suggests that these response errors exist and likely affect uses of data, especially with the more recent studies employing trajectory analysis. Little more can be said definitively about these response errors until more methodological work is conducted to estimate the magnitude and direction of those errors.

Interim Strategies for Estimating the Effects of Response Errors

While we await this "state of the art" survey, researchers have other means to learn more about response errors in self-report surveys of victimization and offending and to protect their substantive analysis from likely errors. The following three options suggest ways to learn more about response errors. First, researchers can look across surveys and exploit the existing variation in offender survey methodologies to learn more about the effects of survey design on the resulting responses. The NYS, for example, has much more extensive cuing in their screening procedures than other high quality offender survey. Comparing the offense distributions in the NYS to those found in the PRCCD surveys such as RYDS or PYS can provide some idea of the importance of these design differences for reporting in the surveys. These comparisons will be very blunt assessments for investigating the effects of response errors because the

designs of these surveys differ in many ways, which could be offsetting. In addition to design differences, the surveys vary in the samples and populations addressed, which could affect the results as well. To minimize the effect of population and sample differences, a second option is to assess response errors within a single survey. Cuing density in the NYS may differ across the types of crime such that comparisons of reporting by cue and type of crime could reveal the effects of cuing on reporting. Similarly, changes in survey methodology over the course of these longitudinal surveys provide another opportunity to understand the effects of method on reporting. This type of quasi-experimental design has been used extensively in the NCVS to study various response errors (Biderman et al. 1985; Biderman and Cantor 1984; Reiss 1982; Addington 2005). Finally, comparisons can be made with surveys of offending other than the dominant surveys discussed in this essay. These surveys often utilize very different methodologies. The survey used by Roberts et al. (2005) that utilized a life events calendar is a good example. In addition, the distributions of offending across population subgroups in these surveys can be compared with those from RYDS, CYS, PYS or NYS to see how sensitive findings are to differences in survey methods. These interim efforts to assess the effects of response errors in self-report offending surveys are not a substitute for the "state of the art" survey, but they can keep the focus on the effects of response errors until that survey is fielded.

Avoiding and Delimiting the Effects of Response Errors

In some cases, we know enough about response errors in the self report tradition and have the requisite information to take specific steps to limit the effect of response errors on estimates. Two examples illustrate how this can be accomplished. The first example concerns responses to questions of "how many times." Here, it is well established that such questions evoke imprecise responses unless the respondent is required to recount specific incidents. Compounding this problem is the well-established effect of extreme values on most statistical methodologies. The effect of these issues on trajectory models, though, is not well understood. Researchers using such questions in trajectory models may want to examine how different treatment of extreme responses might affect their results. For example, offense counts could be treated as an ordinal variable rather than as a continuous variable.[15] Another approach could impose upper limits on responses to the how many times questions and assessing the effects on the resulting analyses.

The second example concerns the type of crime studied. Response errors are known to have greater effect on less serious crimes than more serious crime. Robbery, for example, is particularly unaffected by changes in survey methodology and this is attributed to the fact that it is the quintessential crime that involves violence, theft and occurs among strangers (Cantor and Lynch 2005; Rosenfeld 2007). Assault, on the other hand, is much more sensitive to changes in self-report survey methodology in part because it does not comport with stereotypic definitions of crime.[16] Here, researchers can restrict their analyses to more serious crimes in order to reduce the effects of response errors. Alternatively, researchers can

[15] The authors thank the editors for suggesting this option.

[16] Some debate exists over what "serious" means. Typically, a more serious crime means a more morally objectionable act, but these acts also are more likely to have consequences, such as being report to police or insurance companies and hospital visits, all of which make them more memorable. Still others argue that it is the unambiguous nature of certain crime events that makes them more immune to response errors.

analyze separately more and less serious crimes and to estimate the effects of response errors. If models restricted to serious crimes are different from those for less serious crime in ways that can be attributed to response errors, researchers should be cautious in relying on analyses conducted using less serious crimes as well as the conclusions that are drawn.

