Cumulative Prevalence of Arrest From Ages 8 to 23 in a National Sample
Robert Brame, Michael G. Turner, Raymond Paternoster and Shawn D. Bushway
Pediatrics; originally published online December 19, 2011;
DOI: 10.1542/peds.2010-3710

The online version of this article, along with updated information and services, is located on the World Wide Web at:
http://pediatrics.aappublications.org/content/early/2011/12/14/peds.2010-3710
Cumulative Prevalence of Arrest From Ages 8 to 23 in a National Sample

WHAT’S KNOWN ON THIS SUBJECT: Although there is some older literature examining how arrest prevalence accumulates through adolescence and adulthood, there is no contemporary research examining the arrest histories of a representative sample of American youth.

WHAT THIS STUDY ADDS: Using a contemporary US sample of adolescents and young adults, we estimated the cumulative arrest prevalence through age 23. The results suggest a substantial increase in the cumulative prevalence of arrest since the 1960s.

OBJECTIVE: To estimate the cumulative proportion of youth who self-report having been arrested or taken into custody for illegal or delinquent offenses (excluding arrests for minor traffic violations) from ages 8 to 23 years.

METHODS: Self-reported arrest history data (excluding arrests for minor traffic violations) from the National Longitudinal Survey of Youth 1997 (N = 7335) were examined from 1997 to 2008.

RESULTS: By age 18, the in-sample cumulative arrest prevalence rate lies between 15.9% and 26.8%; at age 23, it lies between 25.3% and 41.4%. These bounds make no assumptions at all about missing cases. If we assume that the missing cases are at least as likely to have been arrested as the observed cases, the in-sample age-23 prevalence rate must lie between 30.2% and 41.4%. The greatest growth in the cumulative prevalence of arrest occurs during late adolescence and the period of early or emerging adulthood.

CONCLUSIONS: Since the last nationally defensible estimate based on data from 1965, the cumulative prevalence of arrest for American youth (particularly in the period of late adolescence and early adulthood) has increased substantially. At a minimum, being arrested for criminal activity signifies increased risk of unhealthy lifestyle, violence involvement, and violent victimization. Incorporating this insight into regular clinical assessment could yield significant benefits for patients and the larger community. Pediatrics 2012;129:21–27

AUTHORS: Robert Brame, PhD,a Michael G. Turner, PhD,a Raymond Paternoster, PhD,b and Shawn D. Bushway, PhDc

aDepartment of Criminal Justice and Criminology, University of North Carolina at Charlotte, Charlotte, North Carolina; bDepartment of Criminal Justice and Criminal Justice, University of Maryland at College Park, College Park, Maryland; and cSchool of Criminal Justice, University at Albany, State University of New York, Albany, New York

KEY WORDS
crime, criminals, arrest, health risk factors

ABBREVIATIONS
MAR—missing at random
NLSY97—National Longitudinal Survey of Youth 1997

Each of the authors made substantive intellectual contributions to this study. Dr Turner procured the data with the approval of the University of North Carolina-Charlotte Institutional Review Board, and Drs Turner, Paternoster, and Bushway collaborated closely on the data analysis with Dr Brame who assumed primary responsibility for that analysis. All authors participated in writing and revising the article through multiple drafts. All authors concur in their approval of the article’s contents and the findings reported herein.

www.pediatrics.org/cgi/doi/10.1542/peds.2010-3710
doi:10.1542/peds.2010-3710
Accepted for publication Sep 9, 2011

Address correspondence to Robert Brame, PhD, Department of Criminal Justice and Criminology, University of North Carolina at Charlotte, 9201 University City Blvd, Charlotte, NC 28223. E-mail: rbrame@uncc.edu

PEDIATRICS (ISSN Numbers: Print, 0031-4005; Online, 1098-4275).
Copyright © 2012 by the American Academy of Pediatrics

FINANCIAL DISCLOSURE: The authors have indicated they have no financial relationships relevant to this article to disclose.
For American youth, involvement in criminal behavior is a well-known risk factor for adverse health, social, academic, occupational, and economic outcomes. Youth with arrest records have lower levels of earnings, longer bouts with unemployment, greater work instability, diminished educational levels, and a greater risk of destructive family conflicts. Once youth are formally processed by the criminal justice system, they may find themselves at an early age with diminished personal, social, human, and cultural capital, a process described by Hagan as “criminal embeddedness.” Often ensnared in criminal social networks as the result of 1 or more arrests, youth may find it hard to escape, leading to additional crime as adults. Indeed, there is a risk that the collateral social and personal damage created by an arrest mortgages the futures of young people as they make the transition to adulthood.

