

‘Keeping the right people on the DNA database’

Response to consultation

Dr Julian Huppert, University of Cambridge

5th August 2009

Introduction

I am an academic at the University of Cambridge, working on DNA and genomics; I therefore have a particular interest in this topic. My work further involves extensive use of statistics to analyze complex data sets, and so I have some expertise in this aspect as well. I also believe very strongly in evidence-based policy making.

Such evidence base is especially important in a context such as this one, where the proposed policy (keeping DNA data) is explicitly a reduction in liberty for people in the UK. I absolutely accept that there are cases where liberty needs to be reduced in exchange for other benefits, but the onus should be on Government (or those proposing the policy) to evidence that the benefits are real are proportionate to the loss of liberty. Essentially, the core of the argument is that people are innocent until proven guilty, and should be treated as such.

This current consultation presents a few statistics to argue for a collection of policies, already more liberal than those thrown out by the ECHR in *S and Marper*. Indeed, some of the proposed policy changes are to be welcomed, but many are poorly argued, and certainly not evidenced by any data in this document. Such data as are shown are in many cases flawed and over interpreted, which is disappointing in a document such as this.

In this response, I will discuss what I see as the logical thread of the argument presented, and discuss my responses both to the statistics presented and the conclusions in each section. For the avoidance of doubt, I understand the argument to run broadly as below, and this is the order I shall discuss the topics. I then conclude with a summary of changed recommendations.

- DNA data are useful for convictions, and so should be retained.
- For those who have not been convicted, keeping the data permanently is not allowed, so it is proposed to keep it for 6 years.
- Those accused of violent, sexual or terrorist offences should have data kept for 12 years.
- Children should have separate rules.
- The samples themselves should be destroyed.

Is the DNA database useful for convictions?

The first item that must be demonstrated is that the DNA database is useful. While DNA data clearly have enormous utility in specific (and limited) circumstances, both in identifying the guilty and clearing the innocent, this is not the same as an argument for storage generally in a database. I am very surprised that there is such a paucity of systematic data on this, especially in this current consultation. There are only four references that refer to this question, all of which are deeply flawed.

First, appendix C, page 27 refers to statistics presented by Lord Steyn, showing that many DNA profiles that may have been destroyed under the former provisions had been matched with crime-scene profiles. However, this leaves entirely unanswered the question as to whether these matches had a significant effect on later proceedings; in how many of these instances could a sample not have been collected from each of a relatively small number of suspects, avoiding the need for a database? It is this marginal usefulness that needs to be presented, and very little can be concluded without such comparators.

Second, Table 1, page 29, shows data from Dunsmuir, apparently describing the statistical association of DNA with charging and clear-up in various cases. However, since the colour scheme so carefully described is not reproduced, it is impossible to read the table – the positive and negative effects have been removed!

Third, on page 29, data are presented (from Smith 2004) to investigate the usefulness of DNA testing in investigations, with a large increase in detection rates when useable DNA is collected. However, these results suffer from a massive confounding effect; there is a sizable proportion of recorded crime that is not seriously investigated by the police, in most cases for good reason; when my bike is stolen in town, or a mobile phone on the underground, the police know that detection is likely to be impossible, and so it would be foolish to invest resources in a futile effort. However, these crimes would still be recorded, and so count towards lowering the overall detection rate. Collecting DNA evidence then merely acts as a poor indicator of the nature of the crime.

Removing this confounding effect is relatively tricky to do; it may be that fundamentally easier-to-detect crimes have DNA at the scene. One possible proxy would be to compare the figures to detection rates given that at least 3 hours of police time were spent on detection, for example, to eliminate the bulk of the impossible cases. There may or may not be some residual contribution from DNA evidence, but as presented this does not argue for its utility.

Fourth, page 29 then discusses the work of Roman *et al.* (2008) to do a proper randomized control trial of DNA samples. This is an ambitious piece of work and mostly well-constructed, although it is focused on cost-effectiveness questions. I am intrigued to note that although this research was carried out entirely in the USA, this is not mentioned in the text, with the implication that it is a UK study. This is particularly relevant because the full study shows significant variation in the utility of DNA samples in different communities, possibly because of legislative differences across the five communities. One thus has to question how transferable the results

are to the UK system, and I would suggest that a study like this should be urgently undertaken in the UK.

