DNA RETENTION POLICY: RE-ARREST HAZARD RATE ANALYSIS

Aim

This paper presents a review of the evidence in relation to a policy of DNA record retention for the population in general, and also in relation to possible differential retention periods for serious offenders (those committing crimes of sex or violence) and juveniles.

We are concerned here with the evidence base for retaining the DNA records of people arrested, but not cautioned, or convicted of any crime. We refer to this group as the “no further action” group or “NFA group” for brevity. An additional characteristic of these individuals is that they have no previous conviction – the DNA record of those convicted will be retained indefinitely and hence are not relevant to this analysis.

The paper is not concerned with the retention of either DNA samples or fingerprints.

The benefits of DNA retention

The ACPO Criminal Records Office (ACRO) has undertaken an analysis of all matches to the DNA database in the 2008/09 financial year for rape and murder and manslaughter, and considered 413 rape and 226 murder and manslaughter matches in detail (ACRO, 2009). The exploration revealed that 46 (11 per cent) and 23 (10 per cent) respectively of the rape and homicide matched samples belonged to individuals who did not have a conviction at the time of the match. If these results can be generalised, they suggest that disposing of profiles of people without a proven offence – that is, a zero retention length – might result in around 10 per cent fewer DNA matches for these serious offences. The research therefore demonstrates the potential value in securing detections of retaining DNA profiles of non-convicted arrestees.

The research can shed some light on how much reducing the retention period towards zero might have reduced the number of DNA matches – and hence possible detections – which were made in 2008/09. However, it could only do this for the time period over which DNA matches have so far been possible, and then only in relation to the specific offence types of murder, manslaughter and rape. Additionally, the approach does not provide a way of estimating how many matches might be achieved in future, and hence the expected future value of retaining a given DNA profile. This information is important to assessing how much longer than zero the retention period should be. Therefore, although providing a prima facie justification for a policy, the ACRO research does not provide evidence to inform the length of any general DNA retention period.1

1 This paragraph is different from the one contained in the original version of this paper published on November 11th 2009, and corrects the earlier description of the way the ACRO (2009) research considered the time aspects of retention.
The remainder of this note describes an approach to considering appropriate DNA retention periods, based on relative risks of offending, for general offence types, for more serious offence types, and for juveniles compared with adults. It presents the results of analysis based on this approach, using data obtained from the Police National Computer (PNC).

**The retention of DNA for those arrested but not cautioned or convicted of any crime**

In addressing this issue we take the following broad approach.

We estimate the risk of criminal behaviour by a member of the NFA group, and examine how this risk varies through time. Once the risk of criminal behaviour by the NFA group equals the risk present in the general population, the case for DNA record retention falls away. There is past that point no stronger an argument for retaining the DNA of the NFA group than there is for holding a DNA record for the entire population, as a DNA record would be equally likely to have value in securing a conviction either case.

*The risk of criminality in the “No Further Action” (NFA) group compared with the general population*

(i) The use of proxies for offending behaviour

The risk of criminal behaviour by an individual is realised when an offence is committed. However, not all illegal activities are detected or reported, so conviction is only a partial measure of the risk of criminality. However, for the same reasons, a comprehensive measure of offending is impractical, and a proxy measure is needed.

The best proxy measure of the risk of criminal behaviour by the NFA group is the probability that a group member is arrested in relation to a subsequent offence and receives a *conviction* for that offence. Our judgement of the optimal retention period for a DNA record would then be based on the risk of conviction subsequent to arrest and NFA – an “arrest-to-conviction” rate. However, data constraints described below mean that this “arrest-to-conviction” risk cannot be calculated with any confidence.

The Annex to the Impact Assessment included in the consultation document, *Keeping the right people on the DNA database*, presented analysis based on a “conviction-to-conviction” risk – the chance of an individual being convicted subsequent to their most recent conviction. The analysis shed light on the general pattern of risk in the offender population in comparison to the general population, and provided a general case for the retention of DNA records. As was acknowledged at the time, however, the usefulness of this analysis was limited because an essential characteristic of the group of individuals who are affected by the S and Marper judgement is that they have been arrested but not convicted of an offence.
In this paper we therefore present new analysis based on the “arrest-to-arrest” risk – the chance that an individual, with no previous convictions, who is arrested but NFA, is subsequently arrested again in the years following. Although subsequent arrest is not itself indicative of criminal behaviour, in the absence of the ideal alternative (for reasons described below), we judge it a proxy which is closer to our preferred measure than the previous one based on “conviction-to-conviction”.

(ii) Data

The PNC was used to identify all individuals arrested in April 2006. ACPO guidelines on the retention of criminal records were published in the preceding March (ACPO, 2006). Before this date, PNC arrest data were heavily weeded, and we are unable to say what impact this might have on any analysis of the remaining data. April 2006 is therefore the earliest date available for PNC arrest information suitable for analysis of the DNA record retention period, and maximises the period over which subsequent re-arrest behaviour can be examined.