Moreover, rates of involvement with the criminal justice system after conviction, from probation to incarceration, have dramatically increased (three- to fourfold) since the 1970s (S. Shannon, PhD, C. Uggen, PhD, M. Thompson, PhD, J. Schnittker, PhD, M. Massoglia, PhD, unpublished data). What is not clear is whether the prevalence of arrest for nontraffic offenses has changed over time. The best available benchmark for assessing this question was provided by Christensen in the 1960s. Christensen’s work forecast the future prevalence of arrest for nontraffic offenses assuming that 1965 conditions remained stable. Based on this work, the best available estimate of the percent of the population arrested at least once for a nontraffic offense by age 23 is 22%. Yet there has never been a national estimate of the cumulative prevalence rate of arrest for criminal activity of adolescents and young adults by using individual-level survey data from a nationally representative sample.

**METHODS**

The National Longitudinal Survey of Youth 1997 (NLSY97) is a prospective household-based longitudinal study of American youth between the ages of 12 and 16 years in 1997. It is based on a stratified multistage cluster probability sampling design. The initial stage of data collection included a screening interview to determine whether age-eligible youth resided in selected households. These screening interviews were successfully completed in 94% of the targeted households. Based on the information obtained from the screening interviews, a total of 9808 youth were included in the NLSY97 targeted sample, and a subset of 8984 of these youth actually participated in the first interview.

The targeted sample of 9808 youth comprised 2 mutually exclusive groups: (1) a “cross-sectional” sample ($N = 7335$) and (2) a supplemental oversample of minority youth ($N = 2473$). The cross-sectional sample was designed to be a self-weighting representation of US households with adolescents between the ages of 12 and 16 on December 31, 1996. Because 6% of the selected households did not complete an initial screening interview, it is not possible to definitively say that this representativeness was actually achieved but the sample certainly includes a broad cross-section of American youth in the late 1990s.

We examined self-reports of arrest from the cross-sectional NLSY97 sample of 7335 youth. The vast majority of these youth ($N = 6748; 92\%$) participated in the first round of surveys that were conducted in 1997 and 1998. This leaves a residual sample of 587 individuals who did not participate at the first round. Since the first round of data collection, 11 additional (approximately annual) surveys have been conducted (through 2008) and made publicly available. The NLSY97 survey methodology allowed individuals who participated at the first wave to miss 1 or more waves of data collection and still remain in the study, but individuals who did not participate at the first wave did not participate at any of the subsequent waves.

At the first wave of data collection, each study participant was asked the following question: “Have you ever been arrested by the police or taken into custody for an illegal or delinquent offense (do not include arrests for minor traffic violations)?” Participants who answered “yes” were asked a follow-up question about their age (in years) at the time of their first arrest. Our analysis makes the assumption that these arrest events occurred at the midpoint of the age reported by the respondent.

At each of the follow-up interviews, study participants were asked a slightly different question: “Since the date of last interview on [date of last interview], have you been arrested by the police or taken into custody for an illegal or delinquent offense (do not include arrests for minor traffic violations)?” Participants who answered “yes” were also asked to recall the month and year of the first arrest since the last interview. Based on the participants’ answers to the first-wave and follow-up questions, we compiled a history of what is known about each individual’s “ever-arrested” status at each age from age 8 until 2008 when the most recent available round of data were collected. At each age, we divided the NLSY97 cross-sectional sample ($N = 7335$) into 3 groups: (1) those who have not been arrested yet; (2) those who have been arrested; and (3) those whose arrest status cannot be determined at...
that age (because of missing data). Table 1 summarizes this information from ages 8 to 23 (beginning at age 23, the youngest participants’ arrest status cannot be determined; therefore, after age 23, respondents are progressively censored).