In any event, there is one critical flaw in Roman for our current purposes, and it is quite revealing of the mindset of the study authors. The conclusions describe the number of suspects identified, arrests made and prosecutions brought, but say absolutely nothing about the actual conviction rate; this surely is the important outcome. The (ideal) outcome must be one where the number of successful convictions is as high as possible, but the number of suspects, arrests and prosecutions are as low as possible given that, as each excess suspect, arrest and prosecution above the number of convictions merely represents a cost to individuals, society and the police.

I am very disappointed that there is such a shortage of well-researched UK evidence for the use of DNA in increasing conviction rates. Arguably, in the absence of such evidence, it cannot at all be claimed that keeping any DNA at all is proportionate and evidence-based. However, I accept that there is sufficient anecdotal evidence for the existence of a DNA database in some form to be contemplated.

If such a database is to exist, then clearly those convicted should be more readily placed on it than those not convicted, as they have forfeited some of their rights to liberties as a consequence of conviction. However, even for those convicted, there is no evidence presented as to why their data should be kept indefinitely; indeed, if one accepts the argument made that those not convicted can be treated analogously to those convicted in terms of future offending rates, then the conclusion must surely be that the two groups should be treated the same throughout!

I am particular concerned about the point raised in 6.23 (page 17) that those given **cautions, warnings or reprimands** would be treated as those who have been convicted. This seems excessive and inappropriate. Will people be advised of this when they accept on of these outcomes? I would argue that, since these people have not been through due process of a court, they should be treated as not convicted.

Should DNA data for those not convicted be kept for 6 years?

If it is accepted that DNA data from those not convicted should be kept (and as discussed above I do not think this case has been properly made), then the question that arises is how long this data should be kept for. Ideally, the data to support this decision would be based on the conviction rates of such individuals at various time intervals. However, as is acknowledged in the consultation (6.9-6.11, page 15), these data are not available in the UK. The consultation then uses data from those convicted and their reoffending, as a proxy for the data required.

This is a highly controversial assertion (6.11), and is based on analysis from the Jill Dando Institute (JDI). Before discussing this analysis, it is important to consider the implications of this assertion – that Government believes that those who have never been convicted can be treated the same as those who have been convicted. This

really is an astonishing claim, and stands in stark contrast to the claim (3.1, page 8) that existence of a profile does not indicate innocence or guilt; there is clearly an implied assumption of guilt.

Leaving aside the philosophical implications, the question arises as to whether the claim is statistically valid. The work is very poor for the purposes here. I have always felt that there is an association between care for presentation and care for the underlying data, and am therefore very concerned to note that it is unclear as to whether the data presented relate to 2004,5 and 6 (as in text on page 30) or 1994,5 and 6, as in figure 3, p 15, text, p15, and the figure on page 31. This is clearly very sloppy. Given the footnote on page 31, which itself points to problems with the data, I assume the data must have been from 2004, 5, and 6, as there was no preceding bank holiday in 1996. In this case, the data only became available very recently – the time periods expired in December 2008, and one may wonder therefore how this work was ‘in parallel’ with a doctoral thesis published in 2005.

In any event, we turn to the data itself, and assume, despite presentation, that it has been correctly treated. The claim is that there is no significant difference between the ‘no further action’ (NFA) cases and the non-custodial cases, and this is indeed true. However, it is important in such investigations to consider the *power* of the statistical test –how large a difference is required for the difference to be statistically significant, given the number of cases studied. This dataset is clearly very under-powered, and the sample size is simply too small for anything but gross differences to be detected.

Because the number of cases in each category is not presented, one cannot explicitly calculate the power of this study, although the original author would have been able to do so (and indeed, should have done so). If we take the 2004 sample, as the longest time-window, and assume that the 99 cases were equally spread between the three categories, then even if the re-arrest rate of those with non-custodial sentences was as high as 64% (much bigger than the 40% for NFA), this would not give a statistically significant result!

