Explaining Systematic Bias in Self-Reported Measures: Factors that Affect the Under- and Over-Reporting of Self-Reported Arrests

Marvin D. Krohn, Alan J. Lizotte, Matthew D. Phillips, Terence P. Thornberry & Kristin A. Bell

Published online: 11 Aug 2011.


To link to this article: http://dx.doi.org/10.1080/07418825.2011.606226

PLEASE SCROLL DOWN FOR ARTICLE
Explaining Systematic Bias in Self-Reported Measures: Factors that Affect the Under- and Over-Reporting of Self-Reported Arrests

Marvin D. Krohn, Alan J. Lizotte, Matthew D. Phillips, Terence P. Thornberry and Kristin A. Bell

The self-report method of collecting data on delinquency, crime, and arrests continues to be one of the most popular techniques of examining the causes of such behavior and assessing the bias in the responses to it. However, the

Marv Krohn is a professor in the Department of Sociology and Criminology & Law at the University of Florida. He is primarily interested in developmental and life course approaches to the explanation of delinquency, drug use, and crime. He has pursued these interests as a co-principal investigator of the Rochester Youth Development Study. His co-authored book (with Thornberry, Lizotte, Smith, and Tobin) Gangs and Delinquency in Developmental Perspective, won the 2003 American Society of Criminology’s Michael J. Hindelang award for Outstanding Scholarship. Alan Lizotte is dean and professor in the School of Criminal Justice, The University at Albany. He is co-principal investigator on the Rochester Youth Development Study, a 20 year ongoing longitudinal study of juvenile delinquency and drug use covering three generations of subjects. His substantive interests include illegal firearms ownership and use and developmental criminology. In 2003, together with his RYDS coauthors, he was awarded the American Society of Criminology’s Hindelang Award for the book Gangs and Delinquency in Developmental Perspective. Matthew Phillips is a doctoral candidate at the University at Albany and a research assistant for the Rochester Youth Development Study. His primary research interests include drug offending, violence, drug trafficking, and quantitative methods. Terence P. Thornberry, PhD, is professor in the Department of Criminology and Criminal Justice at the University of Maryland and the principal investigator of the Rochester Youth Development Study, a three-generation panel study begun in 1986 to examine the causes and consequences of delinquency and other forms of antisocial behaviors. In 1995, he was elected a fellow of the American Society of Criminology, and in 2008, he was the recipient of that society’s Edwin H. Sutherland Award. He is the author of a number of books, including Gangs and Delinquency in Developmental Perspective which received the American Society of Criminology’s Michael J. Hindelang Award in 2004. Kristin A. Bell is a doctoral student in the School of Criminology and Criminal Justice at Northeastern University. After receiving her BA in criminology and BS in psychology, she completed her Master’s degree in criminology, law, & society at the University of Florida. She is currently a project manager at the Center for Criminal Justice Policy Research and the Institute on Race and Justice at Northeastern University. Her interests include victim behavior and help-seeking, human rights violations, and the use of mixed research methods. Correspondence to: M.D. Krohn, Department of Sociology and Criminology & Law, University of Florida, 3219 Turlington Hall, P.O. Box 117330, Gainesville, FL 32611-7330, USA. E-mail: mkrohn@crim.ufl.edu
problem of systematic bias in reporting such data continues to cloud its use. Self-report data from a longitudinal study of youth at high risk for serious delinquent behavior are compared with data from official police records to examine systematic bias in both the under-reporting and over-reporting of self-reported arrests. Although under-reporting and over-reporting occur in nearly equal proportions in our data, we find that these two phenomena operate quite differently. Further, we show that systematic bias of self-reported arrests is largely a function of the number of official arrests, and that the effect is non-linear. We offer explanations for these findings, and discuss their implications for the future use of self-report methods.

**Keywords**  
self-report arrests; systematic bias; validity; official arrests; hidden delinquency

### Introduction

The development and expanded use of the self-report survey as a device to collect information on involvement in delinquency and crime has had a significant impact on how we study crime, on how we think about the causes of crime, and how we assess the operation of the criminal and juvenile justice systems. Early studies demonstrated that delinquent behavior was more equally distributed across racial, ethnic, social class, and gender lines than official arrest and court data (Akers, 1964; Clark & Wenninger, 1962; Dentler & Monroe, 1961; Empey & Erickson, 1966; Erickson & Empey, 1963; Gold, 1966; Murphy, Shirley, & Wimer, 1946; Nye, Short, & Olson, 1958; Porterfield, 1943; Slocum & Stone, 1963; Vaz, 1966; Voss, 1966; Wallerstein & Wylie, 1947). This led to the questioning of the operation of the juvenile justice system and the influence of extra-legal factors in enforcement decisions. Later studies that incorporated questions on theoretically derived factors such as family relationships and functioning, school commitment and success, and peer associations (Dentler & Monroe, 1961; Elliott, 1966; Erickson & Empey, 1963; Gold, 1970; Hirschi, 1969; Kelly, 1974; Mathews, 1968; Polk, 1969; Reiss & Rhodes, 1964; Stanfield, 1966; Voss, 1964) led to a shift of emphasis away from macro-level theories such as social disorganization and strain theories to more social process theories like social control and differential association.

Since those early studies, self-report techniques have become more methodologically sound in terms of sample selection, inclusion of a wider range of behaviors, the expansion of response categories to account for the upper limit on the frequency of participation, the mode of administration of the survey, and the application of the self-report methodology in longitudinal studies (see Krohn, Thornberry, Bell, Lizotte, & Phillips, 2011; Thornberry & Krohn, 2000). These advancements have improved the quality of the data collected and its usefulness for examining the myriad of issues that have been considered with self-report data on delinquency and crime.
In spite of these improvements, one nagging question lingers—do respondents tell the truth when they are asked about their participation in delinquency and crime and their involvement with the juvenile and criminal justice systems?

Prior research has generated mixed results regarding the answer to these questions. The purpose of the current study is to provide a more nuanced examination of the issue. Specifically, we examine both over- and under-reporting of self-reported arrests as compared with official arrests, examine factors in addition to race and gender that might account for discrepancies in reporting, use sophisticated statistical techniques to explore the possibility that relationships may not be linear, and determine if the factors that explain the discrepancies are different for adolescents and for adults.

This study examines the discrepancy between data from a well-known longitudinal self-report study and official police records to estimate the validity of the former. We begin with simple, naive models of the relationship between race and gender predicting the discrepancy between self-report and official reports of arrest. We determine the predictors of whether subjects under-report, over-report, or are accurate. We predict the magnitude of under- and over-reporting separately in linear models. Each successive step builds in information on how under-reporting and over-reporting behaves for adolescents and we conclude with a model that integrates all the features of the simpler models. Our final model predicts differences in under- and over-reporting in a non-linear functional form accounting for both across-category and between-category variation that leads to somewhat different conclusions than the earlier naive models. We do this for both an adolescent and adult sample.