It would be possible to use the PNC to examine “arrest-to-conviction” behaviour – to examine the risk of conviction following arrest but NFA – but given the time lapse between arrest and conviction, this would shorten the number of years of data available for analysis, and increase the reliance of the analysis on extrapolated data (see below). For instance, statistics on the time from “offence to completion” for cases passing through magistrates’ courts alone suggest a mean duration of over three months, with a significant ‘tail’ extending towards 12 months and beyond (Ministry of Justice, 2009). Time taken in crown courts, where more serious offences are tried, is likely to be even longer. The analysis would also therefore be biased towards those offences for which conviction follows arrest relatively more quickly, which in general are likely to be less serious offences. These limitations were judged sufficient to make an arrest-to-conviction analysis based on currently available data an unreliable basis for assessing appropriate retention periods at this time.

Arrestees were excluded from the April 2006 data extracted for analysis in the following cases.

First, individuals were deselected if they had any convictions in the previous 11 years. Ideally, we need data on arrestees with no previous conviction at all, because no one with a previous conviction is the subject of this particular of the proposals. However, the process which oversees how old, paper-based conviction records are added to the PNC means that the convictions histories of individuals who only have previous convictions older than 11 years do not appear on the PNC and hence could not be excluded from the arrestee sample. Such individuals are assumed to account for a small enough proportion of the sample to avoid biasing the results.²

² We do not have arrest data from the PNC to test this assumption. However, we do have convictions data which can provide an indication of the likely number of individuals with similar conviction histories. For instance, in 2007, the number of individuals receiving a
Second, if further action was taken against the arrestees following an April 2006 arrest, they were also omitted from the sample, since by definition they do not belong in the NFA group.

The application of these constraints left 17,238 unique arrestees. 6,748 of these were subsequently re-arrested between 1st May 2006 and 1st August 2009. For each of these, the PNC generates the date of the first re-arrest. We need to examine the data on first re-arrest because the policy is that each arrest will “re-set the clock” on the length of time a record may be retained on the DNA database to the full retention period. If we used data on all arrests, we would tend to exaggerate the retention period justified by the analysis, as we would be ignoring the fact that more prolific arrestees would already be established in the database on the basis of earlier re-arrests.

(iii) Hazard rate analysis

In assessing the risk of offending behaviour following arrest but NFA there are two key concepts to be borne in mind:

(a) the absolute risk of criminality – the probability that someone in the NFA group commits an offence (using our proxy: is re-arrested) for the first time in the given period;

(b) the relative risk of criminality – the additional probability that someone in the NFA group commits an offence for the first time since the original arrest, compared with the probability that someone in the general population does so.

We focus on (b), the measure of relative risk. The idea is that in each year, or part-year, we calculate the probability of a first-time re-arrest amongst the group of all who were NFA, where the base population for this probability is all who were NFA who have by that point not been re-arrested. In statistical terms this is called the hazard. The hazard rate is the relevant statistic for comparison with the probability of arrest in the general population. Once the hazard rate for first-time re-arrest reaches the general population rate of arrest, the additional or relative risk of criminality is zero and the case for holding DNA past this point expires.

The hazard rate is calculated from the PNC data extract and presented graphically in Annex 1 (Figure A2). If a member of the NFA group is subsequently arrested, the hazard rate is calculated and he or she is removed from the analysis. The sample size consequently decreases continuously.

In order to perform the analysis we need more data points than those provided by PNC data for 2007-2009, subsequent to arrest in April 2006. For this reason we smooth the curve through the available data points and conviction who had had a conviction 10 years earlier, but not in the intervening period, was zero.
extrapolate it forwards, using a general power-curve formulation of exact specification as follows:

Hazard Rate (time in years) = 0.166*time^{0.686}

The extrapolation allows us to analyse data points up to eight years following the initial arrest and NFA (as shown in Figure A2 in Annex 1) and beyond. By this point the line is already flattening out, especially in comparison to its path in the early years. In the first year following the original arrest the probability of re-arrest for the NFA group starts at a relatively high level (just below 35 per cent). It then declines sharply, ending that year at around 15 per cent, and continues to fall smoothly, year by year, as the risk flattens out around the 5 per cent mark.

The methodologies used to construct: the power curve extrapolation of the known hazard rates; the conversion of those rates to annual rates; and the likely confidence limits around the line are all discussed in Annex 1. The data sources used are described at Annex 2.

(iv) The risk of criminality in the general population

As argued above, the choice of retention period for a DNA record should be influenced by the point at which the risk of offending following arrest and NFA (in this analysis proxied by the risk of re-arrest) approaches the risk of offending in the general population. We therefore need to estimate the general population risk of arrest.

