

Considerations on the UK Re-Arrest Hazard Data Analysis: How Model Selection can alter Conclusions for Policy Development

Houlding, Brett

Trinity College Dublin, Discipline of Statistics

Dublin 2, Ireland

E-mail: brett.houlding@tcd.ie

Wilson, Simon P.

Trinity College Dublin, Discipline of Statistics

Dublin 2, Ireland

E-mail: swilson@tcd.ie

At time of writing, the policy of DNA profile retention for Constabularies within England and Wales is determined by the Association of Chief Police Officer's (ACPO) 2006 Retention Guidelines for Nominal Records on the Police National Computer (PNC), which was developed following the passing of the Criminal and Police Act 2001 and the Criminal Justice Act 2003. The former of these legislations ended the requirement for Constabularies to destroy DNA records relating to persons acquitted or who had their case discontinued, whilst the latter extended powers so as to permit the taking of DNA records without consent from any individual arrested for a recordable offence. These Retention Guidelines detail a governing principle that all records held on the PNC should be maintained until the person in question reaches 100 years of age, regardless of status of conviction, caution, acquittal, or No Further Action (NFA). As such, by 2010 there were over 5 million persons with profiles on the UK National DNA database, with approximately 1 million of these having no record of conviction or caution.

This policy was challenged, first unsuccessfully through the UK judicial system, before successfully being appealed in the European Court of Human Rights (ECtHR) in the case of *S and Marper v. UK* (2008) ECHR 1581, where it was found to violate the data subject's rights under Article 8 of the European Convention of Human Rights:

1. Everyone has the right to respect for his private and family life, his home and his correspondence.
2. There shall be no interference by a public authority with the exercise of this right except such as is in accordance with the law and is necessary in a democratic society in the interests of national security, public safety or the economic well-being of the country, for the prevention of disorder or crime, for the protection of health or morals, or for the protection of the rights and freedoms of others.

As part of the proceedings in the ECtHR, representatives of the UK referred to what was described as impressive statistical reports. In particular, by 30th September 2005 the National DNA database held profiles of approximately 181,000 individuals who had been subject to either a NFA order or acquittal, out of which 8,251 were subsequently linked with crime scene stains involving 13,079 claimed offences. These offences included 109 murders, 55 attempted murders, 116 rapes, 67 sexual offences, 105 aggravated burglaries and 126 offences of the supply of controlled drugs. In response the applicants' representatives argued that such statistics were misleading and that there was a lack of empirical evidence justifying the indefinite retention of DNA profiles from those neither charged nor convicted. They also argued that such statistics did not reveal the extent of any link between being associated with a crime scene sample and any resulting conviction, and no research or figures are available for the number of crimes that are solved where DNA matches included those profiles of individuals never previously convicted.

Despite the arguments put forward by the UK authorities for indefinite retention of DNA profiles, following the European decision the UK Government sought to bring in new legislation so as to conform with its understanding of the requirements made by the European ruling. As a result, the Crime and Securities Act 2010 stated that the maximum duration DNA profiles could be maintained for persons acquitted or subject to a NFA order would be between three and six years; depending on the seriousness of the offence and the age of the person arrested. In the case of an adult who was subject to a NFA order following arrest for any indictable offence, the retention duration is at the upper limit of six years. This six year limit is in line with initial conclusions that may be drawn from the UK Home Office's arrest-to-arrest data analysis, where it is argued that there is an increased risk of re-arrest of a previously arrested person against an all-person comparator for up to approximately 6 years after the initial date of arrest. Nevertheless, the appropriateness of the response of the Crime and Securities Act 2010 remains under dispute, and was opposed by both political parties that subsequently formed the new Coalition Government following the UK general election of 2010. This Coalition Government has so far refused to provide a Commencement Order for the Crime and Securities Act 2010, instead proposing legislation through the Protection of Freedoms Bill. This alternative legislation involves adopting the Scottish system of DNA retention whereby, following acquittal or other discontinuance of a case, DNA profiles are only retained in cases of sexual or violent offences, and even then only for a maximum of three years, though it is disputed by the main opposition party:

“I want finally to turn to DNA, which is another area where we believe that the Government are going too far. My Rt. Hon. Friend ... had already legislated for safeguards on DNA use, including a six-year limit on retention for those who were not convicted. He based those safeguards on analysis of reoffending rates and the benefits in terms of preventing and solving crimes. The Government have decided to reject those safeguards and to go much further in restricting the use of DNA, but not on the basis of evidence.”