Prior Research

The primary way to answer this question is to compare information from self-report questions about behavior and arrests with some alternative measure (criterion validity) (Thornberry & Krohn, 2000). Most frequently, the alternative measure against which to gage the truthfulness of self-reported delinquent behavior and arrests is official arrests (for an excellent discussion, see Brame, Fagan, Piquero, Schubert, & Steinberg, 2004). Those who have been officially arrested should report engaging in some delinquent behavior, although there could be some slippage due to false perceptions of arrests or from missing arrests from other jurisdictions. If there is a disparity between these measures, then the self-reported behavior is called into question.

Typically, the focus is on under-reporting behavior or arrests because we are concerned that respondents will be reluctant to indicate that they have done something wrong. However, prior research comparing the two measures of arrest has noted that some respondents over-report arrests. The most probable reason for this is that some respondents may mistakenly view any contact
with law enforcement agents as an arrest even if they were not actually officially arrested.¹

To examine both under-reporting and over-reporting in self-reports, self-reported arrests rather than self-reported behavior must be used. If respondents have been arrested, and do not indicate that they have committed a delinquent behavior, we can assume that they are under-reporting delinquency. However, if they report committing a delinquent behavior, but have not been arrested, we cannot assume that they are over-reporting delinquency.

Discrepancies between self-reported behavior or arrests and official arrests can be randomly distributed among a sample population or they can be disproportionately evident among certain subgroups of that sample (systematic bias). While any discrepancy is a problem in terms of accurately representing the prevalence of a behavior or arrest, systematic bias is of particular concern to researchers. Systematic bias may impact the relationships found among self-reports and critical (theoretical) variables. Therefore, most studies examining the validity of self-report data try to determine if discrepancies with an alternative measure of the behavior (official arrests) are more evident among certain sub-groups (i.e. are African-Americans or males more likely to under-report crime?).

Huizinga (1991) stated that there had been few studies that compared official data with self-report data. While there have been a number of studies since Huizinga made that statement, (e.g. Farrington, Loeber, Stouthamer-Loeber, Van Kammen, & Schmidt, 1996; Kirk, 2006; Maxfield, Weller, & Widom, 2000; Thornberry & Krohn, 2000), given the importance and growing use of self-report studies, there is still a need to explore this relationship. Indeed, Maxfield et al. (2000, p. 108) conclude their study by calling for "researchers to resurrect the analysis of official records and self-reports of arrests..." and urging others to explore the validity of self-reports among different subpopulations.

Validity of Self-Reported Arrests

In an extensive review of the use of self-reports in the study of crime and delinquency, Thornberry and Krohn (2003, p. 57) state that the problem with assessing the criterion validity of self-reported measures is "...there is no gold standard by which to judge the self-report measure." By this, they are refer-

¹. It is also possible that some over-reporting is due to respondents being arrested outside the jurisdictions in which data are collected. We collect arrest data from the city of Rochester and from the NY state Department of Criminal Justice Services (DCJS) which obtains data from all jurisdictions in the state of New York. However, we do not have data for other states. So if respondents were arrested in other states, they would not appear in the official arrest data used in this study. However, the vast majority of RYDS subjects remained in New York state. Specifically, at Wave 9, 98% of subjects lived in New York. At Wave 12, 87% of subjects lived in New York state.
ring to the lack of an objective criterion that does not have its own set of flaws. For example, self-report measures are most often compared with official measures of arrests. Ironically, the imperfections of official measures were largely responsible for the development of self-report measures. However, the logic of using official measures of arrest as a benchmark for self-reported arrests or delinquent and criminal behavior is straightforward and persuasive. If someone has been arrested, in most cases they will have engaged in some form of delinquent or criminal behavior and should respond affirmatively to at least one of the self-reported items. And, if the official data indicate that they have been arrested, if asked on a self-report instrument, they should so indicate.

Most of the research comparing an official measure of arrests and either self-reported arrests or delinquent behavior generates consistent results concerning the overall relationships among the measures. Hindelang, Hirschi, and Weis (1981), in their classic study on measuring delinquency, reported that the correlation between a measure of official arrests and self-reported delinquency was around 0.60 for all subjects. As expected, the correlations between official arrests and self-reported arrests were substantially higher, ranging from 0.70 to 0.83. This finding is consistent with a number of other studies (Brame et al., 2004; Farrington et al., 1996; Huizinga & Elliott, 1986; Kirk, 2006; Maxfield et al., 2000; Thornberry & Krohn, 2000).

While researchers using a self-report methodology may be encouraged by the strength and consistency of the observed relationships, the key issue in determining the usefulness of self-report measures is whether there is systematic bias in the reporting of either arrests or delinquent behavior. The findings in this regard are not as consistent and, as such, raise concerns.

Concerning the race by gender analysis, Hindelang et al. (1981) indicate that there may be systematic bias in either official or self-report data. For all females, and for white males, the correlations between official records and self-reported behavior ranged from 0.58 to 0.65. However, for African-American males, the correlations average only 0.30. Huizinga and Elliott (1986), using a different measurement strategy in which they differentiate between a tight match versus a broad match of the type of offense in police records and on the self-report instrument, come to a similar conclusion regarding the discrepancy between official and self-report data for African-American males.

Farrington et al. (1996) examined more recent data from the Pittsburgh Youth Study. They found that African-American males were no less likely to self-report delinquent behavior than were white males. Essentially, they found that some odds-ratios were higher for whites and some were higher for African-Americans. They conclude that there is no evidence in this data-set for differential validity (i.e. systematic bias).

Maxfield et al. (2000) provide an in-depth examination of reporting patterns among young adults with a history of child abuse or neglect. Focusing on the comparison between self-reported arrests and official arrests, they found that abused and neglected subjects were no more or less likely to under-report
delinquent behavior when compared to those who were not abused or neglected. Interestingly, abused subjects did over-report arrests. That is, abused subjects were more likely to report an arrest when there was no official record of an arrest.

Maxfield et al. (2000) also investigated differential validity by both race and gender. The authors investigated whether race or gender affects the probability of ever reporting an arrest to an interviewer. In this study, the indicators of self-reporting an arrest, of under-reporting, and over-reporting were all dichotomous measures. In other words, they were measures of prevalence, rather than incidence. Similar to the earlier studies, African-American subjects were more likely to conceal an arrest (i.e. to under-report arrests). Additionally, females were more likely to conceal an arrest than were males. In terms of over-reporting arrests, males and white subjects were more likely to report an arrest when official records did not have one recorded.

Although Maxfield et al. (2000) explored other potential correlates of under-reporting and over-reporting, the most important finding for the current focus concerned differential reporting by the number of arrests. The discrepancy between self-reported arrests and official arrests was much lower for those who had multiple arrests compared to those with fewer arrests. Among those with five or more arrests, race was not related to self-reporting. Maxfield et al. (2000, p. 107) conclude by stating that “...it appears that subjects with more recorded official contacts (numbers of arrests, convictions, and arrests as juveniles and adults) more often self-report arrests, regardless of ethnicity.” In other words, subjects with more official arrests are more likely to self-report ever being arrested. However, not having the data to count the number of self-reported arrests, the authors were unable to compare the counts of official and self-reported arrests.

