An Analysis of the Offenders Index

Sources of Data

In this book we are concerned with criminal careers. In general we will consider only the documented record of a criminal career as seen in formal convictions in an offender’s criminal record. The most complete criminal records in England and Wales are held in the Criminal Record Office (CRO) in New Scotland Yard and on the Police National Computer (PNC). However these records are maintained for the police and other agencies in the criminal justice system and are only rarely available for research purposes. The Home Office and Ministry of Justice distribute a ‘cut-down’ version of the PNC to researchers that excludes vital variables such as co-offenders. It is often unclear, in the distributed version, whether a person found in the PNC search is the same person who was submitted. The uncertainty can only be resolved if the individuals are interviewed.

Also as an operational database the PNC is subject to weeding and periodic reconstruction, and consequently the early cohort samples are incomplete. Offenders who have not offended since May 1995 (when the microfiche collection was discontinued) are not included in the current PNC database unless their offences were very serious. Thus, it is impossible to use the PNC retrospectively for valid criminal career research, although it can be used validly in prospective longitudinal surveys with repeated searches of criminal records over a 40-year period, such as the Cambridge Study in Delinquent Development (Farrington et al 2006).

As an alternative to the PNC, the Research, Development, and Statistics Directorate (RDS) of the Home Office maintained a database of all ‘standard list’ convictions in England and Wales.

1 The Offenders Index and the offences included in the Standard List are described in The Offenders Index: A User’s guide and The Offenders Index: Codebook (Home Office 1998a, 1998b).
This database, the Offenders Index (OI), was created in 1963 and is based on records obtained from courts in England and Wales of each court appearance resulting in a conviction for one or more ‘standard list offences’. The ‘standard list’ includes all offences which may be tried at the Crown Court (so called ‘indictable’ and ‘either-way’ offences) as well as the more serious summary offences which can only be tried in the magistrates’ courts. The definition of standard list has changed during the period covered by the OI, offences being added or more rarely removed from the list, but our analyses are based on the definition used in the early 1990s.

The records of the different convictions of each offender obtained from the courts must be matched to form the OI criminal career record. This is done by a combination of automatic and manual methods. The details on each offender include name, date of birth, gender, and date of conviction. The date of the offence itself is not recorded but there is clearly a relationship between the dates of offence and conviction. Offence classification, sentence and disposal for each conviction are also recorded on the database. The OI was created to facilitate the study of criminal justice interventions and has been the source of data for many statistical and research studies conducted by the Home Office and others over many years. While the CRO and PNC have changed a lot over time (eg from paper to microfiche in 1979, from microfiche to computer in 1995), and many earlier conviction records have been deleted (weeded out) from them, the OI has never changed and is complete from its inception in 1963 to its demise in 2006. The size and completeness of the OI data set allows the extraction of subsets of data conditioned on any of the recorded information.

In extracting the subsets considered in this book, considerable pains (rigorous manual matching) were taken to ensure that all convictions relating to each individual were collated. Although this process can never be perfect, every effort had been made to ensure that there were relatively few cases where the history was incomplete or where two individuals had been erroneously linked together. All but one of the birth cohort samples used in our analysis were extracted from the Offenders Index, in 1992 or 1993, prior to its redevelopment in 1997. As part of this redevelopment the matching system changed to an automatic computerized system. Updated cohort samples were extracted from the OI in 1999 using the new matching rules. The cohorts were updated and Prime, White,
Liriano, and Patel (2001) revised the estimates of criminal career parameters reporting reduced criminality (the proportion of males convicted up to age 46) and increased recidivism after each conviction number when compared with previous estimates. These inconsistencies were reported by Prime et al as due to improvements in the data. However, we carried out analyses on the original and updated samples which indicate that the new matching rules may have introduced serious errors into the OI. For example, before the changes the measured recidivism probability of offenders with uncommon names was essentially identical to that of offenders with common names. With the new matching procedures the recidivism of offenders with uncommon names has scarcely changed whereas that for offenders with common names has increased. This suggests that the new matching rules are ‘overmatching’, that is combining the records of offenders with similar names and dates of birth. In order to extend the follow-up period of the 1953 cohort we disentangled the over matching by collating convictions prior to 1992 between the original and updated samples. The disentangled 1953+ sample provided results consistent with the original sample without the differential recidivism estimates for common and uncommon names. Despite having voiced our concerns at the time and judging from the male criminality estimate (33.2 per cent; much the same as Prime et al, 2001) for the 1953 cohort up to age 52 (Ministry of Justice 2010, p 8), these problems have not been resolved. The use of the OI as an operational research tool has subsequently been phased out in favour of the incomplete and unsatisfactory cut-down version of the PNC database. The cohort samples used in our analyses are available to researchers via the ESRC Data Archive (SN 3935).