As Table 1 shows, some of the arrest data are missing at each age covered by the survey. It is possible, however, to obtain interval estimates that definitely contain the in-sample prevalence rate at each age.24 At each age, we estimated the upper bound of the in-sample prevalence rate by a 2-step process: (1) add the number of people who had been arrested by that age to the number of missing cases at that age, and (2) divide that sum by the total number of people (N = 7335). This calculation is based on the extreme assumption that all of the missing people have been arrested. The lower bound of the in-sample prevalence rate is estimated by dividing the number of people who had been arrested by that age by the total number of people (N = 7335). Symmetrically, this calculation makes the assumption that none of the missing people have been arrested.

It is, of course, unlikely that the missing cases behave in such uniform ways. For this problem, many researchers would be willing to invoke the assumption that the missing cases are missing at random (MAR).24–26 The MAR estimates are obtained by dividing the number of cases who have ever been arrested at a particular age by the number of valid cases at that age (ever arrested + never arrested, excluding the missing cases).

The final step of the analysis involves the estimation of 95% confidence limits around the outer bound estimates and the MAR point estimates. We estimated these intervals by drawing 1000 bootstrap samples and calculating the SD of each parameter’s bootstrap distribution.26 One potential problem with this approach is that the NLSY97 is based on a cluster sampling design. Although the point and interval estimates obtained from the cross-sectional sample are valid, the SEs and confidence intervals obtained under the assumption of simple random sampling will generally be too small. To correct for this problem, we conservatively assumed a SE design effect multiplier of 2.0.23 The resulting design-effect corrected confidence intervals are only trivially different from those obtained by conventional bootstrap SEs. Confidence intervals presented in this article are based on the design-effect corrected SEs.

### RESULTS

Figure 1 displays the main analysis results from the NLSY97 (ages 8–23) alongside the 1965 point estimates (indicated by black squares) presented by Christensen20 (ages 10–23). It is important to note that Christensen20 graphed the population male and female growth curves separately. We obtained his point estimates by visually inspecting his graph and assuming that male and female participants each comprise 50% of the population. The MAR point estimate of the cumulative prevalence of arrest is documented by a solid black circle at each age. Moving outward, the next set of lines represents the 95% confidence interval for the MAR estimate (our uncertainty due to sampling error if the usual MAR assumption is valid). The next set of lines represents the outer bounds of the in-sample prevalence estimates based on the assumptions that the missing cases were either all arrested (upper bound) or all not arrested (lower bound). Finally, the interval endpoints represent the design-effect corrected 95% confidence limits for the outer bound prevalence estimates.

It is useful to compare the MAR estimates from the NLSY97 to Christensen’s20 estimates from 1965. The 1965 estimates are higher than those in the NLSY97 during early adolescence (through age 15). At ages 16 to 18, there is considerable parity between the 2 measurements; the 1965 estimates lie within the 95% confidence limits of the MAR estimate. From ages 19 to 23, the 1965 estimates appear to increase at a much slower pace than the NLSY97 MAR estimates. By age 23, the NLSY97 MAR estimate of 30.2% is over a third higher than the 1965 estimate of 22%. And the 1965 estimate lies clearly outside the 95% confidence interval for the MAR estimate.

There are 2 potential concerns with this comparison: (1) we do not know what

### TABLE 1 Summary of Ever-Arrested Status (N = 7335)

<table>
<thead>
<tr>
<th>Age, y</th>
<th>Yes</th>
<th>No</th>
<th>Missing</th>
</tr>
</thead>
<tbody>
<tr>
<td>8.0</td>
<td>0</td>
<td>6176</td>
<td>619</td>
</tr>
<tr>
<td>8.5</td>
<td>4</td>
<td>6172</td>
<td>619</td>
</tr>
<tr>
<td>9.0</td>
<td>4</td>
<td>6172</td>
<td>619</td>
</tr>
<tr>
<td>9.5</td>
<td>10</td>
<td>6078</td>
<td>619</td>
</tr>
<tr>
<td>10.0</td>
<td>10</td>
<td>6076</td>
<td>619</td>
</tr>
<tr>
<td>10.5</td>
<td>25</td>
<td>6691</td>
<td>619</td>
</tr>
<tr>
<td>11.0</td>
<td>25</td>
<td>6691</td>
<td>619</td>
</tr>
<tr>
<td>11.5</td>
<td>60</td>
<td>6656</td>
<td>619</td>
</tr>
<tr>
<td>12.0</td>
<td>60</td>
<td>6656</td>
<td>619</td>
</tr>
<tr>
<td>12.5</td>
<td>146</td>
<td>6568</td>
<td>620</td>
</tr>
<tr>
<td>13.0</td>
<td>151</td>
<td>6555</td>
<td>629</td>
</tr>
<tr>
<td>13.5</td>
<td>262</td>
<td>6438</td>
<td>635</td>
</tr>
<tr>
<td>14.0</td>
<td>286</td>
<td>6408</td>
<td>641</td>
</tr>
<tr>
<td>14.5</td>
<td>448</td>
<td>6232</td>
<td>655</td>
</tr>
<tr>
<td>15.0</td>
<td>495</td>
<td>6167</td>
<td>673</td>
</tr>
<tr>
<td>15.5</td>
<td>638</td>
<td>6006</td>
<td>691</td>
</tr>
</tbody>
</table>
the confidence intervals for the 1965 estimates are, and (2) the MAR assumption is untestable.\textsuperscript{25,26} Concerning the first issue, although we know Christensen’s\textsuperscript{20} estimates are subject to some degree of error, we are uncertain how much error there is. It is worth noting, however, that if the error in Christensen’s\textsuperscript{20} estimates was as large as the sampling error for the MAR estimate in the NLSY97, there would still be a significant difference between the 2 estimates.