Kirk (2006) used data from the Project on Human Development in Chicago Neighborhoods to compare self-report and official data. While the main thrust of the article was to examine these data across the life course, he provides analyses that are directly relevant to the current analysis. Similar to Maxfield et al., Kirk finds that there is about an equal amount of under-reporting and over-reporting when self-reported data are compared with official arrests. More importantly, he concludes that African-Americans are more likely to under-report arrests and are most likely to do so at the peak arrest level.

Current Research

Although the evidence is mixed, the majority of studies do find differential validity when comparing self-report data on participation in delinquent behavior

2. To our knowledge, these are the only two studies that consider whether or not respondents over- or under-reported arrests. Unlike the analysis to follow, neither considered the amount of over- and under-reporting.
and self-reported arrests with official arrests. The focus has been on under-reporting among African-Americans males. Few studies have examined the phenomenon of over-reporting. Even fewer studies have attempted to explore the reasons for either under-reporting or over-reporting of arrests and participation. Maxfield et al. (2000) began this process by examining a number of covariates of both under-reporting and over-reporting. They find race/ethnicity, sex, and the number of arrests matter in the probability of a discrepancy existing in the two data sources. Recognizing that their focus was on those who have a history of child abuse, they encourage researchers to explore the relationship between self-report and official data with other samples. Kirk (2006) also explores the prevalence of both under-reporting and over-reporting, but indicates that African-Americans with higher arrest rates are more likely to under-report being arrested.

We take our lead from these previous studies in comparing self-reported arrests and participation in delinquent behavior with official arrests for a general youth population. Specifically, we are interested in the degree of discrepancy and in what variables might account for this discrepancy if it exists. The typical exploration of this relationship only focuses on the under-reporting of delinquency. In this study, we focus on both under-reporting and over-reporting of delinquent behavior. However, we go considerably beyond prior research that typically uses prevalence measures and linear estimation. Our strategy is to move from simple and less sophisticated comparisons to more complex and meaningful ones. We begin by describing cases with an official arrest that have no delinquent acts reported to show the strengths and limitations of this comparison.

Then, we estimate a multinomial logistic model to predict dichotomous categories of over-reports, under-reports, correct reports, and no reports to determine if there are differences in types of reporting. This analysis provides baseline information. However, it is naive, as it does not consider the amount of under- and over-reports, does not control for other potential covariates, and is linear in functional form. We go on to examine how well a limited number of variables can explain the discrepancy in the number of arrests between official and self-reported data. We do this in a multivariate context holding constant race/ethnicity and gender. In order to compare under-reporters, over-reporters, and those accurately reporting arrests, we also include in the model the reference group of those accurately reporting no arrests. So, we model under-reporting, over-reporting, and both types of accurate reporting in the same equation to avoid censoring.\footnote{By censoring, we mean we are using the entire data-set to estimate a single equation, instead of partitioning the data.} To accurately reflect the behavior of the data, we model discrepancies in under-reporting and over-reporting each nonlinearly and separately in that equation. In other words, we do not assume a linear relationship between the number of official arrests and the discrepancy in self-reports as has been done in previous research. Instead we allow
for different nonlinear effects for each because there could be limits to the amount of under- and over-reporting. In other words, one would not expect the discrepancies between official and self-reports to increase infinitely. Further, we show that nonlinear discrepancies in under-reporting and over-reporting in fact do reach practical limits as a function of the number of official arrests. Finally, we estimate equations predicting these differences for adolescents and young adults separately.

Methods

Data

To examine the concordance between self-reported arrests and official arrests, we use data from the Rochester Youth Development Study (RYDS), an ongoing, longitudinal investigation of antisocial behavior. The RYDS project has followed a panel of juveniles from their early teenage years through their adult years. The study began in 1988, and selected 1000 seventh and eighth graders in the Rochester (New York) Public School System to be interviewed, along with a parent or guardian. To date, RYDS has completed 14 interviews for the panel, spanning ages 14 through 31. For our adolescent analysis, we use data from Waves 4 through 9, when respondents were between the ages of 14 and 18. During this phase, we interviewed each subject nine times (or waves) and their parent or guardian eight times at six-month intervals. Prior to Wave 4, the format for the follow-up questions concerning arrest varied slightly. Additionally, there were very few official arrests recorded in the early waves. We considered it prudent, therefore, to begin with Wave 4. For the young adult analysis, we use data from Waves 10 to 12, when subjects were between ages 21 and 23. During this phase, subjects were interviewed annually, rather than biannually.

The original RYDS sample was stratified on two dimensions in order to select subjects who were at high risk for violence and serious delinquency. First, males were oversampled (75% versus 25%), as they are more likely than females to engage in serious and violent offenses (Blumstein, Cohen, Roth, & Visher, 1986; Huizinga, Morse, & Elliott, 1992). Second, students living in areas of the city with high residential arrest rates were also oversampled. This was based on the assumption that adolescents who live in such areas are at greater risk for offending than students living in areas with lower residential arrest rates. High-crime areas were identified by assigning each census tract in Rochester a resident arrest rate which reflected the proportion of the tract’s total adult population arrested by the Rochester Police Department in 1986.

The subject panel is 68% African-American, 17% Hispanic, and 15% white. These proportions are quite close to what was expected given the population

4. Sample selection criteria have no impact on the results, and weighting the data is not necessary.
characteristics of Rochester schools and the decision to oversample high-risk youth. Subject attrition in the RYDS is quite low when compared to other longitudinal studies. From Waves 2 to 12, we experienced only about 1% attrition per year. At Wave 12, 85% (846) of the original 1000 subjects were re-interviewed. A formal test of subject attrition within RYDS revealed that the subjects retained did not significantly differ from those not retained on multiple dimensions, including gender, social class, family structure, drug use, delinquency, property crime, and violent crime (see Krohn & Thornberry, 1999).

Table 1 presents basic characteristics of the total panel, the Wave 9 sample, and the Wave 12 sample.

### Table 1  Demographic characteristics of RYDS sample

<table>
<thead>
<tr>
<th></th>
<th>Total panel (%)</th>
<th>Wave 9 (%)</th>
<th>Wave 12 (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Gender</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>72.9</td>
<td>72.4</td>
<td>71.5</td>
</tr>
<tr>
<td>Female</td>
<td>27.1</td>
<td>27.6</td>
<td>28.5</td>
</tr>
<tr>
<td><strong>Race/ethnicity</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>African-American</td>
<td>68.0</td>
<td>68.2</td>
<td>68.1</td>
</tr>
<tr>
<td>Hispanic</td>
<td>17.0</td>
<td>16.7</td>
<td>16.5</td>
</tr>
<tr>
<td>White</td>
<td>15.0</td>
<td>15.1</td>
<td>15.4</td>
</tr>
<tr>
<td><strong>N</strong></td>
<td>1000</td>
<td>881</td>
<td>846</td>
</tr>
</tbody>
</table>

Measurement

Subjects were asked to indicate if they had engaged in a wide variety of violent, property, and drug offenses since the last interview (i.e. in the last six months). If the subject indicated they had committed the offense since the date of their last interview, they were then asked if they were arrested for the offense. If the subject reported they were arrested, they were asked how many times they were arrested for such an offense. Immediately following this portion of the interview, subjects were asked if they were arrested in the last six months for any other reason, and how many times. This pattern of questions continued through Wave 9. From these responses, we created several measures of self-reported arrests. The same process is used for adults in Waves 10-12, except that status offenses for juveniles are removed.