Of particular use in the context of modelling the criminal process are the cohort samples drawn from the OI. The cohort samples consist of all records on the database with a date of birth included in one of four sample weeks (spread throughout the year) for each of the cohort years: 1953, 1958, 1963, 1968, and 1973. Because the OI began in 1963, and because the minimum age of conviction is 10, the first birth cohort that could be followed up were children born in 1953. In addition to the automated matching procedures, manual matching of court appearances (using the old matching rules) has also been carried out on these cohort samples to give the maximum possible assurance that they are complete records of unique individuals. In general the ideas to be discussed here were
based on analyses of the 1953\(^2\) and 1958 cohorts, which provide the longest follow-up periods and hence capture the most complete criminal careers. The ideas were then tested on the later cohorts, taking at least partial account of the censoring effects.

For the purpose of the analyses, the event considered is conviction for the most serious offence at each court appearance, the ‘principal conviction’. There are between one and 25 convictions (different offences) per court appearance with an average of about 1.5. The distribution of convictions per court appearance is highly skewed towards the low values with a little over 70 per cent of court appearances resulting in only one and a little less than 20 per cent resulting in just two convictions. A small proportion of offenders are convicted of a disproportionately large number of offences and it might be reasonably assumed that the allocation of crimes to criminals is similarly skewed. To avoid confusion, throughout this book the words ‘conviction’ and ‘appearance’ will be used interchangeably to mean ‘a court appearance resulting in conviction for one or more offences’. Thus, one conviction in this book means one occasion of conviction at one court appearance.

### Recidivism

We begin with an analysis of recidivism. Figure 2.1 indicates a typical graph of the proportion reconvicted versus the number of previous convictions. The data happens to come from the 1953 cohort but similar graphs are obtained from any data source containing the required information. The main feature of the graph is clear: the recidivism probability starts at about 40 per cent after the first conviction and increases with each subsequent conviction to around 84 per cent after six or seven convictions. A commonsense hypothesis would be that the risk of further offending increases with each conviction. It is important to note that for the 1953 cohort we are effectively seeing the lifetime recidivism probabilities rather than the more usual ‘two year’ reconviction rate (the proportion reconvicted within two years).

There is a more useful way of looking at this data by following procedures first outlined in Grove, MacLeod, and Godfrey (1998)\(^2\) In order to extend the follow-up period of the 1953 cohort to the end of 1999 the authors disentangled the overmatching by collating court appearance data to identify where erroneous mergers had taken place.
Recidivism and MacLeod (2003). This is a kind of ‘survival’ curve. Assuming that we had in fact a lifetime recidivism probability of 80 per cent from offence to offence, then starting with the number of offenders with at least one conviction in their lifetime we can plot the number surviving (continue to offend and be reconvicted) after each consecutive conviction. If we assume a cohort size of 100 offenders with recidivism probability ‘\( p \)’ then the number of offenders surviving to the \( n \)th conviction ‘\( y(n) \)’ is given by:

\[
y(n) = 100 \times p^{(n-1)}
\] (2.1)

With the recidivism probability equal to 80 per cent after each conviction (\( p = 0.8 \)) the graph would look something like Figure 2.2a.

A logarithmic transformation (base 10) of the \( y \) axis results in a straight line graph if there is a constant probability of recovation. This is shown in Figure 2.2b. The slope of the line is directly related to the recidivism probability and is simply \( \log(p) \): the steeper the line the lower the recidivism probability. Because \( p \) is a probability and therefore less than one the slope is negative. Figure 2.3 shows what we get if we plot the actual data for the 1953 cohort on such a graph. The ‘+’ symbols represent the number of individuals in the

![Figure 2.1 Proportion reconvicted for given previous conviction count](image)

Source: 1953 cohort, Offenders Index.
Note: The error bars show ±1 standard deviation about the data points.
cohort with at least $n$ court appearances (convictions). The solid line is the ‘best fit’ to the data for ‘$n > 6$’, given by the equation:

$$y(n) = 2786 \times 0.84^{(n-1)}$$ (2.2)

This equation is of the form of Equation 2.1, suggesting that for the higher appearance numbers the probability of reconviction is constant (84 per cent) as illustrated in Figure 2.1.

