Published in final edited form as:

J Quant Criminol. 2011 June 1; 27(2): 151–171. doi:10.1007/s10940-010-9101-y.

Reliability and Validity of Prisoner Self-Reports Gathered Using the Life Event Calendar Method

James E. Sutton,
California State University at Chico

Paul E. Bellair,
The Ohio State University

Brian R. Kowalski,
Ohio Department of Rehabilitation and Correction

Ryan Light, and
University of Oregon

Donald T. Hutcherson
Ohio University, Lancaster

Abstract

Data collection using the life event calendar method is growing, but reliability is not well established. We examine test-retest reliability of monthly self-reports of criminal behavior collected using a life event calendar from a random sample of minimum and medium security prisoners. Tabular analysis indicates substantial agreement between self-reports of drug dealing, property, and violent crime during a baseline interview (test) and a follow-up (retest) approximately three weeks later. Hierarchical analysis reveals that criminal activity reported during the initial test is strongly associated with responses given in the retest, and that the relationship varies only by the lag in days between the initial interview and the retest. Analysis of validity reveals that self-reported incarceration history is strongly predictive of official incarceration history although we were unable to address whether subjects could correctly identify the months they were incarcerated. African Americans and older subjects provide more valid responses but in practical terms the differences in validity are not large.

The life event calendar (LEC) method is commonly employed to assess relationships between changing social circumstances and criminal behavior over the life course.¹ High incarceration rates add momentum to this trend as scholars try to understand how lives unravel in the months leading to incarceration (Horney et al. 1995) and to identify the characteristics that facilitate successful reentry. Criminologists use the LEC method to study homeless youth (Hagan and McCarthy 1998; Whitbeck, Hoyt, and Yoder 1999), life course trajectories (Laub and Sampson 2003), victimization (Wittebrood and Nieuwebeerta 2000; Yoshihama et al. 2002; Yoshihama et al. 2005), arrestees, probationers, and prisoners (Kruttschnitt and Carbone-Lopez 2006; Lewis and Mhlanga 2001; MacKenzie and Li 2002; Morris and Slocum 2010; Yacoubian 2003), and drug and alcohol use (Day et al. 2004; Sobell et al. 1988). Although the LEC method is gaining acceptance, whether prisoners or other difficult to reach subjects provide reliable and valid self-reports remains a critical and under-researched question.

*The second author is principal investigator.

¹The life event calendar (LEC) method is alternatively referred to as the event-history calendar method, the icon calendar method, the timeline method, the timeline-follow back method, and the life-history calendar method.

It is known that subjects generally provide reliable and valid responses to delinquency items in self-report surveys (Hindelang, Hirschi, and Weiss 1981; Thornberry and Krohn 2000). However, the reference period in most surveys reflects the previous year whereas the LEC we employ seeks greater specificity and assumes that respondents can further identify the months in which their criminal behavior or life changes occurred.² Relative to conventional surveys the LEC method requires greater accuracy and therefore may be more susceptible to memory lapse and hence inconsistent response. In addition the LEC method is often employed to study subjects that lead chaotic lives which poses additional challenges to consistency. And despite assertions that offenders are minimally deceptive in self-report research, many if not most non-criminologists need substantial convincing that self-reported data collected from prisoners is reliable, valid, and hence meaningful.

The most common strategy employed to evaluate reliability of the LEC method is comparison of item responses during administration of the LEC to those provided in a previous prospective interview. Using that standard the literature suggests that the LEC method produces consistent data (see Caspi et al. 1996; Engel, Keifer, and Zahm 2001; Freedman et al. 1988, but casts some doubt on its reliability and validity among criminally involved subjects. For instance, Roberts et al. (2005) compared life event calendar data collected from mentally disordered patients with a history of violence to data collected from those subjects in a previous, prospective interview. They found that respondents substantially under-reported violence during calendar interviews relative to prospective interviews thus calling the reliability of the LEC approach into question. Yet, acknowledging that their test was extremely conservative because the subjects exhibited disorders and “were specifically chosen for their significant histories of substance use and violence,” they concluded that “drastically different results could be found in a different sample (e.g., prison inmates, college students)” (Roberts et al. 2005: 189).

More recently, Morris and Slocum (2010) addressed the validity of data collected using the LEC approach by examining correspondence between monthly self-reports of arrests gathered from incarcerated women and their official arrest records. They found that “the LEC elicits valid data on prevalence and frequency of arrests, while the self-reported timing of arrests is recalled with less accuracy” (Morris and Slocum 2010: 210) However, in general, the timing of arrests was reported with greater accuracy during the year preceding incarceration than it was two to three years prior, and this was not impacted by measures of the saliency of the arrest or the subject's self-reported drug use. Further, they found that arrests were more accurately placed in time when the arrest coincided with a period of incarceration (a memorable cognitive landmark).

This study addresses somewhat different issues and contributes to the literature by examining test-retest reliability of monthly self-reports of criminal behavior collected using the LEC method from a random sample of minimum and medium security prisoners. Virtually all previous reliability tests of the LEC method focus on memorable life-course markers and transitions such as school attendance, employment, marriage, or child bearing rather than criminal incidents which may be more difficult to place in time. Given the consequences of growing prison populations for state budgets and the debates over how to reduce their size, this research addresses reliability and validity issues at a time when more research is being focused on prisoners. Moreover, surveying prisoners presents an excellent opportunity to study a drug and alcohol abusing, criminally active population because subjects are much less likely to be under the influence of alcohol or drugs relative to

²The focus in this paper is monthly self-reports, but the LEC method is not defined by a particular calendar length. Rather, the calendar is a function of research questions and prior theory. Thus, calendars may appropriately be constructed to document changes occurring hourly, daily, weekly, monthly, yearly, or over multi-year periods across the life course.

interviews completed on the street. We also address concurrent validity by comparing the frequency, but not timing of self-reported prison terms with benchmark data collected from official prison records. In the following sections foundations of the LEC method, reliability, and validity are reviewed.

The Life Event Calendar (LEC) Method: An Emerging Research Strategy

The fundamental premise of life course theory, that human lives can be understood as a set of experiences and events that are interconnected and mutually reinforcing (Wheaton and Gotlib 1997), is a guiding principle in the development of the LEC (Freedman et al. 1988). The LEC method is informed by the insight that memories are organized within the brain in memory structures that are both patterned and interrelated (Belli 1998; Bradburn, Rips, and Shevell 1987; Caspi et al. 1996; Sudman, Bradburn, and Schwarz 1996). More formally, Belli (1998: 385) states that “personally experienced events are structured in autobiographical memory by hierarchically ordered types of memories for events that vary in prevalence, frequency, and timing, and this structure is organized along temporal and thematic pathways which guide the retrieval process.” LEC's are specifically designed to tap into these temporal and thematic pathways, making them advantageous over traditional survey methods (Belli, Shay, and Stafford 2001).

The LEC method entails interviewers working with respondents to fill out monthly calendars that map the occurrence of events (Freedman et al. 1988). This mode of administration enables researchers to efficiently collect complicated longitudinal data from respondents (Axinn, Pearce, and Ghimire 1999). Given that survey respondents may “telescope,” or inaccurately report that an event happened during the researcher's time frame when it in fact occurred earlier or later (Sudman and Bradburn 1974), an advantage of using the LEC method is that it facilitates respondents' ability to more accurately remember the timing of key events in their backgrounds (Axinn, Pearce, and Ghimire 1999; Lin, Ensel, and Lai 1997). Drawing on this method an interviewer may, for instance, begin an interview by asking the respondent to indicate memorable events on the calendar such as a birthday, dates of schooling or employment, or childrens' birthdays. These events might then be used to trigger memories of mundane or taken-for-granted occurrences and activities that occurred in proximity to more memorable events (Caspi et al. 1996).

