5

Data and Methods

This chapter details the data and methods to be used in the analyses to be presented in Chapters 6, 7, and 8. The first section of this chapter describes the data sources, the specific variables to be used in the data sets, and the limitations of these data. This section also contains a description of how the final analytic samples were constructed. We conclude this chapter with a brief discussion of the statistical methods to be employed in our study. Readers interested in a thorough discussion of these methods are directed to Appendix A, which contains a technical description and comparison of the Poisson model, the standard negative binomial model, the finite mixture or semiparametric random effects Poisson model of Nagin and Land (1993; Land and Nagin 1996; Land, McCall, and Nagin 1996), and the parametric random effects negative binomial model.

The Three Release Samples

As mentioned previously, the data used in this study consist of three samples of California Youth Authority wards. Norman Skonovd and Rudy Haapanen of the CYA’s Research Bureau had previously collected the data for the two earliest samples (1981–2, 1986–7) (Skonovd and Haapanen 2000); the data for the 1991–2 sample were collected by the authors (with funding for the data collection provided by a grant from the National Institute of Justice). To maintain consistency across the three samples, the 1991–2 sample was coded according to the same rules and procedures used in collecting the two prior samples. The only differences in the coding rules pertained to the data on prior criminal history (as described below).

---

1 NIJ Grant Number #98—CE—VX—0026.
The two earliest samples are based on data collected on 2,086 wards released from the California Department of the Youth Authority institutions in fiscal year 1981–2 (1 July 1981 to 30 June 1982) and 2,078 wards released in fiscal year 1986–7 (1 July 1986 to 30 June 1987).\(^2\) Initially, a random sample of 2,200 wards was drawn from the 4,425 wards released from Youth Authority institutions in fiscal year 1981–2 and another random sample was obtained for 2,200 wards from the 3,048 wards released in fiscal year 1986–7 (Skonovd and Haapanen 2000). Of the 2,200 wards in the 1981–2 sample, 114 cases were removed from the study because their records were ‘court-ordered seals,’ which prevents any access to their files.\(^3\) This resulted in a 2,086 ward data file for this baseline data set. Of the 2,200 wards in the 1986–7 sample, 122 cases were removed for the same reason. This resulted in a 2,078 ward data file for the second data set. Together, these two samples provide records on 4,164 individual wards.

The arrest data for these two samples originally were only available through 31 December 1991 for the 1981–2 sample and through 31 December 1990 for the 1986–7 sample. We have, however, updated the arrest data for these two samples through 30 June 2000. The fact that the coded cases did not appear to be significantly different from the initial sample on any critical variables is not surprising given that 70 per cent of the initial sample was coded and that the cases were strictly coded according to their random numbers.

The 1991–2 sample is based on data collected on 1,527 wards released from the California Department of the Youth Authority institutions in fiscal year 1991–2 (1 July 1991 to 30 June 1992). Initially, a random sample of 2,198 wards was drawn from the total of 4,030 wards released from Youth Authority institutions in that fiscal year.\(^4\) Of the 2,198 cases, 13 of the sample members were

---

\(^2\) Throughout this study, the term ‘sample stay’ is used to refer to the incarceration period in the CYA that resulted in the ward being released during the sampling time frame. The date of release for the sample stay was the key defining element that resulted in the ward having been included in one of the samples.

\(^3\) When a CYA ward’s records are ordered ‘sealed’ by the court, the ward’s CYA Master Files are sealed and the ward’s 5-digit ‘YA number’ is replaced with an S and 4 digits (e.g. S0001). Law then prohibits access to the ward’s juvenile offending history and CYA Master File.

\(^4\) The sample initially consisted of 2,200 cases, but two individuals were sampled twice because they had been released twice during the sampling time frame. We only
subsequently removed because their records were ‘court-ordered seals’ (resulting in an initial file of 2,185 cases).

Importantly, prior to beginning the formal coding on the 1991–2 sample, concerns arose over possible time and budget limitations given that cases were taking longer to code by coders at the culmination of the training sessions than previously estimated. The decision was made that guaranteeing accuracy of the coded cases would be of greater value compared to the speedy collection of data for the entire 2,198 cases. Accordingly, a random number was assigned to each case in the sample at the outset of the formal coding process, and cases were then coded according to their random number (rank ordered from lowest to highest). This ensured that at the end of the available time allotted and available financial resources, the resulting sample would still be a random subset of the initial sample. At the point in time when both the time and financial resources had been exhausted, 1,527 of the original 2,198 (70%) cases had been coded. Thus, the final file for the 1991–2 sample consisted of a total of 1,527 wards. Comparisons of the wards coded with the original entire sample indicated no significant differences in terms of ethnicity, county of commitment, commitment offense, court of commitment, and the probability of arrest after release from the CYA. The fact that the coded cases did not appear to be significantly different from the remainder of the initial sample on any critical variables is not surprising given that 70 per cent of the initial sample was coded and that the cases were strictly coded according to their random numbers; thus, the ‘randomness’ of the file was maintained even though the entire sample initially drawn could not be coded.

We have also updated mortality data for the two earliest samples and collected original mortality data on the 1991–2 sample though allowed each of these individual’s one record in the data file, and we selected the later release date for these two individuals. The later release data was selected so as to not artificially create a very rapid failure time, which would have been assured if we had chosen the earlier release date.

