

Considerations on the UK re-arrest hazard data analysis

BRETT HOULDING[†]

AND

SIMON P. WILSON

Discipline of Statistics, Trinity College Dublin, Dublin, Ireland

[Received on 14 October 2010; revised on 22 May 2011; accepted on 11 June 2011]

The offence risk posed by individuals who are arrested, but where subsequently no charge or caution is administered, has been used as an argument for justifying the retention of such individuals' DNA and identification profiles. Here we consider the UK Home Office arrest-to-arrest data analysis, and find it to have limited use in indicating risk of future offence. In doing so, we consider the appropriateness of the statistical methodology employed and the implicit assumptions necessary for making such inference concerning the rearrest risk of a further individual. Additionally, we offer an alternative model that would provide an equally accurate fit to the data, but which would appear to have sounder theoretical justification and suggest alternative policy direction. Finally, we consider the implications of using such statistical inference in formulating national policy, and highlight a number of sociological factors that could be taken into account so as to enhance the validity of any future analysis.

Keywords: probabilistic causality; statistical inference; reliability modelling; DNA profiling.

1. Introduction

The issue of profiling a person's deoxyribonucleic acid (DNA) following arrest, but where subsequently no further action (NFA) is taken, divides the opinions of society. Those who are against such a practice argue that it discriminates members of society from others who are similarly entitled to a presumption of innocence, leading to a risk of stigmatization, and that all non-convicted or non-cautioned individuals should share the same right to a private life. On the other hand, those in favour of DNA profiling claim that, other than if a suspect were to commit an offence in the future, the profiling of their DNA would have no direct consequence for them. Furthermore, they cite studies such as the UK arrest-to-arrest data analysis,¹ which are argued as indicating that those persons who are subject to a NFA order are statistically more likely to be rearrested than the population of individuals never previously arrested, and that as they are more likely to be rearrested, they constitute a greater risk of committing, or at least being connected with, a future indictable offence.

Here we do not explore the ethics of DNA profiling *per se*, but instead focus on the statistical evidence used in support of such a practice. In doing so, we offer our own considerations concerning

[†] Email: brett.houlding@tcd.ie

¹ Home Office Economics and Resource Analysis Group (2009) "DNA Retention Policy: Re-Arrest Hazard Rate Analysis" Available at: <http://webarchive.nationalarchives.gov.uk/20100418065544/http://www.homeoffice.gov.uk/documents/cons-2009-dna-database/dna-retention-evidence-paper2835.pdf?view=Binary>

the conclusions that can be formulated following the UK Home Office's arrest-to-arrest data analysis, and review the underlying implicit assumptions that are required for making inference regarding risk of rearrest of a newly arrested individual. In particular, we note that only in limited and seemingly unreasonable circumstances will the statistical analysis provided constitute sufficient statistical evidence of a greater risk of future offence for individuals subject to a NFA decision.

The remainder of this paper is as follows. In Section 2, we detail the background to current UK policy on DNA profiling within England and Wales, and discuss how statistical arguments have contributed to the formulation of this policy and how they have been previously presented. In Section 3, we detail the statistical analysis performed by the UK Home Office Economics and Resource Analysis Group regarding the rearrest hazard rate for individuals who were subject to a NFA order in April 2006. In particular, we review the implicit assumptions that are required in making such inference and offer an alternative analysis that would similarly fit the data, but which would result in different conclusions concerning rearrest risk. Finally, in Section 4, we conclude and offer a critique concerning the use of statistical modelling assumptions when contributing to policy formation and suggest that any attempt to determine actual rearrest risk would require a more appropriate analysis that explicitly takes into account a number of important sociological factors.

2. UK policy of DNA profile retention

At the time of writing, the policy for the processing of DNA and identification details 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,² which were 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 and fingerprint records relating to persons acquitted at court or who otherwise had their case discontinued, while the latter further extended policing powers so as to permit the taking of DNA and fingerprint records without consent from any individual arrested for a recordable offence. 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, caution, reprimand or final warning.³

The Retention Guidelines detail a governing principle that all records held on the Police National Computer should be maintained until the person in question reaches 100 years of age, regardless of status of conviction, caution, acquittal, or NFA. There is, however, an 'exceptional case procedure', but as the guidelines recognize, such cases will be by definition rare. Effectively, this procedure requires the data subject to either prove that their arrest had been unlawful or to establish beyond doubt that no offence existed. This changes the burden of proof concerning the existence of an offence from that of it being the State's responsibility to that of it being the suspect's, for in criminal justice the presumption of innocence or *ei incumbit probatio qui dicit, non qui negat* (the burden of proof rests on who asserts, not on who denies) requires that the State prove both that a crime had been committed, and that it was the suspect who had committed that crime. Hence, the exceptional case procedure requires the arrestee to disprove the commission of any offence, when any conviction

² ACPO (2006) "Retention Guidelines for Nominal Records on the Police National Computer". Available at <http://www.acpo.police.uk/documents/policeCertificates/SubjectAccess/Retention%20of%20Records06.pdf>

³ House of Commons Home Affairs Committee, Eighth Report of Session 2009-2010 "The National DNA Database". Available at <http://www.publications.parliament.uk/pa/cm200910/cmselect/cmhaff/222/222i.pdf>.

would have required the State to prove beyond reasonable doubt that an offence had indeed been committed.

This previous policy was challenged, first unsuccessfully through the UK judicial system,⁴ before successfully being appealed in the European Court of Human Rights (ECtHR),⁵ where it was found to violate the data subject's rights under Article 8 of the European Convention of Human Rights (the right to a private life):

ARTICLE 8

- 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.

The ECtHR ruled that the UK's DNA retention policy violated this fundamental right, and criticized the blanket nature of the guidelines whereby, irrespective of the offences involved or a person's circumstances, DNA profiles were being maintained for what was effectively an indefinite period, a procedure that was considered by the Strasbourg court as being unnecessary in a democratic society.

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 8251 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, which was a view that was also advocated in the conclusions of the Nuffield Report.⁶ The Nuffield Report in particular referred to a lack of empirical evidence justifying the indefinite retention of DNA profiles from those neither charged nor convicted. It 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. Moreover, such statistics did not reveal the use of the crime scene sample in obtaining any conviction that might otherwise not have occurred, and for the majority of the cases mentioned the DNA records were only matched with earlier crime scene stains retained on the database, meaning that such matches could have been made without the continued processing of DNA profiles following either acquittal or a NFA order.

It would appear then, that while such historical statistics do not establish that the course of justice could not have been achieved without the retention of DNA profiles, it is often of benefit to obtain a

⁴ The citations are *S, R (on the application of) v. Chief Constable of South Yorkshire & Anor* (2002) EWHC 478 (Admin), *Marper and Anor, R (on the application of) v. Chief Constable of South Yorkshire & Anor* (2002) EWCA Civ 1275, and *LS, R (on the application of) v. South Yorkshire Police (Consolidated Appeals)* (2004) UKHL 39.

⁵ *S and Marper v. UK* (2008) ECHR 1581.

⁶ Nuffield Council on Bioethics (2007) "The Forensic Use of Bioinformation: Ethical Issues". Available at <http://www.nuffieldbioethics.org/sites/default/files/The%20forensic%20use%20of%20bioinformation%20-%20ethical%20issues.pdf>.

DNA profile at point of arrest and to review that sample against historic crime scene stains. Indeed, there does not appear to be any non-governmental organization that does not accept this procedure as being supportive in the detection of criminal behaviour. What is less clear, however, is whether it is necessary to continue retention beyond such an initial check against the crime scene database.