Considering the validity of the MAR assumption, it is clearly possible that individuals who are missing or who choose not to answer some or all of the arrest questions may not be comparable to those who provide valid data. To address this concern, we consider the assumption that the MAR estimate is a lower bound measure of the true arrest prevalence rate. Because research suggests that active offenders are underrepresented in broad population surveys such as the NLSY97, this assumption seems plausible.\textsuperscript{14,22,27,28} Figure 2 presents this comparison assuming that the MAR estimate and lower 95% confidence limit capture the lower bounds for what we would accept as a credible arrest prevalence rate in the NLSY97. From age 19 on, the

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure1.png}
\caption{Analysis results from the NLSY97 and the 1965 point estimates.}
\end{figure}

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure2.png}
\caption{Arrest Rates and 95\% confidence intervals assuming the lower bound is equal to MAR.}
\end{figure}
NLSY97 is clearly pointing to a higher arrest prevalence rate than what was estimated in 1965.

A second feature of both Figs 1 and 2 is the growth in the cumulative prevalence of arrest that occurs during adolescence and into early adulthood. An observable pattern of growth begins around age 12 when the MAR estimate is virtually 0. By age 18, the MAR rate is 17.8% and is very similar to the 1965 estimate. This is a steep rate of growth and reflects the relatively common experience of criminal involvement and arrests for criminal involvement during the adolescent years in the United States (both in the 1960s and today).

Finally, although growth continues into adulthood, the rate of growth compared with the adolescent years becomes slower. Visually, the figure appears to be taking on a flatter trajectory as individuals enter their 20s. In fact, the rate of change of the MAR estimate between ages 16 and 17 ([(14.1 – 11.0)/11.0 = 28.2%] is greater than the rate of change of this same estimate between ages 22 and 23 ([(30.2 – 28.4)/28.4 = 6.3%]. But the rate of growth in early adulthood appears to be on steeper trajectory in recent years than it was in 1965.

**DISCUSSION**

Early arrest research conducted by Christensen in 1965 estimated that if 1965 conditions remained stable, ~22% of the US population would be arrested for a nontraffic offense by age 23. More recent research has been based on individual criminal history searches for populations in particular locales. But estimation of cumulative arrest prevalence rates based on these methods is problematic. Calculations based on aggregate statistics from government agencies rely on strong and hard-to-test assumptions. And, criminal history searches also have well-documented ambiguities and difficulties including record accuracy, satisfactory name-matching, jurisdictional boundaries, change in jurisdiction as a result of a family move, and confidentiality of juvenile records. There are, therefore, 3 significant gaps in the extant literature: (1) lack of contemporary evidence about cumulative arrest prevalence among young people; (2) estimation of cumulative arrest prevalence for a national sample of specific individuals rather than aggregated data sources; and (3) estimates based on measurement methods that are not marred by the inadequacies of criminal history searches.

This study examined self-reports of arrest from a broad, contemporary US sample during adolescence and young adulthood to address these gaps. Arrest experiences were measured via self-reports of study participants. The analysis was designed to answer the following 2 questions: (1) what proportion of American youth is arrested by age 23, and (2) how does that proportion accumulate as the population moves through adolescence and into early adulthood? Of course, the self-report method is accompanied by its own difficulties including whether perceptions of arrest are accurate, whether timing can be adequately recalled, and whether respondents are making an effort to be truthful in reporting their experiences.

