MEASURING THE EFFECTIVENESS AND EFFICIENCY OF FORENSIC DNA-DATABASES

Dr. Ir. C.P. (Kees) van der Beek MBA
Custodian Dutch DNA-database
Netherlands Forensic Institute
PO Box 24044, 2490 AA The Hague, The Netherlands
+31 6 18302465; k.v.d.beek@nfi.minvenj.nl

INTRODUCTION

Many countries operate forensic DNA-databases to identify owners of crime related stains. Politicians, who provide the funds for the operation of these DNA-databases, of course want to know if this is an effective and efficient instrument in the fight against crime. This summary of my ISHI 2015 presentation describes how the effectiveness and efficiency of forensic DNA-databases can be measured. Parameters for the mutual comparison of forensic DNA-databases are derived from a theoretical model and are illustrated by examples from existing forensic DNA-databases. In addition a model is presented by which the efficiency of a forensic DNA-database can be compared with other ways which the police uses to identify possible suspects. Last but not least an interesting case will be presented which only could be solved by a DNA-database match and not by any other means and hence a priceless match.

DEFINITIONS

If we want to know if an instrument is effective we have to investigate if the instrument does what it is supposed to do. If we want to know if an instrument is efficient we have to investigate how well the instrument performs in relation to similar or alternative instruments.

EFFECTIVENESS OF A FORENSIC DNA-DATABASE

The main purpose of a forensic DNA-database is to identify unknown donors of crime related DNA-material. To achieve this, a DNA-database should contain DNA-profiles of known persons as well as DNA-profiles of crime related samples. This sounds self-evidently, but in the semi-annual overview of European DNA-databases complied by the ENFSI DNA Working Group shown below, we can see that in December 2014, the DNA-database of Greece did not yet contain DNA-profiles of known persons and hence was not an effective DNA-database.

Country	Population size	Persons				Stains	Matches					Date
		A	S	CO	T		Person/Stain			Stain/Stain	Total	
							S	CO	T			
Albania	3.600.000											
Armenia	3.000.000											
Austria	8.100.000				187.331	72.277			20.380	9.137	29.137	dec-14
Belgium	10.400.000		617	30.723	31.340	39.151	540	2.645	3.185	6.268	9.453	dec-14
Bosnia & Herzegovina	4.400.000											
Bulgaria	7.900.000											
Croatia	4.300.000				31.199	5.320					6.452	apr-14
Cyprus	772.000			389	389	13.053	115	129	244	49	293	dec-14
Czech Republic	10.528.500		3.181	134.294	137.475	22.486			8.716	3.103	11.819	dec-14
Denmark	5.500.000				105.824	36.201			24.176	4.509	28.685	dec-14
Estonia	1.315.819				46.494	10.551			4.913	998	5.911	dec-14
Finland	5.475.866				150.188	18.057			21.288		21.288	dec-14
France	64.300.000	135.412	2.280.448	472.505	2.752.953	254.038	91.594	24.971	116.565	15.897	132.462	dec-14
Georgia	4.700.000		746		3 342	197						dec-15
Germany	80.200.000				832.695	264.847			148.728	38.464	187.192	dec-14
Greece	10.600.000					12.789				977	977	dec-14
Hungary	9.982.000		100.610	20.155	120.765	5.412			917	264	1.181	dec-14
Iceland	315.000											
Ireland	4.200.000											
Italy	58.000.000											
Kosovo	1.800.000											
Latvia	2.400.000		41.305	10.061	53.327	5.853			1.847	294	2.141	dec-14
Liechtenstein	37.000											
Lithuania	2.960.000				76.895	4.522			2.375	438	2.813	dec-14
Luxembourg	550.000		147	1.970	2.117	3.075	82	163	235	2.379	2.614	dec-14
FYR Macedonia	2.100.000				12.363	3.815			1.229	183	1.412	dec-14
Malta	400.000		30		30	423			1		1	dec-14
Montenegro	650.000											
Netherlands	16.100.000				201.328	61.854			46.652	5.216	51.868	dec-14

THEORY

In 2010 Simon Walsh[1] and his colleagues have published a formula which describes the output of a DNA-database:

$$H = \frac{\alpha N}{M} \times \omega C$$

H = number of hits/matches
N = number of person profiles in the DNA-database
M = active criminal population
C = number of crime related profiles on the DNA-database
α = quality factor (person sampling)
ω = quality factor (crime/exhibit sampling)

From this formula two efficiency parameters can be derived:
*   H/C: the number of stain-to-person-matches relative to the number of stains included in the DNA-database.
*   H/N: the number of stain-to-person-matches relative to the number of persons included in the DNA-database

H/C

This parameter expresses the chance that a stain profile which is included in the DNA-database will match a subject profile (% stains matching a person). This parameter is often called the hit-rate of a DNA-database. Rewriting the formula above leads to: H/C=αωN/M. So H/C is expected to increase if the number of persons in the DNA-database (N) increases (given that α remains the same). In the tables below which are taken from the annual reports of the DNA-databases of the United Kingdom and the Netherlands, it can be seen that as more persons are added to the DNA-database over the years, also H/N increases.