In the analysis of “conviction-to-conviction” rates for Keeping the right people on the DNA database, we used a simple comparison between the re-conviction hazard rate and the national population conviction rate. This comparison is open to the criticism that the demographic profiles of the offender group and the general population are quite different. The general population is much older, on average, than the majority of offenders, and offenders are far more likely to be male. A better comparison would allow for such age and sex differences.

We accordingly calculate the risk of arrest for the general population, where that population is weighted to give it the same age and sex characteristics as the NFA group. This risk turns out to be about five per cent. The comparison line is shown in Figure A2 in Annex 1.

In principle, the comparison rate should be calculated for members of the (weighted) general population who have no previous convictions. The question we are addressing is the time period over which the risk of re-offending in the NFA group falls to that exhibited by a member of the general population for whom no DNA record is retained. Unfortunately, data were not available on the proportions by age of the general population with no convictions. This is likely to mean that our estimate of the general population arrest rate is higher than that for the actual population of interest. This issue is considered further under “caveats”.
The exact construction of the general population arrest rate is set out in Annex 3.

(v) Determining a retention period

The two proxy estimates of the risk of offending – for the arrest and NFA group and for the weighted general population – are equalised at the six year point as shown in Figure A2 in Annex 1. This result provides a justification for a retention period up to six years in length, because that is the point at which the risk of offending in the NFA group is judged to have fallen to the risk in the general population. Only when that risk is higher is there a strong evidential case for retaining a DNA record on the database.

(vi) Caveats

This analysis is subject to important caveats.

The six-year point represents the intersection between the NFA hazard curve and the arrest risk comparison line for the weighted general population. Both these lines are measured with some degree of uncertainty.

Although our view is that the hazard rate lies within reasonably tight error margins (see Annex 1), the intersection lies in the zone within which the hazard rate curve is extrapolated, rather than being based on real data. To the extent that the extrapolation is an inaccurate representation of the true hazard rate for the April 2006 NFA group, the intersection will be measured with error. The margins we have calculated to estimate the effect of this error, which are presented in Figure A1 in Annex 1, suggest upper and lower bounds of around seven and five years respectively.

The comparison arrest rate for the general population is also not known precisely but estimated. In particular, our inability to estimate the probability of arrest for the general population who have no previous convictions may affect the exact position of the comparison line. As already argued, to the extent that those with previous convictions are more likely to be arrested than those with no such convictions, this will tend to make our estimate of the general population risk too high. This means that the re-arrest hazard curve would tend to cross a general population risk curve excluding people with previous convictions later than we have estimated, implying a longer retention period.

The extrapolated risk curve for the NFA group is quite flat when it cuts the general population line. This means that the intersection of the two, around the six-year point, is quite sensitive to errors in either the extrapolation or the estimated risk of criminality for the general population. For example, if, as we expect, our estimate of the general population risk of arrest is higher than the true risk, this could have a significant impact on the intersection point. If the general population risk were, for instance, three per cent instead of our current estimate of five, the risks of criminality would equalise after 11 years rather than six.
Finally, as discussed in section (ii), the analysis is based on a proxy measure of offending risk, namely arrests, not convictions. Because we have not been able to observe any conviction outcomes associated with the initial arrest and no further action of individuals with no previous convictions, it is possible that there are individuals in our sample who did not receive any convictions over the period, and who are, hence, innocent. We make an attempt to estimate the proportion of these “false positives” in the next section. However, as stated at the beginning of this paper, the approach we have adopted to analysing possible retention periods is based on a consideration of the relative offending risk between the general population and individuals with no previous convictions who are arrested but not convicted or cautioned. As a result, even if our proxy measure of offending risk is inaccurate, so long as it does not significantly bias the comparison of risks between the two groups, it is still a useful and valid measure. For instance, because a subsequent conviction is contingent on a subsequent arrest, the “arrest-to-arrest” measure will tend to be an overestimate of the true “arrest-to-conviction” risk. However, this overestimation will apply to the general population risk estimate as well as the hazard curve, leaving the broad relativity between them – and any general conclusions we draw in relation to appropriate retention periods – unaffected.

In summary, therefore, there are a number of sources of uncertainty in the current analysis, which means that our basic estimate of six years for the maximum retention length is subject to variation. On balance, it seems likely that the uncertainty points towards a maximum which is longer rather than shorter than six years. However, it should be remembered that this is an estimated maximum which is based solely on a consideration of relative risks. Other considerations could result in an actual policy which is longer or shorter than six years – the evidence we present in this paper is not able to say what this final decision should be.