Despite the arguments previously put forward by the UK authorities for indefinite retention of DNA profiles, following the Strasbourg decision, the UK Government sought to bring in new legislation so as to recognize, and bring policy into line with, its understanding of the requirements made by the European ruling. As a result, the Crime and Securities Act 2010 included an extract (S.14) stating that the maximum duration DNA profiles could be maintained for persons acquitted or subject to a NFA order would be between 3 and 6 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 6 years (though the police would still be permitted to appeal for further extensions on a case by case basis). This upper limit of 6 years is in line with initial conclusions that may be drawn from the UK Home Office's arrest-to-arrest data analysis, where it is claimed that there is an increased risk of rearrest 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 went on to form the new coalition government following the UK general election of 2010. This coalition government has so far refused to provide a Commencement Order for S.14 of the Crime and Securities Act 2010, instead specifying intentions to alter UK legislation through a proposed Protection of Freedoms (Great Repeal) 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 3 years (again with permission for police to appeal on a case by case basis), though it is currently being debated in Parliament having passed its Second Reading stage, as indicated by the following:

I want finally to turn to DNA, which is another area where we believe that the Government are going too far. My right hon. Friend the Member for Kingston upon Hull West and Hessle 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, Shadow Home Secretary), in the 2nd Parliamentary Reading of the Protection of Freedom Bill, 1 March 2011)

Moreover, concurrent with parliamentary debate over legislative change, there has recently been a further judicial challenge to the existing DNA retention policy following the decision made by the ECtHR. A recent application for judicial review⁷ resulted in a ruling that the Administrative Court was bound by the decision of the UK Supreme Court (which had subsequently assumed the judicial functions of the House of Lords), rather than the ECtHR. As a result, the two lead cases

⁷ *GC and C v. Commissioner of Police of the Metropolis* [2010] EWHC 2225 (Admin).

were subject to ‘leapfrog’ appeal and have subsequently been heard in the Supreme Court.⁸ The Supreme Court found that, on the basis of the ECtHR ruling, the previous ruling in the House of Lords could no longer hold nor be authoritative. Furthermore, the Supreme Court found that the existing ACPO guidelines were unlawful, though stopped short of proposing a ‘lawful’ alternative, instead permitting opportunity for ACPO or Parliament to find one that is compliant with the ECtHR ruling. Of particular interest here is the finding of Lady Hale, who noted that:

The Equality and Human Rights Commission argue, in their intervention in this case, that the premise on which such data are kept, that people who are arrested are more likely than the general population to be involved in future offending, is “unsustainable” ... It is not clear that the underlying premise is indeed that people who have been arrested but not charged or convicted are more likely than the general population to commit crimes. . .

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. Certainly the question of necessity in the retention of DNA records remains, and unless or until legislation is altered, it is evident that the implementation of the current policy will continue to be tested in both the UK and European courts.

3. UK arrest-to-arrest data analysis

3.1 Home office rearrest hazard rate analysis

To determine the statistical basis for retention of DNA profiles beyond an initial check against historical crime scene samples, an investigation was conducted by the UK Home Office Economics and Resource Analysis Group. 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, but who also had not been subject to any prior conviction (for in such instances the DNA records would be maintained indefinitely).

The data concerned all suitable individuals who were arrested in April 2006, hereafter the ‘NFA group’, and while the analysis accepts that an individual’s arrest does not equate to their committing an indictable offence (which as discussed above, also does not prove that an indictable offence has even been committed), it acknowledges that a conviction for a recordable offence will necessarily be preceded by an arrest. As such, and despite whether or not it is ethical or even appropriate to do so, the analysis argues that arrest can be considered an indicator or proxy for offending risk.

Alternative measures for offending risk such as arrest-to-conviction rate, or conviction-to-conviction rate were considered, but rejected. In the former instance, this was because of a time lapse arising from the necessity for any conviction to be approved through court procedures, reducing the number of events that could be observed by the end of the data monitoring, i.e. where a first arrest was followed by a subsequent second arrest and successful conviction. This would have significantly

⁸ *R(on the application of GC) (FC) (Appellant) v. The Commissioner of Police of the Metropolis (Respondent)* [2010] UKSC 0173 and *R (on the application of C) (FC) (Appellant) v. Commissioner of Police of the Metropolis (Respondent)* [2010] UKSC 0186.

reduced the amount of available information concerning the NFA group of April 2006 and would have increased the reliance on extrapolating results.

In the case of a conviction-to-conviction rate, such an analysis would not have supported any government response to the ruling in the ECtHR, as that decision related to the retention of DNA records for individuals who do not hold a caution or conviction. Hence, with such an understanding concerning the definition of offending risk, i.e. based on arrest-to-arrest rate, the analysis sought to determine when the risk of offence by the NFA group would equal the estimated risk in the general population, as it could be claimed that at such a point the case for DNA record retention for the NFA group was no stronger than the case for DNA retention of the general population.

The NFA group consisted of 17 239 eligible individuals arrested in April 2006, of which 6748 were subsequently rearrested between 1 May 2006 and 1 August 2009 (the latter being the end point before the Home Office Economics and Resource Analysis Group reported their analysis). The data analysed concerned the date of first rearrest, but as mentioned above, did not take into account whether or not any of these rearrests resulted in any further action. Hence, there are no numbers available detailing how many of the NFA group who were rearrested would have constituted a counterfactual claim from any suggestion they had committed an offence.

The analysis focuses on relative offending risk, which was defined to be the additional risk that an individual in the NFA group is subsequently rearrested for the first time since the original arrest, compared to the risk that an individual in the general public is arrested. To do this, attention is drawn to the rearrest hazard rate, here denoted by $h(t)$. Denoting a time interval as Δt , the hazard rate $h(t)$ is defined to be the limit as Δt tends to 0 of the ratio of the conditional probability that, given rearrest has not occurred by time t , it will occur within the next interval Δt , to the time interval Δt itself. Mathematically this is expressed as $h(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t < T < t + \Delta t | T > t)}{\Delta t}$, where T is the actual time of rearrest. The hazard rate is thus an event rate at time t conditional on the event not having occurred prior to time t (where here the event of interest is arrest), and as such, it measures ‘tendency’ for the event to occur, or in a loose sense, the greater the hazard rate the greater the likelihood of imminent occurrence of the event.

While the above description of a hazard rate is the formal mathematical definition, it is of little use for practical application when the objective is to estimate the hazard rate from observed event data. Instead it can be noted that for small values of Δt , the hazard rate can be approximated via the relationship $\Delta t \times h(t) \approx P(t < T < t + \Delta t | T > t)$, with the term $P(t < T < t + \Delta t | T > t)$ being the conditional probability that, given the event of interest has not occurred by time t , it will occur in the next Δt units of time. Crucially, however, the conditional probability $P(t < T < t + \Delta t | T > t)$ can be empirically evaluated from observed event data, and hence an approximation for the hazard rate may be obtained.

To estimate the hazard rate, the proportion of the NFA group rearrested who had not previously been rearrested was calculated for each year or part year from time of initial arrest, i.e. April 2006. This was done on a monthly basis for the first year and then on a quarterly basis for a further two and a quarter years. Once an individual in the NFA group was rearrested, they were removed from the analysis, leading to the sample size of the NFA group being reduced accordingly. This empirical proportion was then used to represent the probability of rearrest as a function of time lapse from initial arrest. Table 1 provides the raw data obtained by the Home Office concerning the quarterly numbers of rearrests for the NFA group.

The quarterly observed hazard of Table 1 is only indicative for the quarter under consideration and is easily calculated as the number of rearrests observed within a quarter as a percentage of the

TABLE 1 *Home Office data detailing the number of rearrests per quarter for the NFA group following an initial arrest in April 2006*

Years	Arrested	Sample	Q. Obs. Hzd (%)	Ann. Hzd (%)
0.25	1500	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

The columns refer to the time in years from first arrest (years), the number of the NFA group rearrested within that quarter (arrested), the size of the NFA group without rearrest at the beginning of the quarter (sample), the resulting observed hazard on a quarterly scale (Q. Obs. Hzd) and the annualized value of the hazard (Ann. Hzd)

sample size of the NFA group at the beginning of that quarter, i.e. the arrested column of Table 1 as a percentage of the sample column. 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 ‘annualized’. This was achieved by assuming the observed hazard for the quarter under consideration remained constant over the next successive three quarters, allowing the annualized hazard rate to be found by subtracting the fourth power⁹ of the observed success rate from 1 (where the success rate is 1 minus the hazard rate), i.e. the Ann. Hzd of Table 1 is $100(1 - (1 - \text{Q. Obs. Hzd}/100)^4)$.

For example, at the beginning of the fourth quarter from initialization of the study there were 13 977 members of the NFA group who had not been rearrested, while it was also observed that within the fourth quarter 618 members of the NFA group were rearrested. This results in an observed hazard rate of $618/13\,977 = 0.0442$, or approximately 4.4%, and hence the annualized hazard rate is found to be $1 - (1 - 618/13\,977)^4 = 0.1654$, or approximately 16.5%.

For the first year after arrest, the estimated hazard rate of the NFA group was taken to be the annualized version of the proportion of the NFA group who were found to have been rearrested within a given quarter, as explained above. However, beyond this initial year, the estimated hazard rate was calculated by fitting a power-curve regression to the data. This required assuming that as a function of time elapse t from initialization of the study, the hazard rate $h(t)$ was of the parametric form $h(t) = at^b$, where $a \geq 0$ and $b \leq 0$ are unknown parameters estimated from the data.

⁹ This formula results because the probability of arrest not occurring in a given year is equivalent to the product of the probabilities of arrest not occurring in each of the four quarters of that year, each of which is calculated as the observed hazard rate on the quarterly scale, hence requiring the calculation of the fourth power of the observed hazard rate.