The comparisons were made based on the fact that some variables were available regardless of whether the case was coded. Certain information is available on all wards committed to the CYA through the OBITS (Offender Based Institutional Tracking System) computer system. Further, since the arrest data after the sample release was obtained electronically, we obtained the post-release criminal history data on all available cases.
31 December 1999. The possible mortality of the subjects in this study is important for two reasons. First, we did not want a ward to appear as ‘arrest-free’ simply as a function of his death. Thus, wards who had died were removed from the risk of arrest after the point in time at which they were found to have been deceased. Second, as argued by Gottfredson and Hirschi (1990: 94), serious chronic offenders should experience ‘death at higher rates than the general population.’ Analyses by Dobrin (2001) showed that individuals with criminal arrests in their backgrounds have a greater chance of dying by homicide compared to individuals with no criminal arrests in their backgrounds. Preliminary analyses of the previously collected mortality data on the two earliest data sets analyzed herein lend some credibility to this assertion as well (see Lattimore, Linster, and MacDonald 1997).

**Data Sources and Variables**

The sources of information that were used to collect the data varied according to both the type of data element and whether the data element was referring to a characteristic or behavior prior to or after the date of release from the CYA that resulted in the ward being included in the sample. We will refer to these two different segments of the data collection process as the ‘pre-release’ and the ‘post-release’ periods. The key division point is the date of release for the ‘sample stay,’ which refers to the specific incarceration or ‘stay’ at the CYA that resulted in the ward’s inclusion in one of the three samples employed in our analyses. The ‘pre-release’ period refers to time prior to the sample release, while the ‘post-release’ period refers to time after the date of parole release.

**Pre-release Data: Case Characteristics Information**

Data for the pre-release period on the characteristics of the cases were collected from two sources: (1) Youth Authority’s electronically stored information on the ward and (2) Youth Authority’s ‘hard copy’ ward Master Files. From various computer files within the CYA and the CYA’s Offender Based Institutional Tracking System (OBITS), data were obtained for the following variables:

- Dates of admission and release for sample stay
- Base commitment offense (e.g. adjudication for murder, forcible rape, burglary, robbery)
• Admission status (first commitment, parole violator, recommit-
ment)
• Date of birth
• Gender
• Ethnicity (White, African-American, Hispanic, Asian, Native American, Filipino, Pacific Islander, Other)
• Court of commitment (juvenile court or adult criminal court)
• M Case (CDC ‘Housing’ Case)
• County of commitment (e.g. Los Angeles, Sacramento, San Diego, San Francisco)
• Major CYA infractions (DDMS violations) for such things as fighting and gang behavior.

The second data source is the individual, hard copy Master File completed for each ward. The Master File contains all available prescribed program and parole data, as well as data pertaining to the ward’s entire medical, educational, psychosocial, and criminal history up through the date of discharge from Youth Authority’s jurisdiction.

Trained coders reviewed and coded relevant data from the following types of documents: police, probation, and court reports, Youth Authority staff reports and documents, consultant reports and evaluations, and letters and appeals. Information from these sources was coded according to uniform guidelines. The Master File was the major source of information regarding the prior behavior and characteristics of the cases. Information regarding prior criminal record, as well as family background, substance abuse, gang activity, and prior placement information is either not available in OBITS, or it is not as complete or accurate as that contained in the Master File. The following variables were coded from the detailed information contained in the Master Files:  

• Family violence: Evidence of violence among the family members (not including the ward).
• Parental alcohol or drug dependence: Evidence that the ward’s parents or guardians had an alcohol or drug dependence problem (e.g. arrests for drug offenses, been in treatment for

6 In the following descriptions, ‘parental’ indicates the ward’s parents unless the parents are no longer the guardians of the ward. In that case, parental refers to the ward’s guardians.
drug/alcohol problems). Social drinking or occasional marijuana use was not recorded as evidence.

- **Parental criminality**: Evidence the ward’s parents had been involved in criminal activity (e.g. prior arrests or incarcerations). A single arrest for drunk driving was not recorded as such evidence.

- **Sibling criminality/delinquency**: Evidence the ward’s siblings had been previously involved in criminal or delinquent activity (e.g. prior arrests or incarcerations). One arrest for drunk driving was not recorded as evidence.

- **Parental lack of supervision/neglect**: Evidence that the ward was not adequately supervised or ward was neglected by his or her parents (e.g. ward was removed from custody of parents due to the behavior of the parents; parents do not know where the ward is usually).

- **Ineffective parental control**: Evidence that the parents had ineffective or inconsistent control over the ward (e.g. ward arrested for being ‘out of control’ or noted in probation records as being ‘out of control’). Naturally, arrests for criminal offenses were not recorded as evidence of being beyond the control of the parents.

- **Physical abuse**: Evidence that the ward was subject to either extreme punitiveness or physical abuse (e.g. parent arrested for abuse of the ward; severe whippings; spankings that cause injury). Spanking alone (without injury) was not recorded as evidence.

- **Sexual abuse**: Evidence that the ward was subjected to sexual abuse by others (e.g. molestation).

- **Drug abuse**: Evidence that the ward abused drugs (not including alcohol). Experimental drug or chemical use was not recorded as evidence of drug abuse. Frequent use of hard drugs such as cocaine, PCP, and heroin, and ‘sniffing’ were recorded as such evidence.

- **Gang member/association**: Evidence that ward associated with gang members, participated in gang activities for self-protection, or was a fully participating ‘gang member’ who engaged in ‘gang banging.’ Often association could be identified by the presence or nature of a ‘moniker’ (e.g. ‘Little OG’), previous arrests for gang-related activities, or the presence of gang tattoos that denote affiliation.
• *Previous violent behavior:* Evidence that ward had previously been violent, including a record of assaultive behavior and/or arrests for violent offenses.

• *School dropout:* Evidence the ward had dropped out of school. Evidence for this variable included the ward not being enrolled in school, ward had not attended school for six months while free on the street, ward had been expelled, or the ward was absent without excuse more than they were present.