First, we created a dichotomous measure indicating whether or not a subject reported he or she had been arrested at a particular wave, from Wave 4

---

5. Less than one quarter of 1% of the total sample (23) were missing on the arrest measures and were excluded from the valid sample. The valid sample size for adolescents in the current study is 858 and 790 for young adults.

6. Phrasing the question this way allows us to accurately match the dates of official and self-reported arrests.
These measures were coded "1" if the subject reported at least one arrest and "0" if the subject reported no arrests. We also created a dichotomous measure indicating whether or not a subject reported ever being arrested during the period lasting from Wave 4 to Wave 9. This measure was coded "1" if the subject reported at least one arrest at any wave (Waves 4-9) and "0" if the subject reported no arrests between Waves 4 and 9. A continuous variable was created which counted the number of arrests a subject reported during each particular wave, from Waves 4 to 9. Finally, we created a continuous variable, which counted the total number of arrests a subject reported from Waves 4 to 9. This measure was simply the sum of the wave-specific arrest counts above. Recall that most of the items that asked subjects to report arrests were follow-up questions to offense items. If a subject indicated he or she had not engaged in a given offense, the subject was scored as not being arrested for the same offense, and thus having no arrests for the same offense (i.e. subject received a "0" score for both the dichotomous and continuous measure). Although this leaves the possibility of a subject being arrested for an offense which he or she did not commit, our subsequent item regarding being arrested for any other reason would capture the event.

In addition to the self-reported measures of arrests, the RYDS contains information reflecting official arrest data. These data were provided by the Rochester Police Department (RPD) and the New York DCJS. These agencies reported the number of times each subject was arrested during a given wave of the study. Using these counts, we were able to obtain parallel official measures to our self-reported measures described above. Specifically, we created a dichotomous variable indicating whether or not the subject was officially arrested at each wave, from Waves 4 to 9. These measures were coded "1" if the police agencies reported the subject had one or more official arrests during the wave, and "0" otherwise. Another dichotomous measure was created to indicate whether the subject was ever officially arrested between Waves 4 and 9. This measure was coded "1" if the police reported the subject experienced one or more official arrests during the period between Waves 4 and 9, and "0" otherwise. Finally, we summed the wave-specific official arrest counts to create a measure of the total number of official arrests reported for a subject during the period lasting from Waves 4 to 9.

Results

We report the results for adolescents and adults in two stages. For adolescents, we examine the bivariate results first, then report the results of the multivariate analysis. For adults, we report only the results of the multivariate analysis.

7. The same process was also used throughout for data drawn from Waves 10 to 12 in the adult analyses.
Bivariate Analyses of Adolescents

To analyze the agreement between the number of arrests that are self-reported and officially reported in the RYDS project, we employ a series of cross-tabulation tables. The tables that follow all use measures spanning the entire period from Waves 4 to 9. While it is possible to create similar tables for each wave, the cell counts diminish rapidly. Although not shown, the results shown in wave-specific tables do not differ from those in the overall (Waves 4-9) tables. Because interpretation of the overall tables is more straightforward, we devote our attention to these tables in the present analysis.

Table 2 depicts three cross-tabulations between our dichotomous measures of self-reporting an arrest and being officially arrested between Waves 4 and 9. Perhaps the first finding to mention is that 688 subjects (80.2% of the valid sample) are in agreement on these two measures. A $\chi^2$-test estimated on the first section of Table 2 yields a value of $\chi^2 = 208.4$ and a $p$-value of $p < 0.001$. This confirms what the table appears to show at face value, that subjects with zero official arrests are most likely to self-report zero arrests. In other words, if subjects were asked if they were ever arrested during the period, their answers would agree with the answer given by police. Of the 20% of the sample not in agreement, 78 subjects (9.1%) self-reported at least one arrest when the official police data reflected the subject as having zero arrests. These subjects are our over-reporters. Conversely, 92 subjects (10.7%) self-reported zero arrests, despite official police data reporting at least one. These subjects are our under-reporters. The difference in these two proportions is not substantial.

McNemar’s test provides a formal test of whether the proportion of over-reporters significantly differs from the proportion of under-reporters. Imagine that in Table 3, $a$ represents the cell count for subjects with a “0” value on both arrest measures. Further, $d$ represents the cell count for subjects with a value of “1” on both arrest measures. The cell count for subjects with a value of “1” on the official arrest measure and “0” on the self-reported arrest measure (the under-reporters) is represented by $b$. Finally, the cell count for sub-

8. For the sake of completeness, we began our preliminary analyses with the traditional method of comparing self-reported offenses to official arrests. We began by searching for the maximum logical discrepancy between self-reported offending and official arrests; subjects who reported no offenses, but were arrested at least once. There were only 13 such cases. With such a small pool, analyzing systematic bias in the difference between offenses and arrests is uninformative. In the interest of space, we do not report these results, but they are available upon request. The current study begins by reporting the results from the more fruitful comparison of self-reported arrests and official arrests. Furthermore, we do not mean to imply that the police always justifiably arrest individuals or do so systematically across races and genders. Rather, our goal is to determine how systematically official arrests are related to self-reported arrest. We are interested in the truthfulness of the respondents and not the motivations or appropriateness of police actions.

9. This test is fundamentally different from the chi-square test above. The chi-square tests if respondents in one category of the self-reported measure are likely to be in the same category of the official measure.
projects with a value of ”0” on the official arrest measure and ”1” on the self-reported arrest measure (the over-reporters) is represented by $c$. This new table is shown as Table 3 for illustration. In McNemar’s test, one tests the null hypothesis that $b=c$ against the alternative hypothesis that the two are significantly different. Applying McNemar’s test to Table 2, the test statistic of 1.153 has an associated $p$-value of $p=0.319$. Here, we fail to reject the null hypothesis and conclude that the proportion of over-reporters does not significantly differ from the proportion of under-reporters. In other words, when the subject self-report measures disagree with official police measures, the disagreement is as likely to occur in either direction. Recall that Maxfield et al. (2000) had a similar finding.