Let us now project the line, $y(n)$, back to appearance numbers less than or equal to 6 and in so doing assume that high recidivism offenders form a homogeneous category with a constant recidivism

---

**Figure 2.2a** Hypothetical number of offenders with at least $n$ convictions given 80% recidivism on a linear $y$ scale

**Figure 2.2b** Hypothetical number of offenders with at least $n$ convictions given 80% recidivism on a logarithmic $y$ scale
Recidivism probability from the first conviction onwards. We can now calculate the residuals, by subtracting the value of $y(n)$ for $n \leq 6$ from the corresponding data. If we now plot the result we get the ‘square’ symbols of Figure 2.3. Quite remarkably, these also fall on a straight line given by the equation:

$$y(n) = 2786 \times 0.84^{n-1}$$

This suggests that there is a second category of offenders, in addition to those identified above, with a constant probability of recidivism. The probability of reconviction for this second category is much lower at 0.313 (31 per cent) and the category size is much higher at 8884 individuals. Thus, this simple graphical method shows convincingly that the conviction data can be fitted very well by assuming that there are only two risk categories of offenders with constant but different recidivism probabilities.

This is the first critical point of our analysis. What at first sight looks like evidence that the recidivism probability for individuals increases in a complicated way depending on the number of previous convictions, can also be explained quite simply, by the existence of two categories of offenders, with each category having its own constant recidivism probability.\(^3\)

\(^3\) In this context, constant probability implies that when members of one of the categories are convicted then a constant proportion will go on to sustain at least one more conviction irrespective of the numbers of previous convictions.
The two fitted equations can be combined into a single equation, *the dual risk recidivism model*, with the general form:

\[ Y(n) = A \cdot \left( a \cdot p_1^{\lfloor a^{-1} \rfloor} + (1 - a) \cdot p_2^{\lfloor a^{-1} \rfloor} \right) \]  

(2.4)

Where:
- ‘A’ is the total number of individuals in the cohort with at least one conviction (11642 in the 1953–cohort),
- ‘a’ is the proportion of offenders in the high-risk (of reconviction) category (0.237 in the 1953–cohort),
- ‘p_1’ is the high-risk probability of recidivism (0.84 in the 1953–cohort), and
- ‘p_2’ is the low-risk probability of recidivism (0.313 in the 1953–cohort).

More technically and arguably more precisely than the graphical approach, the values of the three parameters, ‘a’, ‘p_1’, and ‘p_2’, were obtained using a ‘joint iterative maximum likelihood procedure’ (see the Appendix). However this is no more than a sophisticated way of carrying out the graphical analysis described above. Figure 2.4 shows the result of the fit of the model to the 1953–cohort data.

The formal statistical properties of the fit of the model to the data are impressive. The model accounted for over 99.9 per cent of the variance in the data with the correlation coefficient \( R = 0.9994 \); that is, it described almost all the recidivism seen at each conviction.

![Figure 2.4 Dual-risk recidivism model fit to the 1953 cohort data](image)

*Figure 2.4 Dual-risk recidivism model fit to the 1953 cohort data*

Note: The data point at appearance number = 1 corresponds to the total number of convicted offenders in the cohort. The data point at appearance number = 2 is the number of convicted offenders with at least two convictions, etc.
It might be argued that the very high value of $R$ is due the data points not being independent; an individual with $n$ appearances will also have contributed to each of the previous appearance number counts. However, a similar analysis of a sentencing sample,\(^4\) in which the appearance number counts of separate individuals were used, provided similar parameter estimates\(^5\) and an $R$ value of 0.9990. In this sentencing sample all the data points are independent.

Although the 1997 Sentencing sample is essentially cross-sectional, longitudinal information on each of the included offenders is available. Both the appearance number of the current conviction and the time since the previous conviction are known. Indeed, all the longitudinal information on each of the offenders is known back to 1963, the creation date of the Offenders Index, but only current conviction information is used in the cross-sectional analysis. The estimated parameter values for all of the cohort samples and the 1997 Sentencing sample are shown in Table 2.1.

We see the same ‘dual risk’ characteristics in all the cohorts. Very similar graphs are obtained for the 1958, 1963 and 1968, and 1973 cohorts which all have the same shape, but the slopes are progressively steeper as the cohorts become more recent. This is as expected as we are not seeing lifetime reconviction probabilities but only

\(\begin{array}{cccccccc}
53+ \\
53 & cohort & 58 & cohort & 63 & cohort & 68 & cohort & 73 & cohort & 97 & sentencing \\
\hline
a & 0.237 & 0.274 & 0.365 & 0.452 & 0.444 & 0.592 & 0.217 \\
p_{1} & 0.840 & 0.822 & 0.799 & 0.779 & 0.771 & 0.696 & 0.879 \\
p_{2} & 0.313 & 0.276 & 0.238 & 0.183 & 0.196 & 0.068 & 0.276 \\
\end{array}\)

Note: 53+ cohort followed up to 1999. 53, 58, 63 cohorts followed up to 1992 and the 68 and 73 cohorts followed up to 1993.

\(^4\) The sample consisted of individuals sentenced during the first weeks of alternate months from February to December 1997, six weeks in all. Each individual offender was included only once in the sample. The total sample size was 58,916.