**Pre-release Data: Arrest History Information**

The arrest history of each ward prior to the sample stay was compiled using the information contained in the Master Files. The Master File contains all of the previous probation reports and court records of wards because that information is required to be submitted with the court order of commitment to the CYA. Using all of the available police reports and records, probation reports, court records, and CYA parole performance summary information (for the wards previously released from the CYA prior to the sample stay), the following variables were coded and checked for accuracy for each arrest event:

• *Date of arrest event:* The date the ward was arrested by law enforcement personnel. In the rare situation (<1% of the time) when that arrest date was not known, the date of the offense was coded.

• *Arrest charges:* Up to 3 arrest charges per arrest event were coded (i.e. some arrest events involved multiple charges against the offender). Only behaviors that reflect distinct law violations were coded as separate charges and only the most serious charge per behavior was coded; ‘lesser-included’ offenses were never coded. For example, if a ward was arrested for evading the police in a stolen car after a robbery, the three arrest charges would reflect each behavior (robbery, auto theft, evading the police). In the cases where a ward is arrested for grand theft auto, occasionally the ‘lesser-included’ offenses of possession of stolen property, and unlawful taking of a motor vehicle were also filed. We only coded the most serious arrest charge per behavior, and thus this arrest would reflect a single arrest charge (grand theft auto). Similarly, if a ward was arrested for attempted murder, as well as assault with a deadly weapon,
and assault and battery, we only coded the attempted murder arrest charge (e.g. you cannot attempt to kill someone with a firearm without committing both assault with a deadly weapon and assault and battery). Allowing multiple arrest charges per arrest date is a more accurate way of cataloguing an individual’s prior record (Geerken 1994).

As noted above, the only difference in coding procedures and rules between the two earlier samples and the most recent sample was in the coding of the criminal history data. In actuality, all distinct arrest charges for a given arrest event were coded for the 1991–2 sample, not just the three most serious charges. This allowed us to ascertain if there were any biases associated with using only the three most serious charges rather than all arrest charges. First, 93 per cent of the arrest events had three or fewer arrest charges; 98 per cent of the arrests only involved four or fewer arrest charges. Second, of the charges that were dropped as a result of our three-charge limit, over 70 per cent involved only charges for being drunk in public, possession of alcohol, giving false information to a police officer, being under the influence of a controlled substance, and other ‘miscellaneous’ minor charges. There was a precipitous drop-off in the seriousness of the arrest charges after the third arrest charge, and in no case did dropping these records result in a ward being misclassified as a non-violent or non-serious offender. Third, among the 1,460 males that were coded in the 1991–2 sample, the mean number of charges was 9.62 using only the three most serious charges, whereas it was only 9.8 if we allowed for the inclusion of all of the arrest charges. Thus, using only the three most serious charges seems to accurately depict the arrest histories of this sample with little possibility of bias.

To make the prior arrest data of the 1991–2 sample equivalent to the two prior samples, we employed the same process used in coding the earlier two samples to arrive at the three most serious charges per arrest event. Although that data were coded manually and then entered into a computer database, we automated the selecting of the three most serious charges using a computer program that pulled out the three most serious charges. First, all offenses were classified according their corresponding OBITs offense category (which ranges from 1 to 100), as was performed in the two prior samples. Then, the computer program pulled out the
three most serious charges according to the seriousness hierarchy that was programmed into an algorithm. Any charges that were ranked fourth or lower according to the algorithm were then dropped. In all of the analyses in this study, only the three most serious charges per arrest event were employed for the 1991–2 sample.

Briefly, the algorithm always considers violent offenses the most serious charges, then serious property offenses (e.g. burglary, auto theft), followed by major drug offenses (e.g. sales and trafficking), and, finally, the least serious miscellaneous charges (e.g. petty theft, drunk in public, trespassing). A table listing the seriousness hierarchy of the offenses is available via the web (see Appendix C).

Post-release Data: Arrest History Information

The source of data for arrests that occur after release from the CYA for the ‘sample stay’ is the Automated Criminal History System maintained by California Bureau of Criminal Statistics and Criminal Identification of the California Department of Justice (CDOJ). This data source was used to obtain the post-release criminal history data because neither OBITS nor the Master Files contain any relevant arrest data subsequent to each ward’s discharge from the Youth Authority (usually after a period of parole). The data from this third source are known as the California Information and Identification ‘CII rap sheet’ information. A list of CII identification numbers for the wards was submitted to the CDOJ, who then compiled a data file containing all of the information in the CII rap sheets (including arrest records) of the wards in our samples.

Whenever an individual is committed to the CYA, the ward is assigned a CII identification number and a computerized (automated) CII rap sheet file is initiated and maintained by the CDOJ. When an adult is arrested in California, the arrest is reported by the arresting law enforcement agency to the California Department of Justice (which houses the state repository for arrest data). If a ward is released while still a minor (under age 18), the CYA reports any criminal arrests of the ward while he or she is a minor to the CDOJ.

The files of the California Department of Justice were searched in late November of 2000. We permitted five months of ‘lag time’ to allow sufficient time for any arrests to be entered by the Department of Justice into the case’s ‘rap sheet’ file. Thus, the arrests were
considered censored on 30 June 2000 and any arrests occurring from that date forward were not included in the analyses for this study. The post-release exposure periods for the samples were between 18–19 years (depending on the date of release) for the 1981–2 sample, 13–14 years for the 1986–7 sample, and 8–9 years for the 1991–2 sample. The average ages at the end of follow-up period were 37, 33, and 27 for each of the release samples, respectively.