In the case of the Home Office rearrest analysis, the parameters a and b were found to be approximately 0.166 and -0.686 , respectively.¹⁰ This means that, e.g. an individual in the NFA group who had not been rearrested for a period of 2 years, would be considered as having a probability of $0.166 \times 2^{-0.686} = 0.103$ of being arrested in the upcoming year, or 10.3% (note that in this instance Table 1 provides the same result), while an individual who had not been rearrested within a 6-year period would have probability 0.049 or 4.9% of arrest in the seventh year from initial arrest.

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 and assumptions 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 rough and crude estimate for a constant hazard rate of 4.9% was suggested as the Home Office's all-person comparator value, and as such, as the baseline target for declaring equality of risk between the NFA group and the general unarrested population (further details on how the value of 4.9% was obtained are provided in Appendix A, though in subsequent arguments we will outline why this approach, and hence the actual value of the all-person comparator, may not be relevant).

The model for the hazard rate of the NFA group was then fitted using data up to 3.25 years beyond April 2006, and in particular, using the 11 quarterly annualized estimated hazard rates concerning the proportion of rearrests in the NFA group between 0.75 and 3.25 years after arrest (the quarterly rearrest proportions for 3 and 6 months after initial arrest were found to be incompatible with the smooth power-curve assumption for the hazard rate). The fitted curve was then extended to estimate the rearrest hazard for the NFA group for a further 4.75 years, resulting in an estimated hazard rate for a total of 8 years following initial arrest.

The results of the analysis are shown in Fig. 1, where it can be observed that there was a high risk of initial rearrest of the NFA group that dropped sharply in the first year (dropping from approximately 33–17%). There was then found to be a diminishing reduction in rearrest hazard over the subsequent 5 years until equality with the all-person comparator is found approximately 6 years after initial arrest (though the actual intersection in the hazard rate between the NFA group and the all-person comparator occurs at a part of the fitted hazard curve that is relatively flat, making the precise intersection point a sensitive conclusion).

Beyond 6 years after initial arrest, the extrapolated hazard curve indicated that the NFA group were of reduced risk in comparison to the general population. No data were available to indicate what proportion of those rearrested within the first year had been rearrested because of a match of their DNA profile with historic crime scene stains or because of administrative offences such as the violation of bail conditions.

3.2 *The assumptions of a parametric hazard rate*

While it would appear that tentative conclusions can be drawn concerning heightened arrest risk of the NFA group for up to approximately 6 years following initial arrest, this form of analysis does

¹⁰ Reproducing the analysis led to 95% confidence intervals of (0.160, 0.172) and $(-0.748, -0.625)$, respectively. The analysis was run using the function `nls` 'nonlinear least squares' in the free statistical software R: R Development Core Team (2009), R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org>.

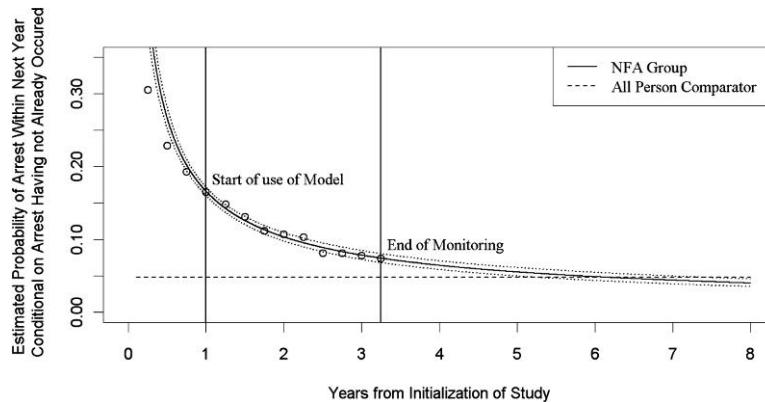


FIG. 1. The Home Office estimated hazard rate of future arrest of members of the NFA group (solid) and the allperson comparator value (dashed). The circle points indicate the observed annualized rearrest rate, while the vertical lines mark the beginning of inferring the hazard rate through a parametric model and the end time of observing the NFA group, respectively. Finally, the dotted lines indicate approximate 95% confidence intervals for the estimated NFA group hazard rate (the process of evaluating these is described in Appendix B).

require agreement with a number of implicit statistical assumptions if it is to be used in inference. In particular, the parametric forms representing the hazard rate for the NFA group and the all-person comparator are different.

Recall that in the case of the NFA group, a parametric form $h(t) = at^b$ was assumed for the hazard rate, while the all-person comparator value is constant and independent of t (we further explain why this is the case in the next section). The former is a generalization 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 probability of observing the data if this were the case, the ‘p-value’, is approximately nil), meaning that there is little if no evidence that the hazard rate in the case of the NFA group should be constant. Nevertheless, while the data support this conclusion for the NFA group, there is no empirical justification for the assumption that the alternative situation occurs in the case of the all-person comparator.

Either of the assumptions of a constant hazard rate or of a parametric power curve is common within reliability, survival, failure, or extreme event analysis, but they arise in differing situations.¹¹ The study of reliability or survival analysis is a branch of statistics that is regularly used to predict time to failure in mechanical systems, or the future lifetime of a biological specimen etc. (an alternative name sometimes used in the social sciences is duration analysis). While this type of study is less common in determining policy implementation, it has recently been considered for estimating the compensation that should be awarded to successful plaintiffs in employment discrimination cases, where the use and effects of various survival analyses was highlighted.¹²

In general, this topic seeks to determine the time to a future ‘event’, where here the event of interest is arrest, and the fundamental object that enables such a prediction is known as a survival

¹¹ SINGPURWALLA, N.D., *Reliability and Risk: A Bayesian Perspective* (Wiley) (2006).

¹² PAN, Q. AND GASTWIRTH, J.L., “Issues in the use of Survival Analysis to Estimate Damages in Equal Employment Cases”, *Law, Probability and Risk* 8 (2009).

function. A survival function details the probability that the event of interest will occur beyond some specified time t for all $t > 0$. Alternatively, and equivalently, interest may lie in the lifetime function, which specifies the probability that the event of interest will occur prior to the specified time t , and as such, is 1 minus the survival function. The associated hazard rate can then be derived mathematically from either the survival function or the lifetime function.¹³

In the case of a constant hazard rate, as was used for the all-person comparator, this arises from the assumption that the time to event (arrest) follows an exponential distribution, which is an appropriate assumption for systems or objects that do not degrade or wear over time, and arises naturally when the times between events occur continuously and independently at a constant average rate. The parametric hazard rate assumed for the NFA group, however, results from the assumption that time to arrest follows a Weibull distribution,¹⁴ which under extreme value theory, is an appropriate assumption when a failure occurs as a result of the first of a number of competing and comparable independent processes failing.

In an engineering context, the Weibull assumption would be appropriate for predicting the future lifetime of a machine that consists of a series of independently functioning components, and where the machine would only continue to operate while all of its components continued to be in working order. If the lifetime of any particular component was independent of the lifetime of any other component and if the lifetimes of those components were not too dissimilar, then the Weibull modelling assumption would be an appropriate representation of uncertainty in the future lifetime of the machine. In the context of failure being equated with arrest, an analogy might be suggested whereby arrest will occur following suspicion of the commission of any number of crimes that might be committed, and so long as the collection of expected times to arrest for the range of possible suspected offences are not too dissimilar, and that suspicion of offence is independent of suspicion of any other offence, then the use of the Weibull distribution might hold justification. It should be noted, however, that the assumption of independent suspicions concerning the commission of a crime, and that there is similarity in the expected times until being suspected of any of a collection of possible offences, are both questionable and without empirical support.

Nevertheless, the use of two different modelling assumptions would imply that the underlying process leading to arrest differs between the NFA group and the general population and would appear to suggest that the act of being arrested without subsequent caution or conviction fundamentally alters the assumed behaviour of that person (we return to this consideration later when discussing probabilistic causality). This is despite such individuals as represented by the NFA group being entitled to the description of holding 'good character' within court scenarios.

Formally, setting the present as time 0 and letting T denote the unknown future arrest time of an individual, the assumption that T follows an exponential distribution, as is the case in the all-person comparator, is equal to stating that the probability density function¹⁵ for T is $f(T = t) = \alpha e^{-\alpha t}$,

¹³ The precise formula states that the hazard rate is minus the ratio of the derivative of the survival function to the survival function itself.

¹⁴ WEIBULL, W., "A Statistical Distribution Function of Wide Applicability", *Transactions of the American Society of Mechanical Engineers, Journal of Applied Mechanics* 18 (1951).

¹⁵ Without referring to a technical mathematical definition, a 'probability density function' allows calculation of the probability that a continuous random variable will fall in a given interval. While the actual probability of any specific value will be 0 (as there are an uncountable infinity of such possibilities for a continuous variable), the probability of falling in a specified interval is found by integrating the probability density function over that interval.