[1] S.J. Walsh et al (2010) Modeling forensic DNA-database performance. J. For. Sci. 55(5) 1174-1183
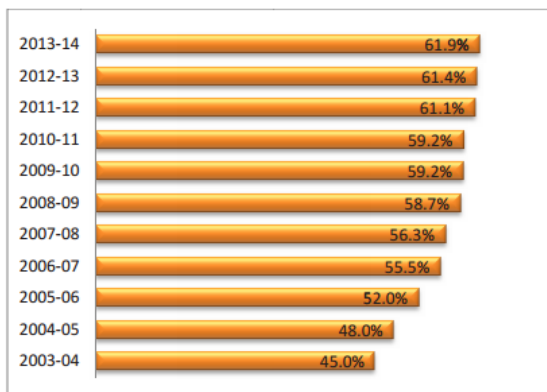
Year	%
2006	32
2007	36
2008	40
2009	42
2010	44
2011	47
2012	50
2013	52
2014	54

H/C in the United Kingdom   H/C in the Netherlands

In the Netherlands about 30% of the stains immediately matches a person when the stain profile is included in the DNA-database. The remaining 24% matches at a later date (sometimes many years later) when the matching person is included in the DNA-database. At some point in time the increase in the number of persons will be neutralized by the number of persons that stop committing crimes but are still on the DNA-database. H/N will then be determined by ω (the % of stains that will never match a person). The UK seems to have reached this stage as the increase of H/N is much less in the past five years.

H/N

This parameter expresses the % of persons that ever has caused matches. For H not the total number of matches should be taken but the number of clusters (stains matching the same person) In the Netherlands about 11% of all persons in the DNA-database has ever matched one or more stains. In the UK 1,7 (of over 7 million) person profiles had to be removed from the DNA-database after the introduction of the "Protection of Freedoms Act". So N was decreased by 24% and it was interesting to see what the effect would be on H (the number of matches). In the quarter from 1 April to 30 June 2014, the UK-DNA database produced 37 matches to murder, 127 to rapes and 6,111 to other crime scenes. In the same quarter of 2013, before 1,7 million DNA-profiles of persons were removed, it produced 37 matches to murder, 103 to rapes and 6,141 to other crime scenes. So apparently the persons of which the DNA-profiles were removed did not contribute much to the number of matches. As a result of the removal the percentage of people on the UK DNA-database that can match stains (α) has increased.

WARNING

H/C and H/N can be used to compare the efficiencies of different DNA-databases. However the data which are semi-annually collected by the ENFSI DNA Working Group on the numbers of profiles and matches in European DNA-databases cannot be used for this purpose for the following reasons:

- There are countries that add more than one DNA-profile of a person to their DNA-database
- There are countries that add more than one unique DNA-profile of a stain per crime scene to their DNA-database
- Countries use different removal regimes for DNA-profiles of stains: never or immediately after a match with a person or after the authorities have dealt with the match with the person
- Countries use different removal regimes for DNA-profiles of persons: After some storage time and/or if a person is not prosecuted or convicted
- Stain-to-person matches can be so-called "cold hits" (matches for which no suspect was known) or matches where both a stain and a suspect were added to the DNA-database
- One person can match more than one stain

COMPARISON TO OTHER INSTRUMENTS

A DNA-database match means that an as yet unidentified suspect is identified. So the cost of an identification by means of the DNA-database could be compared with the cost of other ways which the police uses to identify suspects. The cost of a DNA-database identification can be calculated by dividing the total annual cost of a DNA-database operation by the annual number of stain-to-person matches. Apart from the DNA-database operation itself also other costs may be included in the calculation:

- The cost of making the DNA-profiles of persons for the DNA-database
- The cost of making the DNA-profiles of stains for the DNA-database
- The cost of making all DNA-profiles
- The cost of the police related to the collection of stains and reference samples

In the Netherlands the annual cost of the DNA-database plus the production of person profiles is about 2.500.000 euro and there are about 5000 stain-to-person matches per year. So in the Netherlands the price of a DNA-database identification is about 500 euro. But some matches are priceless as is shown in the example below.

A PRICELESS MATCH

In 2013 an abandoned baby boy was found in the city of Roermond in the Netherlands. From the towel in which the baby was wrapped a DNA-profile from its mother could be obtained. Her profile was included as a stain in the Dutch DNA-database because abandoning a child is punishable by a 4,5 year jail term. The Netherlands compares its DNA-profiles also with a number of other European countries. So the profile was also sent to these countries and a match with a stain was obtained in the German DNA-database. This DNA-profile also proved to come from a towel in which an abandoned baby girl was found in 2011 in Germany. So by means of this match it was shown that the abandoned boy from the Netherlands is the brother of the abandoned girl from Germany. As a result of the media attention, a person claiming she was the mother came forward and further DNA-testing proved that she indeed was the mother.