A methodological advantage of the LEC method over standard techniques is that the format facilitates interaction between the interviewer and respondent. Traditional surveys employ a formal style of interviewing characterized by unilateral exchanges that do little to establish rapport (Cannell and Kahn 1968: 527; Fontana and Frey 2003). In contrast, in the LEC method interviewers read introductory statements that frame sets of interrelated questions and then work together with respondents to fill out calendars (Belli, Shay, and Stafford 2001). This creates a conversational dynamic that is non-threatening, making it easier to pose follow-up questions aimed at clarifying responses (Belli, Shay, and Stafford 2001; Caspi et al. 1996). Engel, Keifer, and Zahm (2001) compared the LEC method to a traditional self-report survey and found that subjects put more effort into providing correct information and were more cooperative and patient during the LEC administration.

Assessment of Reliability

Social science is concerned with data quality, which is typically evaluated by assessing reliability and validity (Babbie 1995: 124). Reliability is defined as “the extent to which a measuring instrument produces consistent results” (Kinnear and Gray 2006: 548). There are four primary methods of assessing reliability, each of which has strengths and weaknesses. The split halves and internal consistency methods are typically utilized when a study is unable to collect repeated measures. In contrast the alternative form and test-retest methods

are preferred because they entail collection of repeated measures and hence directly assess the issue of consistency.

The test-retest method uses the same instrument in both administrations and thus a lack of reliability cannot be attributed to use of alternative items (Singleton and Straits 1999: 117). For this reason it is the most commonly used reliability test in the social sciences (Litwin 1995) and is widely considered the best strategy (Thornberry and Krohn 2000). When administering the test and retest it is important that researchers leave enough time between contacts with respondents to minimize conditioning effects (Litwin 1995). Thornberry and Krohn (2000) advise researchers to use intervals of one to four weeks when developing test-retest designs in criminological research. Moreover, they recommend that reliability and validity coefficients (i.e., gamma) should exceed a minimum threshold of .7. We adopt that threshold. We also report values of kappa, a measure commonly used to address agreement between observers in medical research. For kappa we adopt a threshold of .6 given reviews that classify values above that level as an indication of “substantial agreement” (Viera and Garrett 2005: 362, Table 2).

Assessment of Validity

According to Kinnear and Gray (2006: 550) “a test is said to be valid if it measures what it is supposed to measure.” Validity is affected by nonrandom error that occurs when research instruments capture processes other than the ones they set out to study (Carmines and Zeller 1979). Assessment of validity can be seen as a continuum beginning with content validity and progressing through construct and criterion validity. Content validity refers to whether items designed to measure a phenomenon seem to reasonably and logically capture it. Construct validity is evident when items intended to measure a construct correlate with other constructs in theoretically expected ways (Carmines and Zeller 1979).

Junger-Tas and Marshall (1999) assert that establishing criterion validity is more convincing than construct validity. There are two forms of criterion validity: predictive and concurrent (Litwin 1995). Predictive validity refers to how well an indicator predicts future outcomes (Carmines and Zeller 1979). Besides prediction, criterion validity is also concerned with the consistency of findings across multiple data sources (Huizinga and Elliott 1986; Weis 1986). This form is known as concurrent validity (Carmines and Zeller 1979), and it entails comparison of self-reported data with data collected from a different source (Northrup 1997). A true gold standard does not exist (Hindelang, Hirschi, and Weiss et al. 1981), but criminologists traditionally compare adolescents' self-reported criminal justice contacts, such as being stopped by police or frequency of arrest, to their official record, such as police records of arrests or court records of convictions to determine whether the accounts match up and draw conclusions about the validity of data (Thornberry and Krohn 2000).

Hypotheses

An unresolved and central issue is whether there are race differences in the reliability or validity of self-reports. For instance, some literature suggests that African American youth provide less valid self-reports compared to whites (Hindelang, Hirschi, and Weiss 1981; Fendrich and Vaughn 1994; Mensch and Kandel 1988). Those findings were produced using traditional survey methods and cross-sectional designs. However, some scholars report that validity of African-American self-reports is high and comparable to white respondents (Jolliffe 2003; Farrington et al. 1996; Morris and Slocum 2010). Thornberry and Krohn (2000: 58) state that “this is perhaps the most important methodological issue concerning the self-report method and should be a high priority for future research efforts.” In addition, prior research indicates that low educational achievement is related to lower reliability in

self-report surveys (Golub et al. 2002; Mensch and Kandel 1988), though other researchers find that education does not affect reliability (Chaiken and Chaiken 1982).

Subjects with little or no legitimate income, who have greater illegal income, or are young may provide less reliable responses. For instance, subjects with little or no legitimate income are less structured by conventional pursuits such as employment and perhaps less committed to mainstream society. Low conventional bonding to employment may undermine their willingness to think carefully and recall the details of life circumstances during the interview. Likewise, subjects with greater illegal income may be less motivated to accurately report criminal involvement due to lower levels of commitment. They may also have less confidence that researchers will keep their responses confidential, especially if they think honesty would jeopardize a potentially ongoing stream of illegal income. Given the age-crime curve, young inmates may have higher rates of participation in criminal behavior at higher frequency than older inmates and this may reduce their level of trust. Younger inmates are more likely to be in prison for the first time and thus may have less experience discussing their lives with researchers, which likewise may reduce trust.

Reliability may also decline over the course of the recall period or as a function of testing procedures. For instance, subjects may not enumerate criminal events that occurred several months prior as reliably as events that occurred more proximally due to memory decay (see Morris and Slocum 2010). Finally, prior research indicates that the strength of reliability coefficients decrease as the lag between the test and retest increases (Bachman 1970; Farrington 1973). Here the issue is addressed by testing whether the stability of self-reports declines with an increase in the number of days between the initial test and the re-test. We test the following hypotheses pertaining to reliability:

H₁: The relationship between self-reported criminal behavior during the initial test and responses provided in the re-test equals or exceeds a gamma value of .7 and a kappa value above .6.

H₂: The relationship between self-reported criminal behavior during the initial test and responses provided in the re-test is not conditioned by race, younger age, low education, diminished legitimate income, greater illegal income, memory decay over time, or a greater time lag between tests.

We examine concurrent validity by examining the strength of the relationship between frequency of self-reported prison terms and frequency of official prison terms revealed in a search of official prison records. The following hypotheses state expectations regarding concurrent validity:

H₃: The relationship between self-reported criminal history and criminal history measures derived from official prison records equals or exceeds a gamma value of .7 and a kappa value above .6.

H₄: The relationship between frequency of self-reported prison terms and frequency of official prison terms is not conditioned by race, younger age, low education, diminished legitimate income, and greater illegal income.

Data and Measures

The data are derived from life event calendar interviews with 250 prison inmates randomly selected from four minimum/medium security Ohio Department of Rehabilitation and Correction (ODRC) state prisons during 2005-2007.³ The initial goal was to conduct

³Our protocol initially specified interviews with female prisoners. This part of our request was denied by the ODRC institutional review board (IRB) because of the large number of ongoing research projects being conducted in the Ohio prisons that house females.

interviews at the ODRC reception centers. All inmates entering the ODRC system are routed through them, and thus executing the study at reception centers would produce a representative sample of all state prisoners without requiring extensive travel to multiple institutions. The ODRC review board, however, denied our request because most prisoners transfer relatively swiftly from the reception center to their parent institution. Thus, completion of a re-test for those subjects was deemed problematic. They suggested that the best chance of success was to conduct interviews in a small number of minimum/medium security institutions.⁴ Approximately 70% of the ODRC prison population is comprised of minimum and medium security prisoners and thus our results generalize to a large proportion of the prison population. To maximize variability and to minimize our intrusion at each site the interviews were spread across four institutions.