Table 4 depicts cross-tabulations of the same arrest measures, but by gender. The top panel shows the values for male subjects, and the bottom panel shows the accompanying values for female subjects. Chi-square tests for the two tables provide statistics of $\chi^2=142.924 \ (p<0.001)$ for males and $\chi^2=49.594 \ (p<0.001)$ for females. These tests suggest the most likely outcome for both males and females is agreement between the official arrest and self-reported arrest measures. McNemar’s test is again used to determine whether males are more likely to over-report or to under-report being arrested. The test statistic of 0.180 has an associated $p$-value of $p=0.73$. Therefore, we fail to reject the null hypothesis. Males are no more likely to over-report than to under-report. Conducting the test on female subjects yields a McNemar’s test statistic of 2.613 and a $p$-value of $p=0.15$. Like their male counterparts, females are no more likely to over-report than to under-report arrests. Gender appears to have no effect on whether a subject over- or under-reports being

---

### Table 2  Official arrest and self-reported arrest measures, Waves 4-9

<table>
<thead>
<tr>
<th></th>
<th>Officially arrested</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No</td>
</tr>
<tr>
<td>Self-reported</td>
<td>No</td>
</tr>
<tr>
<td>Ar rested</td>
<td>Yes</td>
</tr>
</tbody>
</table>

Notes. $\chi^2=208.431$, $df=1$, and $N=858$.

### Table 3  Hypothetical $2 \times 2$ table to illustrate McNemar’s test

<table>
<thead>
<tr>
<th></th>
<th>Officially arrested</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No</td>
</tr>
<tr>
<td>Self-reported</td>
<td>No</td>
</tr>
<tr>
<td>Ar rested</td>
<td>Yes</td>
</tr>
</tbody>
</table>
arrested. Maxfield et al. (2000) reported that males were more likely to over-report an arrest, while females were more likely to under-report an arrest. It should be noted again, however, that Maxfield et al. used a limited sample, whereas we use a representative sample. This may be responsible for differences in results.

Finally, we conduct similar tests for each racial group in our sample. The top panel of Table 5 shows the official arrest and self-reported arrest measures for white subjects only. The chi-square value of $\chi^2 = 30.180$ ($p < 0.001$) again shows that agreement between official and self-reported arrest measures is the most likely result. However, this section of Table 5 shows marked differences in the number of over-reporters and under-reporters. While only two white subjects (1.6%) under-reported being arrested, 19 subjects (14.8%) over-reported an arrest. McNemar’s test statistic of 13.762 ($p < 0.001$) confirms this difference to be statistically significant. Our data suggest whites are more likely to self-report at least one arrest when the official arrest data show no arrests than they are to self-report zero arrests when official data show at least one official arrest.

The middle panel of Table 5 shows the official arrest and self-reported arrest measures for African-American subjects only. As with the white subjects, agreement between official and self-reported arrest measures for African-American subjects is expected ($\chi^2 = 158.284$, $p < 0.001$). Cell counts show an apparently large disparity between over-reporting and under-reporting in this section of Table 5. Although 75 African-American subjects (12.8%) under-report an arrest, only 38 (6.5%) over-report an arrest. McNemar’s test statistic of 12.115 ($p < 0.001$) confirms this difference to be statistically significant as well. African-American subjects in our data are more likely to under-report than they are to over-report arrests. This is consistent with Maxfield et al. (2000) and Kirk (2006).

Lastly, the bottom panel of Table 5 shows the official arrest and self-reported arrest measures for Hispanic subjects only. Yet again the chi-square
test ($\chi^2 = 28.385, p < 0.001$) leads us to conclude that with Hispanic subjects, agreement between the self-reported and official arrest measures is the most likely outcome. Unlike white and African-American subjects, however, Hispanic subjects over-reported and under-reported arrests in nearly equal numbers (21 subjects, or 14.5%, over-report, and 15 subjects, or 10.3%, under-report). This small difference generates a McNemar test statistic equal to 1.000, with a $p$-value of $p = 0.405$. Hispanic subjects in our data are equally likely to over-report being arrested as they are to under-report being arrested.

Although not shown here, we also created cross-tabulation tables for each race-gender combination. Unfortunately, expected cell counts became too small (less than five) to allow estimation. Overall patterns of results did not differ from those described above.

### Multivariate Analyses for Adolescents

The bivariate analyses above are useful for determining the level of agreement between official and self-reported arrest measures in the RYDS sample. But as noted above, the bivariate analyses are naive, as they do not account for other covariates, and they do not allow under- and over-reporting to operate differently. To understand the disagreement between official and self-reported measures, therefore, more involved analyses are necessary. To begin, we used the official and self-reported arrest measures described previously to create a difference score (DS) for each subject. Specifically, this DS expresses the difference between the subject’s total number of official arrests and total number of self-reported arrests.
With the DS, we separated subjects into four categories. Subjects who had zero self-reported arrests and zero official arrests were placed into the “Never Arrested” category. Subjects whose self-reported arrests exceeded their official arrests were placed into the “Over-reporting” category. Subjects whose official arrests exceeded their self-reported arrests were placed into the “Under-reporting” category. Finally, subjects whose self-reported arrests matched their official arrests went into the “Arrested” category. As mentioned above, most of our subjects belong to the Never Arrested category, and McNe- mar’s test confirmed that the proportion of under-reporters did not significantly differ from the proportion of over-reporters.

A multinomial logistic regression equation was estimated to predict membership in each of the four categories described above that a subject was most likely to belong to, based on gender and race. By estimating a multinomial logistic regression, we are able to obtain all the relevant coefficients without excluding any of the different types of reporting. So, we can determine the predictors of membership in each category without censoring. This model provided a test of across-category differences by gender and race. Later, we consider within-category differences by gender and race as well. Results are shown in Table 6. Being a male increases a subject’s likelihood of being an over-reporter or an under-reporter. That is, males are more likely to be incorrect. This is not surprising given that males have higher rates of arrest compared to females, a common feature of delinquency studies (see Brame et al., 2004; Krohn et al., 2011). Being black or Hispanic increases the likelihood of under-reporting. Hispanics are less likely to over-report arrests. Finally, Hispanics are more likely to belong to the Arrested category. These results generally confirm the results of the cross-tabulations above. This model is useful for

### Table 6  Multinomial logistic regression, G2 arrest status, Waves 4-9

<table>
<thead>
<tr>
<th></th>
<th>Over-reported</th>
<th>Under-reported</th>
<th>Arrested</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>1.3104***</td>
<td>1.2606***</td>
<td>0.6460</td>
</tr>
<tr>
<td></td>
<td>(0.28)</td>
<td>(0.25)</td>
<td>(0.40)</td>
</tr>
<tr>
<td>Black</td>
<td>0.0852</td>
<td>1.8615***</td>
<td>1.3843†</td>
</tr>
<tr>
<td></td>
<td>(0.27)</td>
<td>(0.44)</td>
<td>(0.75)</td>
</tr>
<tr>
<td>Hispanic</td>
<td>0.7386*</td>
<td>1.6569***</td>
<td>1.8008*</td>
</tr>
<tr>
<td></td>
<td>(0.31)</td>
<td>(0.49)</td>
<td>(0.80)</td>
</tr>
<tr>
<td>Intercept</td>
<td>−2.6104***</td>
<td>−3.9486***</td>
<td>−4.4618***</td>
</tr>
<tr>
<td></td>
<td>(0.35)</td>
<td>(0.49)</td>
<td>(0.81)</td>
</tr>
<tr>
<td>N</td>
<td>135</td>
<td>141</td>
<td>37</td>
</tr>
</tbody>
</table>

Notes. Total N = 858.
"Arrested" refers to respondents who report the correct number of arrests greater than zero. Reference category for the equation is Never Arrested (respondents who accurately reported not being arrested). Standard errors in parentheses.
†p < 0.10, *p < 0.05, **p < 0.01, ***p < 0.001.
predicting who is more likely to belong to what category (i.e. predicting if the DS is positive, negative, or zero), but it does not address the magnitude of the discrepancy between official and self-reported arrests. That is, it does not tell us how far off a subject’s count of self-reported arrests is likely to be from their official arrest count. It also does not examine the effect the number of arrests might have on the DS. It makes the naive assumption that the number of arrests has no impact on the reporting category.