Our primary conclusion is that arrest experiences are common among American youth (most likely on the order of ~1 out of 3 by age 23). In fact, our MAR estimate of 30.2% is substantively higher than the 22% previously forecast by Christensen in the 1980s. And there are a number of compelling reasons to believe that the prevalence of arrest may have increased over this time period. The criminal justice system has clearly become more aggressive in dealing with offenders (particularly those who commit drug offenses and violent crimes) since the 1960s (S. Shannon, PhD, C. Uggen, PhD, M. Thompson, PhD, J. Schnittker, PhD, M. Massoglia, PhD, unpublished data, 2011). In addition, there is some evidence that the transition from adolescence to adulthood has become a longer process; more youth are involved in postsecondary education, whereas marriages, childbearing, and the beginning of careers are occurring later in life, perhaps contributing to a longer period of “adolescence” today than in years past.

**CONCLUSIONS**

Given the considerable risks signaled by arrest experiences and the developmental handicaps that arrest may create for youth and young adults, these findings raise important questions about consequences and opportunities for intervention by pediatricians. For example, there is little doubt that pediatricians are concerned about the long-term developmental health of youth, including their involvement in violent and antisocial behavior. We have identified more than 400 articles published in *Pediatrics* since 1948 that pertain to delinquency and problem social behaviors (cigarette smoking, drinking, drug use, bullying, and various forms of abuse). However, although we have prevalence estimates for the risk of victimization, exposure to violence, insomnia, and sun exposure until now there simply was no contemporary national prevalence estimate of the risk of a criminal arrest for American youth. This is in spite of the fact that having an arrest record is
known to be an important risk marker for violence involvement, violent victimizations, and an unhealthy and unsafe lifestyle.

In addition, the experience of being arrested may be more than a marker insofar as it has its own effects both at the time of the arrest and in the months and years after. The primary routes for youth to be successful today are to cultivate conventional social networks and social capital through education and securing stable employment. Youth with an arrest record, however, may fail to secure this long-term beneficial form of capital and as a result may be effectively shut out of educational and employment opportunities. Youth with arrest records have been shown to have unstable and abbreviated employment histories, are less likely to stay in high school and enroll in college, are at greater risk of failing to obtain other markers of adult success such as having their own home and a stable relationship with a partner, and are more likely to have medical problems and adult drug and alcohol abuse.\(^{43,44}\)

There are early risk factors that appear for delinquency (risk factors that can easily be identified by pediatricians). For example, we know that the following factors put youth at risk for later and ongoing delinquent behavior if they appear before age 12: hyperactivity or poor concentration, delayed language development, low academic performance, poor relationships with parents and general home discord, antisocial parents; a broken home, harsh, punitive, or inconsistent discipline from parents; family violence; child abuse or neglect; diminished executive function or other cognitive deficits; sleep disorders; low birth weight or other perinatal complications; teenage parenthood; and early aggressive or bullying behavior.\(^{13,37,38}\)

These are the very types of risk factors that any pediatrician could easily identify with a risk assessment examination in any normal office visit, and physicians may be the first noncaregiver to view both the composite landscape of risk factors confronting a child and the medical consequences of those factors.\(^{37,38}\) The early efforts of pediatricians could play a pivotal role as early intervention has repeatedly been shown to be the most effective avenue for dealing with delinquent behavior. What Yancy\(^{45}\) noted 15 years ago is no less true today: “Because the pediatrician has a continuous relationship with children and their families, he or she can direct them to the appropriate facility and encourage and support them in carrying out the treatment.” Based on our findings, a significant percentage of American youth will experience at least 1 arrest for a nontraffic offense by age 23, and the greatest increase in the risk of this experience occurs during late adolescence. Timely intervention by pediatricians in the lives of these youth may be an important opportunity to move young people onto a path toward safer, healthy, productive, and successful lives.

REFERENCES

41. Johnson EO, Roth T, Schultz L, Breslau N. Epidemiology of DSM-IV insomnia in adolescence: lifetime prevalence, chronicity, and an emergent gender difference. Pediatrics. 2006;117(2). Available at: www.pediatrics.org/cgi/content/full/117/2/e247