(vii) Comparison with previous research

As was discussed in *Keeping the right people on the DNA database*, a small number of previous studies have considered the question of relative risks of offending following initial criminal justice system events. Two US studies have considered the relative risk of arrestees *versus* comparable population groups. Kurlychek et al (2006) considered the hazard rates for a cohort from Philadelphia, over a five-year period. It found that the risk of re-arrest immediately following an initial arrest is high but falls off quickly. By the end of the five-year period, there is still a statistically significant difference between the arrestees and the comparable non-arrested group with the same age profile. Nakamura (2008) considered the hazard of re-arrest for a cohort of first-time arrestees in New York. It investigated the effect of being arrested for different crime types on the time for the arrestees’ hazard rate to reach a demographically-adjusted population rate, concluding this was in the range four to 10 years, depending on the offence type and age of first arrest. Again, this was preceded by an initial ‘spike’ in the risk of re-arrest, followed by a rapid fall-off. These findings are in line with the analysis discussed in this paper, therefore.
We can also compare the approach and results presented in this paper with the research presented in *Keeping the right people on the DNA database* by the Jill Dando Institute. This considered the proportion of total criminal activity committed by a group of offenders at different points in time after a criminal justice system event. As was pointed out by a number of consultation respondents, this approach did not take account of the fact that a smaller group of offenders are responsible for a disproportionate volume of offences, and hence would be “captured” by the clock-resetting which occurs as a result of re-arresting people.

The possibility of “false positives” associated with a six-year retention policy

A different approach to exploring the impact of a six-year DNA retention period is to estimate what proportion of the NFA group might be expected to be re-arrested and convicted over the course of that period. Some of the group will not offend during this time. There will have been, in retrospect, no need to have retained these records for the purposes of securing a conviction. These are the “false positives” thrown up by the retention policy. The lower the proportion of false positives, the stronger is the case for a longer retention period.

Because we do not have access to data on arrest with NFA and subsequent conviction, we can only roughly estimate the relationship between arrest and conviction. These estimates suggest that at the end of a six-year retention period following arrest and NFA we would, broadly:

(a) retain 40 per cent of profiles because a proven offence occurred in that period;

(b) retain a further 10 per cent because of re-arrest; and,

(c) delete 50 per cent at the end of the retention period.

In other words, we estimate that some two fifths of individuals whose DNA samples are retained following arrest with NFA will have received a conviction by the end of a six-year retention period. Three fifths of individuals would therefore constitute “false positives”.

This needs to be compared against the number of convictions which might be expected in the general population over the same period of time. As with the hazard rate analysis, this comparison needs to be made on the basis of similar demographic characteristics. We have attempted to estimate this figure using data from the PNC on convictions, information on the demographic profile of arrestees (Home Office, 2009) and national population statistics. On this basis we estimate that around 10 per cent of individuals, with the same demographic profile as the NFA group, would be expected to receive at least one conviction for a notifiable offence over a six-year period. This should only be regarded as a preliminary estimate, but provides some context to the 40 per cent figure we estimate for the NFA group, and reflects
the additional risks of re-arrest (and possible subsequent conviction) as measured by the hazard curve in Annex 1.

Conclusions on the retention of the DNA of those arrested but not convicted or cautioned

Our basic estimate of the maximum retention period for DNA profiles is six years. However, most of the incremental offending risk of the NFA group is dissipated some time before this point, as indicated by the relative flatness of the hazard curve after, for example, three years. This means that our estimate of the maximum retention period is sensitive to errors and variations in assumptions that might cause either the hazard rate curve or the population offending risk curve to change. As described above, we have examined two of the major possible reasons for such changes, which serve to increase and decrease (respectively) our estimate.

The shape of the hazard rate curve indicates how significant offending risk is in the years immediately following an initial arrest. This means that the choice of shorter retention periods is quite robust to the sensitivities introduced by these errors and assumptions. However, to what extent these errors are acceptable, and hence what retention period is actually chosen, is ultimately a matter of judgement.

The retention of DNA of those arrested for more serious offences

It is generally accepted in the academic crime literature that offenders tend to be relative generalists across the course of their careers, rather than specialists.\(^3\) If they do specialise in forms of crime – such as more serious violence – they will tend to do so for short periods, rather than an entire career. This result implies that the evidential case for retaining for a longer period the DNA of someone in the NFA group arrested for a more serious crime is less strong. There is no evidence that that person will have a noticeably higher risk of committing a similarly serious crime than any other NFA individual in that period.

McGloin et al (2009) presented evidence that offenders might favour certain offence types over short periods, largely because of the way opportunities present themselves. Over longer periods, however, because of natural changes to the contexts in which they live their lives, their offending profiles will overall exhibit more versatility. Soothill and Francis (2009) have argued that the justification for longer retention periods for selected offence types should be linked to empirical evidence that those arrestees have significant crime-free periods, for otherwise they will be on the database anyway. There is little (to our knowledge) formal evidence of individuals offending in an intermittent and specialised way, suggesting no strong case for a longer retention period for more serious crimes.