CONCLUSION

Self-report surveys of offending behavior were developed to identify predictors of criminal behavior as well as to distinguish high rate offenders from low rate offenders and to explain why these people offended at different rates. More recently, these data have been used with sophisticated statistical procedures to characterize offense trajectories. These methods have many features to recommend them over the relatively simple and more subjective methods of data reduction and classification previously used (Nagin 1999, 2004). At the same time, these new trajectory methods have been shown to be sensitive to the nature of the data used (Eggleston et al. 2004). Inaccurately dating an offense, for example, or substantially under-reporting the number in a given period may not have mattered when the data from offending surveys were only used to distinguish high- and low-rate offenders, but these errors may have much greater effects on trajectory analysis. Victimization surveys, in contrast, have always had high demands for exhaustive reporting of crime events and for accurate dating of these events. This chapter reviewed some of the methodological work done on self-report victimization surveys in an effort to suggest where and why self-report offending survey data may not be as complete and accurate as these new uses require. In particular, we focused on response errors introduced by recall bias, high volume crime events, and temporal placement.

Comparisons of self-report offending survey and self-report victimization survey methodology suggest that several types of response error in self-report offending surveys may affect the results obtained. Confirming these suspicions is not simple given the difficulties in validating self reports of victimization or offending. Further exploration of these sources of response error may only be possible by comparing "state of the art" survey to lesser surveys to determine the adequacy of the latter survey. The state of the art survey would employ the best procedures as defined by survey research and cognitive psychology. To the extent that the results other surveys comport with the results of the state of the art survey, they would be considered valid. In the interim, researchers can utilize existing self-report surveys to obtain a better understanding of how response errors affect crime estimates, especially for particular uses such as trajectory modeling.

REFERENCES

Addington LA (2005) Disentangling the effects of bounding and mobility on reports of criminal victimization. J Quant Criminol 23:321–343

Biderman AD (1966) Social indicators and goals. In: Bauer R (ed) Social indicators. MIT Press, Cambridge, MA

Biderman AD, Cantor D (1984) A longitudinal analysis of bounding respondent conditioning and mobility as sources of panel bias in the National Crime Survey. In: American Statistical Association 1984 proceedings of the Section on Survey Research Methods. American Statistical Association, Washington, DC

Biderman AD, Lynch JP (1981) Recency bias in data on self-reported victimization. American Statistical Association 1981 proceedings of the Social Statistics Section. American Statistical Association, Washington, DC

Biderman AD, Lynch JP (1991) Understanding crime incidence statistics: why the UCR diverges from the NCS. Springer, New York

Biderman AD, Moore J (1982) Report on the Workshop on Cognitive Issues in Retrospective Surveys. Bureau of Social Science Research and U.S. Census Bureau, Washington, DC

Biderman AD, Reiss AJ Jr (1967) On exploring the "dark figure" of crime. Ann Am Acad Pol Soc Sci 374:1–15

Biderman AD, Johnson LA, McIntyre J, Weir AW (1967) Report on a pilot study in the District of Columbia on victimization and attitudes toward law enforcement. President's Commission on Law Enforcement and Administration of Justice, Field Surveys no. 1. U.S. Government Printing Office, Washington, DC

Biderman AD, Cantor D, Reiss A (1985) A quasi-experimental analysis of personal victimization by household respondents in the NCS. Paper presented at the annual meetings of the American Statistical Association, Philadelphia

Biderman AD, Cantor D, Lynch JP, Martin E (1986) Final report of the National Crime Survey redesign. Bureau of Social Science Research, Washington, DC

Biemer PP, Groves RM, Lyberg LE, Mathiowetz NA, Sudman S (2004) Measurement errors in surveys. Wiley, New York

Budd T, Sharp C, Mayhew P (2005) Offending in England and Wales: first results of the 2003 Crime and Justice Survey. Home Office Research Study No. 275. London Home Office

Bushery J (1981) Recall bias for different reference periods in the National Crime Survey. Proceedings of the American Statistical Association Survey Methods Research Section. American Statistical Association, Washington, DC, pp 238–243