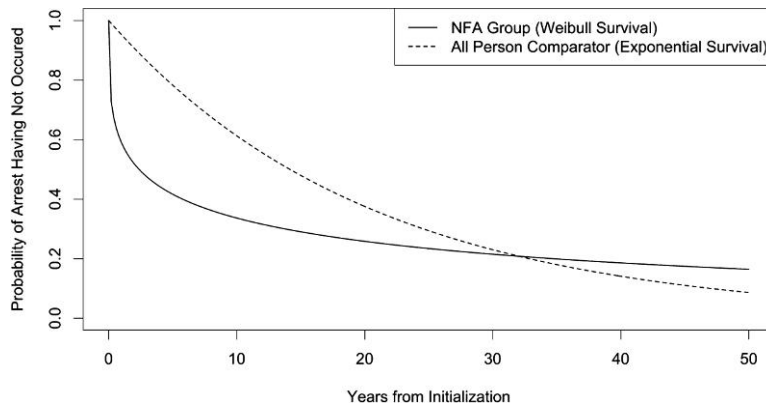


FIG. 2. The survival functions arising out of the estimated parameters in the case of the Weibull distribution as used for the NFA group (solid) and the exponential distribution as used for the all-person comparator (dashed).

with $\alpha > 0$ a rate parameter to be specified,¹⁶ or analogously, that the probability of future arrest occurring prior to time t (the lifetime function) is $P(T \leq t) = 1 - e^{-\alpha t}$. In the case of the Home Office analysis, the value of α was assumed to be 0.049 (though we again refer to the derivation in Appendix A and how it is a rough and crude assumption that is inflated to match arrest rates of age and sex profiles where the numbers of arrestees is more prevalent than the general population). Hence, if this model were true, the expected value of T , or in other words the expected time to future arrest, would be 20.4 years (the reciprocal of the value of α), and that $P(T \leq 15) > 0.5$, i.e. it would be more likely than not that an individual would be arrested within a 15-year period.

In the case of the NFA group, the power-curve assumption for the hazard rate is equivalent to stating that T follows a Weibull distribution, which in turn implies that the probability of arrest occurring before time t is $P(T \leq t) = 1 - e^{-\left(\frac{t}{\gamma}\right)^\beta}$, with $\beta > 0$ and $\gamma > 0$ parameters to be specified. While this equation may appear complicated, it allows calculation of the expected time to arrest. In the case of the Home Office analysis, the parameters β and γ are estimated to be 0.314 and 7.614, respectively, leading to an expected future arrest time of 57.9 years. This is the case despite the model suggesting $P(T \leq 2) > 0.5$, i.e. at point of NFA decision, it is more likely than not that an individual in the NFA group would be rearrested within 2 years, while the probability for rearrest within the next 6 years is approximately 0.66. Figure 2 plots the assumed Weibull and exponential survival functions that result in the fitted hazard rates for the NFA group and the all-person comparator, respectively.

This would at first appear a rather remarkable result. In effect, the fitted hazard rates for the NFA group and the all-person comparator, as seen from the initialization of the experiment in April 2006, indicate that the expected (average) future arrest date of a member of the NFA group is over twice that for an individual represented by the all-person comparator. Hence, if offending risk was not associated with probability of arrest in only the next year, but instead took into account the probability of arrest in all future years, it could be argued that the NFA group constituted a reduced risk of offending. Furthermore, it indicates that, from the time point of initial arrest, more than half

¹⁶ The term e refers to a well-known mathematical constant, namely Napier's constant, and its value is 2.718 to three decimal places (though because it is an irrational number, it cannot be expressed precisely in a finite decimal representation).

of the NFA group would be rearrested within 2 years of initial arrest,¹⁷ in which case their DNA profile would again be obtained, a consideration that appears not to have been taken into account when suggesting a blanket retention period of 6 years.

This result is obtained because using the mathematical formula for the estimated hazard rate beyond 6 years the hazard rate of the NFA group continues to decline, while for the all-person comparator it remains constant and higher than that of the NFA group. This of course requires use of the fitted model for all future times, which only worsens a key and fundamental problem of the Home Office's analysis, i.e. that it is heavily reliant on extrapolating hazard curves beyond periods in which data were actually monitored. Yet, they are the logical conclusions that can be drawn from the assumed form for the estimated hazard rates, and it is noted that, when formulating conclusions concerning equality of hazard at approximately 6 years into the future, the Home Office analysis does extrapolate the fitted NFA group hazard rate for a further period greater than that for which there are data available. Indeed, in discussing any hazard rate beyond 2.25 years following initial arrest, it must be highlighted that (because of the approximation between the hazard rate and the conditional probability of arrest) we are discussing the conditional probability of an individual who had not been rearrested for 2.25 years being subsequently arrested in the following year, yet no person in the NFA group was monitored for more than 3.25 years.

3.3 *Modelling time to arrest*

To determine an appropriate distribution model for future lifetime (here time until arrest), various considerations should be taken into account. For example, are there physical or theoretical reasons why future lifetimes should follow a given distributional form, or does a particular model offer a reasonable fit to the data? In the case of the Home Office analysis, a parametric assumption is required for two particular reasons. First, the analysis only monitored the NFA group, and did not track a selection of individuals similar to that group who had no prior arrest record. This is likely to have been the case due to logistical, privacy and expense concerns, but it means that any elicited hazard rate for the all-person comparator would have to be independent of the time since initialization of the study, and hence a constant.

Secondly, the monitoring of the NFA group was only over a 3.25-year period, and so any discussion of changes in this rate beyond that time requires extrapolation of a defined curve. Common alternative approaches in reliability analysis that do not assume that either the survival function or the hazard rate follows a defined family of mathematical forms are known as non-parametric analyses. The most common of these is the Kaplan–Meier estimator,¹⁸ where only the observed data are taken into account, and the lifetime variable is not assumed to follow any particular mathematical form. Another notable example is that of nonparametric predictive inference (NPI).¹⁹

Generally, such non-parametric approaches result in a series of horizontal and vertical steps of declining survival probability over time that would repeatedly intersect any smooth curve resulting

¹⁷ Only approximately a third were actually observed to be rearrested during this period, and this discrepancy arises because the fitted parametric hazard curve in the Home Office analysis was only applied for the hazard rate beyond the first year of arrest, while the conclusion reported here is based on using the fitted hazard rate for both the first and second year following arrest.

¹⁸ KAPLAN, E.L. AND MEIER, P., "Nonparametric Estimation from Incomplete Observations", *Journal of the American Statistical Association* 53 (1958).

¹⁹ COOLEN, F.P.A AND YAN, K.J., "Nonparametric Predictive Inference with Right-Censored Data", *Journal of Statistical Planning and Inference* 126 (2004).

from a parametric assumption for survival probability. Unfortunately, however, while such non-parametric analyses allow the ‘data points to speak for themselves’, their use in predicting future survival probability would be of limited use for determining a rearrest hazard rate. In the case of the Kaplan–Meier estimator, beyond the end of the observation period, the survival function would be estimated to remain as a constant, meaning that the derived hazard rate for any time beyond the monitoring period would be 0.

In the case of NPI, this is a generalization of the Kaplan–Meier estimator that only provides bounds on what the true survival probability would be by explicitly taking into account any absence of additional information. While the lower bound for survival probability beyond the monitoring period would be 0 (meaning that all individuals have been arrested), the upper bound would again be a positive constant. Hence, NPI would suggest that the derived hazard rate could be any value between 0 and 1, and as such, is simply stating that in the absence of any data concerning the hazard rate beyond 3.25 years following initialization of the study, all that can be said is that it can take any valid value, and that nothing else should be stated concerning the likelihood of any particular value.

The actual Weibull distribution that was used for the NFA group is a distributional form that is so flexible it is likely to provide a reasonable fit to many small sample failure data sets (note here that although the NFA group consisted of 17 239 individuals, it was actually fitted using 11 quarterly failure proportions, and as such, should be considered as a data set consisting of only 11 data points). However, the conclusions that can be drawn over the monitoring period would not be very different if alternative parametric forms were considered, e.g. a log-normal distribution, but these may lead to different conclusions when extrapolating into the future or when considering the underlying assumptions concerning the offending behaviour of the NFA group. Indeed, ignoring the first year of rearrests (which may be due in part to matches of DNA profiles with the crime scene database or because of administrative offences such as violations of bail conditions), even the assumption that the hazard rate declines in a linear fashion would not provide a very unreasonable fit to the data and would indicate that equality between the NFA group and the all-person comparator would occur much earlier than the estimated 6 years following initial arrest.

Hence, because many different models may well provide a reasonable fit to the observed data, but would result in very different conclusions when extrapolating beyond the monitored period, it becomes very important to focus on the theoretical justification for any proposed parametric form, especially if the alternative models would have different implications for policy development. From a sociological perspective, this may be a challenging problem, but if two or more curves would fit the data equally well, yet would result in vastly different outcomes for policy development, then the choice as to which policy is more appropriate will have to rely on the theoretical justifications or strengths of the competing models.