To make the post-release data equivalent to the prior arrest data, we included only the three most serious charges per arrest event. We extracted the three most serious charges using the same process described above for the 1991–2 prior arrest data. The following variables were obtained from the CII rap sheet data files:

- *Date of arrest event:* The date the ward was arrested by law enforcement personnel.
- *Arrest charges:* The three most serious arrest charges per arrest event.

**Post-release Data: Mortality Information**

Mortality data on the cases in the release samples were extracted from the Death Statistical Master Files (DSMF) of the California Department of Health and Human Services (DHS). The DSMF files are based on the death certificates completed by either the presiding physician at the time of death, or in the case of sudden or unexpected deaths such as homicide, suicide, or drug overdose, by the coroner or medical examiner investigating the deaths. The DSMF files for 1989–99 were used to obtain the mortality data, and thus the last known possible date of death for our data would be 31 December 1999. Death dates prior to 1 January 1989 for the 1981–2 and 1986–7 samples were obtained from data previously compiled by Skonovd and Haapanen (2000). Mortality data are crucial for the topics of this study, since they allow us to remove an individual from being ‘at risk’ of arrest when he is no longer alive. Thus, any cases that died after release will not be counted as individuals who were ‘arrest-free’ at any age purely as a result of their mortality. As will be shown in Chapter 6 of this study, a sizeable number of cases died (usually of homicide) after release from the CYA.
From the information contained in the DSMF files, we retrieved the following two variables:

- *Date of death*
- *Cause of death*: International Classification of Death (ICD) codes were used to identify the major cause of death (e.g. homicide, suicide, drug overdose, auto accident, and AIDS).

**Post-release Data: Adult Incarceration Information**

Due to the fact that the CII ‘rap sheet’ files only contains accurate reports of the dates of *intake* into the state penal system (California Department of Correction—CDC), we also obtained the adult incarceration records related to all of the stays in the CDC subsequent to release from the CYA for all of our sample members. Information on the following variables was made available to us:

- *Date of intake*
- *Date of release*
- *Commitment offense*: Criminal offense that resulted in the incarceration in the CDC.
- *Second strike*: Whether sentenced as a ‘Second Strike’ case.
- *Third strike*: Whether sentenced as a ‘Third Strike’ case.

**Deriving the Analytical Samples**

In order to be included in the analyses in this study, there were several conditions a case had to satisfy. Table 5.1 details the effects that adhering to the conditions for inclusion in the final analytical sample had on the final sample size.

The first constraint used in deriving the analytical sample was the gender of the ward; the analytical sample was limited to only male wards. This constraint was imposed for two primary reasons. The major reason the female wards were excluded was that there were simply too few females in each of the data sets to allow for separate models or reliable estimation of model parameters. Table 5.1 indicates that females constituted only about 4 per cent of each sample (n = 88, 81, and 87, for the 1981–2, 1986–7, and 1991–2 samples, respectively).  