Prison populations, particularly minimum/medium security institutions, are dynamic and constantly churning with new inmates arriving and departing weekly. To achieve a random sample we employed consecutive sampling beginning with the most recently admitted inmates. The procedure yields a random sample because the temporal flow of inmates into prison approximates a random process. The sampling frame is comprised of all prisoners between 18 and 32 years old at each institution admitted to ODRC up to one year prior to our study. We focused on that age group to avoid including older prisoners who had worked their way down to minimum security with good behavior after having spent several years in a more secure facility. It was assumed that those subjects would have more difficulty recalling the 18 month calendar period.

Recruitment was a two step process. Small groups of inmates drawn from the sampling frame were issued passes to meet with project staff in a semi-private setting, such as a classroom, Chapel, or visiting room. The prisoners were told that the study was focused on previous criminal behavior, drug use, their life history across several domains such as family and education, and that Ohio state law specifically prohibits compensating prisoners for participation in research. 250 out of 468 prisoners drawn from the sampling frame volunteered to be interviewed yielding a 53% response rate. A consent form was administered at the beginning of the interview, including a discussion of subject's rights, confidentiality procedures, and the voluntary nature of the study.⁵ With few exceptions two interviewers were present during each interview. The typical interview spanned an hour and ten minutes while the re-test averaged thirty minutes.⁶ Because of the time and expense of conducting a second interview the re-test was limited to the first 110 subjects.

In Table 1 subjects that completed the first interview and a re-test (column 1) are contrasted with the entire sample (column 2), with refusals (column 3), with the sampling frame (column 4), and with the statewide population of male prisoners admitted to ODRC in 2005 (column 5) across race, age at admission, and number of previous times incarcerated. Perhaps the most crucial comparison, given the 53% response rate, is between the characteristics of our sample and the refusals, and between the sample and the sampling frame from which it is drawn. The sample is not significantly different from the refusals and with one small exception is representative of the sampling frame, mirroring racial composition and number of prior incarcerations. The subjects that completed the test and retest are slightly older than the subjects comprising the sampling frame. The sample also favorably mirrors the racial composition of the statewide prison population. However, there

⁴ODRC's decision about which institutions we could visit was based largely on how many research projects had been recently approved, and which institutions those projects were conducted in. The institutions we were granted access to had not fielded as many recent studies.

⁵Confidentiality is protected by a Certificate of Confidentiality from the National Institutes of Health.

⁶The retest was much shorter because it contained less than half the questions included in the test, with a focus on criminal behavior and other topics collected with the calendar.

are noteworthy differences in the frequency of prior incarceration and age at admission. Yet, those differences are expected because this study includes minimum and medium security prisoners whereas the state prison population also includes prisoners at higher security levels who are, on average, older and have more extensive criminal histories.

Format of the Study

The instrument was developed and extensively pre-tested in mock interviews for about one year prior to data collection by the same staff that conducted interviews.⁷ During that time all components of the data collection including the recruitment script, informed consent, and the questionnaire were extensively rehearsed. After the project went into the field new interviewers were recruited and extensively trained via a graduate-level research practicum taught by the principal investigator. Interviewers were carefully observed and critiqued during training to ensure consistency. After each completed interview the interviewers were debriefed by the principal investigator to ensure that any confusion or problems were identified and addressed.

After the consent form was signed at the outset of the interview, subjects were asked to identify the month in which they were arrested for the offense that led to their incarceration. The month immediately prior was treated as month 18 in the data collection. The questionnaire used month 18 as the focal point for most questions but for a sizeable subset of topics, including criminal behavior, subjects were asked whether there were changes during the preceding 17 months. If so the monthly changes were recorded on the calendar. After making sure that the subjects understood the calendar format they were asked if there were months during the calendar period when they were incarcerated or in residential treatment (or off the street for some other reason). If so those months were blocked off on the calendar to avoid inadvertently entering other life events.⁸

Next a series of questions about life events that should be readily recalled were asked including: date of birth, residential addresses and neighborhood conditions, marital status, divorce, child birth, school history, employment history, legal and illegal income, social support networks, stressful events such as death in the family, et cetera. The criminal behavior items followed those questions, but not before the subjects were again reminded that their participation and responses are voluntary and confidential.

Variables

Reliability—Analysis of reliability is within individual, addressing whether subjects accurately recall the specific months in which they committed crimes across two interviews. H_1 is examined using cross-tabulation and examination of gamma and kappa coefficients. H_2 is addressed with a two level “coefficient as outcome” hierarchical model. In those models (described below) the level 1 unit of analysis is street months, which are nested within persons (level 2). The dependent variable, *criminal behavior (re-test)*, is a scale ranging from 0 to 3 constructed by summing subject's participation in three broad (single item) categories of crime in each month including drug dealing, property, and violent crime and is formed using data collected during the re-test.⁹ Thus, in any particular month a

⁷We paired male and female interviewers together in the majority of interviews. Two out of eight primary interviewers are African American, and half are female. The remaining interviewers are White males. We also attempted to balance age when assigning interviewers.

⁸Each subject's responses were recorded simultaneously on a paper calendar that was kept in front of the subject for reference and on the electronic version of the calendar maintained by the second interviewer (the laptop was positioned so the subject could see the calendar screen).

⁹We focus on whether subjects correctly specified the months in which they committed crimes. Out of necessity to reduce the length of the interview, we did not collect frequency data for each month crimes were reported. Rather, if subjects reported that they committed crime in any month a follow-up question asked subjects to report the frequency of offending in the typical month.

subject may have refrained from drug dealing, property, or violent crime (coded 0), engaged in one of the three types (coded 1), in two of the three (coded 2), or all three (coded 3).¹⁰ We treat the outcome as a Poisson sampling distribution with constant exposure and over-dispersion.¹¹ The key independent variable in the analysis is *criminal behavior (test)* reported during the initial interview and is constructed using the same procedures used to form the dependent variable. To the extent that the data are reliable, criminal behavior reported during the initial interview should correspond closely to, and be a strong predictor of, criminal behavior reported during the re-test. The descriptive statistics indicate that subjects participate in just over 1 of 3 types of crime for every two months on the street during the retest (i.e., $.59 \times 2$), and a comparable number during the test (i.e., $.63 \times 2$). In the final step of the reliability analysis an interaction effect at level 1 is formed between *Month* and criminal behavior (test) to assess whether the reliability of self reports decline as they are reported backwards in time. Descriptive statistics for the reliability and validity analysis are presented in Table 2. Two hundred and eighty months are excluded from the analysis due to the subject being off the street (and hence not at risk for street crime) resulting in 1,700 street months of data nested within 110 subjects.

The within-individual relationship between criminal behavior (test) and criminal behavior (re-test) is modeled as an outcome that varies as a function of person-level characteristics in a hierarchical analysis. Several person-level characteristics that literature suggests may pattern test-retest reliability are considered. Each subject's self-reported racial group is captured in dummy variables that are contrasted with whites. *Black* is coded one if the subject is African American, while *other* primarily reflects multi-racial origins although a few Hispanic subjects are also included in this category. Approximately 43% of the sample is black, 55% are white, and roughly 2% are multi-racial or Hispanic. *Age* is measured in years at the time of the initial interview. The typical respondent is 24.6 years old. *Education* is measured as the number of years of completed schooling, and averages just over 11 years. *Legal monthly income* reflects the average monthly income earned by subjects across the 18 month calendar period, and averages approximately \$1,201 per month. *Illegal monthly income* is similar to the legal income measure but tallies income that was generated by criminal behavior. On average, subjects produced about \$554 of illegal income per month, typically from selling drugs or stolen goods. The final variable included in the reliability analysis is the lag in days between administrations of the initial interview and the re-test (*test/re-test lag*). The typical subject completed a retest approximately three weeks (i.e., 22.9 days) after the initial interview was administered, although there was variation in the timing of the retest because of scheduling difficulties that occurred for a variety of reasons.