To predict the magnitude of the DS and determine the impact of number of arrests, two separate Ordinary Least Squares (OLS) regression models were estimated. We did this because we suspected the processes of under-reporting and over-reporting would behave differently. Results from these two models are shown in Table 7. Using gender and race as predictors in these models tests for within-category gender and race differences (as opposed to the across-category differences tested in the multinomial logistic model above). In addition to using gender and race as predictors, we include a measure of the number of official arrests. Inclusion of this variable erased almost all gender and race effects from the previous model (although being male is associated with a reduction in the DS for over-reporters). So truly there is no effect of race in our sample, and only a minor effect of gender. In other words, the number of arrests is correlated with gender and race. Their effects are spurious until the effect of the number of arrests is controlled.

As can be seen in the table, the size of the effect of the number of arrests is approximately 0.5 times larger (in absolute value) when predicting under-reporting than when predicting over-reporting (0.659, compared to −0.427). In addition, the $R^2$ of the model predicting under-reporting (0.80) is substantially larger than that predicting over-reporting (0.15). In other words, under-reporting would appear to be more systematic and we appear to do a much better

<table>
<thead>
<tr>
<th></th>
<th>Under-reported DS</th>
<th>Over-reported DS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>−0.002 (0.05)</td>
<td>−0.189* (0.09)</td>
</tr>
<tr>
<td>Black</td>
<td>0.047 (0.07)</td>
<td>0.117 (0.11)</td>
</tr>
<tr>
<td>Hispanic</td>
<td>−0.067 (0.08)</td>
<td>−0.031 (0.13)</td>
</tr>
<tr>
<td>Number of arrests</td>
<td>0.659*** (0.01)</td>
<td>−0.427*** (0.04)</td>
</tr>
<tr>
<td>Intercept</td>
<td>−0.023 (0.08)</td>
<td>−0.186 (0.11)</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.80</td>
<td>0.15</td>
</tr>
<tr>
<td>$N$</td>
<td>723</td>
<td>698</td>
</tr>
</tbody>
</table>

**Notes.** *p < 0.05, **p < 0.01, ***p < 0.001.*
job predicting under-reporting than over-reporting. However, this model assumes linear effects of the number of arrests on the DS. That is, it assumes that the effects are constant over the range of the possible numbers of arrests, and therefore assumes the DS could increase without limit. As will be shown below, modeling under-reporting and over-reporting non-linearly tells a different story.

Table 8 presents results from an OLS model estimated to predict the overall DS. Whereas the model in Table 6 tested for across-category differences, and the models in Table 7 tested for within-category differences and the impact of number of arrests, the model in Table 8 tests for all of the above. In this model, we include a term for the number of official arrests for the under-reporters separately from the number of official arrests for over-reporters. Additionally, we suspected a non-linear effect of the number of arrests on the DS (i.e. the more arrests they have, the more they have to forget). To test this suspicion, we included the squared number of official arrests for both under-reporters and over-reporters. This allows for several possibilities, testing for a linear effect, a limited effect, or an explosive effect of number of arrests. To account for multicollinearity and variable skew in the model, the official number of arrest variables had to be centered and logged. Unfortunately, this makes interpretation of the unstandardized coefficients somewhat troublesome. For this reason, Table 8 includes standardized coefficients. The form of the equation is as follows:

\[
\ln(DS) = \alpha + \beta_1(\text{Male}) + \beta_2(\text{Black}) + \beta_3(\text{Hispanic}) + \beta_4(\ln(\text{UR})) + \beta_5(\ln(\text{OR}))
\]

\[
+ \beta_6(\ln(\text{UR}))^2 + \beta_7(\ln(\text{OR}))^2 + \epsilon
\]

where DS is the overall difference score,

Black and Hispanic are control variables for respondent race/ethnicity,

UR is the number of official arrests for under-reporters, and

OR is the number of official arrests for over-reporters.

Due to skewness, we log the number of arrests for under- and over-reporting and the DS for under- and over-reporters. We center all variables to remove collinearity. We allow over- and under-reporting to have a non-linear relationship to the DS and we model the across-category and within-category differences of race and gender simultaneously. All these things were ignored in previous analyses, both in this paper and by others.

Results show that a higher number of official arrests for under-reporters is associated with a larger positive DS, just as we expected. This indicates that more arrests are associated with larger discrepancies in under-reporting. Similarly, a higher number of official arrests for over-reporters is associated with a larger negative DS. So, more arrests lead to more of both types of error. Further, the significance of the squared terms indicates that these effects are non-linear. This means that the impact of the number of arrests on the DS get
<table>
<thead>
<tr>
<th></th>
<th>Adolescent model (Waves 4-9)</th>
<th></th>
<th>Young adult model (Waves 10-12)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Unstandardized coefficient</td>
<td>Standardized coefficient</td>
<td>Unstandardized coefficient</td>
<td>Standardized coefficient</td>
</tr>
<tr>
<td></td>
<td>(standard error)</td>
<td></td>
<td>(standard error)</td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>-0.146* (0.07)</td>
<td>-0.039 (0.06)</td>
<td>-0.134* (0.06)</td>
<td>-0.034 (0.06)</td>
</tr>
<tr>
<td>Black</td>
<td>0.166 (0.09)</td>
<td>0.045 (0.08)</td>
<td>0.027 (0.08)</td>
<td>0.007 (0.08)</td>
</tr>
<tr>
<td>Hispanic</td>
<td>-0.015 (0.11)</td>
<td>-0.003 (0.10)</td>
<td>-0.029 (0.10)</td>
<td>-0.006 (0.10)</td>
</tr>
<tr>
<td>Number of arrests for</td>
<td>0.922*** (0.07)</td>
<td>0.384 (0.03)</td>
<td>1.293*** (0.03)</td>
<td>0.763 (0.03)</td>
</tr>
<tr>
<td>under-reporters</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Squared arrests for</td>
<td>0.910*** (0.06)</td>
<td>0.413 (0.04)</td>
<td>0.787*** (0.04)</td>
<td>0.342 (0.04)</td>
</tr>
<tr>
<td>under-reporters</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of arrests for</td>
<td>-2.900*** (0.33)</td>
<td>-0.460 (0.41)</td>
<td>-5.770*** (0.41)</td>
<td>-0.609 (0.41)</td>
</tr>
<tr>
<td>over-reporters</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Squared arrests for</td>
<td>0.543* (0.24)</td>
<td>0.120 (0.32)</td>
<td>2.840*** (0.32)</td>
<td>0.376 (0.32)</td>
</tr>
<tr>
<td>over-reporters</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>-0.348 (0.14)</td>
<td></td>
<td>-0.125 (0.12)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>R² 0.72</td>
<td>N 839</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Notes. *p < 0.05, **p < 0.01 ***p < 0.001.
smaller with each additional arrest. Effectively, the DS for under- and over-reporters reach limits quickly. The nature of these limits are discussed below. The $R^2$ of the final model is still quite large for individual-level data, at 0.72.