Cantor D, Lynch JP (2000) Self-report surveys as measures of crime and criminal victimization. In: Duffee D, McDowall D, Mazerolle LG, Mastrofski SD (eds) Criminal Justice 2000: measurement and analysis of crime and justice. National Institute of Justice, Washington, DC

Cantor D, Lynch JP (2005) Exploring the effects of changes in design on the analytical uses of the NCVS data. J Quant Criminol 21:293–319

Dodge R (1970) The Washington DC Recall Study. Reprinted in Robert G. Lehnen and Wesley G. Skogan (1981) The National Crime Survey: Working papers, volume 1: Current and historical perspectives. Washington, DC: U.S. Department of Justice

Dodge R, Balog F (1987) Series victimization: a report on a field test. Bureau of Justice Statistics, Washington, DC

Eggleston E, Laub JH, Sampson RJ (2004) Methodological sensitivities to latent class analysis of long-term trajectories. J Quant Criminol 20:1–42

Elliot D (2008) National Youth Survey [United States]: Wave V, 1980. [Computer File]. Inter-university Consortium for Political and Social Research, Ann Arbor, MI

Fisher BS, Cullen FT (2000) Measuring the sexual victimization of women: evolution, current controversies and future research. In: Duffee D, McDowall D, Mazerolle LG, Mastrofski SD (eds) Criminal Justice 2000: measurement and analysis of crime and justice. National Institute of Justice, Washington, DC

Gold M (1972) National Survey of Youth, 1972 [Computer file]. Inter-university Consortium for Political and Social Research, Ann Arbor, MI

Horney J, Marshall I (1992) An experimental comparison of two self-report methods for measuring lambda. J Res Crime Delinq 29:102–121

Hubble D Wilder BE (1988) Preliminary results from the National Crime Survey CATI experiment. In: American Statistical Association 1988 proceedings of the Section on Survey Research Methods. American Statistical Association, Washington, DC

Huizinga D, Elliot DS (1985) A preliminary examination of the reliability and validity of the National Youth Survey self-reported delinquency indices. National Youth Survey Project Report No. 2. Behavioral Research Institute, Boulder, CO

Huizinga D, Elliot DS (1986) Reassessing the reliability and validity of self report delinquency measures. J Quant Criminol 2:293–327

Jabine TB, Straf ML, Tanur JM, Tourangeau R (eds) (1984) Cognitive aspects of survey methodology: building a bridge between disciplines. National Academy Press, Washington, DC

Junger-Tas J, Ineke Marshall (1999) The self-report methodology in crime research. In: Tonry M (ed) Crime and Justice: a review of research, vol 25. University of Chicago, Chicago, pp 291–359

Kazemian L, Farrington DP (2005) Comparing the validity of prospective retrospective and official onset for different offending categories. J Quant Criminol 21:127–147

Kindermann C, Lynch JP Cantor D (1997) The effects of the redesign on victimization estimates. U.S. Department of Justice, Washington, DC

Kobelarcik E, Alexander C, Singh R, Shapiro G (1983) Alternative reference periods for the National Crime Survey. Proceedings of the American Statistical Association Survey Methods Section, 1983. American Statistical Association, Washington, DC