Rt. Hon Yvette Cooper, 2nd Reading of the Protection of Freedom Bill, March 1st 2011.

Moreover, concurrent with parliamentary debate, there are also judicial challenges to the existing policy, and a recent application for judicial review resulted in a ruling that the lower Courts were bound to the previous decision of the House of Lords, rather than the ECtHR, and hence the new UK Supreme Court is now reconsidering the decision. Yet, whether or not legislation is indeed further altered, interest remains in the actual risk that is posed by an individual arrested, but where there is neither a caution nor a conviction, and the statistical basis for this.

UK Arrest-to-Arrest Data Analysis

To determine the statistical basis for DNA retention an investigation was conducted by the UK Home Office Economics and Resource Analysis Group [1]. The aim of the investigation was to establish the likelihood or probability of future arrest as a function of time elapsed since initial arrest, as this could then be argued as constituting statistical evidence for retaining DNA records of individuals subject to a NFA order. The data concerned all suitable individuals who were arrested in April 2006, hereafter the ‘NFA group’, and consisted of 17,239 eligible individuals of whom 6,748 were subsequently re-arrested between 1st May 2006 and 1st August 2009 (the end of the investigation).

The analysis concerned relative offending risk, defined as the additional risk that an individual in the NFA group is re-arrested compared to the arrest risk of the general non-arrested population. To do this attention is drawn to the re-arrest hazard rate $h(t)$. Denoting a time interval as Δt and the time of event as T , a hazard rate $h(t)$ for time $t \geq 0$ is defined to be $\lim_{\Delta t \rightarrow 0} P(t < T < t + \Delta t | T > t) / \Delta t$. A hazard rate is thus an event rate at time t conditional on the event not having occurred prior to time t (where here an event is equated with arrest), and as such it measures the ‘tendency’ for

the event to occur. However, whilst this is the formal definition of a hazard rate, it is of little use for practical application when the objective is to estimate the hazard rate from observed event data. Nevertheless, for small values of Δt , the hazard rate can be approximated via the relationship $\Delta th(t) \approx P(t < T < t + \Delta t | T > t)$, with the expression $P(t < T < t + \Delta t | T > t)$ being empirically evaluated from event data.

To estimate the hazard rate the proportion of the NFA group re-arrested who had not previously been re-arrested was calculated for each quarter from time of initial arrest, and Table 1 provides the raw data obtained by the Home Office concerning the quarterly numbers of re-arrests for the NFA group. The quarterly observed hazard of Table 1 is calculated as the number of re-arrests observed within that quarter as a percentage of the sample size of the remaining NFA group at the beginning of that quarter. However, to determine a hazard rate for when time is measured in units of whole years, rather than on a quarterly basis, this quarterly observed hazard was ‘annualised’. This was achieved by assuming the observed hazard for the quarter under consideration remained constant over the next successive three quarters, allowing the annualised hazard rate to be calculated as $100(1 - (1 - \text{Q.Obs.Hzd}/100)^4)$.

Table 1: Quarterly re-arrests and Annualized Hazard for NFA Group

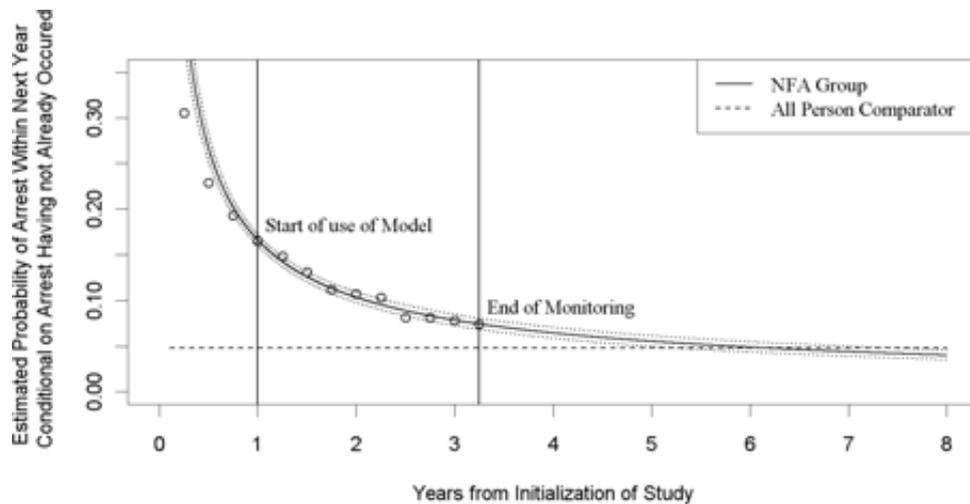
Years	Arrested	Sample	Q. Obs. Hzd.	Ann. Hzd
0.25	1,500	17,239	8.7%	30.5%
0.50	990	15,739	6.3%	22.9%
0.75	772	14,749	5.2%	19.3%
1.00	618	13,977	4.4%	16.5%
1.25	523	13,359	3.9%	14.8%
1.50	441	12,836	3.4%	13.1%
1.75	364	12,395	2.9%	11.2%
2.00	334	12,031	2.8%	10.7%
2.25	313	11,697	2.7%	10.3%
2.50	237	11,384	2.1%	8.1%
2.75	234	11,147	2.1%	8.1%
3.00	218	10,913	2.0%	7.8%
3.25	204	10,695	1.9%	7.4%