\(^5\) Cohort samples and cross-sectional samples are equivalent if the processes generating them are ‘stationary’ (that is the same processes operate over the time period under consideration). The similarity of the estimated parameters suggests that this condition held.
An Analysis of the Offenders Index

those convictions and reconvictions sustained in the time available for each cohort: 24, 19, 15, and 10 years respectively (from age 10 to 1992/3, the extraction dates of the cohort samples). Fitting the dual risk recidivism model to each of the cohorts generates the parameter values shown in Table 2.1 and plotted in Figure 2.5.

The first thing to note is that although they do differ, the measured $p_1$ and $p_2$ vary little from cohort to cohort. In more detail, the parameters $p_1$ and $p_2$ both increase and the proportion of high-risk offenders $a$ decreases as the follow-up period increases from 10 to 36 years and we move from the 1973 birth cohort through to the 1953 (followed up to 1992) and 1953+ (followed up to 1999) cohorts. The 1997 Sentencing sample parameter estimates are broadly consistent with the estimates for the longer follow-up periods 1953 cohorts. This suggests that the reconviction probabilities and the proportions of the population in each of the risk categories had in fact changed very little over the time-span of the cohorts. The 1973 cohort parameters deviate from the trend but we might expect this as the follow-up period is dominated by the teenage years when the prevalence of convictions is changing very rapidly.

To correct for the effect of the ‘censored’ offending lifetimes, we need to understand the rate at which offenders are reconvicted which we will investigate in the next section.

The consistency across cohorts provides a strong indication that all offenders fall into one of our two risk categories, a ‘high-risk’
category and a ‘low-risk’ category, and that each of these categories is homogeneous with respect to the probability of recidivism.

We can use the dual risk recidivism model to calculate the proportion reconvicted for a given number of previous convictions. The proportion is given by:

\[ P(n) = \frac{Y(n+1)}{Y(n)} \quad \text{for } n \geq 1. \]  

(2.5)

Where \( n \) is the number of previous convictions and \( P(n) \) is the proportion of offenders convicted for the \( n \)th time who sustain one or more further convictions. The solid line in Figure 2.6 shows the modelled proportion superimposed on the 1953 cohort data from Figure 2.1. Under the dual risk recidivism model the apparently increasing probability of reconviction is explained by the changing mix of high and low-risk offenders. At the first conviction over 76 per cent of offenders are in the low-risk category and just under 24 per cent are in the high-risk category. The modelled recidivism probability for first offenders is 0.437 compared with 0.405 calculated from the 1953 cohort data. By the second conviction the model predicts that nearly 69 per cent of low-risk offenders will have dropped out (ceased to offend) but only 16 per cent of high-risk offenders will have done so, increasing the modelled recidivism probability at the second conviction to 0.55 compared with 0.61.

**Figure 2.6** The proportion reconvicted by number of convictions

Source: 1953 cohort, Offenders Index.

Note: The error bars show ±1 standard deviation about the data points.
calculated from the data. However by the seventh conviction, less than three in 10,000 low-risk offenders would still be active, giving a combined recidivism probability indistinguishable from the 0.84 of the high-risk category. Above the seventh conviction, all offenders are effectively in the high-risk category (see pp 6–11 for a discussion of similar previous analyses).

In agreement with Blumstein et al (1985), we have shown that the overall reconviction probability changes with the number of previous convictions because the proportions of offenders in our two risk categories change with conviction number and not because the probability of reconviction for any given offender is changing.

Reconviction Rate

Reconviction rates (individual conviction frequencies) can be studied in a similar way to reconviction probabilities. Figure 2.7 shows a graph of data from the 1953+ cohort plotted with a logarithmic scale on the y axis. The graph shows the number of offenders surviving at least the amount of time indicated on the x axis between consecutive convictions. We see that the inter-conviction survival time data falls on a straight line, for times between 7 and 25 years. The equation to that straight line is:

Figure 2.7 Inter-conviction survival time

Source: 1953 cohort, Offenders Index.
Note: The data point at reconviction time = 0 corresponds to the total number of reconvictions sustained by the cohort. The second data point is that total less the number of inter-conviction times less than one year, etc.
Where \( s(t) \) is the number surviving at least \( t \) years between consecutive convictions.

An individual with more than two convictions will have multiple inter-conviction survival times. However, there is no reason to suppose that these multiple measures are not independent samples from the same parent distribution.

On this graph the straight line is characteristic of a Poisson process and indicates that there is a constant rate of reconviction. Here a constant rate means that: the probability of being convicted in a given time period, say one week, is the same whether that time period is now or at some arbitrary time in the future.\(^6\)

For very long survival times the data drops below the fitted line. But, given that we are looking at measurements of the 1953+ cohort, individuals in this cohort would have been convicted from the mid-1960s onwards. By the end of 1999 we might well expect that censoring because of potential convictions beyond age 47, or illness, or death, would become important for time periods of 25 years or more.