For the all-person comparator, a constant hazard rate was considered, and as discussed previously, this is the appropriate model if the likelihood of arrest was not affected by the passage of time. For the Weibull distribution, this is not the case, and the likelihood of arrest will be affected by the passage of time. However, one possibility that does not appear to have been explored, but which might hold greater theoretical justification, is the existence of different groups in society that are subject to differing hazard rates for arrest.

For the all-person comparator, there was an implicit assumption that the hazard rate was independent of the passage of time, and this was generated by assuming that the unknown future time of arrest T followed an exponential distribution with a parameter $\alpha = 0.049$. Hence one possibility is to model the future arrest time T of a member of the NFA group as being exponential with

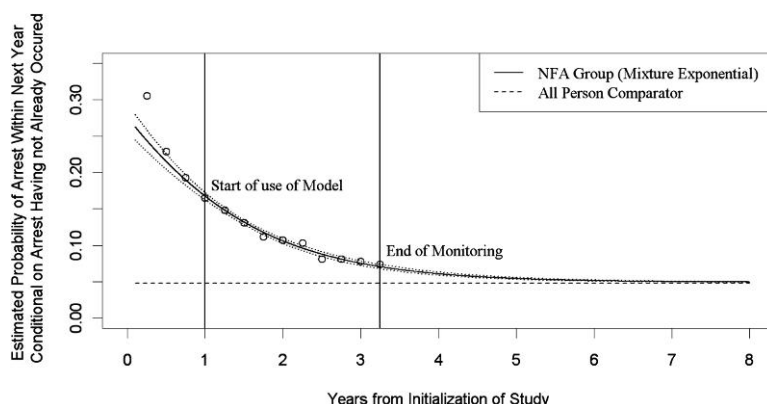


FIG. 3. An alternative estimated hazard rate for the Home Office data that are based on a 'mixture' of exponential distributions (solid) and the all-person comparator (dashed). The circle points indicate the observed annualized rearrest rate, while the vertical lines mark the beginning of inferring the hazard rate through a parametric model and the end time of observing the NFA group, respectively. The dotted lines represent approximate 95% confidence intervals, the derivation for which is discussed in Appendix B.

parameter $\alpha_1 = \alpha$ with probability p , or otherwise as being exponential with parameter $\alpha_2 > \alpha$. In other words, with probability p , a member of the NFA group would constitute no greater offending risk than the members of the non-arrested population (the law-abiding 'good' people in the NFA group), or would otherwise constitute a greater offending risk, but one that is also independent of time (say the 'not-so-good' people or those more inclined to criminal behaviour). For reference, we will label these subgroups as NFA-G1 and NFA-G2, respectively.

Such a 'mixture model' does not generate a constant hazard rate, and so could provide a reasonable fit to the rearrest data that were obtained by the Home Office. For example, the shape of the curve would be expected to follow that of the observed data in that there would be an expected sharp initial drop in the rearrest hazard of the total NFA group due to the rearrests within the subgroup that constituted a greater offending risk (NFA-G2), followed by a diminishing reduction in the rearrest hazard as the number of individuals within NFA-G2 dwindled, leaving only individuals who are no different to the general unarrested population (NFA-G1) remaining in the NFA group.

Moreover, regardless of how far into the future the 'mixture model' is extrapolated, the hazard rate for the NFA group converges asymptotically to the hazard rate for the all-person comparator. This is in contrast to the assumption of a Weibull distribution, where a reduction in risk of the NFA group compared to the all-person comparator occurs after 6 years. Furthermore, the immediate rearrest hazard, i.e. the hazard at time 0, or time 0 plus some small amount, would be more reasonable under the 'mixture model' and is the weighted average of the parameters of the two exponential distributions, i.e. $pa + (1 - p)\alpha_2$. Again, this is in contrast to the Weibull distribution, where the immediate hazard is exceptionally large for small time points, hence explaining why the Weibull model was only fitted using data at least 0.75 years from initial arrest.

A fit of such a 'mixture model' that was also based on using only the 11 quarterly rearrest proportions between 0.75 years and 3.25 years following initial arrest is shown in Fig. 3. The fitted 'mixture model' was found to have parameters²⁰ $p = 0.72$ and $\alpha_2 = 0.87$, resulting in a conclusion

²⁰ These parameters have 95% confidence intervals of (0.714, 0.732) and (0.800, 0.938), respectively.

that 72% of the NFA group would constitute no difference in risk than that assumed for the general unarrested population, while the remaining 28% constitute a nearly 18-fold increase in rearrest hazard. A comparison of goodness-of-fit measures for the models is given in Appendix C and supports the conclusion that the mixture exponential has a (slightly) enhanced fit over the Weibull model. A plot overlaying both the fitted Weibull hazard model and the fitted mixture exponential model is provided in Fig. 4, while the survival functions are provided in Fig. 5.

While we are not seeking to advocate the use of this or any other particular model (in fact a mixture with more than two subgroups would provide an even greater fit), we wish to highlight how the parametric form can be altered so as to still fit the data, but such that very different conclusions would result. Unfortunately, applying such a mixture model in practice would require additional information in order to classify a member of the NFA group as belonging to one subgroup or another, for all that is recorded in the Home Office data are the fact of their historic arrest in April 2006.

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 rearrest 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 the two possible parameters for the exponential distribution (depending on which subgroup the individual belongs to), the probability of rearrest not occurring by time t is determined once subgroup membership is known. However, this relationship can be reversed through Bayes' theorem so as to allow calculation of the probability of subgroup membership given that rearrest has not occurred by time t .

Bayes' theorem concerns the relationship between conditional probabilities: for two events A and B where the conditional probability of B occurring given that A is true is denoted as $P(B|A)$, the conditional probability of A occurring given that B is true is $P(A|B) = \frac{P(B|A)P(A)}{P(B)}$. In the context of subgroup membership, given that arrest has not occurred by time t , the above formula can be used by equating event A as membership of subgroup NFA-G1, and B as the event that re-arrest has not occurred by time t . That is to say, if we are interested in the probability that an individual is of

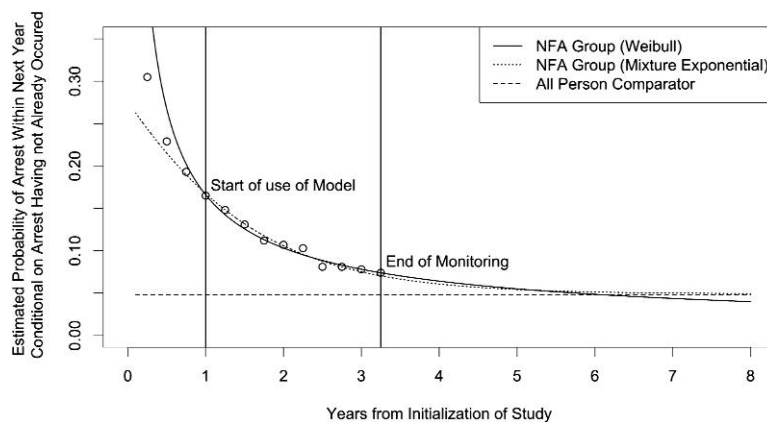


FIG. 4. A comparison of the fitted Home Office hazard rate for members of the NFA group (solid), an alternative model based on a 'mixture' of exponential distributions (dotted) and the all-person comparator (dashed).

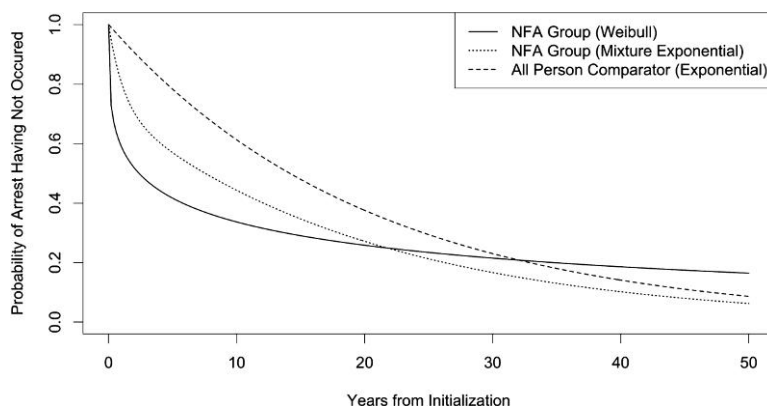


FIG. 5. The survival functions arising out of the estimated parameters in the case of the Weibull assumption for the NFA group (solid), the mixture exponential alternative (dotted), and the all-person comparator exponential distribution (dashed).

subgroup NFA-G1 conditional on rearrest not having occurred by time t , $P(\text{NFA-G1}|T > t)$, then utilizing Bayes' theorem this is $\frac{P(T > t|\text{NFA-G1})P(\text{NFA-G1})}{P(T > t)}$.