\(^3\) See, for example, Smith (2007) for a review.
Keeping the right people on the DNA database presented two related pieces of evidence, although their impact might have been reduced by a printing error which omitted one of the accompanying charts. These are reproduced in Annex 4. First, an analysis of type of conviction in 2002 showed little association with previous offence types in 2001. Second, an analysis focused on serious and sexual offences showed a slightly raised risk of subsequent conviction for a second sexual offence in the following year, but still an overwhelming probability that any such conviction would be for a different type of offence.

As these analyses run over only two years, they are biased towards more minor offences in each category, but still provide evidence consistent with the view of offenders as generalists over long periods, rather than specialists, and consistent with previous research (e.g. Farrington et al, 1988). The implication is that although a longer retention period for someone arrested, but NFA, for a serious offence would increase the possibility of detecting a subsequent offence, it would probably not raise the chance any further of detecting an offence in the same category. In this sense the evidence does not tend to support a longer retention period for someone arrested for a serious offence than any other person arrested, on the basis of increased risk of offending.

Retention of the DNA of juveniles

The previous analysis has been based on data for all ages to establish the general position on DNA retention. In establishing the evidence base that might underpin a retention policy for juveniles specifically, we take account of the following broad considerations: first, the pattern of offending for juveniles in relation to adults, in particular the age of peak offending; second, the evidence on the reconvictions behaviour of juveniles; and third, the evidence on the offending behaviour of juveniles in relation to sexual and violent crimes.

The age of peak offending

Peak offending, in terms of the largest volume of convictions, occurs in the age range 16-20, with a peak at 18, based on PNC data for 2007. (Other cohorts might generate peaks at marginally different ages, but we would expect these to lie within the age range specified on the basis of the data we have examined.) Previous analysis in terms of the rate of conviction (e.g. Soothill et al (2002)) has found peaks at similar ages, although the pattern can vary across crime types. (For instance, the peak age tends to be higher for drug offending).

These are findings which have been replicated in other countries and at different points in time (Gottfredson and Hirschi, 1990). They reflect a range of factors. For instance, progressively more young people are drawn to offend as they age until, as they mature and move into adulthood, desistance sets it for many. Second, early onset offenders will develop criminal careers that make repeat offending progressively more important in the data on crimes by age.
In general, the concentration of crime volumes at younger ages, at or around the age of majority, implies not only that the decision on retention of DNA records for juveniles merits serious attention, but also that it is not *prima facie* obvious that juvenile retention periods should necessarily be shorter than adult periods, on the basis of the evidence on offending patterns.

*The reconvictions behaviour of juveniles*

Separate arrest-to-arrest hazard rates for juveniles and those aged 18 and over were not estimated from PNC data, because it was judged that this would result in error margins to our simulations which would be too wide to be useful for analytical purposes. We are, however, able to calculate conviction-to-conviction rates by age. Although these do not align directly with the question of the risk of offending following arrest but NFA, they do shed some indicative light on the risk of repeat criminality in cohorts of different ages.

The conviction-to-conviction hazard rates are set out in Annex 5. From these it is clear that the risk of re-conviction is higher for 16-17 year-olds than for both the general population and for adults (18-plus). It is also the case that the re-conviction hazard is even higher for 10-15 year olds.

However, although higher than the general population, these reconviction hazard rates are still low enough to mean that many individuals might not be expected to receive a reconviction. The distribution of convictions within the juvenile age group is relevant to the question of whether there is a case for differential treatment on the grounds that an arrest or conviction might represent an unfortunate, one-off event which “deserves a second chance”. This question can be examined by considering the proportion of the 2001 juvenile first convictions for non-serious offences receiving no further conviction in the following years. This reveals that, five years following their first conviction as a juvenile, 44 per cent of individuals have still not received another conviction, with 40 per cent still in this position after seven years. The full statistics are presented below.

<table>
<thead>
<tr>
<th>Juvenile first convictees</th>
<th>Number/proportion remaining un-reconvicted</th>
</tr>
</thead>
<tbody>
<tr>
<td>2001</td>
<td>87,317</td>
</tr>
<tr>
<td>2002</td>
<td>61,647</td>
</tr>
<tr>
<td>2003</td>
<td>52,661</td>
</tr>
<tr>
<td>2004</td>
<td>46,434</td>
</tr>
<tr>
<td>2005</td>
<td>41,905</td>
</tr>
<tr>
<td>2006</td>
<td>38,747</td>
</tr>
<tr>
<td>2007</td>
<td>36,398</td>
</tr>
<tr>
<td>2008</td>
<td>34,693</td>
</tr>
<tr>
<td>71%</td>
<td>60%</td>
</tr>
<tr>
<td>53%</td>
<td>48%</td>
</tr>
<tr>
<td>44%</td>
<td>42%</td>
</tr>
<tr>
<td>40%</td>
<td></td>
</tr>
</tbody>
</table>