---

7 Removing the females with missing arrest data further reduced the small sample size. Further, recall that only 1,527 cases were actually coded in the 1991–2 sample. Of these 1,527 cases, only 67 of them (4%) were females.
<table>
<thead>
<tr>
<th>Step 1: Sex</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Original</td>
<td>Starting</td>
<td>Removed</td>
<td>Original</td>
<td>Starting</td>
<td>Removed</td>
<td>Original</td>
<td>Starting</td>
<td>Removed</td>
</tr>
<tr>
<td>Female</td>
<td>88 (4%)</td>
<td>88 (4%)</td>
<td>88</td>
<td>81 (4%)</td>
<td>81 (4%)</td>
<td>81</td>
<td>87 (4%)</td>
<td>67 (4%)</td>
<td>67</td>
</tr>
<tr>
<td>Male</td>
<td>1,998 (96%)</td>
<td>1,998 (96%)</td>
<td>0</td>
<td>1,997 (96%)</td>
<td>1,997 (96%)</td>
<td>0</td>
<td>2,098 (96%)</td>
<td>1,460 (96%)</td>
<td>0</td>
</tr>
<tr>
<td>Total</td>
<td>2,086</td>
<td>2,086</td>
<td>88</td>
<td>2,078</td>
<td>2,078</td>
<td>81</td>
<td>2,185</td>
<td>1,527</td>
<td>67</td>
</tr>
<tr>
<td>Step 2: M Case status</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
<td>0</td>
<td>208 (10%)</td>
<td>203 (10%)</td>
<td>203</td>
<td>220 (10%)</td>
<td>0 (0%)</td>
<td>0</td>
</tr>
<tr>
<td>No</td>
<td>2,086 (100%)</td>
<td>1,998 (100%)</td>
<td>0</td>
<td>1,870 (90%)</td>
<td>1,794 (90%)</td>
<td>0</td>
<td>1,965 (90%)</td>
<td>1,460 (100%)</td>
<td>0</td>
</tr>
<tr>
<td>Total</td>
<td>2,086</td>
<td>1,998</td>
<td>0</td>
<td>2,078</td>
<td>1,997</td>
<td>203</td>
<td>2,185</td>
<td>1,460</td>
<td>0</td>
</tr>
<tr>
<td>Step 3: Missing pre-arrest data</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
<td>0</td>
<td>206 (10%)</td>
<td>151 (8%)</td>
<td>151</td>
<td>—</td>
<td>0 (0%)</td>
<td>0</td>
</tr>
<tr>
<td>No</td>
<td>2,086 (100%)</td>
<td>1,998 (100%)</td>
<td>0</td>
<td>1,872 (90%)</td>
<td>1,643 (92%)</td>
<td>0</td>
<td>—</td>
<td>1,460 (100%)</td>
<td>0</td>
</tr>
<tr>
<td>Total</td>
<td>2,086</td>
<td>1,998</td>
<td>0</td>
<td>2,078</td>
<td>1,794</td>
<td>151</td>
<td>—</td>
<td>1,460</td>
<td>0</td>
</tr>
<tr>
<td>Step 4: Missing post-arrest data</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>8 (0.4%)</td>
<td>7 (0.4%)</td>
<td>7</td>
<td>18 (12%)</td>
<td>200 (12%)</td>
<td>200</td>
<td>71 (3%)</td>
<td>26 (2%)</td>
<td>26</td>
</tr>
<tr>
<td>No</td>
<td>2,078 (99.6%)</td>
<td>1,991 (99.6%)</td>
<td>0</td>
<td>1,830 (88%)</td>
<td>1,443 (88%)</td>
<td>0</td>
<td>2,114 (97%)</td>
<td>1,434 (98%)</td>
<td>0</td>
</tr>
<tr>
<td>Total</td>
<td>2,086</td>
<td>1,998</td>
<td>7</td>
<td>2,078</td>
<td>1,643</td>
<td>200</td>
<td>2,185</td>
<td>1,460</td>
<td>26</td>
</tr>
<tr>
<td>Step 5: Missing death date</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>3 (0.001%)</td>
<td>2 (0.001%)</td>
<td>2</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
<td>0</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
<td>0</td>
</tr>
<tr>
<td>No</td>
<td>2,083 (99.9%)</td>
<td>1,989 (99.9%)</td>
<td>0</td>
<td>2,078 (100%)</td>
<td>1,443 (100%)</td>
<td>0</td>
<td>2,185 (100%)</td>
<td>1,434 (100%)</td>
<td>0</td>
</tr>
<tr>
<td>Total</td>
<td>2,086</td>
<td>1,991</td>
<td>2</td>
<td>2,078</td>
<td>1,443</td>
<td>200</td>
<td>2,185</td>
<td>1,434</td>
<td>26</td>
</tr>
<tr>
<td>Final analytical sample</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1,989</td>
<td></td>
<td></td>
<td>1,443</td>
<td></td>
<td>1,434</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: The numbers in bold indicate the frequency (percentage) of cases that are available at the ‘start’ of the next limiting step.
A second reason why the females were excluded is that there is much empirical evidence that shows that male and female offending patterns are not equivalent (see, e.g., D’Unger, Land, and McCall 2002). Thus, we preferred to exclude the female cases entirely. It was clear during the coding of the 1991–2 sample that there was a marked division between the offending patterns of the males and females, with the offending patterns of the male wards exhibiting significantly more frequent, more serious, and more violent behavior. This is not to say that the females in the original samples were not serious and/or violent offenders, but just that compared to the males in the sample they were not as violent. Comparing all of the males in the samples to all of the females, the mean number of prior criminal arrest charges was 9.80 for the males, and it was 6.8 for the females. Similarly, 90 per cent of the males in these three samples were rearrested in the post-period, whereas 76 per cent of the females were rearrested. However, the offending patterns of the females in these samples compared to typical females in the general population are certainly both much more frequent, more serious, and more violent; this was especially true for the female wards who were gang members. Nonetheless, the female cases were removed at this point, and the sample sizes at the end of this step were 1,998 (1981–2), 1,997 (1986–7), and 1,460 (1991–2), respectively.

The second constraint imposed for inclusion in our study was that only the cases that were ‘regular’ CYA cases (i.e. directly committed to the CYA) would be included in the final analytical samples. Thus, M Cases were not included in the final analytical samples. Recall that M Cases are the California Department of Corrections commitments ‘housed’ in Youth Authority facilities (i.e. juveniles who have been ‘waived’ to the adult criminal justice system or select young adult offenders). Due to the fact that M Cases are not subject to the YOPB control, they do not have extensive clinical summaries and Youthful Offender Parole Board-related records in their Master Files. Since they were not subject to overview by the YOPB, their Master Files are generally missing the reports and documentation that contain the necessary information for many of the variables previously described. Especially problematic was the fact that their previous offending patterns were not described in detail like those of the regular CYA cases. In sum, it was impossible to code these cases with
the same detail and attention that was given to the CYA regular cases.  

As shown in Table 5.1, roughly 10 per cent of the initial 1986–7 and 1991–2 samples were M Cases released from the CYA institutions during the respective fiscal years. As noted above, both budget and time constraints were a problem with the coding of the 1991–2 sample, and as a result the decision was made that the M Cases would only be coded pending available time and budget resources after coding of all regular CYA cases. This decision was made after discussions with CYA Research Bureau personnel with considerable experience with CYA Master Files and after the review of the CYA Master Files of 10 randomly selected M Cases. The review of those 10 cases indicated that it was impossible to code the M Cases with the same accuracy and detail as the regular CYA cases. Thus, none of the 220 M Cases in the original 1991–2 sample was coded. As shown in Table 5.1, after removing the M Cases (that were coded) from the 1986–7 sample that entered this step, the resulting sample size of the 1986–7 sample was now 1,794 (1986–7); the samples for the 1981–2 and 1991–2 samples were unaffected by this step either because there were no M Cases in the CYA population (1981–2) or because the M Cases were not initially coded (1991–2).

