

Paul Honigmann

request-15021-4295049b@whatdotheyknow.com

Our Reference: 46983

25th January 2010

Dear Mr Honigmann,

I am writing with reference to your request for information regarding the DNA database, dated 25th November 2009 made under section 1(1) of the Freedom of Information Act.

You asked for the following information:

When I asked "what was the sample size", I meant, not: "how many matches were there to THIS particular duplicate" but: "how many replicates have been checked to see if they are actually more than one person, rather than simply double-entries of the same person".

Perhaps it would help if I laid out what I'm trying to determine. Basically, it is not clear to me that the oft-stated statistic of "one in a billion" is based on fact and I'm trying to reassure myself that it is realistic. It seems to have been plucked out of the air as a "best guess" based on assumptions of independent markers according to the 2004-05 NDNAD report p.9, which refers to a "large exercise currently in progress" to confirm this which never seems to have been completed (according to your reply Reference: 46851 dated 26 Aug 2009). Thereafter the statistic appears to have been unquestioned: perhaps when responsibility for the NDNAD passed to a new body (NPIA) any concerns over the statistical integrity it is founded on were lost in the transfer. So, another way to cross check this is to determine if any of the 13% of replicates in the NDNAD are, in fact, different people rather than simply double-entries of repeat offenders etc. You would expect that if the chances of duplicates were truly 1 in a billion, there would be about 16,000 duplicates in the NDNAD (note these are true duplicates, different human beings with identical SGM+ fingerprints, not replicates). This may seem unintuitively high, but google "Birthday Paradox" and you'll see why they were

concerned on p.9 of the 2004-05 report.

So, simply tabulating the number of replicates in the DQIT's weekly report, or stating this year's total of replicates in the annual NDNAD report, tells us little. The replicates need to be physically checked by, say, the DQIT to find out if any are different people. Checking 747,000 replicates is obviously unfeasible, but I'm hoping SOME such checking has been done. For example, an alert could be automatically generated to the DQIT if the same full SGM+ profile appears in 2 different police reports in the same week. The 2004-05 report page 25 mentions that [someone] investigates if the names for a given profile are different. THIS is the sample size I am enquiring about. How many replicates have been checked to confirm they are just the same person?

1. I am writing to advise you I have established that the NPPIA holds some of the information you have requested.
2. As outlined to you in your previous responses when a subject profile is loaded to the NDNAD it will be compared against all other subject profiles held on the NDNAD.
3. The National DNA Database Unit's Data Quality and Integrity Team (DQIT) investigate matching SGM Plus profiles that are shown on different Police National Computer (PNC) records. The investigation is carried out to see if the DNA samples relate to identical siblings, the same individual who may have provided alias details or if an adventitious match has occurred. Each record on PNC has a unique number called a PNC ID; this number is also held on the NDNAD. DQIT therefore only investigate matching SGM Plus profiles with different PNC ID's, as replicates with the same PNC ID can be identified as being from the same person.
4. The investigation into each replicate will vary but may consist of requesting fingerprint comparisons, comparing PNC records, requesting photographs that were taken at the same time as the DNA sample, requesting DNA/PACE cards and/or speaking to the Police Force that owns the DNA sample to gain any additional information.
5. If there is enough evidence to confirm that the DNA samples were taken from the same individual the DQIT will request the owning Police Force to merge the 2 PNC records. An example of this evidence would include matching fingerprints on the two PNC ID records. If there is less evidence but there is a strong indication that the samples were taken from the same individual then an information marker is put onto PNC by the DQIT to alert any officer viewing the records that the two PNC records may relate to the same individual. If the records relate to identical siblings DQIT will put an information marker on PNC to that effect.

6. If the samples are not believed to be taken from the same individual or identical siblings they are re-analysed at additional areas of DNA that are not examined with the SGM plus test type to eliminate the match or provide more assurance of association.

7. Detailed records of the number of replicate profiles DQIT have checked have been recorded since January 2007.

8. In the period 01/01/07 – 10/12/09 DQIT have investigated a total of 7165 replicate profiles. This figure includes some profile records with the same PNC ID which would be known to be the same individual. However they would be included in investigations because another sample in the same replicate group has a conflicting PNCID. From these 7165 replicate profiles they have identified 1092 sets of twins, 3 sets of triplets and 1183 sets of profiles which are from the same individual (for which the records have been merged on PNC).

9. In your previous response, reference 46851, we provided details that at that point one match had been confirmed, where two siblings have the same DNA profile, when the SGM Plus test type has been used. This match was identified prior to 2007.