Validity—The analysis of validity is conducted at the person level, investigating the relationship between the frequency of self-reported incarceration history and official incarceration history derived from ODRC records. The validity analysis, therefore, does not bear directly on issues specific to the LEC method such as the timing of incarceration but addresses a more general issue concerning the validity of prisoner self reports.¹² Previous research indicates that prisoner self-reports are reasonably valid (for a review of research that assesses the validity of prisoner self-reports see Blumstein et. al. 1986). However, as several scholars note (see Thornberry and Krohn 2000), analyses that examine differential

¹⁰There is a substantial body of literature positing a relationship between the salience of a criminal incident and accuracy of recall. We attempted to model the three components of the criminal behavior scale independently as binary outcomes, but the HLM program was unable to compute robust standard errors for those models. Tabular results for each component are presented in Appendix 1, indicating minimal difference in reliability among the component items of the criminal behavior scale.

¹¹At level-1 we model: $\log(\hat{\mu}_{ij}) = \log(\mu_{ij})$, where μ_{ij} is the event rate reflecting the number of self-reported crimes during the retest and $\hat{\mu}_{ij}$ is the log of the event rate. Note that while μ_{ij} is constrained to be non-negative, $\log(\hat{\mu}_{ij})$ can take on any value. The predicted log event rate can be converted to an event rate by generating $\hat{\mu}_{ij} = \exp(\log(\hat{\mu}_{ij}))$.

¹²Subjects were not asked to provide the dates of incarceration in state prisons, thus precluding examination of the timing issue.

validity by race are especially important and a high priority for future research. H_3 is addressed with cross-tabulation and examination of gamma and kappa coefficients. H_4 is addressed using poisson regression models.

The highest quality criminal history information documented in ODRC records is each subject's previous history of incarceration in ODRC prisons.¹³ Incarceration history is a salient life event that is likely to be recalled with more accuracy than less salient events such as arrests or police stops. Thus, while our validity analysis is not as conservative as it might be if events less salient than incarceration were included, it does provide a gauge of whether prisoners provide accurate information to researchers and whether it varies, for instance, by race. Data are available for all 250 subjects, and were hand coded. We therefore model the number of previous *ODRC prison terms* as a function of the number of *self-reported prison terms* during the initial interview. We next assess whether the relationship between the number of previous *self-reported prison terms* and the number of *ODRC prison terms* varies across person-level characteristics suggested as important in prior literature (including race, age, education, legal income, illegal income, and an interaction between month and self-reported crime during the test) by assessing interaction effects. To facilitate the analysis product terms are formed after first centering each continuous variable about its mean (Aiken and West 1991).

Results

Reliability

Analysis of test-retest reliability begins in Table 3 with a cross-tabulation of self-reported criminal behavior during the initial interview (test) with behavior reported during the re-test. A cross-tabulation of the individual items that comprise the crime scale is presented in Appendix 1. Of particular interest is whether cases fall along the diagonal, which indicates correspondence between self-reports. Cases that fall above the diagonal reflect street months in which more criminal behavior was reported during the retest than in the initial interview, while cases appearing below the diagonal indicate street months in which less criminal behavior was reported during the retest than the first interview. The data indicate that there are 1,429 (651+731+47) street months out of 1700 in which self-reported criminal behavior reported during the first interview precisely corresponds with behavior reported in the retest corresponding to 84% agreement. Both measures of association, gamma (.914) and kappa (.709), indicate a substantial association between responses that is well above the threshold of acceptable reliability. There is some evidence of under-reporting criminal behavior from the first interview to the retest – i.e., there are 172 cases below the diagonal versus 99 above it. However, the descriptive statistics indicate a minimal difference in the mean number of crimes reported between the test and retest and similar variance. Overall, the tabular evidence indicates that the data are reliable indicating support for H_1 . The cross-tabulations presented in Appendix 1 are also supportive.

A limitation of the tabular approach that compares test and retest responses is that it does not account for the nesting of street-months within individuals. Further, test-retest correspondence may be influenced by time invariant person-level characteristics such as biological and social-psychological traits (i.e., intelligence, low self-control, etc.). These issues are addressed in Table 4 with a 2-level hierarchical analysis that provides a more

¹³Arrest data maintained by ODRC (our data source) is less consistently recorded as a result of missing pre-sentence investigation (PSI) paperwork from the criminal history records of a significant proportion of cases. Collection and analysis of arrest data also requires significant resources due to the amount of time required to hand code official data from ODRC data bases into a file matching the 18 month calendar. We hope to have resources to collect and compare monthly self-reports of arrests with monthly official arrests in the future.

rigorous approach. The level 1 model is expressed as follows: $y_{ij} = \alpha_{0j} + \alpha_{1j}X_{1j} + \epsilon_{0j}$ where y_{ij} is the logged number of crimes self-reported during the retest; α_{0j} is the within-individual intercept for person j ; X_{1j} is a time varying covariate reflecting the number of self-reported crimes during the initial interview with α_{1j} reflecting its association with crime reported during the retest; X_{1j} is group mean centered so that time invariant traits are held constant; and ϵ_{0j} is a level 1 random effect that represents prediction error.

The key question addressed is whether the relationship between self-reported crime during the test and retest varies across person-level characteristics. Thus coefficient α_{1j} is modeled as a function of person level characteristics and random error while the level 1 intercept α_{0j} is conceptualized as randomly varying between persons but is not modeled. More specifically, two level 2 equations are estimated: (1) $\alpha_{0j} = \gamma_{00} + \mu_{0j}$ and (2) $\alpha_{1j} = \gamma_{10} + \gamma_{1j}W_j + \mu_{1j}$. α_{0j} is the level 1 intercept modeled at level 2 as a function of an intercept γ_{00} and a random error term μ_{0j} . The relationship between self-reported crime during the first interview and the retest (α_{1j}) is modeled as a function of an intercept γ_{10} , several predictors W_j including race, age, education, legal and illegal income, the lag between test and retest, and a level 2 random effect μ_{1j} reflecting prediction error (for an overview of the statistical procedures see Luke 2004 10, or Raudenbush and Bryk 2002).

The results of hierarchical modeling are presented in Table 4. Model 1 is the unconditional model which indicates that self-reported crime during the retest varies significantly across persons. The intra-class correlation is .976 indicating that approximately 97.6% of the variance is between persons, whereas 2.4% is within-individual. The percentage of variance that is within individual is small, reflecting the small proportion of each prisoner's life (i.e., 1.5 years out of 24 years on average) that is captured in the 18 month calendar period. In model 2, self-reported crime from the first interview is added as a predictor of self-reported crime during the retest. To the extent that the data are reliable self-reported crime from the first interview should correspond strongly and significantly with crime reported during the retest. This is clearly the case and, because of the group mean centering, the results can not be attributed to traits that might increase or decrease the chances of reliable responses such as intelligence or self-control. In model 3 the relationship between test and retest responses, α_{1j} is specified as a level 2 outcome and regressed on several person level predictors. Consistent with previous research reviewed above, results indicate that the lag (in days) between administration of the test and retest undermines reliability – i.e., as more days elapse between the test and retest the reliability of responses decreases. None of the other predictors, including age, race, education, and legal and illegal income pattern reliability. In model 4 an interaction effect between month and criminal behavior (test) is included to assess whether the reliability of self reports decline as they are reported backwards in time. The interaction is not significant indicating that reliability does not decay within the 18 month recall period. The results in Table 4 support the hypothesis that LEC data collected from prisoners yields reliable data, and thus we accept H_2 .