For the first year after arrest, the estimated hazard of the NFA group was taken to be the annualized hazard as detailed above. However, beyond this the estimated hazard was calculated by fitting a power-curve regression to the data. This assumed that, as a function of time t , the hazard rate was of the form $h(t) = at^b$, with a and b estimated parameters. To compare this hazard rate with that of the general population an all-person comparator was estimated using data on national arrest rates and survey responses. However, to take into account differences between the demographic profiles of the NFA group and the general population (where on average the NFA group is younger and more likely to be male), the arrest risk of the general population was weighted so as to give the same age and sex characteristics. Additional approximations also had to be taken into account as the available data only listed number of arrests, not numbers of unique individuals arrested, but once such issues were taken into consideration, a crude estimate for a constant hazard rate of 4.9% was suggested as the all-person comparator value, and as such, as the base-line target for declaring equality of risk.

The model for the hazard rate of the NFA group was fitted using data up to 3.25 years beyond April 2006, and in particular, the 11 quarterly annualized estimated hazard rates between 0.75 years and 3.25 years after arrest (the quarterly re-arrest proportions for 0.25 and 0.5 years after initial arrest were incompatible with the power-curve assumption for the hazard rate). This resulted in a

fitted model of $h(t) = 0.166t^{-0.686}$, and Figure 1 plots the estimated hazard rate with 95% confidence intervals. It can be seen that there would appear to be a high risk of initial re-arrest that drops sharply in the first year, followed by a diminishing reduction in re-arrest hazard over the subsequent 5 years until equality with the all-person comparator is found approximately 6 years after initial arrest.

Figure 1: Home Office estimated re-arrest Hazard rate for NFA Group



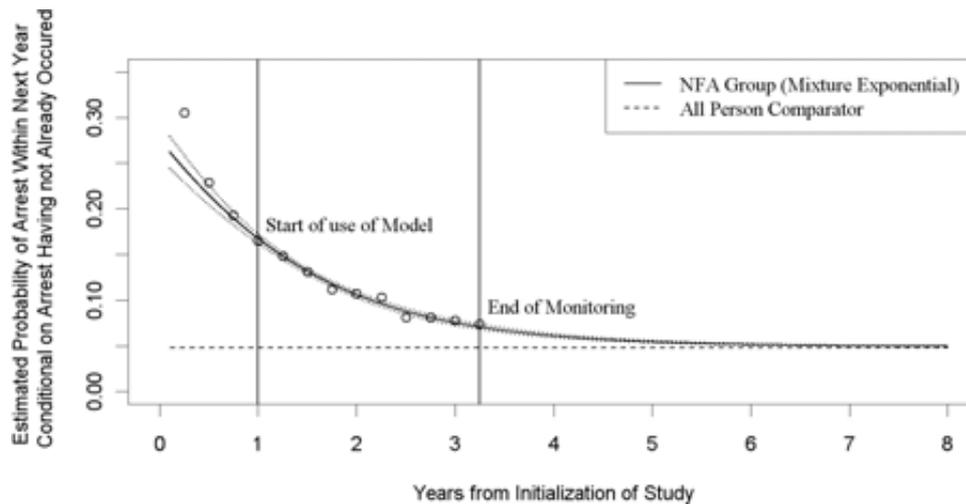
Whilst tentative conclusions indicate heightened arrest risk of the NFA group for up to 6 years following initial arrest, this analysis requires acceptance of the assumptions for the parametric forms of the hazard for the NFA group and the all-person comparator. Whilst for the NFA group a parametric form $h(t) = at^b$ was assumed, the all-person comparator value is constant and independent of t . The former is a generalisation of the latter and would also reduce to a constant form if it was estimated that the parameter b was 0. However, based on the assumption that the parametric model is indeed correct, a statistical hypothesis test rejects the null-hypothesis that $b = 0$.