The next two constraints required for inclusion in the final analytical sample concerned whether any of the ‘pre-release’ and ‘post-release’ arrest data were missing. Cases were not included if they were missing the prior criminal arrest history. This turned out to be a major problem only for the 1986–7 sample. Of the 1,794 male, regular CYA cases that entered this step for that sample, 151 (8%) were found to be entirely missing their arrest histories. Since we did not code those data, it is beyond our speculative powers to assess why those cases are missing their arrest data. Our analysis of the probability of arrest in the post-release period

---

8 It bears noting that there are not only selection effects that determine which cases end up in the CYA as M Cases, but since all M Cases are subject to transfer to the CDC if they misbehave and threaten institutional security, it is likely that there are even heavier selection effects determining which M Cases are available to be released from the CYA in any given year. Comparatively, for the male regular CYA cases released in 1986–7, 91 per cent of them had been arrested once by the end of the follow-up, whereas 85 per cent of the male M Cases in that sample had been arrested at this time point. For the 1991–2 cases, the corresponding arrest percentages as of 30 June 2000 were 89 per cent (male regular CYA cases) and 80 per cent (male M Cases).
indicated that the cases missing their prior arrest histories were more likely to remain arrest-free in the post-release period than were those cases with available prior arrest data. Eighty-six per cent of the cases missing their prior arrest data were arrested after release, whereas 92 per cent of the cases not missing their data were arrested. Thus, both of these groups experienced what might be called ‘excessive failure rates’, but one group was marginally more excessive than the other.

Cases also were not included in the final analytic samples if we did not have access to their CII ‘rap sheet’ arrest data for the post-release period. There are a variety of reasons why an individual’s CII identification number would not be available at a point in time after their release. Sometimes a ward’s CII Identification Number is not entered into the OBITs system because it is not known or not available, and occasionally records are purged from OBITs (as a result of discharge from the jurisdiction of the CYA) prior to retrieving their CII number. The OBITs system was used to obtain the CII identification numbers that were submitted to the CDOJ, and thus if the CII identification numbers were missing in the OBITs system, the CII rap sheets for those wards could not be obtained. Missing CII identification numbers were only a problem in the 1986–7 sample, with 200 cases (out of the 1,643 that entered this step of the construction of the analytical sample) dropped because we had no access to their post-release arrest information. The number of cases dropped for the 1981–2 and 1991–2 samples was 7 and 26, respectively.

It appears that CII numbers were unavailable for some wards in the 1986–7 sample because their OBITs records had already been purged when the CII identification numbers for this sample were obtained. This occurred when the sample was initially drawn (Skonovd and Haapanen 2000). We compared the prior arrest histories of those missing arrest data and those not missing this information and found that the cases missing their CII numbers (and thus missing their post-release arrest data) had a higher mean number of prior arrest charges. The cases missing (n = 200) their post-release data had an average of 11.79 prior arrest charges for criminal offenses, whereas the group (n = 1443) with known CII numbers averaged 10.22 prior arrest charges. Further, a comparison of parole performance between the two groups indicated a similar finding: 90 per cent of those missing their CII numbers
were given a ‘Dishonorable Discharge’ (and only 1 earned an ‘Honorable Discharge’), whereas 78 per cent of those with valid CII data were ‘Dishonorably Discharged’ (9% of them earned an ‘Honorable Discharge’).

Since it is impossible to reliably impute a longitudinal pattern of arrest charges (over an extended segment of the age distribution), cases missing either of these portions of their criminal arrest histories were excluded from the final analytic samples. Missing arrest data proved not to be a problem for either the earliest (1981–2) or the latest (1991–2) release samples. For the 1986–7 sample, however, we suspect that, on average, there is little bias that results from missing data because an examination of those in the sample with and without missing arrest data were equally distributed among the highest and lowest parole failure risk offenders. Perhaps more importantly, it bears noting that even the lowest risk cases in these samples still have incredibly high failure rates. Complete arrest information was available in both the pre- and post-arrest periods for over 80 per cent of the male CYA regular cases. After removing the cases missing arrest data in either the pre- and/or post-release periods, the resulting sample sizes were 1,991 (1981–2), 1,443 (1986–7), and 1,434 (1991–2), respectively.

Two of the males in the 1981–2 sample were reported to have died, but the dates of their deaths were not recorded. Neither of these individuals had any arrests in the post period; these two cases were dropped, since it was unknown how long they were on the street before their deaths.

After removing the cases that failed one of the five steps in the hierarchical process, we obtained the final analytic samples that are used in the analyses presented herein.

**Data Limitations**

Before concluding this section on the nature of the data used in this study, a discussion of the possible limitations of the data is necessary. The first limitation of this study is that both the arrest and mortality data only use records from the state of California. To the degree the wards migrated outside of California and either died and/or were arrested elsewhere, the data will undercount the extent of these outcomes. We do not feel, however, that this is a fatal limitation within the serious youthful offender population studied here. Most of these wards (over 95%) show up in one of the sources
of data in the post-release period (i.e. they died, were arrested, and/or were incarcerated in California at some point during the post-release period). Nonetheless, it is necessary to keep in mind that these data refer only to records from the state of California and thus may underestimate arrest (and mortality), particularly if they occurred outside California.

A second limitation of this study is that in the analyses that follow, we are unable to control for time spent in secure custody (time incapacitated), or stated differently, we are unable to adjust for differential rates of ‘time on the street.’ Although, as noted above, we did obtain adult incarceration records related to all of the stays in the CDC for all of our sample members, we did not have access to complete secure custody movements over the entire age distribution studied here. More specifically, recording time spent in secure custody during the juvenile years was extremely difficult based on the records available in the ward’s Master File. While wards from certain select counties had detailed records allowing for an easy determination of street time, for wards from other counties it was extremely difficult to figure out when the ward was in and out of secure custody.\textsuperscript{9} Also, we did not have access to subsequent movements in and out of the CYA after the ward was released for the ‘sample stay.’ Finally, during the adult years, we only had access to the state-level CDC movements, but we did not have any access to local-level movements in and out of jail. As such, in the analyses that follow, we are unable to enter a street time ‘offset term’ in the equations we estimate. We will complete, to the best of our ability, a descriptive analysis of the time spent in the CDC in an effort to ascertain whether that incarceration time could be responsible for the results we will report.\textsuperscript{10}

\textsuperscript{9} For the offense the ward was committed to the CYA for committing, the calculation of time spent incarcerated (even at the local level) was not problematic due to the required calculation of time spent incarcerated for the offense.