For survival times less than seven years the slope of the data is somewhat steeper than the straight line from the equation. However, if we assume that, in the straight line modelled by Equation 2.6, we are now seeing a homogeneous rate category of offenders who have a constant rate of offending, we can, as before, extend this line backward to lower survival times and calculate the residuals by subtracting the line from the data. If we do this we discover that the residuals (square data points on the graph) fall on a second straight line given by:

\[
s_r(t) = 10401 \times e^{-0.847t}
\]  

(2.7)

The simplest explanation of this is that it also indicates a category of offenders who have a constant rate of reconviction, though higher than that of the first category. The equations to these lines can be combined to form a dual rate survival time model of the general form:

\[
S(t) = B \times (b \times e^{-\lambda_1 t} + (1 - b) \times e^{-\lambda_2 t})
\]  

(2.8)

\(^6\) The Poisson process and what we mean precisely by a constant rate of conviction is described in the Appendix.
Where \( S(t) \) is the number surviving at time \( t \) from the previous conviction, \( B \) is the total number of inter-conviction times in the data, \( \lambda_1 \) and \( \lambda_2 \) are the mean numbers of convictions per year for the high-rate and low-rate categories respectively, and \( b \) is the proportion of inter-conviction times attributed to the high-rate category.

As before, the parameters in Equations 2.6 and 2.7 can be more precisely jointly estimated using a ‘least squares iterative procedure’, formalizing the graphical method used above, resulting in a correlation coefficient of \( R = 0.9999 \) between the model and the data, indicating that the model describes almost all the shape of the graph. The fitted function \( S(t) \) is shown as a dotted line in Figure 2.7. The dotted line is coincident with the solid line for survival times greater than six years.

The same dual rate survival time model structure is seen in all the OI cohorts. However, in the later cohorts there are, necessarily, fewer long inter-conviction intervals simply because of the shorter follow-up periods, and the consequent censoring effects are increasingly apparent. Table 2.2 and Figure 2.8 show how the best fit parameter values change with the cohort samples. The data are taken from the 1953 to 1973 cohorts and the updated 1953+ cohort.

Again the most important point to notice is the trend in parameter values as the follow-up period increases, from the 1968 cohort to the 1953 cohort.

As expected the mean conviction rates, \( \lambda_1 \) and \( \lambda_2 \), for the high and low-rate categories respectively, tend to reduce as the follow-up period increases. Also, as the follow-up period increases the proportion \( b \) of high-rate inter-conviction survival times initially

**Table 2.2 Parameter values for the dual rate survival time model by OI cohort**

<table>
<thead>
<tr>
<th>Cohort</th>
<th>( b )</th>
<th>( \lambda_1 )</th>
<th>( \lambda_2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1973 cohort</td>
<td>0.992</td>
<td>1.235</td>
<td>0.163</td>
</tr>
<tr>
<td>1968 cohort</td>
<td>0.679</td>
<td>1.026</td>
<td>0.467</td>
</tr>
<tr>
<td>1963 cohort</td>
<td>0.519</td>
<td>1.035</td>
<td>0.413</td>
</tr>
<tr>
<td>1958 cohort</td>
<td>0.542</td>
<td>0.956</td>
<td>0.315</td>
</tr>
<tr>
<td>1953 cohort</td>
<td>0.531</td>
<td>0.911</td>
<td>0.248</td>
</tr>
<tr>
<td>1953+ cohort</td>
<td>0.565</td>
<td>0.859</td>
<td>0.212</td>
</tr>
</tbody>
</table>

Note: \( b \) = proportion high-rate; \( \lambda_1 \)=high-rate; \( \lambda_2 \)=low-rate (convictions per year)
reduces, from close to 100 per cent in the 1973 cohort to 52 per cent in the 1963 cohort and it then increases slowly to 56.5 per cent in the updated 1953+ cohort. Again the 1973 cohort parameters deviate from the trend, due mainly to the rapidly changing prevalence of convictions during the teenage years and the very short follow-up period for most of the offenders. In particular the low-rate offenders, who have more than one conviction before age 20, would have inter-conviction times similar to the high-rate offenders with the consequent difficulty of separating the categories.

There is, however, a consistent pattern in the trends of both the recidivism probabilities and rate parameter values. What we would expect given the different follow-up periods of the various cohort samples is that, for later cohorts, our measured recidivism probabilities would be lower than for earlier cohorts and the rates of offending would be higher. This is precisely what the slopes of the risk and rate parameter value plots indicate in Figures 2.5 and 2.8 respectively.