Validity

Concurrent validity is addressed in Table 5, where the number of previous ODRC prison terms (row) is cross-tabulated with self-reported prison terms (column). Of particular interest is whether cases fall along the diagonal because they indicate correspondence between ODRC records and self-reported prison terms. Cases that fall above the diagonal reflect subjects that reported more state prison terms than are reflected in the ODRC records, while cases that fall below the diagonal reflect subjects that reported fewer state prison terms than are reflected in their records. The data indicate that 210 (143+47+15+ 4+1) subjects out of 250 (84%) accurately reported the number of times they served state prison terms. Both gamma (.964) and kappa (.715) indicate substantial correspondence between self-reports and ODRC's records, and both are well above the established threshold. For the

most part there are few cases of underreporting (i.e., cases that fall below the diagonal). Rather, most discrepancies are the result of over-reporting state prison terms served during the interview. Over-reporting prison terms served is, in our view, most likely attributable to confusion of jail sentences subjects may have served for state prison terms, or to self-reports of state prison terms served in another state. In any event there are very few cases in which there is a discrepancy of greater than one when self-reported and ODRC prison terms are contrasted suggesting greater confidence that the interviews yielded valid responses. We therefore accept H₃.

Concurrent validity is further examined in Table 6 with poisson regression of ODRC state prison terms on self-reported prison terms and interaction effects that assess whether that relationship varies by subject's socio-demographic characteristics. As we noted previously, the continuous independent variables modeled in Table 6 were mean centered prior to forming product terms. In model 1 self-reported prison terms and socio-demographics are included, whereas model 2 includes the aforementioned and further specifies interaction effects between self-reported prison terms and socio-demographics. Model 1 indicates that the number of self-reported prison terms is strongly predictive of ODRC terms, and that none of the socio-demographic covariates included are significant.

In model 2 interaction effects are included, revealing significant interactions between self-reported prison terms $\times$ black and self-reported prison terms $\times$ age. The former indicates greater validity of self-reports provided by blacks relative to whites, conflicting with older yet well-known results reported in Hindelang, Hirschi, and Weiss's (1981) monograph *Measuring Delinquency*. It is more consistent with recent research reported by Farrington et al. (1996) although they found no differences in validity by race. We interpret the finding as an indication that the conversational dynamic and cooperation fostered in the administration of the LEC breaks down racial barriers and leads to greater validity. Despite statistical significance, in practical terms the findings of differential validity are less meaningful. For instance, the gamma between self-reported prison terms and ODRC prison terms is .99 among blacks and .94 among whites, well above the threshold of acceptability. The latter interaction between self-reported prison terms and age suggests that older inmates provide more accurate accounts of their past. Perhaps older inmates have more experience with researchers and possibly adjust more quickly to the chaotic prison environment, both of which may boost the accuracy of self-reports. However, as was the case with the interaction with race, when age is partitioned into quartiles the relationship (gamma) between self-reported prison terms and ODRC prison terms within each quartile ranges from .90 (lowest) to .97 (highest). Thus, within each subgroup the validity coefficient is sufficiently high and substantially above the minimum threshold to suggest acceptable validity despite some variation between subgroups. We accept H₄.

Discussion

Previous research is consistent with the hypothesis that important life course markers such as marriage or childbearing are reliably measured by the LEC method. However, there is very little research that examines whether rare events such as self-reported criminal behavior can be reliably measured. And while it is known that offenders generally provide reliable and valid responses to delinquency items in social surveys the reference period in most surveys reflects the previous year. The LEC method, in contrast, seeks greater specificity and assumes that respondents can identify the months in which their criminal behavior or life changes occurred. Relative to conventional surveys the LEC method requires greater precision and therefore may be more susceptible to memory lapse and hence inconsistent response. In addition the LEC method is often employed to study subjects that lead chaotic lives, in our case prisoners, which poses additional challenges to consistency.

A recent study by Roberts et al. (2005) evaluated the LEC and found that respondents with psychiatric issues substantially under-reported their violence involvement. They noted that their more pessimistic results may not generalize to other populations such as prison inmates. Consistent with that supposition, our research suggests that the LEC method generates more reliable data among prisoners than subjects suffering from psychiatric issues. I should be recognized, however, that there are important methodological differences between the execution of the Roberts et. al. (2005) study and the execution of this study. We re-interviewed subjects three weeks after the initial test, on average, whereas the follow up period in the Robert et. al. (2005) study was one to three years. Reliability coefficients are known to decline as the time lag between interviews lengthens.

We find over 80% agreement between criminal behavior self-reported during the test and retest and high values of gamma and kappa, and that reliability is not impacted by race, age, education, legal income, or illegal income. Consistent with prior research stability of response is impacted by the amount of time that elapses between the test and retest. This is important because it suggests that the threshold level of reliability considered acceptable in a study should probably be adjusted downward at some point as the average time lag between the test and retest increases. We also find over 80% agreement between the number of self-reported prison terms and the number revealed in official ODRC records, and high values of gamma and kappa which suggest substantial concurrent validity. Subsequent analysis indicates that responses provided by Blacks have greater validity relative to Whites and that responses provided by older subjects have greater validity. Although statistically significant, in both cases the substantive impact is less meaningful – white prisoners and young prisoners still provide acceptably valid responses.

The main implication of this study is that the LEC method appears to yield reliable self-reports of criminal behavior. The validity analysis reflects favorably as well but does not address whether subjects can accurately recall the months in which their criminal justice contacts occurred. Recent research by Morris and Slocum (2010) suggests that self-reported timing of arrest does not correspond as well to the timing of arrest derived from official records. More LEC analyses of the reliability and validity of self-reported criminal justice system contacts (arrests, time periods on probation, time periods incarcerated, etc.) are needed to unravel why criminal justice system contact appears to be less validly measured in LEC studies than other time-varying measures appear to be.

The positive portrayal of the LEC method in previous research is consistent with our experience. The stage was set for the interviews during the recruitment process. We were very careful to treat our subjects with respect and to accommodate them during the interview process in order to stimulate participation. During recruitment we greeted inmates with a handshake, a forthright and honest explanation of the study, and a genuine display of appreciation for taking time out from their day to meet with us. We also paid attention to balancing the gender and race of interviewers to the extent permitted by circumstance. Although difficult to quantify, our procedures seemed to pay off. Prisoners were, in general, easily able to grasp the LEC concept when it was explained to them at the outset of the interview. The very act of explaining it helped to break down social distance between the subjects and interviewers. In the vast majority of cases the conversational dynamic fostered by the LEC piqued subject's interest and enhanced cooperation, particularly when sensitive questions regarding participation in crime were asked. Hughes (1945) observed long ago that individuals often unwittingly expect others to possess auxiliary traits based on their most visible social identities and thus may associate traits such as dishonesty and insincerity with prisoners. The analysis indicates the fallacy of assuming that prisoners are inherently dishonest in interview settings.

There are several limitations which we acknowledge. First, it was decided early on to use an 18-month reference period. We considered a longer calendar but opted for 18 months after pre-testing indicated that the length of the calendar substantially lengthens the time required to complete the interview – a central consideration in this study given that the time available to conduct interviews is constrained by meal times, visitors, program participation, and prisoner counts throughout the day. Thus, it is unclear how far backwards in time the calendar can probe before reliability begins to suffer. Based on the non-significant interaction between criminal behavior (test) and month, our sense is that the calendar period could be expanded. Precisely how far back in time data can be reliably gathered is an important issue that can only be unraveled with more research. Morris and Slocum's (2010) results suggest that our broad conclusion about the LEC may not generalize to arrest outcomes.