\textsuperscript{10} The limited set of studies using finite mixture models that compare models with and without controls for time incarcerated indicate to us that the primary consequences of controlling for time incarcerated seems to be its effects on the scale of the dependent variable, the peak absolute age of offending, and determining the percentage of ‘desistors,’ rather than generating completely distinct trajectory shapes (Eggleston, Laub, and Sampson 2004; Laub and Sampson 2003; Piquero \textit{et al.} 2001; Sampson and Laub 2003). Eggleston \textit{et al.}'s (2004) comprehensive results do indicate that some of the trajectory attributes (e.g. peak age, peak rate) can be affected by controls for incarceration time. Nagin’s (2004: 30) discussion of Eggleston \textit{et al.}'s
A third limitation of this study is that the analyses are based entirely on official criminal justice data (i.e. arrest records). There are no self-report data available for these samples. The strength of self-report data is that it may allow for the investigation of hidden criminal activity patterns that do not depend on the offender being apprehended by law enforcement officials. Official arrest data, on the other hand, are entirely dependent on apprehension, and thus offenders who are actively offending but never get arrested do not appear in the official data records as an offender. To the degree that the wards in this sample were committing criminal acts and were not being arrested for them, the analyses here would understated the extent of their criminal activity. There is little doubt that the individuals in our samples committed many crimes for which they were not arrested, but there are three points that counter the argument that this limitation is a serious impediment to our analyses. First, the majority of studies comparing self-report data to official arrest data find that those offenders who report the most frequent and serious offenses are also consistently the most frequent and serious offenders in the official data (see, e.g., Farrington 1989; Hindelang, Hirschi, and Weis 1979, 1981; Huizinga and Elliott 1986). As Farrington (1989: 418) notes, self-report and official data generate ‘comparable and complementary results on such important topics as prevalence, continuity, versatility, and specialization in different types of offenses.’ Second, self-report data are not without criticism, especially with respect to the topics of interest in this study. Recent criticisms of self-report data include issues surrounding both the validity of these data (Piquero, MacIntosh, and Hickman 2002), the reliability of using self-report data to examine within-individual changes in the relationship between age and crime (Lauritsen 1998,
1999), and about how conclusions regarding the nature of the age–crime curve are likely to be ‘compromised’ when using longitudinal self-report data as a result of the selective attrition (when not properly accounted for) of the high-rate offenders (Brame and Piquero 2003).

Third, collecting self-report data that would allow one to examine the issues addressed in this study and to make reliable generalizations concerning the population of serious youthful offenders would be both economically and practically infeasible. Recall that the wards of the CYA represent less than 5 per cent of the known (arrested and processed) delinquents in the entire state of California. If one only wanted to study issues related to the development of criminal offending patterns of these serious offenders as they age, the initial sample size that would be required to encapsulate a considerable number of serious offenders (that are comparable to the CYA wards) in a sample would have to be so large that the research would be economically infeasible. This is especially true when you take into account that in order to reliably record the self-reported offense patterns at any given age, the interviews, beginning in early childhood, would have to be conducted across the entire state of California.

Certainly, some researchers would not agree with this conclusion and hold the belief that serious violent offenders can be accurately studied through self-report data found in samples of the general population. For example, Elliott (1994: 17) has argued that ‘truly serious violent offenders are included and retained in longitudinal general population studies. In fact, persons with arrest histories and incarceration experience are among the most easily tracked, and seldom are lost in longitudinal studies.’ Our experience with the CYA data used in the present study does not lead us to the same

\[\text{11} \quad \text{There no doubt is a qualitative divide between what Elliott (1994) calls serious violent offenders in his sample and the youthful offenders in our CYA samples. To make this point clear, consider the following. Elliott (1994: 18) noted that by age 27 there was a 30 per cent cumulative prevalence rate of serious violent offending (i.e. admitting that they had done something seriously violent such as ‘attacking someone with the idea of seriously hurting them’) in the National Youth Survey data. In the samples used in our study, by the end of the follow-up period, 82 per cent of the cases had been arrested for a serious violent offense (and averaged three serious violent arrest charges such as homicide, aggravated assault, armed robbery, rape, and sodomy); further, an almost astounding 10 per cent of each sample had at least one homicide arrest charge in their records.}\]
conclusion. The serious youthful offenders in our samples are the hardest of all offenders to track for interviewing because they are often literally and figuratively ‘under the gun’ of law enforcement officials, and thus often cannot be found at home. They are not hard to track in the sense that they have high migration rates across state lines, but this does not mean they are easily tracked within their neighborhoods. Let us suppose that you could easily locate them. These individuals both by their nature and legal status would probably not be open to the intrusive questioning by researchers about the patterns of their criminal activity, even after assurances of anonymity. Some of these individuals are the very people that were contained in metal cages in order to be taught by a CYA teacher (wearing a flak jacket). These individuals often have problems maintaining scheduled appointments with their parole officers under the threat of a loss of their freedom (and they are often AWOL for periods of time while on parole). Thus, the idea of scheduling a self-report interview and actually obtaining accurate data (that deals with very sensitive information regarding the offending behavior of individuals often under some form of supervision) seems an unrealistic expectation in our opinion. In our judgement, there simply is no better (actually feasible) way to study this population than through the use of official data.