There is no collinearity and there are no influential cases. We can conclude that most of the variance is accounted for by the number of arrests and the quadratic terms. Our model is a very effectual means of predicting if subjects will under-report or over-report, and the magnitude of the DS. Obviously, this is a complex equation differentially modeling non-linear logged and centered variables. However, it very accurately reflects Figure 1. It also shows us that there is no impact of race/ethnicity on the DS, after controlling for the number of official arrests, while there is a slight impact of gender.

Over-reporting is not less systematic than under-reporting, as appeared to be the case in the linear model shown in Table 7. Rather, the relationship is non-linear, but quite systematic. In the linear models above, we explained about 80% of the variability in under-reporting and 15% of the variability in over-reporting. Since there are roughly equal numbers of under-reporters and over-reporters, one might average the two $R^2$’s to reach an overall $R^2$ of 0.475 ($(0.80 + 0.15)/2 = 0.475$). Contrary to that expectation, and taking into account the non-linear relationships, we explain 72% of the variance overall, indicating that like under-reporting, over-reporting is intensely non-linear and non-linear modeling dramatically improves the explained variance. This finding goes well beyond prior studies’ explanations of the phenomena.

10. In logged units.
Because the DS are functions of both linear and quadratic effects, clear interpretation is not straightforward. To more easily illustrate what the total effect of the number of arrests on the predicted DS looks like (net of any race and gender effects), we include a graph of this effect in Figure 1. While the curve for under-reporters bends slightly, the predicted DS continues to grow with the number of arrests.

However, for the over-reporters, the curve drops quickly with each additional arrest (recall the DS for over-reporters was negative by definition) until about seven arrests. At this point, the curve’s growth slows down, and the curve eventually reaches a minimum value. In other words, the curve flattens out, indicating that the DS for over-reporters is as large as it will ever be after about seven arrests.\(^{11}\) One can see that the over-reporters’ curve does slowly approach the axis after it reaches its minimum value at 14 arrests. But, as shown in Figure 2, no over-reporter in our sample actually experienced more than 10 arrests. As the number of arrests increases, the number of subjects experiencing many arrests falls off quickly. Although there are more subjects, a similar distribution is formed by the under-reporters. For this reason, prediction becomes unstable at the tail end of the distributions.

Over-reporting appears to be less stable than under-reporting, but it approaches its practical limit much faster. Because a likely scenario that could lead to over-reporting would be a subject’s confusion about what actually constitutes an official arrest, we suspected that we might see the “bottoming out” simply because as the subjects get older, they learn what is and is not an arrest. For this to be plausible, the number of arrests would need to be highly

\(^{11}\) Unlogged arrests.
correlated with the subject’s age. Of course, since our measures cumulated
the number of arrests across six waves of data, obtaining a single measure of
age is not straightforward. Instead, we created 22 measures indicating the
mean age of subjects at their first arrest, their second arrest, third arrest, and
so forth, up to the mean age at the 22nd arrest. Figure 3 shows the mean age
at arrest for both under-reporters and over-reporters together. As we sus-
ppected, the number of arrests is highly correlated with the mean age at arrest
\((r = 0.98)\). That is, as the subject gets older, they accrue more arrests.

Figure 4 shows the same variables, but separating under-reporters and over-
reporters. Again, the two variables are highly correlated for both under-report-
ers and over-reporters \((r = 0.97 \text{ and } 0.92, \text{ respectively})\). The mean age at the
tenth arrest is less than the mean age at the ninth arrest for over-reporters. How-
ever, looking back at Figure 2, shows that only one over-reporter experi-
enced 10 arrests. In fact, a comparison of Figures 2 and 4 shows that mean
age at arrest for both over-reporters and under-reporters increase nearly iden-
tically (in Figure 4) from 1 to 7 arrests, where the frequencies drop off sub-
stantially (in Figure 2). After this point, the prediction becomes somewhat
erratic, because we have so few cases on which to base our predictive models.

**Multivariate Analysis for Adults**

Since age has such a profound effect on number of arrests, this might lead one
to question if the same processes operate for young adults. The second panel
on the right hand side of Table 8 estimates Equation 1 for our young adult
sample independent of the adolescent model. The results indicate that the

![Figure 3](image)

**Figure 3** Mean age at arrest, under-reporters, and over-reporters together.
young adult model is extremely similar to the adolescent model. The model explains somewhat more variance ($R^2 = 0.81$) than the adolescent model ($R^2 = 0.72$). The same variables are significant, identical in sign, and similar in size with one exception. The effects for over-reporters are substantially larger for adults than for adolescents. The shape of the effects shown in Figure 5 is nearly the same as Figure 1. The under-reporting has the same general shape and limit as the adolescent curve. This is shown in Figure 5. Figure 6 shows there are only 28 over-reporting adults, compared to 194 under-reporters, indicating that adults are much more likely to understand when they are arrested than adolescents. While the effect is statistically significant at the $p<0.001$ level, there are relatively few cases of over-reporters having three or more official arrests. The tail of the effect for over-reporting (three or more official arrests) is built on very few cases. We suspect the small sample size has the effect of making the tail of over-reporters much less stable. Overall, the two models (for adolescents and young adults) are remarkably similar, so one can conclude the process operating for adolescents continues to operate for adults.

Discussion

Since the 1960s, the self-report method of collecting data on crime and delinquency has been a major tool in understanding those behaviors and their consequences. In spite of the popularity of this methodology, there has been some trepidation in its use. The key concern is the accuracy of the obtained data when respondents are asked to report their own participation in such behavior
and whether they have been arrested. While there has been much work on exploring the reliability and validity of these data (see Thornberry & Krohn, 2000), the issue continues to be a major concern. More specifically, criminologists continue to ask whether there is systematic bias in the reporting of criminal behavior and arrests.