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, $P(\text{NFA-G1}|T > t) = \frac{0.72e^{-0.049t}}{0.72e^{-0.049t} + 0.28e^{-0.87t}}$, and Fig. 6 plots the probability of belonging to subgroup NFA-G1 (the subgroup that has the same arrest risk as that estimated for the un-arrested population) as a function of time elapsed since NFA decision without subsequent rearrest.

Figure 6 demonstrates that the longer the time delay until rearrest, the more likely the individual is to be of 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, of particular interest may be the time taken before an individual has a 95% chance of being a member of NFA-G1 (as this is often considered a statistically significant probability) and this is found to be 2.42 years. After 6 years without rearrest, the individual has a 99.7% chance of being of subgroup NFA-G1, while if we wished to know when, on the balance of probability, an individual is more likely to be of subgroup NFA-G1 than of NFA-G2, it should be noted that even at time of NFA decision, the probabilities are 0.72 and 0.28, respectively.

Finally, we wish to highlight that the fitting of any parametric curve requires making the assumption that the future times to arrest for all individuals within the NFA group are exchangeable. Note that two random variables X and Y are exchangeable if and only if, for all values x and y , the relation $P(X = x, Y = y) = P(X = y, Y = x)$ holds (the concept generalizes straightforwardly for more than two random variables). As such, exchangeability is an assumption of symmetry that is often suitable in situations where draws are randomly taken from a population (which can be finite or infinite), hence making it an appropriate assumption in many situations. Indeed, the common statistical assumption of sampling independent and identically distributed random variables would clearly result in exchangeability, but so would sampling without replacement, which is not independent.

However, to assume exchangeability of the future arrest times of the NFA group, or of any further individual that the model is used to predict future arrest time of, requires ignoring any additional information relating to that person. While exchangeability is common within reliability analysis,

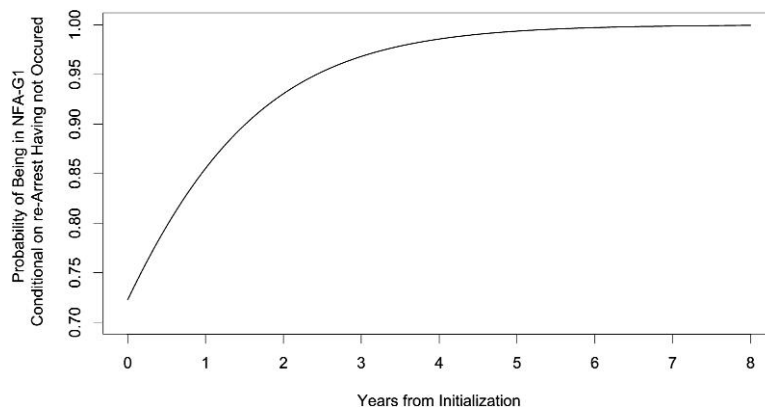


FIG. 6. The probability arising from the mixture exponential model of a member of the NFA group belonging to subgroup NFA-G1 (the subgroup that behaves like the unarrested population) as a function of time elapsed since NFA decision without subsequent rearrest.

there is often good cause for this. For example, exchangeability would be reasonable when considering the future lifetime of a collection of light bulbs all manufactured in the same process by the same machine. However, in the context of arrest, it is unlikely that additional information cannot be obtained which would be informative of when a future arrest is likely to occur, and so when predicting the future arrest of a new individual who is subject to an NFA order, the use of the Home Office model would require ignoring all aspects of that person other than the fact they had been previously arrested and NFA was taken.

3.4 Probabilistic causality

Combining the Home Office models for the NFA group and the all-person comparator leads to a generic model for the hazard rate of the entire population of non-convicted individuals. Such a generic model can be generated by including an indicator event A that takes value 1 if a person has been previously arrested without caution or conviction, and takes value 0 otherwise. The hazard rate can then be considered a function of not only time t from initialization of the study, but also a function of the value of the indicator A that is relevant for the person under consideration.

Remembering that the hazard rate was approximated by a conditional probability, i.e. the probability that arrest will occur in the next unit of time conditional on it not having occurred by time t , this generalized version also incorporates historic arrest status by conditioning on the binary event A . Hence, in the generalized version, the conditional probability of arrest/rearrest will be greater for all t between 0 and 6 if $A = 1$, while it would be reduced for all $t > 6$ if $A = 1$.

The status of a probability being greater depending on whether or not a conditional event has occurred is the subject of the theory of probabilistic causality. Traditional causality can be generally thought of as constituting one of four general notions,²¹ i.e. ‘materialist’ (based on the laws of physics), ‘spiritualist’ (the act of a supernatural being), ‘rationalist’ (the relation between reason and consequence), and ‘phenomenalist’ (empirical observation or association between successive

²¹ LINDLEY, D.V. AND SINGPURWALLA, N.D., “On Exchangeable, Causal and Cascading Failures”, *Statistical Science* 17 (2002).

events). This latter phenomenalist approach to causality underlies the support the Home Office Analysis provides in any formulation of policy, but it is also the approach that supports a probabilistic quantification of causality.

Following Suppes,²² an event *Y* is said to be a *prima facie* (at first sight) probabilistic cause of another event *Z* if event *Y* occurs prior to event *Z*, if the probability of event *Y*'s occurrence is not zero, and if the probability of event *Z* occurring is greater if event *Y* did occur. In other words, an event *Y* will be a *prima facie* probabilistic cause of event *Z* if there is a temporal relation of *Y* occurring before *Z* and if there is a positive association (correlation) between them. Hence, under this definition, we see that using the model employed in the Home Office analysis a historic arrest resulting in NFA (event *Y*) would constitute a *prima facie* probabilistic cause of future arrest (event *Z*) for up to 6 years after the occurrence of the initial arrest.

Nevertheless, a *prima facie* probabilistic cause may be due to a confounding factor and may not be a genuine probabilistic cause. Confounding occurs when correlation is mistaken for causation, and such occurrences lead to spurious causal relationships. A typical example would be the relationship between alcohol consumption and lung cancer, where it can be observed that lung cancer rates increase as alcohol consumption increases. Yet, the consumption of alcohol is not a causal effect but is instead a confounding factor. Actually the relationship is explained by the act of smoking, where an increase in alcohol consumption is correlated with whether or not an individual smokes, while it is the act of smoking that is a causal effect of contracting lung cancer.

To account for this, Suppes defines a spurious *prima facie* probabilistic cause as follows. Let events *Y* and *Z* be as before. The event *Y* is said to be a spurious cause of *Z* if there exists an additional event *X* that occurs prior to *Y*, is such that *X* and *Y* can both occur, and that the probability of *Z*'s occurrence is independent of the occurrence of *Y* if *X* does occur. That is to say, the association between *Y* and *Z* is completely explained by their mutual association with another event *X*. Hence, in the case of the Home Office analysis, prior arrest with NFA (event *Y*) would be a spurious cause of future arrest (event *Z*) if taking into consideration an alternative factor (event *X*) accounts for the probability of future arrest just as effectively. For example, it was highlighted in the Home Office analysis that arrest rates were higher for men than for women, and so a model that takes into account gender, and possibly also age and socio-economic factors, might show a prior NFA decision as a confounding variable. Furthermore, as there are a vast number of such latent factors that could render a prior NFA decision as spurious, previous arrest with no caution or conviction is unlikely to be a genuine probabilistic cause of future arrest.

As an example, various studies have considered whether or not time spent under a custodial sentence increases offence risk upon release,²³ i.e. whether the act of having served a custodial sentence is a genuine probabilistic cause of future offence? Although there would initially appear to be an increase in risk, more detailed research indicates that time spent in custody is more likely to be a confounding factor, and that the offence risk is instead more closely associated with age, gender, nature of previous conviction etc. Indeed, the Home Office itself has considered the effect of confounding variables in this instance, and in a 2002 Prison Statistics report²⁴ noted that:

²² SUPPES, P., *A Probabilistic Theory of Causality* (North-Holland) (1970), Chapter 2.

²³ COWEN, N. "Does Custody Cause Crime?", *CIVITAS: Institute for the Study of Civil Society* (2010). Online Report Available at: <http://www.civitas.org.uk/crime/custodycrime.pdf>.

²⁴ Home Office Research Development and Statistics Directorate (2002) "Prison Statistics England and Wales", Chapter 9. Available at: <http://www.archive2.official-documents.co.uk/document/cm59/5996/5996.pdf>.

The proportion of prisoners reconvicted following discharge from custody is mainly associated with the characteristics of those offenders, rather than the impact of custody. The main predictors of reconviction are: the number and rate of previous convictions, age at sentence, whether they are male or female, and the type of offence for which they were imprisoned...