The general implication of this analysis is that there is no case for a shorter retention period for DNA records for juveniles than for adults, and, if anything, there is a case for a longer period. However, there is evidence that a significant proportion of juveniles are not reconvicted of an offence over a significant period after their first conviction for a non-serious offence. This could be seen as justification for differential treatment of these individuals in line with the views on the appropriate treatment of young offenders expressed in the S and Marper judgement and by consultation respondents.
The evidence on the propensity of juveniles to commit different types of crime

In proportionate terms, male offenders are more dominant in offences related to violence than they are in offences as a whole, and in the case of sexual crimes there are scarcely any female offenders. We therefore focus on male offending when considering the relative offending patterns of juveniles and adults for these crimes.

The convictions profile by age for violent crimes is not very different from the offending profile for all crimes. In each case in 2007, the number of convictions peaks at around 18, and the shape of the curves is similar, although the numbers of offences falls away slightly more sharply for violence than for offences overall.

The convictions profile for sexual crimes is, however, quite different. It rises more slowly than the all-age profile to reach a high point, again around 18, but then plateaus at a fairly constant level for around another 30 years, before tailing away slowly.

These graphs are reproduced at Appendix 5. They again suggest no strong evidence in favour of more lenient treatment of juvenile offenders according to the severity of the crime of which they are convicted.

Conclusions on the retention of the DNA of juveniles

The evidence suggests that offending risks and profiles do not provide a strong justification for establishing a shorter retention period for juveniles than for adults.

From age 10, numbers offending climb sharply to a peak at or around 18 for males, and slightly earlier for females. At juvenile ages, not only are volumes of offences higher – for crime types in aggregate as well as for crimes of violence (but not sex) – but the chances of re-offending seem higher as well.

In this sense a case can certainly be made for retaining the DNA records of juveniles arrested but not convicted or cautioned for more serious crimes for as long as for adults, if not longer. A similar case may be made for all crimes.

The benefits of such a retention regime must be weighed against the impact on the young of establishing a definite linkage to the workings of the criminal justice system for a fixed period at an early age, and the desire of the Court of Human Rights for a differential approach towards the young. In this respect, the evidence that a significant proportion of juveniles receive no reconviction over an extended period following their first conviction could be seen as justification for more lenient treatment of this group.

Economics and Resource Analysis Group
Home Office
References


Annex 1 The hazard curve

Fitting a curve

Rate of change calculations are particularly sensitive to blips in observed data and no sample size is sufficient to eliminate these entirely. Fitting a curve to the calculated offence (re-arrest) hazard rates allows us to smooth out observed blips and produce the type of results associated with a far larger sample size.

Including a first-month data point shows that the early part of the curve produces an “S” shape but that this is too short lived to be of use in fitting a full curve. Excluding the first two data points omits the “S” shaped portion of the curve and produces a better fit for the remainder. Fitting a power curve to the annual rate for the remaining 11 data points produces the curve:

\[
\text{Hazard Rate (time in years)} = 0.166 \times \text{time}^{-0.686}
\]

Figure A1 below shows the observed re-arrest hazard values along with the fitted rate used and a fit based on all the observed values. The latter was discarded for reasons explained below.

Figure A1

The discarded fit provides a much better match to the early observed values. This is followed by a run of four observed values that are above the fitted line and the final four observed values are below the fitted line. This kind of
correlation in the error terms suggests that this is not a good fit, despite a $R^2$ value of 97.4 per cent.

Excluding the first two data points produced a much better fit for the remainder of the curve. This is important as the analysis extrapolates beyond the 3½ years of data available and the discarded fit is likely to consistently over-estimate these values. Any analysis requiring data before the 12-month point uses observed data rather than fitted data points.

The observed and fitted values, with and without the first two data points, are as follows:

<table>
<thead>
<tr>
<th>Year</th>
<th>Data Point</th>
<th>Arrests</th>
<th>Sample</th>
<th>Re-arrest Rate</th>
<th>Annual Rate</th>
<th>Fitted Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>First month</td>
<td>574</td>
<td>17238</td>
<td>3.3%</td>
<td>33.4%</td>
<td></td>
</tr>
<tr>
<td>½</td>
<td>3-6m</td>
<td>990</td>
<td>15738</td>
<td>6.3%</td>
<td>22.9%</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>9m-12m</td>
<td>618</td>
<td>13976</td>
<td>4.4%</td>
<td>16.5%</td>
<td>16.6%</td>
</tr>
<tr>
<td>2</td>
<td>22m-24m</td>
<td>334</td>
<td>12030</td>
<td>2.8%</td>
<td>10.7%</td>
<td>10.3%</td>
</tr>
<tr>
<td>3</td>
<td>34m-36m</td>
<td>218</td>
<td>10912</td>
<td>2.0%</td>
<td>7.8%</td>
<td>7.8%</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td>6.4%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td>5.5%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td>4.9%</td>
<td></td>
</tr>
</tbody>
</table>

**Estimating a confidence interval**

A general adopted approach to calculating a confidence interval for unbiased samples is called bootstrapping. This involves building a large number of datasets by randomly sampling from the observed data points, calculating results from each dataset and then using the range of the results to estimate the uncertainty. Using this approach here would require repeating the hazard rate and curve fitting calculations multiple times. This was too computationally complex to be practical for the current work, but the principles behind the bootstrapping approach should still be used.