Methodology
This section briefly describes the analytical methods employed in this study. Readers interested in a thorough and technical discussion of the methods employed and the reasoning behind their selection are invited to turn to Appendix A. Appendix A covers the statistical methods used to model count variables and includes discussions of the Poisson model, the negative binomial model, the parametric random effects negative binomial model, and the semiparametric mixed Poisson model of Nagin and Land (1993). Our discussion in this chapter contains a summary description of the actual analytical strategies we undertake in Chapters 7 and 8.

The Dependent Variable
The dependent variable assessed in Chapters 7 and 8 is the count number of criminal arrest charges at each age, with each individual’s panel data beginning at age 7. This count variable does not
include arrest charges for probation violations (e.g. out of control), parole violations (e.g. positive drug test), or traffic offenses (e.g. speeding, driving without a license). Instead, the dependent variable only counts arrest charges regarding the more ‘garden variety street crime offenses’ (e.g. homicide, robbery, burglary, theft, possession of drugs).

In the statistics literature, data like that used herein are known as unbalanced panel data sets because the cases in the samples have varying numbers of records in the final analytic files. The longitudinal offending sequence for each of the wards included in our samples began at age 7, and the sequence ended with the final age at which the case was known to be ‘at risk’.\footnote{About five cases in each sample experienced their first arrest event at age 6. To keep the absolute size of the data sets to a minimum, those arrests charges were included in the age 7 counts for these cases.} For most of the cases, the final age at risk was determined by the end of the follow-up period, but for some cases the final age was determined by the age at death. Given that the wards were of varying ages at the time of release from the CYA for the ‘sample stay’ (i.e. the stay at the CYA that resulted in the ward’s inclusion in the sample), the maximum age at which each ward’s criminal arrest history was available also varied. Full ages during which the wards were incarcerated in the CYA for the sample stay were removed from consideration of the risk of arrest. Appendix B contains a table with a detailed description of the percentages of each sample that had available criminal arrest histories at each age; here we simply present a brief description of the number of ‘periods’ or ‘age years’ (hereafter referred to as ‘data points’) that were available for analysis in each sample.

For the 1981–2 sample, age 43 was the maximum age at which a respondent’s criminal arrest history was available, and roughly 50 per cent of the sample was observed through the age of 37. The number of data points used in the panel analyses varied from a minimum of 9 to a maximum of 37, and the average number of data points was 30.

For the 1986–7 sample, the oldest age at risk by the end of the follow-up period was 38, and 50 per cent of the sample was available for study after the age of 33. The minimum number of data points available for analysis within this sample was 11, the maximum was 32, and the average number of data points was 26.
The oldest age at which a case’s criminal arrest history was available for study in the 1991–2 sample was 33, and the arrest histories for 50 per cent of the sample were only followed through the age of 27. The minimum number of data points available for analysis in this sample was 10, the maximum was 27, while the average number was 21.

**Analytical Approach Employed in Chapter 7**

As discussed in Chapter 3, the focus of Chapter 7 concerns the relationship between age and crime among latent classes of serious youthful offenders. In Chapter 7, we first estimate the following semiparametric mixed Poisson model using the final analytic sample constructed for each of the three samples.

Models allowing up to eight latent classes will be tested, and the BIC statistic along with empirical testing regarding whether the solutions are global maximums will be used to select the optimal number of latent classes or components in the mixture distribution. Here we will also test the hypothesis of Gottfredson and Hirschi that there is uniformity in the shape of the age–crime distribution by testing the statistical significance of allowing the magnitude of the age parameters to vary across the latent classes. Following the arrival at the optimal number of latent classes, the sample members will then be assigned to the latent class to which they have the highest probability of belonging. At this point, the basic descriptive features of the offending patterns of the latent classes will be discussed (e.g. age at first arrest, mean number of arrest charges), and then the observed and predicted offending trajectories across age will be graphed and compared. Finally, a descriptive analysis will be undertaken to examine what role, if any, adult incarceration time may have had on the decline in criminal offending among these three groups of serious youthful offenders.

In brief, the latent class analysis models developed by Nagin and Land (1993) and Land, McCall, and Nagin (1996) allow researchers to group individuals according to the similarity in their trajectories of arrests (or some other specific event type) over time. Latent class models may be specified to include varying numbers of ‘points of support’, which divide the sample into multiple categories of offenders. In other words, a model with two points of support would divide the sample into two categories of offenders, a model
with four points of support would divide the sample into four categories of offenders, and so on. Since age variables (age and age-squared) are included in the estimation process, the model generates age-specific offending patterns for each of the latent classes over time. The group-specific parameter estimates for the random effects, age, and age-squared terms determine the shape of the arrest trajectories.

**Analytical Approaches Employed in Chapter 8**

In Chapter 8, the relationship between past and subsequent criminal activity will be examined. In that chapter we will employ the use of the multi-method approach of Bushway, Brame, and Paternoster (1999), which is essentially the ‘compare and contrast strategy’ first described by Heckman and Singer (1984). More specifically, we test the robustness of any observed effect of past and subsequent criminal behavior by employing several different methods of analysis. The relationship between past and subsequent criminal activity will be investigated in four stages. These stages involve the use of the semiparametric random effects model, the parametric random effects model, and the negative binomial model. We present more details on the precise specification used when we introduce each of those four stages in Chapter 8.

Before moving on to the main results of this study to be presented in Chapters 7 and 8, in Chapter 6 we will present a descriptive summary of the data used in this study. This description is important for documenting the basic facts concerning the criminal offending patterns of the sample members, including the nature of their criminal offending behaviors (i.e. the types of offenses for which they were arrested, age at first arrest), certain behavioral characteristics of the subjects (e.g. gang membership, drug abuse), and specifics regarding incidents of mortality and adult incarceration. For background information purposes, we also present a description of the trends in the crime, arrest, and incarceration rates in California from 1960 to 2000.