When comparing the impact of custody on reconviction rates over time it is therefore necessary to control for changes in the characteristics of offenders being given custodial sentences.

Unfortunately, however, such data are not yet publicly available for the reoffence risk of members of the NFA group, but a likely conjecture is that similar findings might also be made, indicating that the act of being arrested with NFA should not be considered as either a cause, or an indicator, for reoffence risk.

3.5 *Summary*

To summarize, we return to the focus of the analysis, i.e. does the Home Office Re-Arrest data analysis actually provide statistical evidence that there is increased offending risk of individuals holding a historic arrest status, but where no further action was taken, even if this is for only 6 years after the occurrence of that arrest? It is our view that the Home Office analysis falls short of achieving this goal, and that any conclusions drawn should be considered in relation to the appropriateness and reasonableness of the underlying assumptions that are either explicitly or implicitly required.

The necessary assumptions of equating arrest risk with offending risk, of individuals subject to an NFA order having a future arrest model that is dependent on the time lag since previous arrest, of extrapolating estimated hazard curves for durations longer than that for which data were actually monitored, and that the hazard curve follows the specific form derived from a Weibull survival function that is fitted under an assumption of an exchangeable population etc. severely limits the use and applicability of the analysis. Moreover, we have demonstrated that the same data would support an alternative model where arrest risk does not have to change over time, but where subgroups of populations have differing rates of arrest. Furthermore, while it is not our intention to specifically advocate the use of any particular model, it should be noted that if the mixture exponential model is true, the vast majority of the NFA group are being inappropriately labelled as constituting an enhanced risk of committing a future offence.

4. Possibilities for the future

The collection and appropriate analysis of data in formulating national policy or guidelines are undoubtedly important; however, the use of such a statistical analysis as provided by the Home Office in suggesting heightened offending risk of an individual who holds no caution or conviction is controversial, especially as the retention of DNA is a contentious issue. In the case of the retention of DNA profiles for individuals who have not been cautioned or convicted, current legislation only specifies that constabularies may retain the DNA profiles, yet other than guideline policy, there is no legal obligation for such data processing.

Nevertheless, and as has occurred previously, any judicial review against a decision to enforce this policy in a particular setting is likely to focus on the statistical arguments justifying the evidence base for retention, for Article 8 of the European Convention of Human Rights specifies

that there shall be no interference by a public authority with the exercise of the right to respect for the private life of an individual except such as is 'necessary' in a democratic society in the interests of (in this instance) the prevention of crime. So the question remains whether such retention is necessary, and whether it strikes a fair balance between the competing public and private interests?

There has recently been a steady rise in academic interest concerning the use and interpretation of probabilistic and statistical arguments by objective organizations such as Government departments, criminal justice agencies, law courts etc. Notable examples for court scenarios include the interpretation of statistical evidence,²⁵ the misuse of assumptions in expert analysis,²⁶ the evidence of guilt following a positive DNA match,²⁷ the meaning of 'reasonable doubt' and whether this can be quantified,²⁸ and the appropriateness of jury size and conviction rule considering the probability of the same jury result if another jury were selected from the overall population.²⁹

Such issues regarding the application of statistical arguments by objective review panels arise because of their unique context whereby statistical evidence should be both empirical and logical. The authors agree with the following definition³⁰ for these norms:

Empirical: Objective inferences should not disagree with empirical evidence.

Logical: If there is no information suggesting that one possible outcome is more likely than another, then this should be reflected by identical uncertainty quantifications for these outcomes.

As such, the use of popular statistical techniques such as the inclusion of a subjective prior in the case of Bayesian inference,³¹ or the use of common distributional forms because of ease of simplicity or a subsequent reduction in computational burden, would be inappropriate, despite being either grounded with a solid axiomatic foundation or having an applicability in approximating a wide variety of situations.

Unfortunately there is as yet no universally agreed mathematical method of expressing uncertainty that is objective, and despite numerous attempts throughout the history of statistical study, this remains the fundamental goal of research in statistics. Various solutions have been proposed, e.g. frequentist analysis,³² objective Bayesian analysis³³ or imprecise probabilistic analysis³⁴; yet, all

²⁵ AITKEN, C.G.G., "Interpretation of Evidence, and Sample Size Determination". In GASTWIRTH, J.L. (Ed.), *STATISTICAL SCIENCE IN THE COURTROOM* (Springer) (2000).

²⁶ KADANE, J.B., "Misuse of Bayesian Statistics in Court" *Chance* 19 (2006).

²⁷ SJERPS, M. AND KLOOSTERNAM, A.D., "Statistical Aspects of Interpreting DNA Profiling in Legal Cases" *Statistica Neerlandica* 57 (2003).

²⁸ TILLERS, P. AND GOTTFRIED, J., "Case Comment—United States v. Copeland, 369 F. Supp. 2d 275 (E.D.N.Y. 2005): A Collateral Attack on the Legal Maxim that Proof Beyond a Reasonable Doubt is Unquantifiable?" *Law, Probability and Risk* 5 (2006).

²⁹ COOLEN, F.P.A., HOULDING B. AND PARKINSON S.G., "Considerations on Jury Size and Composition using Lower Probabilities" *Journal of Statistical Planning and Inference* 141 (2011) (doi:10.1016/j.jspi.2010.06.014).

³⁰ COOLEN, F.P.A., "On Nonparametric Predictive Inference and Objective Bayesianism" *Journal of Logic, Language and Information* 15 (2006).

³¹ BERNARDO, J. M. AND SMITH, A. F. M., *Bayesian Theory* (Wiley) (1994).

³² NEYMAN, J., "Outline of a Theory of Statistical Estimation Based on the Classical Theory of Probability", *Philosophical Transactions of the Royal Society of London A* 236 (1937).

³³ JEFFREYS, H., *Theory of Probability* (Oxford University Press) (1939).

³⁴ WALLEY, P., *Statistical Reasoning with Imprecise Probabilities* (Chapman and Hall) (1991).

fall short of achieving the goal of providing a statistical framework that is objective, can be readily applied, does not suffer from paradoxes following irrelevant transformations of the data or guarantees in providing a precise answer (though the aforementioned NPI might be one approach that can be successfully applied so long as it is accepted that only vague, or even vacuous, inferential statements might result because that is all that can be derived from the actual data). In the case of policy development or legal interpretation, the problem is further expanded in that the details of such statistical arguments should be of a nature that they can be readily understood by ministers, civil servants, judges, legal teams, jurors etc.

In the present case, it is observed that an exchangeability assumption was implicitly used to fit a parametric model. This required acceptance that the parametric model is an adequate description of the uncertainty in the future arrest time of a further individual, and that the parameters of this model could be estimated from the rearrest times of alternative individuals. Yet, neither of these assumptions would appear either empirical or logical, and certainly it is known that additional factors such as gender and age effect arrest rates. As such, there are a number of ways in which a further study could be specifically designed so as to remove doubt in its conclusions, and we now list some of these.

An initial rearrest ‘spike’ was problematic for the assumed Weibull distribution, yet this appears to be a common finding in studies concerning rearrest hazard.³⁵ However, we have mentioned in passing that such a spike may be caused 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 very useful to know the proportion of early rearrests that are due to such effects so as to discount them and diminish any resulting bias.

Moreover, a recent study³⁶ has provided evidence that police forces target certain individuals or groups who are known to them so as to enforce measures of social control, even if such individuals hold no historic record of caution or conviction. While in our discussion of probabilistic causality we mentioned that the act of being arrested is likely to be a spurious probabilistic cause of future arrest, it does appear to be *prima facie* probabilistic cause. Yet, rather than relying on a supposed connection that is explained because of an increase in offending risk following arrest, the relationship might be because once an individual has been arrested, the Police will target, monitor and seek to control the future social behaviour of that individual.

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 subpopulations that share the same probabilistic rationales as the all-person comparator, but with differing hazard rates, it would be appropriate to monitor a ‘control’ population of unarrested individuals that shared common socio-demographic profiles as that of the NFA group.

Moreover, detailed knowledge of the socio-economic profiles of those subject to an NFA order would allow investigation into whether there do exist subpopulations with an offence rate being a function of covariates drawn from those profiles, e.g. age, class, ethnic profile, educational attainment etc., rather than assuming all members of the NFA group behaved in a similar way. Indeed,

³⁵ KURLYCHEK, M.C., BRAME, R. AND BUSHWAY, S.D., “Scarlet letters and recidivism: Does an old criminal record predict future offending?”, *Criminology & Public Policy*, 5(2006).