Splitting the observed data into half in three different ways produces six alternative datasets. Each of these datasets produces a slightly different fitted curve. The upper and lower estimates included in the results are based on the highest and lowest value from any of these six curves. The use of only six datasets is likely to underestimate the true range, but our opinion is that greater uncertainty is introduced by the use of datasets of half the original size, meaning that the final confidence intervals are wider than we would expect from a full bootstrapping exercise, and hence conservative.

**Converting results into annual rates**

The comparison points are based on annual rates, so the hazard rate needs to be converted into annual equivalents.

Scaling quarterly hazard rates up into annual rates requires an adjustment to allow for the fact we are only counting the first re-arrest. This means that only those not arrested in the first quarter can be arrested in the second quarter and so on. The calculation approach first calculates the probability that
someone is not arrested in any quarter and then subtracts it from one to find
the probability that someone is arrested in at least one quarter:

\[ \text{Annual hazard rate} = 1 - (1 - \text{Quarterly hazard rate})^4 \]

The estimated hazard curve

Figure A2 shows the full hazard curve, drawn with the data from the above
table, along with the general population comparison lines. The dashed lines
either side of the hazard rate curve show the upper and lower estimates which
were obtained from the bootstrapping-style analysis previously described.

Figure A2

![Re-arrest hazard rates compared with national arrest rates](image)

Annex 2 Data sources

The hazard rate curve is based solely on data from the Police National
Computer (PNC). The PNC holds details of people, vehicles, crimes and
property that can be electronically accessed by the police and other criminal
justice agencies. It is a national information system maintained and delivered
by the National Police Improvement Agency (NPIA). This analysis would not
have been possible without direct support from the NPIA.

National arrest figures by age and gender are sourced from the Policing
bulletin. The corresponding population data is sourced from the Office for
National Statistics.
The number of unique arrestees per arrest and a full age profile of arrestees are calculated from the 2003-06 Arrestee Survey’s raw data. This was the first nationally representative survey of drugs and crime among the population of individuals arrested in England and Wales. It was conducted by the National Centre for Social Research, on behalf of the Home Office.

Annex 3 The comparator line

The construction of the comparison line involves national data for population, arrests and convictions as well as the Arrestee Survey and can be broken down into the three stages described below.

Estimating age breakdowns for national arrest figures

National arrest figures are available by gender and in four age brackets: 0-9; 10-17; 18-20; and 21 and over. This stage deals with how each of these brackets is broken down into single ages.

The 0-9 age bracket only includes 163 arrests. These are assumed to be nine year-olds as the numbers are too small to merit anything more sophisticated.

The 10-17 year-old age bracket is broken down using 2007 conviction data to assign proportions to each age. This produced the proportions below:

<table>
<thead>
<tr>
<th>Proportion of arrests by age</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>16</th>
<th>17</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>1.3%</td>
<td>3.4%</td>
<td>6.0%</td>
<td>9.5%</td>
<td>14.5%</td>
<td>18.9%</td>
<td>22.3%</td>
<td>24.1%</td>
</tr>
<tr>
<td>Female</td>
<td>0.7%</td>
<td>2.9%</td>
<td>6.7%</td>
<td>12.9%</td>
<td>19.1%</td>
<td>20.9%</td>
<td>19.5%</td>
<td>17.2%</td>
</tr>
</tbody>
</table>

The 18-20 year-old age bracket is broken down using responses to the Arrestee Survey. The survey explicitly excluded anyone under 17 years of age, so it could not be used to break down the younger age groups.

The 21 and over age bracket was based on Arrestee Survey results with the 18-20 year-old bracket providing the ratio of survey responses to national arrests. This approach could have been used for 17 year-olds and produces similar results to the conviction-based breakdown. Starting from the Arrestee Survey figures makes it easier to convert from arrests to unique arrestees.