We may therefore conclude from these graphs that offenders from each birth cohort can be split into two rate categories, each with a constant rate of conviction, as well as two risk categories, each with constant lifetime probabilities of recidivism. As well as being constant over time for each member of a cohort, the parameters also seem essentially constant from cohort to cohort. The best estimates that we have for the lifetime recidivism probability and
rate parameters are given by the cohort with longest follow-up period, the updated 1953+ cohort, which we will now refer to simply as the 1953 cohort.

The rate analysis above has been conducted using survival curves in which each point represents the number of individuals surviving for at least the time indicated on the $x$ axis. Thus successive points on the curve are not independent, since the number surviving for any given time period have also survived in all times less than that given period. However, the survival curves have the advantage of clarifying the structure of the data by averaging out the expected random variations. The fitted survival equations have a direct relationship with the distribution of independent inter-conviction times; this relationship is given by Equation 2.9:

$$-rac{dS}{dt} = B \left( b \cdot \lambda_1 \cdot e^{\lambda_1 \cdot t} + (1 - b) \cdot \lambda_2 \cdot e^{\lambda_2 \cdot t} \right)$$  \hspace{1cm} (2.9)$$

The curve for Equation 2.9 is plotted in Figure 2.9. The parameter values for $\lambda_1$, $\lambda_2$ and $b$, are those estimated above for the survival Equation 2.8. Inter-conviction time frequency data from the 1953 cohort is also plotted in Figure 2.9. The frequency counts are for inter-conviction times falling in three-monthly intervals from zero to 35 years. The dotted curves are the $\pm 2\sigma$ (two

![Figure 2.9](image_url)

**Figure 2.9** frequency of time to reconviction

Source: 1953 cohort, Offenders Index.
Note: Each data point represents the number of reconvictions occurring in a 3-month time interval.
Reconciling the Risk and Rate Categories

We have identified two categorizations of offenders from the OI cohorts. The obvious question to ask next is: are the high-risk recidivists (where risk = recidivism probability) the same as the high-rate offenders, and are the low-risk recidivists the same as the low-rate offenders? The dual risk recidivism model, Equation 2.4, enables us to calculate the expected number of reconvictions for both the high and low-risk categories in the 1953 cohort (see the Appendix for details). The estimate for the total number of reconvictions is within 2 per cent of the observed value, but the estimates for high and low-risk categories do not correspond with the numbers derived from the high and low-rate elements of the fitted dual rate survival model. There are many more low-rate reconvictions than can be accounted for by the low-risk recidivists, which implies that some of the high-risk recidivists are convicted at the low-rate. The risk and rate categories overlap but are not coincident. Table 2.3 shows the proportion of offenders in the 1953 cohort allocated to each of the composite categories.

In total, 7 per cent of offenders have been allocated to a low-rate/high-risk of recidivism category.

Although the above analysis indicates the existence of homogeneous categories of offenders, with each offender having the particular recidivism and rate characteristics of his or her category,

<table>
<thead>
<tr>
<th>Table 2.3 Allocation of offenders between the categories for the 1953 cohort</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Offenders</strong></td>
</tr>
<tr>
<td>High-rate of conviction</td>
</tr>
<tr>
<td>Low-rate of conviction</td>
</tr>
<tr>
<td>Total</td>
</tr>
</tbody>
</table>
allocating individual offenders to the categories suggested by Table 2.3, purely on their conviction statistics, is problematic. Knowing the number of offences committed and the inter-conviction times for an individual does not permit unequivocal allocation to a specific category. For example some 16 per cent of high-risk offenders will have only one conviction and could be allocated to any one of the categories. Similarly an offender with more than six convictions spread over say 10 to 15 years could be allocated to either of the high-risk categories but would be very unlikely to be a member of the low-risk category. In Chapter 6 we investigate whether the psychological characteristics of an offender can be used to help make the allocation to risk/rate categories. From the above analysis there is no evidence of the existence of a low-risk/high-rate category. It is also the case that subsets of offenders, conditioned on characteristics like gender, custody, or a specific offence type, will retain the structure of the risk and rate distributions but may have different parameter values. The effects of gender are discussed below and the subset of offenders given custodial sentences is analysed in Chapter 4.

In the analysis above we have been concerned with those offenders who are eventually reconvicted rather than those who are not. At the time of a conviction, although within each category the probability of each outcome is known, it is difficult to predict whether a particular individual will recidivate or desist. However, as time progresses, for an individual who has not been reconvicted, the probability that he or she has in fact desisted increases. From the mathematical properties of the survival processes evident in the OI cohort data, we can calculate this probability for any time since the previous conviction. Equation 2.10 uses the recidivism probability and the survival time function to make the calculation.

\[ p_{\text{desisted}}(t) = 1 - p \cdot e^{-\lambda t} \quad (2.10) \]

Where \( t \) is the time since the previous conviction, and \( p \) and \( \lambda \) are the recidivism and rate parameters for the category in question.