³⁶ MCARA, L. AND MCVIE, S., “The usual suspects?: Street-life, young people and the police”, *Criminal Justice*, 5 (2005).

knowledge of the suspicion of the original offence leading to arrest would be of benefit so as to determine if certain types of suspected offences lead to increased rearrest risk, and certainly the previous ruling in the ECtHR criticized the blanket nature of the application of the guidelines to DNA retention irrespective of the offences involved or a person's circumstances.

Finally, it is well reported that there exists a distinctive relationship between the age of an offender and the commission of a crime (known as the 'age-crime' curve), with a peak age of conviction at approximately 18 years of age. While it was accepted that the all-person comparator in the Home Office analysis sought to take into account this consideration by estimating an arrest rate for a group of individuals without historic arrest but which had the same age and sex profiles of the NFA group, the rearrest times of the NFA group did not report or indicate the age of the member of the NFA group who was rearrested. In particular, it was noted that:

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.

As such, it would be of benefit to observe the effect of constructing a hazard rate for each of a number of age-related individuals, e.g. for a pre-teen group, for a teenage group, for a young adult group, for a senior adult group etc., though it is noted that this is likely to result in a suggestion that policy not only distinguish between adults and children but also subcategories of these classes.

It would appear then, that a major difficulty in formulating any statistical evidence base for DNA profile retention of persons subject to an 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 would take into account such factors as gender, age, socio-economic background, location of residence etc., though they are indeed known to influence arrest hazard, would not only require 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

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

Disclaimer

The views and opinions expressed in this paper are the authors' own. They are not intended to represent any position or policy of the UK Home Office or any other Public Body or Non-Government Organisation.

Acknowledgements

The authors wish to express their gratitude to the UK Home Office for making available the original data used in its analysis. We are also grateful for the supportive comments and suggestions in

the completion of this work that were provided by Frank Coolen of Durham University, by three anonymous referees, and by Colin Aitken (Editor).

Appendix A: Estimating a hazard rate for the all-person comparator

Annex 3 of the Home Office report provides details on how the all-person comparator value was obtained; however, to aid the reader we here provide a summary description. In particular, we note that the estimated value of 4.9% would appear to have been subject to a rough and crude approximation due to a lack of recorded data. In particular, the following three steps were taken using conviction data and responses from the 2003–2006 Arrestee Survey³⁷:

- 1) Estimate the age breakdown in national arrest data for males and females.
- 2) Calculate the number of unique arrestees.
- 3) Relate arrest proportions (over age and sex) to population proportions.

In estimating the age breakdowns in national arrest data, national conviction data were used for the arrest rates of 10- to 17-year olds (where, e.g. 24.1% of convicted males in this age range were 17 at the time of arrest, while only 1.3% were of age 10), while the results of the 2003–2006 Arrestee Survey were used to calculate the proportion of arrests in the age ranges of 18–20 and over 21, respectively.

Next, a method of converting number of arrests to number of unique arrestees was required, as the available data could not distinguish if two arrests were for the same individual or for different individuals. The 2003–2006 Arrestee Survey did, however, note how many times an interviewee had been arrested. Hence, anyone with a single arrest counted one unique arrest, while for those with multiple arrests the number of unique arrestees per arrest event was calculated using the formula $\frac{1}{0.5 + \text{Number Arrests in Previous Year}}$. The Home Office analysis explained the 0.5 term as representing the arrest leading to interview, and its treatment as half an arrest is claimed to adjust for bias caused by surveying at the point of arrest. This value was subsequently averaged for each age and sex group and was found to steadily increase from 0.47 unique arrestees per arrest for 17-year-old males to 0.60 unique arrestees per arrest for 42-year-old males, indicating that younger males are likely to be arrested more frequently. It was also noted that the ratio of unique arrestees per arrest was substantially higher for women than for men.

Finally, dividing the number of unique 18- to 20-year-old male arrestees per year by the number of 18- to 20-year-old males in England and Wales gave the likelihood of an 18- to 20-year-old male being arrested in a year. This process was repeated for males and females for each of the age groups for which arrest figures were available, and each was subsequently weighted by the proportion of total arrestees within that group.

Appendix B: Approximating a confidence interval for the hazard rate

There is no straightforward method of determining actual confidence intervals of the estimated hazard rate without making additional modelling assumptions on underlying parameters. To account

³⁷ Home Office Statistical Bulletin (2007) “*The Arrestee Survey 2003-2006 2nd Edition*”. Available at: <http://rds.homeoffice.gov.uk/rds/pdfs07/hosb1207.pdf>.

for this, the Home Office analysis (Annex 1) considered a 'bootstrap'³⁸ type technique where the members of the NFA group were subsampled and a hazard rate was recalculated:

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 boot-strapping exercise, and hence conservative.

The results of the approximate confidence interval produced by the Home Office analysis were similar to those found here, however, we have used a slightly alternative means of estimating confidence intervals that allows a greater number of hazard curves to be generated by assuming the actual parameter values for the models follow Normal distributions. In particular, the method utilized here consisted of the following steps:

- 1) The 95% confidence intervals for the estimated parameters for either of the Weibull or mixed exponential survival functions were considered as providing bounds on regions where 95% of the probability mass would be located under the assumption that these parameters were Normally distributed.
- 2) If a parameter a has 95% confidence interval of (a_L, a_U) , then the mean μ_a and the variance σ_a^2 of the normal distribution are calculated as $\mu_a = \frac{a_L + a_U}{2}$, and $\sigma_a^2 = \frac{(a_U - a_L)^2}{16}$.
- 3) 1000 simulations were drawn from the normal distributions representing uncertainty in the parameter values, and for each simulation, the resulting hazard curve calculated.
- 4) An approximate 95% confidence interval for the hazard curve was then estimated by calculating the 2.5 percentile and 97.5 percentile of the value of the simulated hazard curves for each time point.

Appendix C: Comparing goodness of fit for the Weibull and mixture Exponential models

Both the Weibull and mixture exponential models require estimation of two parameters, and so are equally complex in a statistical sense. Furthermore, while both appear to fit the observed data reasonably well, the mixture exponential model would appear to slightly outperform that of the Weibull model. We thus now illustrate this through a few common techniques for assessing model performance.

³⁸ EFRON, B. AND TIBSHIRANI, R., *An Introduction to the Bootstrap* (Chapman & Hall/CRC) (1993).

First we consider the Akaike information criterion (AIC) value, which assesses goodness of fit by considering the probability or ‘likelihood’ of observing the data under the assumption that the fitted model was true. The score is, however, penalized the more complex the model, where complexity is measured by the number of parameters that must be fitted in the model. The value is calculated as $AIC = 2k - 2\log(L)$, where k is the number of parameters included, and L is the likelihood of the observed data under the assumption of the fitted model. A smaller value indicates a preferred model, especially if two competing models have the same number of estimated parameters. The AIC value for the mixture Exponential is -84.9 , while for the Weibull model it is -80.5 (hence indicating that the mixture exponential slightly outperforms the Weibull).

A residual plot also indicates the prediction errors arising as a result of using a specified model to predict the data points already obtained. Ideally, such residuals should be close to zero, but moreover, there should not be a pattern present in their deviations around zero (known as residual heterogeneity), as this would indicate a poor fit to the data and that there may be additional influences that are relevant but which are ignored within the model. Figure A1 provides residual plots for both the Weibull and mixture exponential models and demonstrates that there does not appear to be any evidence of residual heterogeneity in either model. This plot also indicates that the residuals for the mixture exponential model tend to be closer to zero than the corresponding residuals for the Weibull model (though this was to be expected due to the smaller AIC value for the mixture Exponential model), again indicating the slight improvement of the mixture Exponential over the Weibull.

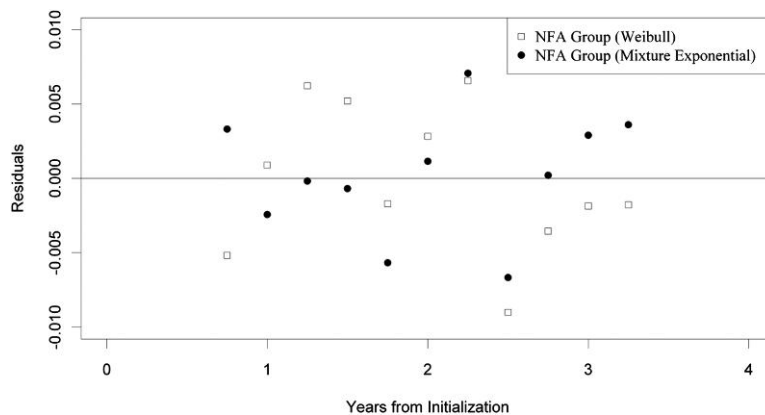


FIG. A1. A plot of the residuals from the fitted Weibull (squares) and mixture exponential (solid circles) distributions used to derive a model for the hazard rate of the NFA group. Both models indicate a good fit to the data with no evidence of residual heterogeneity.