Koss M (1996) The measurement of rape victimization in crime surveys. Crim Justice Behav 23:55–69
Lauritsen J (1998) The age crime debate: assessing the limits of longitudinal self-report data. Soc Forces 77:127–155
Lauritsen J (1999) Limitations on the use of longitudinal self-report data: a comment. Criminology 37:687–694
Lehnen RG, Skogan WG (1981) The National Crime Survey: Working papers, volume 1: current and historical perspectives. U.S. Department of Justice, Washington, DC
Lepkowski J (1981) Sample design issues from the National Crime Survey. Survey Research Center, University of Michigan, Ann Arbor, MI
Loftus E, Marburger W (1983) Since the eruption of Mt. St. Helens has anyone beaten you up? Mem Cogni 11: 114–120
Lynch JP (2006) Problems and promise of victimization surveys for cross-national research. In: Tonry M (ed) Crime and Justice: a review of research. University of Chicago Press, Chicago, pp 229–287
Martin E with Groves R, Maitlin J, Miller P (1986) Report on the Development of Alternative Screening Procedures in the National Crime Survey. Bureau of Social Science Research, Inc. Washington, DC
Mennard S, Elliot D (1990) Longitudinal and cross-sectional data collection and analysis in the study of crime and delinquency. Justice Q 7:11–54
Miller PV, Groves RM (1986) Matching survey respondents to official records: an exploration of validity in victimization reporting. Public Opin Q 49:366–380
Murphy LR, Cowan CD (1984) Effects of bounding on telescoping in the National Crime Survey. In: Lehnen R, Skogan W (eds) The National Crime Survey: Working Papers, vol II. U.S. Department of Justice, Washington, DC
Nagin D (1999) Analyzing developmental trajectories: a semi-parametric group based approach. Psychol Methods 4(2):139–157
Nagin D (2004) Response to "methodological sensitivities to latent class analysis of long term criminal trajectories". J Quant Criminol 20:27–35
Penick BKE, Owens M (1976) Surveying crime. National Academy Press, Washington, DC
Persley C (1995) The National Crime Victimization Survey redesign: measuring the impact of new methods. In: American Statistical Association 1995 proceedings of the Section on Survey Research Methods. American Statistical Association, Washington, DC
Piquero AR (2008) Taking stock of development trajectories of criminal activity. In: Lieberman A (ed) The long view of crime: a synthesis of longitudinal research. Springer, New York
Planty M (2007) Series victimization and divergence. In: Lynch JP, Addington LA (eds) Understanding crime statistics: revisiting the divergence of the UCR and the NCVS. Cambridge University Press, Cambridge, UK
Rand M, Cantor D, Lynch JP (1997) Criminal victimization, 1973–95. Bureau of Justice Statistics, Washington, DC
Reiss AJ (1967) Measurement of the nature and the amount of crime: studies in crime and law enforcement in major metropolitan areas. President's Commission on Law Enforcement and Administration of Justice, vol. 1 of Field Surveys no. 3. U.S. Government Printing Office, Washington, DC
Reiss AJ (1982) Victimization Productivity in Proxy Interviewing. Institution for Social and Policy Studies, Yale University, New Haven, CT
Rennison CM, Rand M (2007) Introduction to the National Crime Victimization Survey. In: Lynch JP, Addington LA (eds) Understanding crime statistics: revisiting the divergence of the NCVS and UCR. Cambridge University Press, Cambridge, UK
Roberts J, Mulvey E, Horney J, Lewis J, Arter M (2005) A test of two methods of recall for violent events. J Quant Criminol 21:175–194
Rosenfeld R (2007) Explaining the divergence between UCR and NCVS aggravated assault trends. In: Lynch JP, Addington LA (eds) Understanding crime statistics: revisiting the divergence of the NCVS and UCR. Cambridge University Press, New York
Sampson R, Laub JH, Eggleston E (2004) On the robustness and validity of groups. J Quant Criminol 20:37–42
Skogan WG (1981) Issues in the measurement of victimization. U.S. Department of Justice, Bureau of Justice Statistics, Washington, DC
Thornberry TP (1989) Panel effects and the use of self-reported measures of delinquency in longitudinal studies. In: Klein MW (ed) Cross-national research in self-reported crime and delinquency. Kluwer, Los Angeles
Thornberry TP, Krohn MD (2000) Self-report surveys as measures of delinquency and crime. In: Duffe D (ed) Criminal Justice 2000: measurement and analysis of crime and justice. United States Department of Justice, Washington, DC, pp 33–84
Tjaden P, Thoennes N (1998) Prevalence, incidence, and consequences of violence against women: findings from the National Violence Against Women Survey. U.S. National Institute of Justice, Washington, DC
Turner A (1972) San Jose Methods Test of Known Crime Victims. U.S. Department of Justice, Law Enforcement Assistance Administration, National Institute of Law Enforcement and Criminal Justice, Washington, DC