This impact of very small numbers of cases can be clearly seen even just within the re-arrest rates of those given non-custodial sentences. Since the time period considered increases down the table, the proportions re-arrested should increase. In fact, we see that the re-arrest rate of those convicted in 2005 over 42 months is much higher than those convicted a year earlier, with an extra year to offend! This is clearly unreasonable, and this variation must be due to the small sample size in the study.

Lastly on this part of the study, the results consider only the proportion of cases **rearrested**, as opposed to the proportion of cases **convicted** after the previous arrest. The conclusion on page 31, that the NFA group is ‘roughly as criminal’ as the comparison groups, is inaccurate – no evidence of comparative criminality is presented, merely the fact that those arrested are likely to be arrested again! This may be through socioeconomic factors, or a variety of other causes, but the existence of people who are repeatedly arrested **but not convicted** is not proof of criminality.

I note that evidence on the types of offence has not yet been analysed, but that there 'is a conclusion anticipated with confidence' (p 32). This does not imply that the analysis is being done with an open mind.

Regardless of these concerns, the study then considers the subsequent criminality based on data from the doctoral thesis of Kazemian. While I have not yet finished going through the thesis itself, I note that the majority of the data are taken from French Canadian sources, rather than UK ones (see for example Kazemian 2007). One must therefore question whether it is directly applicable. It is also unclear what definition of 'criminality' has been used for this work, whether it is based on arrest data, as the previous study, or on actual conviction data. The sample size involved is also not included, making the results hard to interpret.

Home office data is presented to look at the conviction rates in the UK, (figure 1 on page 88). In contrast to the JDI study, this study correctly looks at reconviction rates. There is one simple conclusion from this nice study, which is that the likelihood of conviction or caution for those who have been convicted decreases rapidly, such that after four years there is no difference between their hazard rate and that of 16-20 year old males. As I assume there is no proposal to collect DNA for all 16-20 year old males, this then surely represents a level of convictions below which it would be disproportionate to keep DNA data, as otherwise one would be keeping DNA data on unconvicted people, who had a lower hazard rate than other unconvicted people, on whom one does not wish to keep DNA data. This clearly cannot be right.

I would therefore argue that the Home Office data argues quite clearly that a **four-year retention period is the maximum that would be proportionate**. This specifically applies for those who were convicted, and only applies to those who were arrested but not convicted if one accepts the argument from the JDI paper.

Violent, sexual and terrorism-related cases

The proposal is that these cases should have double the normal retention period, to 12 years. As is highlighted on page 15, 6.13, 'the evidence ... is unclear'. I agree entirely with this, and argue that in the absence of evidence to the contrary, such crimes should be treated in the same manner as other crimes. Following the Home Office comparison of 16-20 year old males, it is clear from figure 2 (p 89) that the hazard rate for sexual offences reaches this benchmark even faster than other offences, and that violence reaches the benchmark at almost exactly the same time.

It is essential to highlight here that we are not discussing those who have been **convicted** of such crimes, where I accept that the risk of reoffending in the same category may be higher (although interestingly the data in figure 5 on page 92 does not support this conclusion for violent offences, although it does for sexual offences). Instead, we are specifically discussing those **not** convicted, and in many cases, not even charged. I see no reason to believe why someone not charged with assault, for example, is *prima facie* more likely to offend. Until such evidence is forthcoming, I believe these cases should be treated as all other non-convicted cases.

It is instructive to compare the proposals in this consultation with that adopted in Scotland. Surprisingly, and despite the praise for the Scottish approach in *S and Marper*, there is virtually no mention of this approach. It is particularly useful to compare the approach in the case of these 'severe' cases. Scottish law provides for three-year retention (much less than proposed here), but for extensions on appeal to a Sheriff (with further appeal to Sheriff Principal), so that if the case for keeping the data for longer can be made, it is possible to keep such data for individual cases. Surely such a system could be implemented in the rest of the UK,

Children

I welcome the fact that the rules for children are to be more relaxed than for adults; this is clearly a sensible approach, and will assist with rehabilitation of such people. I am pleased that the consultation rejects the comments in the JDI report, page 33, which argues for retention of youth DNA data. Their argument is based on a trivial truth; that early onset of criminality is associated with longer criminal careers – it is very hard for a 60-year old first time criminal to have a long criminal career!