Second, one criticism of the test-retest approach to reliability is that subjects may remember what they previously reported and simply replicate it during the retest. We see little merit in that argument due to the volume and relatively rapid pace of questioning during the interview. In addition, the test and retest were separated by about 22 days on average. In our view it strains credibility to argue that subject's could recall their precise responses to well over one hundred questions three weeks later, particularly given the situational demands and burdens of daily prison life.

Third, we acknowledge that our response rate of 53% is lower than that achieved in many social surveys of the general population. The fundamental difficulty of recruiting prisoners for a study of criminal involvement is that subjects are asked to reveal their histories soon after reaching the end point of processing through the criminal justice system. Most prisoners are dealing with negative emotions related to their conviction, their family's response to it, and are in the process of acclimating to an environment that is disorienting and at times volatile (Clemmer, 1940). Without the capacity to receive compensation many prisoners are understandably unwilling to cooperate with strangers. We think that the ability to offer some form of compensation would likely have increased the response rate substantially. Those issues aside, one advantage of the present study over most others is the ability to compare socio-demographic and criminal history characteristics of the sample to refusals, the sampling frame, and to the statewide prison population. This facilitates examination of the consequences of the refusal rate (see Table 1). The sample is not significantly different from the refusals and is clearly representative of the sampling frame from which it is drawn across race, age at admission, and number of prior state prison terms served. Although we would have preferred to reduce the refusal rate the response rate does not appear to have generated a selection process that biased the sample in a fundamental way.

Fourth, and finally, the study sampled minimum and medium security prisoners and thus it is unclear whether the results can be generalized to data gathered from higher security prisons. We suspect that reliable data can be collected in maximum security prisons assuming a well-trained staff, although a lower response rate seems likely without the ability to compensate subjects. In the current political and economic context a research focus on minimum and medium security prisoners may be most relevant given debates about how to achieve a reduction in the size of prison populations. To the extent that states have interest in developing strategies to divert non-violent offenders away from state prisons and towards community correction alternatives, devising strategies to understand and constrain the criminal behavior of minimum/medium security prisoners seems particularly urgent.

The LEC method holds substantial promise for testing life course theories of crime. They posit that the probability of outcomes such as drug dealing, property, and violent crime

increase over time as social circumstances change for the worse. Many life course models place substantial emphasis on structural disadvantage, employment conditions, and family bonds among others. Unfortunately, there is very little longitudinal data with which to test those implications among populations with substantial criminal involvement. In most longitudinal data sets children, adolescents, or young adults are randomly selected from the general population, and interviews take place in schools, at home, or by phone. This is a serious limitation because the prevalence of drug selling, property, and violent crime in general population samples is low, and because those most involved often leave school prematurely, are rarely home, and avoid talking about criminal behavior, especially drug use and dealing, on the phone. Also, when subjects with serious behavioral problems enter population studies they are more likely to attrite because they move frequently and are difficult to locate (Thornberry 1989). The lives of serious drug abusers also change rapidly from month to month because of changing life circumstances but most prospective, longitudinal data sets employ a six month or one year reference period and thus fail to capture short term change (Horney et al. 1995). Application of the LEC method in a prison context holds the potential to help minimize the negative impact of the aforementioned issues in a non-superficial way.

The LEC is a useful technique, but we do not mean to cast it as a silver bullet. This is a technique that places a premium on assembling and training a well-qualified group of interviewers that understand the purpose of each question. It is also a more complicated method than the standard prospective survey that requires careful attention to detail during administration. Finally, it is a time consuming method especially for subjects whose circumstances changed frequently, and this is time that could be spent asking other questions or completing additional interviews. Our experience suggests, however, that the method should continue to be considered a reasonable option when longitudinal data is preferred and the population of interest characteristically seeks to conceal its behavior, is hard to contact, and leads chaotic lives. However, more careful research that addresses the reliability and validity of within-individual data derived from LEC interviews is necessary to establish that the method yields data worthy of scientific inquiry.

Acknowledgments

We gratefully acknowledge research support from the Department of Sociology and especially thank Bob Kaufman, then Chair, for listening to our pleas for help, the Criminal Justice Research Center (CJRC), the Initiative in Population Research (IPR), and the Center for Urban and Regional Analysis (CURA) at The Ohio State University. We are thankful for important contributions from several graduate students including: Rachael Gossett, James Hein, Brianne Hillmer, Ross Kaufman, Anita Parker, Grace Sherman, Matthew Valasik, and Shawn Vest. We thank Julie Horney for graciously providing us with a computer assisted version of a previously used event calendar instrument which provided the starting point for this project. We thank the anonymous reviewers and the Editors for their comments which we think helped improve the paper. We also thank the Ohio Department of Rehabilitation and Correction, especially Gayle Bickle and the staff at the Madison, London, Southeastern, and Richland correctional institutions, for facilitating this research. Finally, data collection would not have been possible without the good will and professionalism shown by the prisoners who agreed to participate without compensation.

Appendix 1

Cross-tabulation of components of the self-reported crime measure (drug, property, and violent crime) during test and re-test.

Drug (property) [violent] TEST	Drug (property) [violent] RE-TEST		
	0	1	TOTAL
0	725	88	813

Drug (property) [violent] TEST	Drug (property) [violent] RE-TEST		
	0	1	TOTAL
1	(1547)	(29)	(1576)
	[1624]	[8]	[1632]
	128	759	887
	(45)	(79)	(124)
	[29]	[39]	[68]
TOTAL	853	847	1700
	(1592)	(108)	(1700)
	[1653]	[47]	[1700]

NOTE: Drug: Gamma = .960*; Kappa = .746*; * p<.001. Property: Gamma = .979*; Kappa = .658*; * p<.001. Violent: Gamma = .993*; Kappa = .667*; * p<.001.

References

- Aiken, Leona S.; West, Stephen G. Multiple Regression: Testing and Interpreting Interactions. Newbury Park, CA: Sage Publications; 1991.
- Axinn, William G.; Pearce, Lisa D.; Ghimire, Dirgha. Innovations in Life History Calendar Applications. *Social Science Research*. 1999; 28:243–264.
- Babbie, Earl. The Practice of Social Research. Seventh. Belmont, CA: Wadsworth; 1995.
- Bachman, G. Youth in Transition. Vol. II. Ann Arbor: University of Michigan, Institute for Social Research; 1970. The Impact of the Family Background and Intelligence on Tenth Grade Boys.
- Belli, Robert F. The Structure of Autobiographical Memory and the Event History Calendar: Potential Improvements in the Quality of Retrospective Reports in Surveys. *Memory*. 1998; 6(4):383–406. [PubMed: 9829098]
- Belli, Robert F.; Shay, William L.; Stafford, Frank P. Event History Calendars and Question List Surveys: A Direct Comparison of Interviewing Methods. *Public Opinion Quarterly*. 2001; 65:45–74. [PubMed: 11264054]
- Blumstein, Alfred; Cohen, Jacqueline; Roth, Jeffry A.; Visher, Christy, editors. Criminal Careers and “Career Criminals”. Washington D.C.: National Academy Press; 1986.
- Bradburn, Norman M.; Rips, Lance J.; Shevell, Steven K. Answering Autobiographical Questions: The Impact of Memory and Inference on Surveys. *Science*. 1987; 236:157–161. [PubMed: 3563494]
- Cannell, Charles F.; Kahn, Robert L. Interviewing. In: Gardner, Lindzey; Aronson, Elliot, editors. *The Handbook of Social Psychology*. 2nd. Reading, MA: Addison-Wesley Publishing Company; 1968. p. 526-595.
- Carmines, Edward G.; Zeller, Richard A. Sage University Paper Series on Quantitative Applications in the Social Sciences. Newbury Park, CA: Sage; 1979. Reliability and Validity Assessment; p. 07-017.
- Caspi, Avshalom, et al. The Life History Calendar: A Research and Clinical Assessment Method for Collecting Retrospective Event-History Data. *International Journal of Methods in Psychiatric Research*. 1996; 6:101–114.
- Chaiken, Jan M.; Chaiken, Marcia R. Varieties of Criminal Behavior. Santa Monica, CA: Rand Corporation; 1982.
- Clemmer, Donald. The Prison Community. Boston, MA: Christopher Publishing House; 1940.
- Day, Carolyn, et al. Reliability of Heroin Users' Reports of Drug Use Behavior Using a 24 Month Timeline Follow-Back Technique to Assess the Impact of the Australian Heroin Shortage. *Addiction Research and Theory*. 2004; 12(5):433–443.