Calculating the number of unique arrestees

The arrests data we have do not distinguish between individuals, so we need a way of estimating the number of unique arrestees from a sample of arrest events. The Arrestee Survey included information on how many times an interviewee was arrested in the last year. Anyone with no other arrests in the last year counts as one unique arrestee; otherwise the formula below is used:
The 0.5 in the formula represents the arrest that led to the interview. Treating this as half an arrest helps to adjust for the bias caused by surveying at the point of arrest. Averaging this value for an age/sex group provides the number of unique arrestees per arrest. This enables us to convert the number of arrest events into a number of arrestees. This value increases steadily from 0.47 for 17 year-old males to 0.60 for 42 year-old males, indicating that younger males are likely to be arrested more frequently, and is noticeably higher for women than for men.

Data for 9-16 year-olds are not available and the arrestees per arrest figure for 17 year-olds is used instead.

**Arrest Likelihoods**

Dividing the number of unique 18-20 year-old male arrestees per year by the number of 18-20 year-old males in England and Wales gives the likelihood of an 18-20 year-old male being arrested in a year. We repeat this calculation for males and females for each of the age groups for which arrest figures are available, and weight each value by the proportion of arrestees in each grouping. This generates an all-age arrest probability against which our arrest-to-arrest risks can be compared.

**Annex 4 More serious offending**

Analysis of short-term reconvictions of offenders by different offence groupings demonstrates elements of a lack of specialisation in offence types. Figure A3 shows the probability of being convicted in 2002 for a given offence type, given that an offender was originally convicted of an offence type in 2001 and was reconvicted the following year. (Those who did not re-offend are therefore excluded.) It shows that, for most crime types, an offender is considerably more likely to be convicted for a different crime type the next time they are convicted than for the same crime type. The analysis is limited in that, in focusing on two successive years with a conviction received in both, it is necessarily biased towards lower severity crimes (so that an offender is at liberty to commit another offence, rather than being incarcerated) and more prolific offenders. Nevertheless, it serves to illustrate how offenders can be generalists, even in the short run, and that being initially convicted of a minor crime does not imply a significantly lower risk of going on to commit a more serious one.
Figure A3

Probability of those convicted of a crime type in 2001, to re-offend in a given crime type, given they are reconvicted.

Figure A4

Probability of those convicted of a crime type in 2001, to re-offend in a given crime type, given they are reconvicted in 2002.

Figure A4 extends this analysis to focus on serious violence and sexual offences. Although Figure A3 suggests offenders tend to be generalists, Figure A4 indicates that there is a slightly higher risk over a year for those offenders convicted of sexual crimes to be reconvicted of a sexual crime than those offenders who were convicted of any other crime. A similar effect occurs for serious violent offenders, although the increase in risk is much smaller.
This suggests that, although such offenders are generalists, there is an increased risk associated with the more serious crimes over the short term.

Annex 5  Juveniles

The analysis of juveniles considers only the offending behaviour of those people who have been convicted at least once, due to a lack of age information relating to arrests on the PNC. We would expect that, although the risk of offending and the number of offenders would be different from those who have merely been arrested, we would still expect the differences, and the direction of the differences between the age groups, to be maintained, since those people convicted are a subset of those arrested. We will consider three areas of evidence for the offending behaviour of juveniles; peak offending, conviction-to-conviction hazard rates and the age profile for serious crimes.

Peak Offending

Figure A5 shows the number of people convicted of a crime in 2007, broken down by gender and age. This shows that the male peak in terms of numbers of offenders occurs at 18 for this year, although this tends to vary from year to year. We can also see that the peak offending group for females occurs at a lower age of 15.

Hazard Rates

We next consider the hazard rates of offenders convicted for the first time in 2001 and their behaviour over the next seven years to 2008. Figure A6 shows the hazard rates for three different age groups. We can see that the risk of
Reconviction for 16-17 year-old first time offenders is higher than that for adults (18 plus) and the population as a whole. We can also see that the reconviction hazard rate is even higher for 10-15 year olds. This suggests that the hazard rates of this conviction group and the general population are likely to be equalised after a longer, rather than shorter period, suggesting a case for longer retention.

Figure A6

![Reconviction hazard rate curves for first time convictees in 2001](image)

**Serious and sexual crimes**

In proportionate terms, male offenders are more dominant in offences related to violence than they are in offences as a whole, and in the case of sexual crimes there are scarcely any female offenders. We will therefore focus on male offending when considering the offending behaviour in relation to serious and sexual crimes. Figure A7 shows the number of offenders of each age with a conviction for the first time in 2007. This clearly shows that the sexual offences rises slowly and starts levelling out from age 16 with a minor peak at age 18, and then plateaus for about 30 years before tailing off gradually. The profile for violent offences is very different, with a rapid increase to a substantial peak at 17-19 years followed by an equally rapid decline until an age of around 30. This indicates that the risk of sexual offending does not vary significantly with age once age 16 has been reached, and that in terms of violent offending the peak in offending is still to come.
**Figure A7**

Male number of convictions by age for 2007

- **Sexual Offences**
- **Violent Offences**

16 year olds