10. Since the earlier response was provided another SGM plus match where two people have the same DNA profile, when the SGM Plus test type has been used has been confirmed. The two samples in this second match were again taken from non-identical siblings. Further work carried out by Forensic Provider laboratories on both DNA samples looking at additional DNA markers has shown a difference in the DNA profile of these individuals.

11. To put these figures into context, each subject profile loaded to the NDNAD will be compared against all other subject profiles held on the NDNAD. As at 30/09/09 the database held 5.89 Million records. The records which have the same SGM plus profiles but different PNC ID's will then be investigated as stated in paragraphs 3-6. The total comparisons between all subject profiles loaded to the NDNAD and subsequent investigations to date have resulted in 2 SGM Plus matches where the profiles were obtained from non-identical siblings. Further DNA investigations, as part of our processes (as stated in paragraph 6) have shown differences outside the DNA areas examined in SGM Plus.

12. In your request you have referred to the SGM Plus match probability of '1 in a billion'. The match probability is an estimate of the evidential significance of a match between the DNA profiles of a suspect and a crime stain. This assesses the probability of obtaining the match if the stain did not originate from the suspect but came from another unknown individual with the same SGM Plus profile. The size of the match probability depends on whether the 'unknown' individual is related to the suspect however in most case circumstances, it is normal practice to consider

that the 'unknown' individual is unrelated to the suspect. Under these circumstances it was shown to be fair, reasonable and conservative to assign a match probability of 1 in a billion (1000 million). This figure was initially derived from statistical assessment of an allele frequency database within a controlled sample group by The Forensic Science Service (FSS) '*Statistical analyses to support forensic interpretation for a new ten-locus STR profiling system: International Journal of Legal Medicine (2001) 114:147-155 L. A. Foreman I. W. Evett*'. It has been subsequently adopted by other Forensic Service Providers in the UK providing evidential information to the CJS.

13. Based on this work it is known that the SGM Plus system is an extremely powerful discriminating tool. The NDNAD provides information on matching DNA profiles to the police for evaluation by forensic scientists in the context of all the case information. A SGM Plus DNA profile is not considered to be conclusive proof of identity and any DNA matches between subjects profiles and crime-scene profiles require corroboration by other means.

14. It is predicted that as the size of the DNA database increases, there will be a greater likelihood that an adventitious match between unrelated individuals will be encountered. This is largely due to the effect of trillions of between-profile comparisons which are carried out. It is possible to identify potential adventitious matches between subject samples held by the NDNAD by confirming the identity of the subjects; as such, potential adventitious matches of this type are investigated. It is also possible to further discriminate between DNA samples using an alternative DNA profiling system which tests DNA samples at additional areas to the SGM Plus system. To date no two full SGM Plus (10 STR areas of DNA under test plus sex marker) profiles held by the NDNAD have been confirmed to have originated from unrelated individuals and there are only two occasions where matching SGM Plus profiles have been obtained from non-identical siblings. These observations provide assurance that the SGM Plus DNA profiling system is an extremely powerful discriminating tool.

15. Further statistical information relating to the NDNAD is published in the National DNA Database Annual Reports the latest of these is available for download on the NPPIA web-site at <http://www.npia.police.uk/en/11403.htm> . Previous years reports are available for download from the Home Office web-site at <http://www.homeoffice.gov.uk/science-research/using-science/dna-database/>.

Your right to complain

We take our responsibilities under the Freedom of Information Act seriously but, if you feel your request has not been properly handled or you are otherwise dissatisfied with the outcome of your request, you have the right to complain. We

National Policing Improvement Agency

will investigate the matter and endeavour to reply within 3 – 6 weeks. You should write to:

David Horne
Director of Resources
National Policing Improvement Agency
10-18 Victoria Street
London
SW1H 0NN

E-mail: david.horne@npia.pnn.police.uk

If you are still dissatisfied following our internal review, you have the right, under section 50 of the Act, to complain directly to the Information Commissioner. Before considering your complaint, the Information Commissioner would normally expect you to have exhausted the complaints procedures provided by the NPIA. The Information Commissioner can be contacted at:

FOI Compliance Team (complaints)
Wycliffe House
Water Lane
Wilmslow
Cheshire SK9 5AF

Further information about the NPIA is routinely published on our website at www.npia.police.uk or through our publication scheme. If you require any further assistance in connection with this request please contact us at our address above.

Yours sincerely,

NDNAD Delivery Unit
NPIA