- Engel, Lawrence S.; Keifer, Matthew C.; Zahm, Shelia H. Comparison of a Traditional Questionnaire with an Icon/Calendar-Based Questionnaire to Assess Occupational History. *American Journal of Industrial Medicine*. 2001; 40:502–511. [PubMed: 11675619]
- Farrington, David P. Self-reports of deviant behavior: Predictive and Stable? *Journal of Criminal Law and Criminology*. 1973; 64:99–110.
- Farrington, David P.; Loeber, Rolf; Stouthamer-Loeber, Magda; Van Kamman, Welmoet B.; Schmidt, Laura. Self-reported delinquency and a combined delinquency seriousness scale based on boys, mothers, and teachers: Concurrent and predictive validity for African-Americans and Caucasians. *Criminology*. 1996; 34:493–517.
- Michael, Fendrich; Vaughn, Connie M. Diminished Lifetime Substance Use Over Time: An Inquiry into Differential Underreporting. *Public Opinion Quarterly*. 1994; 58:96–123.
- Andrea, Fontana; Frey, James H. The Interview: From Structured Questions to Negotiated Text. In: Denzin, Norman K.; Lincoln, Yvonna S., editors. *Collecting and Interpreting Qualitative Materials*. Thousand Oaks, CA: Sage; 2003. p. 645–672.
- Freedman, Deborah, et al. The Life History Calendar: A Technique for Collecting Retrospective Data. *Sociological Methodology*. 1988; 18:37–68. [PubMed: 12282712]
- Andrew, Golub; Johnson, Bruce D.; Taylor, Angela; Liberty, Hillary James. The Validity of Arrestees' Self-Reports: Variations Across Questions and Persons. *Justice Quarterly*. 2002; 19(3):477–502.
- John, Hagan; McCarthy, Bill. *Mean Streets: Youth Crime and Homelessness*. New York, NY: Cambridge University Press; 1998.
- Hindelang, Michael J.; Hirschi, Travis; Weis, Joseph G. *Measuring Delinquency*. Beverly Hills, CA: Sage; 1981.
- Horney, Julie; Osgood, D Wayne; Marshall, Ineke Haen. Criminal Careers in the Short-Term: Intra-Individual Variability in Crime and Its Relation to Local Life Circumstances. *American Sociological Review*. 1995; 60:655–673.
- Hughes, Everett Cherrington. Dilemmas and Contradictions of Status. *American Journal of Sociology*. 1945; 50(5):353–359.
- Huizinga, David; Elliot, Delbert S. Reassessing the Reliability and Validity of Self-Report Delinquency Measures. *Journal of Quantitative Criminology*. 1986; 2(4):293–327.
- Jolliffe, Darrick, et al. Predictive, Concurrent, Prospective and Retrospective Validity of Self-Reported Delinquency. *Criminal Behaviour and Mental Health*. 2003; 13:179–197. [PubMed: 14654870]
- Junger-Tas, Josine; Marshall, Ineke Haen. The Self-Report Methodology in Crime Research. *Crime and Justice*. 1999; 25:291–367.
- Kinnear, Paul R.; Gray, Colin D. *SPSS 14 Made Simple*. New York, NY: Psychology Press; 2006.
- Candace, Kruttschnitt; Carbone-Lopez, Kristin. Moving Beyond the Stereotypes: Women's Subjective Accounts of their Violent Crime. *Criminology*. 2006; 44(2):321–352.
- Laub, John H.; Laub, John H.; Sampson, Robert J. *Shared Beginnings, Divergent Lives: Delinquent Boys to Age 70*. Cambridge, MA: Harvard University Press; 2003.
- Lewis, Darren; Mhlanga, Bonny. A Life of Crime: The Hidden Truth About Criminal Activity. *International Journal of Market Research*. 2001; 43(2):217–240.
- Nan, Lin; Ensel, Walter M.; Lai, Wan-foon Gina. Construction and Use of the Life History Calendar: Reliability and Validity of Recall Data. In: Gotlib, Ian H.; Wheaton, Blair, editors. *Stress and Adversity Over the Life Course*. New York, NY: Cambridge University Press; 1997. p. 249–272.
- Litwin, Mark S. *How to Measure Survey Reliability and Validity*. Thousand Oaks, CA: Sage; 1995.
- Luke, Douglas A. *Multilevel Modeling*. Thousand Oaks, CA: Sage; 2004.
- MacKenzie, Doris Layton; Li, Spencer De. The Impact of Formal and Informal Social Controls on the Criminal Activities of Probationers. *Journal of Research in Crime and Delinquency*. 2002; 39(3): 243–276.
- Mensch, Barbara S.; Kandel, Denise B. Underreporting of Substances Use in a National Longitudinal Youth Cohort: Individual and Interviewer Effects. *Public Opinion Quarterly*. 1988; 52:100–124.
- Morris, Nancy A.; Slocum, Lee Ann. The Validity of Self-reported Prevalence, Frequency, and Timing of Arrest: An Evaluation of Data Collected Using a Life Event Calendar. *Journal of Research in Crime and Delinquency*. 2010; 47(2):210–240.