The form of the hazard rate of the NFA group results from the assumption that time to arrest follows a Weibull distribution, whilst a constant hazard rate results from time to arrest following an Exponential distribution. However, the use of two different modelling assumptions implies that the underlying process leading to arrest differs between the NFA group and the general population, and would suggest that the act of being arrested with NFA fundamentally alters the behaviour of a person. Instead an alternative is to model future arrest time of a member of the NFA group as being Exponential with parameter $\alpha = 0.049$ (that estimated for the all-person comparator) with probability p , or otherwise as being Exponential with parameter $\alpha_2 > \alpha$, *i.e.*, with probability p a member of the NFA group constitutes no greater risk than members of the non-arrested population (the law-abiding people in the NFA group), or would otherwise constitute a greater offending risk, but one that is also independent of time (the people more inclined to criminal activity). For reference we will label these sub-groups as NFA-G1 and NFA-G2, respectively.

Such a 'mixture model' does not generate a constant hazard rate, so could model the re-arrest data. The shape of the curve would follow that of the observed data with a sharp initial drop in the re-arrest hazard of the total NFA group due to re-arrests within the sub-group that constituted a greater offending risk (NFA-G2), followed by a diminishing reduction in re-arrest hazard as the number of individuals within NFA-G2 dwindled, leaving only individuals who are no different to the general un-arrested population (NFA-G1) remaining. Moreover, regardless of how far into the future the model is extrapolated, the hazard for the NFA group will converge asymptotically to the hazard for the all-person comparator. This is in contrast to the Home Office Weibull model, where a reduction in risk

of the NFA group compared to the all-person comparator occurs after 6 years. A fit of such a mixture model that was also based on using only the 11 quarterly re-arrest proportions between 0.75 years and 3.25 years following initial arrest is shown in Figure 2. The fitted mixture model was found to have parameters $p = 0.72$ and $\alpha_2 = 0.87$, meaning that 72% of the NFA group constitute no difference in risk then that assumed for the un-arrested population, with the remaining 28% constituting a nearly 18-fold increase in re-arrest hazard. A comparison of goodness of fit (Aikaike Information Criterion) also suggests that the mixture Exponential has a (slightly) enhanced fit over the Weibull.

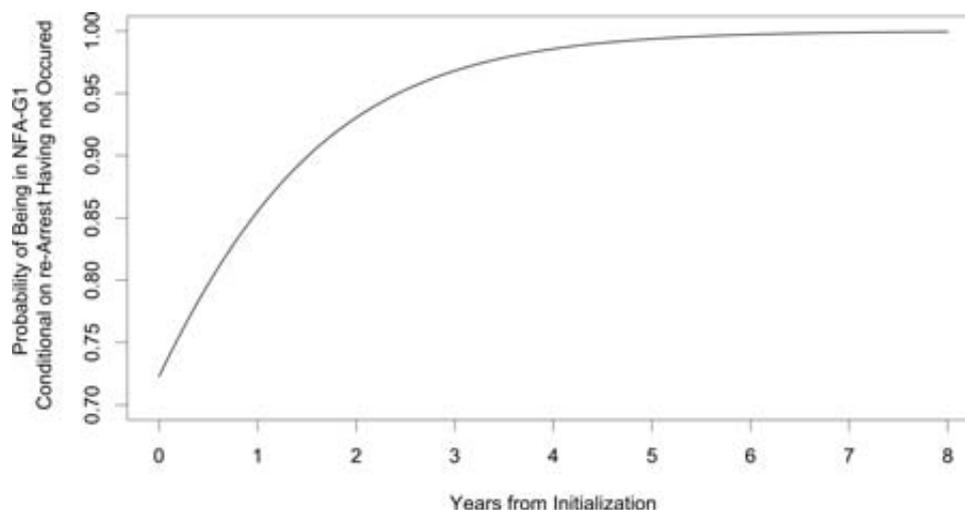
Figure 2: Mixture Exponential model for estimated re-arrest Hazard rate for NFA Group