**Gender**

Repeating the recidivism analysis of the 1953 cohort data for male and female offenders separately yields parameter values for the
dual risk recidivism model (Equation 2.4), which are given in Table 2.4 and the plots and fitted curves are shown in Figure 2.10.

The first point to note is the difference in the offender cohort size between males and females, \( A \) in Table 2.4. Less than 20 per cent of offenders in the 1953 cohort are female, comprising approximately 9 per cent of the total number of females in the birth cohort. Male offenders, on the other hand, comprise over 37 per cent of males in the birth cohort and 80 per cent of the offenders. This result is not too surprising. In self-reports from the 1998–99 Youth Lifestyle Survey (Flood-Page et al 2000), 57 per cent of males and 37 per cent of females, between the ages of 12 and 30, admitted to having committed at least one of the offences asked about. In the Cambridge Study, 40 per cent of the males (born mostly in 1953) were convicted up to age 50 (Farrington et al 2006).

Table 2.4 Dual risk recidivism model parameter estimates for male and female data from the 1953 cohort

<table>
<thead>
<tr>
<th></th>
<th>( A )</th>
<th>( a )</th>
<th>( p_1 )</th>
<th>( p_2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>9399</td>
<td>0.269</td>
<td>0.84</td>
<td>0.35</td>
</tr>
<tr>
<td>Female</td>
<td>2243</td>
<td>0.087</td>
<td>0.81</td>
<td>0.19</td>
</tr>
</tbody>
</table>

Note: \( A \) = No of offenders, \( a \) = fraction with high recidivism probability, \( p_1 \) and \( p_2 \) = high- and low-recidivism probabilities respectively.

Figure 2.10 Male and female recidivism plots

Source: 1953 cohort, Offenders Index.

Note: The data points represent the number of offenders with at least the number of appearances (resulting in conviction) shown on the x axis.
Of greater significance, perhaps, is the difference in value of the parameter $a$. Fewer than 9 per cent of female offenders, compared with almost 27 per cent of male offenders, fall into the high-risk of recidivism category. Not only are females very much less likely to be criminal but, of those who are, a much smaller proportion are in the high-risk category. Interestingly the recidivism probability of high-risk females is very close to that of their male counterparts, 0.81 and 0.84 respectively. The vast majority of female offenders are in the low-risk category and their probability of recidivism is even lower than that for males, 0.19 and 0.35 respectively. Again for both males and females the goodness of fit of the dual risk recidivism model is extremely high with over 99.9 per cent of variation in the data accounted for.

![Figure 2.11 Male and female reconviction survival time plots and fitted curves](image)

Source: 1953 cohort, Offenders Index.

Note: The data points at Inter-conviction survival time = 0 represents the total number of reconvictions sustained by the offenders in the cohort. The subsequent points represent the number of reconviction times longer than the time indicated on the x axis.
Repeating the inter-conviction survival time analysis of the 1953 cohort data for males and females separately produces parameter estimates for the dual rate survival time model, Equation 2.8, given in Table 2.5, and the plots and fitted curves in Figure 2.11. As expected from the recidivism analysis above, the number of recon- victions sustained by female offenders is very much smaller than the number sustained by male offenders, 1,279 and 16,904 respectively. However, the proportion high-rate \( b \) parameters for males and females are not significantly different from each other. The rate parameters, \( \lambda_1 \) and \( \lambda_2 \), are also similar for males and females but the female offenders who do reoffend would appear to do so slightly more quickly than their male counterparts.

Finally we can divide the male and female offenders into the risk/rate categories identified earlier. Table 2.6 replicates Table 2.3 but with each cell broken down by gender.

### Is Criminality Constant over the Cohorts?

We have seen from the analysis of risk and rate across the cohorts that the estimated parameter values are similar and follow a consistent trend with increasing follow-up period. Longer follow-ups tend to increase both recidivism probabilities and mean survival times. The observed trends are consistent with our expectations of the effects of increasing censorship of the data in the more recent cohort samples. This censorship, however, also creates problems in estimating the lifetime prevalence of conviction for standard list offences in the cohorts. To resolve these difficulties fully we need to be able to explain and model the age–crime curve and in particular the distribution of age at first conviction.