- Northrup, David A. *The Problem of the Self-Report in Survey Research*. North York, Ontario, Canada: Institute for Social Research; 1997.
- Raudenbush, Stephan W.; Bryk, Anthony S. *Hierarchical Linear Models*. Thousand Oaks, CA: Sage; 2002.
- Roberts, Jennifer; Mulvey, Edward P.; Horney, Julie; Lewis, John; Arteret, Michael L. A Test of Two Methods of Recall for Violent Events. *Journal of Quantitative Criminology*. 2005; 21(2):175–193.
- Singleton, Royce; Straits, Bruce C. *Approaches to Social Research*. Third. New York, NY: Oxford University Press; 1999.
- Sobell, Linda C., et al. Reliability of a Timeline Method: Accessing Normal Drinkers' Reports of Recent Drinking and a Comparative Evaluation Across Several Populations. *British Journal of Addiction*. 1988; 83(4):393–402. [PubMed: 3395719]
- Sudman, Seymour; Bradburn, Norman M. *Response Effects in Surveys*. Chicago, IL: Aldine Publishing Company; 1974.
- Sudman, Seymour; Bradburn, Norman M.; Schwarz, Norbert. *Thinking About Answers: The Application of Cognitive Processes to Survey Methodology*. San Francisco, CA: Jossey-Bass Publishers; 1996.
- Terrance, Thornberry. Panel effects and the use of self-reported measures of delinquency in longitudinal studies. In: Klein, Malcolm, editor. *Cross-national research in self-reported crime and delinquency*. Los Angeles: Kluwer Academic Publishers; 1989.
- Thornberry, Terence P.; Krohn, Marvin. *Measurement and Analysis of Crime and Justice*. Vol. 4. Washington D.C.: National Institute of Justice; 2000. *The Self-Report Method for Measuring Delinquency and Crime*.
- Viera, Anthony J.; Garrett, Joanne M. Understanding Interobserver Agreement: The Kappa Statistic. *Family Medicine*. 2005; 37:360–3. [PubMed: 15883903]
- Weis, Joseph G. Issues in the Measurement of Criminal Careers. In: Blumstein, Alfred, et al., editors. *Criminal Careers and "Career Criminals,"*. Vol. II. Washington D.C.: National Academy Press; 1986. p. 1-51.
- Wheaton, Blair; Gotlib, Ian H. Trajectories and Turning Points Over the Life Course: Concepts and Themes. In: Gotlib, Ian H.; Wheaton, Blair, editors. *Stress and Adversity Over the Life Course: Trajectories and Turning Points*. New York, NY: Cambridge University Press; 1997. p. 1-25.
- Whitbeck, Les B.; Hoyt, Dan R.; Yoder, Kevin A. A Risk Amplification Model of Victimization and Depressive Symptoms Among Runaway and Homeless Adolescents. *American Journal of Community Psychology*. 1999; 27(2):273–296. [PubMed: 10425702]
- Wittebrood, Karin; Nieuwbeerta, Paul. Criminal Victimization During One's Life Course: The Effects of Previous Victimization and Patterns of Routine Activities. *Journal of Research in Crime and Delinquency*. 2000; 37(1):91–122.
- Yacoubian, George S. Assessing the Efficacy of the Calendar Method with Oklahoma City Arrestees. *Journal of Crime & Justice*. 2003; 26(1):117–131.
- Yoshihama, Mieko; Kimberly, Clum; Alexandra, Crampton; Gillespie, Brenda. Measuring the Lifetime Experience of Domestic Violence: Application of the Life History Calendar Method. *Violence and Victims*. 2002; 17(3):297–317. [PubMed: 12102055]
- Yoshihama, Mieko; Gillespie, Brenda; Hammock, Amy C.; Belli, Robert F.; Tolman, Richard M. Does the Life History Calendar Method Facilitate the Recall of Intimate Partner Violence? Comparison of Two Methods of Data Collection. *Social Work Research*. 2005; 29(3):151–163.

Table 1

Comparison of samples to sampling frame and 2005 state prison admissions (standard deviation in parenthesis).

	(1)	(2)	(3)	(4)	(5)
Race					
African American	43.4%	45.2%	51.4%	47.8%	48.0%
White	54.9%	51.6%	46.8%	49.3%	49.4%
Other	1.7%	3.2%	1.8%	2.9%	2.6%
Age at Admission	24.89 (3.95)	24.32 (3.88)	24.33 (3.81)	23.95 [*] (4.09)	32.25 ^{* # \$ &} (10.18)
# times previously incarcerated	.52 (.88)	.61 (.99)	.53 (.91)	.49 (.91)	1.15 ^{* # \$ &} (1.58)
N	110	250	218	1,789	22,646

NOTE: (1) Completed the initial interview and a re-test (2) Entire sample (3) Refusals (4) Sampling frame (5) 2005 State Prison admissions (male)

^{*} statistically different from corresponding entry in column 1, $p < .05$

[#] statistically different from corresponding entry in column 2, $p < .05$

^{\$} statistically different from corresponding entry in column 3, $p < .05$

[&] statistically different from corresponding entry in column 4, $p < .05$

Table 2

Descriptive statistics.

Variables	MEAN	SD
Level-1 (within-individual)		
Criminal behavior (re-test)	.59	.59
Criminal behavior (test)	.63	.60
N street-months	1,700	
Level-2 (between person)		
Black	.43	.49
Other	.02	.23
Age	24.57	3.92
Education	11.25	1.69
Legal monthly income	1201.28	1703.78
Illegal monthly income	553.92	1087.16
Test/re-test lag	22.91	11.10
ODRC prison terms ^a	.56	.85
Self-reported prison terms ^a	.68	1.04
N persons	110	

NOTE:

^aN persons = 250.

Table 3

Cross-tabulation of criminal behavior (test) by criminal behavior (retest).

Criminal behavior (test)	Criminal behavior (re-test)			
	0	1	2	3
0	651	67	9	0
1	119	731	21	2
2	12	39	47	0
3	0	0	2	0
TOTAL	782	837	79	2
				1700

NOTE: Gamma = .914*; Kappa = .709*; * p<.001.

Table 4

Coefficient as outcome hierarchical models of self-reported crime from re-test on self-reported crime from test and socio-demographic characteristics. ^a

	Criminal behavior (re-test)			
	(1)	(2)	(3)	(4)
Intercept	-1.94 *** (.25)	-2.15 (.27)	-2.15 (.27)	-2.16 (.27)
Level-1 Variables				
Criminal behavior (test)		.98 *** (.23)	.89 *** (.27)	.98 *** (.31)
Level-2 Variables				
Black ^b			.44 (.49)	.45 (.50)
Other			-1.49 (1.34)	-1.39 (1.35)
Age			-.09 (.07)	-.09 (.08)
Education			.15 (.09)	.16 (.10)
Legal income ^c			.02 (.01)	.02 (.01)
Illegal income ^c			-.03 (.03)	-.04 (.03)
Test/re-test lag			-.04 ** (.02)	-.04 * (.02)
Month				.02 (.02)
Criminal behavior (test) * Month				-.01 (.01)
Random Effect	Variance Component			
Level-1, σ_j	.16	.12	.12	.12
Level-2, μ_{0j}	6.38 ***	7.34 ***	7.34 ***	7.42 ***
Level 2, μ_{1j}		1.99 ***	2.03 ***	2.09 ***
N		110		

NOTE:

*
($p < .10$);

**
($p < .05$);

($p < .01$).

^aUnit specific estimates, robust standard errors.

^bWhite is the referent.

^cCoefficients multiplied by 1,000 to reduce places to the right of the decimal.

Table 5

Cross-tabulation of the frequency of official ODRC state prison terms by the frequency of self-reported state prison terms.

Official ODRC prison terms	Self-reported state prison terms					
	0	1	2	3	4	5
0	143	13	1	0	0	0
1	3	47	5	2	0	0
2	0	3	15	7	0	4
3	1	0	0	4	0	0
4	0	0	0	0	1	0
5	0	0	1	0	0	0
TOTAL	147	63	22	13	1	4
						250

NOTE: Gamma = .964*; Kappa = .715*; * p<.001.

Table 6Poisson regression of official ODRC state prison terms on self-reported state prison terms and interactions. ^a

	# of ODRC state prison terms	
	(1)	(2)
Self-reported prison terms	.60 *** (.07)	.78 *** (.11)
Black ^b	.23 (.19)	-.17 (.24)
Other ^b	.05 (.41)	-.49 (.52)
Age	.03 (.03)	.09 *** (.03)
Education	.01 (.05)	-.04 (.06)
Legal income ^c	-.01 (.01)	-.01 (.01)
Illegal income ^c	-.01 (.01)	-.01 (.01)
Self-reported prison × black		.28 ** (.13)
Self-reported prison × other		.77 (.57)
Self-reported prison × age		-.06 *** (.02)
Self-reported prison × education		.04 (.03)
Self-reported prison × legal income ^c		.01 (.01)
Self-reported prison × illegal income ^c		-.001 (.01)
Constant	-1.05 *** (.15)	-.93 *** (.16)
Pseudo R-square	.27	.33
N	250	

NOTE

*
(p < .10);**
(p < .05);***
(p < .01), two-tailed tests.^a All predictors are mean centered, excluding race.^b White is the referent.^c Coefficient multiplied by 100 to reduce places to the right of the decimal.