Unfortunately, applying such a mixture model requires additional information so as to classify a member of the NFA group as belonging to one sub-group or another, for only the fact of historic arrest is recorded. Nevertheless, we can track the probability of group membership over time, and the 72% chance of belonging to type NFA-G1 is only true at the point of NFA decision. Once additional information is available concerning re-arrest not having occurred by a certain time the probability of being of either type NFA-G1 or NFA-G2 will change. Noting that the mixture model assigns one of two possible parameters for the Exponential distribution (depending on sub-group membership) the probability of re-arrest occurring by time t is determined once sub-group membership is known. However, this relationship can be reversed through Bayes' Theorem, allowing calculation of the probability of sub-group membership given that re-arrest has not occurred by time t .

If arrest has not occurred by time t , the probability of membership of sub-group NFA-G1 is found by calculating $P(\text{NFA-G1} | T > t) = P(T > t | \text{NFA-G1})P(\text{NFA-G1}) / P(T > t)$. Using that $\alpha = 0.049$, $p = 0.72$, and $\alpha_2 = 0.87$, we have $P(T > t | \text{NFA-G1}) = e^{-0.049t}$, $P(\text{NFA-G1}) = 0.72$, and $P(T > t) = 0.72e^{-0.049t} + 0.28e^{-0.87t}$ (the mixture Exponential model). Hence we find that $P(\text{NFA-G1} | T > t) = 0.72e^{-0.049t} / (0.72e^{-0.049t} + 0.28e^{-0.87t})$, and Figure 3 plots the probability of belonging to sub-group NFA-G1 (the sub-group with arrest risk equal to the un-arrested population) as a function of time elapsed since NFA decision without re-arrest.

Figure 3 demonstrates that the longer the time without re-arrest, the more likely the person has the same arrest risk as the general population, with the curve asymptotically approaching a probability of 1. In the context of deciding a retention period of DNA profiles, it is of note that an individual has a 95% chance of being a member of NFA-G1 after 2.42 years. After 6 years without re-arrest the individual has a 99.7% chance of being of sub-group NFA-G1, whilst if we wished to know when, on the balance of probability, an individual is more likely to be of sub-group NFA-G1 then of NFA-G2, it should be noted that even at time of NFA decision, the probabilities are 0.72 and 0.28, respectively.

Figure 3: Conditional probability of equality of risk given re-arrest has not occurred

Possibilities for the Future

There are a number of ways in which a further study could be designed so as to remove doubt in its conclusions, and we now list some of these. An initial re-arrest ‘spike’ was problematic for the assumed Weibull distribution, however, this may be due to initial matches on the crime scene database, or possibly because of the commission of administrative offences such as violations of bail conditions. Such occurrences will disturb any fitted parametric model, and it would be useful to know the proportion of early re-arrests that are due to such effects so as to diminish any resulting bias. Furthermore, a key issue with the current analysis is that two different populations have been modelled using two different probabilistic rationales. Though we have highlighted that the observed data for the NFA group could be explained by assuming sub-populations that share the same probabilistic rationale, but with different parameters, it would be appropriate to monitor a ‘control’ population of un-arrested individuals sharing socio-demographic profiles with the NFA group. Indeed, detailed knowledge of the socio-economic profiles of the NFA group would allow investigation into whether there exist sub-populations with an offence rate being a function of covariates drawn from those profiles, *e.g.*, age or gender *etc.*, rather than assuming all members of the NFA group behave in a similar way.

It would appear then, that a difficulty in formulating any statistical evidence base for DNA profile retention of persons subject to a NFA order, is that it should take into account individual circumstances, and incorporate these in a more generalized and formal statistical model. However, to formulate a policy that does take into account such factors as gender or age *etc.*, not only requires a more thorough analysis than that previously performed, but may result in unethical policies where individuals are treated depending on completely innocent aspects of their circumstances, and would likely breach equality laws or further human rights other than that of the right to a private life.

Funding

Brett Houlding is funded by the STATICA project, a Principle Investigator program of Science Foundation Ireland (08/IN.1/I1879).

REFERENCES

[1] Home Office Economics and Resource Analysis Group, *DNA Retention Policy: Re-Arrest Hazard Rate Analysis*, 2009.