In Chapter 3 we will develop a theory of crime and conviction, based on the risk/rate analysis above, which enables us to fit a model to the age at first conviction data from the 1953+ cohort. In particular the model enables us to estimate the number of offenders surviving to a given age prior to their first conviction. The 36-year follow-up period of this cohort (to 1999) ensures that most offenders (an estimated 97.2 per cent of the individuals who will receive a conviction in their lifetime, based on the age at first conviction model of Chapter 3) will have been convicted by age 46. If we assume that the age at first conviction model, and in particular the parameter values estimated from the 1953 cohort data, are valid.
Table 2.6 Proportions of male and female offenders allocated to the risk/rate categories

<table>
<thead>
<tr>
<th>Offenders</th>
<th>Male</th>
<th>Female</th>
<th>Both</th>
<th>Male</th>
<th>Female</th>
<th>Both</th>
<th>Male</th>
<th>Female</th>
<th>Both</th>
</tr>
</thead>
<tbody>
<tr>
<td>High-risk of recidivism</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>19%</td>
<td>7%</td>
<td>17%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>19%</td>
<td>7%</td>
<td>17%</td>
</tr>
<tr>
<td>Low-risk of recidivism</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>8%</td>
<td>1.5%</td>
<td>7%</td>
<td>73%</td>
<td>91.5%</td>
<td>76%</td>
<td>81%</td>
<td>93%</td>
<td>83%</td>
</tr>
<tr>
<td>Totals</td>
<td>27%</td>
<td>8.5%</td>
<td>24%</td>
<td>73%</td>
<td>91.5%</td>
<td>76%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
</tr>
</tbody>
</table>
for all cohorts, then we can estimate the lifetime prevalence of conviction in all of the cohorts. Table 2.7 shows the estimated proportion of each of the birth cohorts expected to receive at least one standard list conviction in their lifetime, defined as the cohort criminality \( q \).

The criminality estimates are all quite close but spread over a wider range than random variation would suggest. The mean of the estimates is 23 per cent with a 2\( \sigma \) (~95 per cent) confidence interval of ±3.4 per cent, as opposed to the expected random 2\( \sigma \) variation, based on the mean cohort size, of ±0.4 per cent. However in making these estimates we have assumed no errors in the estimation process and that nothing has changed over the 30 year period covered by the cohort data. Inspection of the age–crime curves for the individual cohorts suggests that there have been significant changes in convictions for juveniles over the period, which is particularly noticeable in the 1973 cohort. We will return to these changes and their implications in Chapter 3.

Over the period of the cohorts there have also been significant changes in social conditions, lifestyles, education, and employment, all of which might impact on an individual’s decision to engage in crime. The nature and perception of crime has also changed over the period as have policies to deal with it. Another possible explanation is that the cohort size has an amplification effect on criminality (see Maxim 1986). The cohort size (number born) in 1963 was nearly 25 per cent higher than in 1953 and the criminality increased from 22.5 per cent in the 1953 cohort to 24.6 per cent in the 1963 cohort. Over the next decade the cohort size decreased and by 1973 it was 1.2 per cent less than in 1953, while the criminality estimates reduced to 22.3 per cent in the 1968 cohort and 20.4 per cent in 1973 cohort. It could be that, while the young population is increasing, community resources\(^7\) are put under greater strain, creating a more criminogenic environment.

\(^7\) Health, social work, education, police and employment.
Community resources would increase to cope with the increasing demand but would lag behind until the population trend stabilized or reversed. When the young population is in decline the process would be reversed with less strain on community resources, perhaps leading to greater social cohesion and control.

The above may provide explanations for the small variations in criminality observed across the cohorts but we still require an explanation for the relative stability of criminality over time. Criminality, as measured by the proportion of the population with one or more convictions in their lifetime, has barely changed since the 1960s. Thus, at this stage of the argument, we suggest that criminality is broadly constant over the cohorts.

In summary the results of the risk/rate analysis are:

- The data on lifetime reconviction probabilities suggests that there are two categories of offenders each with a constant risk of reconviction after each conviction.
- The data on inter-conviction times suggests that there are two categories of offenders each with a rate of reconviction (convictions per year) which is constant over time.
- The proportion of offenders in a cohort is essentially constant.
- The risk and rate parameters are essentially constant over the different cohorts.
- There is a strong correlation between the high-rate offenders and the high-risk offenders. However the number of high-risk offenders is significantly greater than the number of high-rate offenders, which suggests the existence of a high-risk/low-rate category of offenders. The characteristics of this category will be explored later.
- There is no evidence for a low-risk/high-rate category.
- The proportions of offenders in each of the risk/rate categories is substantially constant across the cohorts.

Because our theory, described in Chapter 3, is derived from these results it will automatically reproduce them. The interesting question is: what other unrelated or more detailed results can we explain? Conversely, no theory or model that cannot reproduce these results is a candidate to describe the large scale structure of criminal careers. We will also discover in Chapter 5 that the assumptions underlying some commonly held views of criminal career analysts cannot reproduce these results.