![Figure 5](image1.png)

**Figure 5** Total effect of the number of arrests on predicted difference score, Waves 10-12.

![Figure 6](image2.png)

**Figure 6** Number of subjects experiencing each number of arrests, Waves 10-12.
Adolescence

Typically, the focus of examinations of self-report data has been on the under-reporting of self-reported arrests or delinquent behavior. Few studies have recognized that there may also be a significant amount of over-reporting of the number of criminal behaviors (although the existence of over-reporting is considered by Kirk (2006), and Maxfield et al. (2000)). As we demonstrate in the current study, this is an important oversight. For adolescents, we find that almost as many respondents over-report being arrested \((N=78)\) as under-report arrests \((N=92)\). This might seem like good news in that, on average, over-reporting and under-reporting would balance each other out in the end. However, the question needs to be reframed so as to examine whether there is systematic bias in either under-reporting or over-reporting or both and, if so, whether different respondent characteristics account for such bias.

Using data from a major longitudinal panel study, we explored the relationship between race and gender and the likelihood of over-reporting and under-reporting arrests. Comparing self-reported arrests with official arrests, we found that males and females were no more likely to under-report or over-report arrests, although males were more likely to be in error in their reporting than were females. When examining errors in reporting by race, we found that Hispanics were no more likely to over-report or under-report arrests. However, African-Americans were more likely to under-report arrests, whereas whites were more likely to over-report arrests.

We further examined the influence of these demographic variables on over-reporting and under-reporting by incorporating a multivariate analysis that included the number of times a respondent had been arrested. We did this because we suspected that some of the differences by race may be explained by the frequency with which respondents were arrested. Blacks and Hispanics are more likely to be arrested, and therefore have more of an opportunity to either under-report or over-report their arrests. So we reason that controlling for number of arrests may eliminate a large part of their discrepancy in reporting, and may eliminate the effect of race in the model.

We estimate nonlinear effects of number of arrests in the models and we estimate both over-reporting and under-reporting separately in the same equation. This compares the different types of reporting to the reference group (accurately reporting no arrests) and allows us to determine if under-reports and over-reports have different nonlinear effects by number of official arrests. They do. So, the simple fact that under-reporting and over-reporting occur at the same rate is not such good news since they do not behave the same in their ranges of discrepancy.

However, the good news is that the effects of race and gender were for the most part no longer significant when we included the number of prior arrests. Those respondents who were arrested more frequently were more likely to
under-report their arrests regardless of race or gender. It is still important to note that race and gender were not significant once the number of arrests was included in the equation. So, race is inconsequential in determining the level of over-reporting and under-reporting once the number and shape of arrests are taken into account. Males are less likely to over-report than females, but they are not different on under-reporting.

However, over-reporting and under-reporting are strongly and differentially related to the level of arrest activity of the individual. The non-linear model does an excellent job of predicting the difference between the frequency of official arrests and the frequency of self-reported arrests. This is much more satisfying than just looking at the prevalence of under- and over-reporting. Frequency measures are more appropriate than prevalence measures once researchers account for the non-linear relationships. The non-linear form shows that the effects are asymptotic at reasonable levels of arrests. At high levels, prediction is not reliable due to low numbers of subjects after about six arrests (see Figure 2). For all practical purposes, data should be concatenated at six arrests for both under- and over-reporters. In addition, the functions reach a limit at this point where the DS stop growing. Furthermore, over-reporting seems to be strongly related to age of the subject. They seem to figure out what arrests are and are not as they gain more experience. These over-reporting and under-reporting effects do not grow ad infinitum as arrests increase. Rather, they reach practical limits bounded by reasonable numbers of arrests. Finally, the models do an extremely good job predicting the DS for both under- and over-reporters, as indicated by the $R^2$ in Table 7.

Adults

The adult equation is remarkably similar to the adolescent equation, as shown in the right panel of Table 7. The conclusions remain largely the same. The one important difference is that adults are much less likely to over-report arrests than adolescents. Only 28 (3.5%) of them do so. In fact, of these 28 adult over-reporters, 13 of them were never arrested as adolescents and experienced their first arrests as adults. Ten of these subjects only experienced one official arrest. This strengthens the idea that these subjects may be naive to what actually constitutes an arrest. Apparently, adults are very likely to know when they are arrested and when they are not.

It is interesting that the adult equation is so similar to the adolescent equation given the methodological disparities in data collection between the two samples. While adolescent data were collected every six months over a two and a half year period, the adult data were collected only once per year, with gaps in interviews ranging between 1 and 3 years. This suggests that recall periods, at least for adults, are not so important in determining under- or over-reporting. What’s more, the nature of arrests changes between these two
phases of data collection. Adults are no longer eligible for status offense arrests and less likely to be treated leniently due to their youth. Adolescents are less likely to travel out of state to be arrested, while adults are more likely, and this would make their arrests invisible to us. This could lead to more slippage and the appearance of over-reporting. This did not occur in our data, as evidenced by the small number of adult over-reporters. Overall, we take this to be very good news for the quality and stability of adult self-reports.

Future Directions

What implications do our findings have for the continued use of self-reported measures of arrests? We interpret our findings as being good news for the proponents of the use of self-reported measures. While we find systematic bias in both the under-reporting and over-reporting of arrests by race and gender, that bias does not appear to be a result of purposeful misrepresentations. Rather, differences in over-reporting and under-reporting appear to be more a result of the level of arrest activity. Future research will need to take this into account in the analysis and interpretation of self-reported measures of arrest, a task that is more directed than accounting for what were unknown reasons for race and gender reporting differences.

Our research shows that our naive models are less accurate in predicting and describing under- and over-reporting than the later, more sophisticated models we estimated. For example, when we estimated the bivariate models, models that only estimated across-category variation or between-category variation, and linear models, we explain less variance and obtain different findings than when we estimate non-linear models that account for both across-category and within-category variation at once.

The question of what our findings may mean for the use of self-report measures of delinquent behavior (as opposed to self-report arrests) is more difficult to determine. In our study, respondents were aware of the fact that we were collecting information from official records (they needed to provide us with permission to do so). It is possible that this knowledge made them more likely to self-report their arrests. Would knowing that official record data were collected also make them more likely to report delinquent behavior if that behavior did not result in an arrest? We think that is probable, but to a lesser extent than for self-reported arrests.

On the other hand, our findings regarding the effect of experience with both juvenile and adult justice systems suggest that there may not be any reluctance on the part of youth to self-report their arrests, but rather it is simply a matter of the difficulty in recalling multiple incidents. A similar factor could be operating in the reporting of delinquent behavior. If so, then those youth who are high-frequency offenders may report a lower percentage of their delinquent behaviors compared to lower rate offenders. The problem is that
without an independent measure of the frequency of offending, the researcher would not be able to observe this.

We believe that future research would benefit from investigating whether discrepancies between official and self-reported measures as an adult are predictable from such discrepancies as an adolescent. The analyses presented herein demonstrate that the processes driving under-reporting and over-reporting functions for adults operate in a remarkably similar way to those driving adolescent reporting. The bias between official and self-reported arrests was extremely systematic, as evidenced by the explained variance measures in the adolescent and adults models shown in Table 8. Using the information in adolescent reporting to explain adult reporting behaviors would likely explain an incredible proportion of variance in adult under-reporting and over-reporting. Exploring these processes along a longitudinal dimension would be a fruitful next step in the research on the accuracy and validity of self-reported measures of crime and delinquency.

References