I support the proposal that DNA data should be removed at the age of 18 for convicted children, although I think this should apply to all types of conviction, possibly with application to a judge for serious cases such as rape. For non-convicted children, I support the proposal for removal of data at 18 or at the expiry of the time limit (although obviously I would like that time limit reduced as previously discussed).

However, there is one section of the recommendation on page 18, that is not discussed in the text and which I do object to. This is the suggestion that if a child is arrested twice (and not convicted on either occasion), then their DNA data would not be removed when they turn 18. I can see no reason for this additional punishment. In particular, it means that if child A is convicted at age 16 of an offence, their DNA will be removed at 18, whereas child B, who has never been convicted but has been arrested twice (possibly with no charges brought), would not benefit from this. Surely it is clear that being convicted is materially different, and worse, than having been arrested and released, even twice!

Other matters

I have focused so far on the areas of disagreement with the proposed recommendations. I should highlight that I support many of the other proposals made, but have not detailed the arguments herein. In particular, I would like to highlight my support for the proposed treatment of the samples themselves, which does go beyond the bare minimum of *S and Marper*. This is a very sensible approach. Similarly, I would endorse the proposed treatment of volunteer samples.

One last area of disagreement concerns those with DNA data already on the system, and when to remove that data. The recommendation on page 18 is that removal should occur six years after the commencement of the regulations. I can see no reason why not to remove them within six years (or whatever time limit is agreed)

from the original arrest, in other words, to treat them as though the new regulations were in force at the time of the original collection. I would accept some minimum time after the regulations come into effect for practical purposes, but can see no reason why someone arrested five years ago should have to wait an additional six years for their data to be removed.

Summary

In the absence of reliable data, policy decisions should be made to support liberty. This ensures that only proportionate decisions are taken, as in the absence of such data it is not possible to justify a claim of proportionality. Having analysed the data presented in this consultation, it is clear that there is not sufficient data to justify the proposals made. This is an important area where research is urgently needed, and would suggest that the Home Office is in an excellent position to fund and support such research.

Many of the objections I have made would be removed if there was a proper study of the usefulness of DNA data from a database in clearing up cases that would otherwise not be cleared up, and proper data on the later conviction rates of those arrested but not convicted, and how long one must wait (if at all) before that level reaches a benchmark, such as the 16-20 year old male conviction rate.

I would therefore propose a series of changes to the recommendations listed on page 18, which are detailed below. I have listed three possibilities for the first bullet point, in decreasing order of preference. In some later cases, the details would depend on the options previously chosen.

Proposed changes to recommendations, page 18.

Changes in bold.

- I: preferred Profiles **not to be retained** for persons arrested for a recordable offence but not convicted, **and for those given a caution, warning or reprimand.**
- I: alternative All profiles to be retained for **four** years for persons arrested for a recordable offence but not convicted, **and for those given a caution, warning or reprimand.**
- I: minimal All profiles to be retained for six years for persons arrested for a recordable offence but not convicted, **and for those given a caution, warning or reprimand.**
- 2 **{Adjust as per 1}**
- 3 Profiles of persons arrested but not convicted for specified violent, sexual or terrorism-related offences, **and for those given a**

caution, warning or reprimand for such offences, to be retained for **{six, four or zero, as in 1}** years.

- 4 Persons over the age of 10 years and under 18 years of age to have profiles deleted at reaching 18 years old whether or not convicted **{delete}** unless **convicted of** a subsequent offence before they reach 18, in which case the rules applicable to adults apply.
- 5,6 Both OK
- 7 Deletion of profiles on the database for persons arrested but not convicted to be applied from **{six, four, or zero, as in 1}** years from the date of **their arrest**.