ESTIMATION OF GENETIC DIFFERENCE BETWEEN SUBPOPULATIONS USING DNA PROFILING DATA FROM CRIME INVESTIGATIONS.

M.J.GREENHALGH AND C.M.BUFFERY.

Metropolitan Police Forensic Science Laboratory
109, Lambeth Road, London, ENGLAND. SE1 7LP

Introduction.

The evidential value of a DNA profile match is usually expressed to the court in terms of the estimated frequency of occurrence of that particular profile. ie. How many <u>other</u> people in the relevant population would be expected to have that profile. The chances of obtaining matching DNA profiles over the commonly used loci are extremely low in all populations that have been examined though estimating the degree of rarity has caused considerable debate.

At a meeting held earlier this year in London (The use of DNA Statistics in crime cases - MPFSL) it was suggested by Richard Nichols and David Balding of Queen Mary and Westfield College (London University) that police enquiries which lead to the arrest of the wrong suspect for a crime are more likely to lead to an individual from the same subpopulation group as the true offender. For example the innocent suspect may physically resemble the criminal or be a member of the same small and possibly inbred group living in that locality.

If there were to be a higher degree of kinship amongst the group containing the true offender, then the probability of finding another matching profile would be greater. Furthermore the distribution of alleles may be different and thus the figure for the chance occurrence of the profile which is calculated from general population data and uses the product rule could be unfair to the accused.

At the Metropolitan Police Laboratory we have collected DNA profiles in the course criminal investigations over the last four years. In two cases, which are the subject of this study, profiles have been obtained from large numbers of suspects which have then been shown not to match the "target" crime profile. These innocent suspects provide data to test whether the Nichols and Balding hypothesis has a significant effect in reality.

Methods.

Two cases were chosen where a DNA profile from the criminal had been prepared and the police had a description of the offender and knew in which area of London he was likely to live. In both cases a large number of suspects were apprehended and these were later shown to be innocent of the crime by DNA profiling. We have treated these individuals as a subpopulation selected by the police on the grounds of physical description and area in which they live.

Case 1:

This case consisted of a series if rapes committed by one individual in a residential area of South London. The assailant was described as Afro Caribbean and approximately 30 years old. Over 200 suspects from the local area were arrested and profiled before a match was found. Additional evidence confirmed the guilt of this individual.

Case 2:

This case consisted of a series of attacks on elderly women in another area of South London. In this instance the offender was Caucasian and again over 200 suspects from the local area have been profiled. No match has been found as yet.

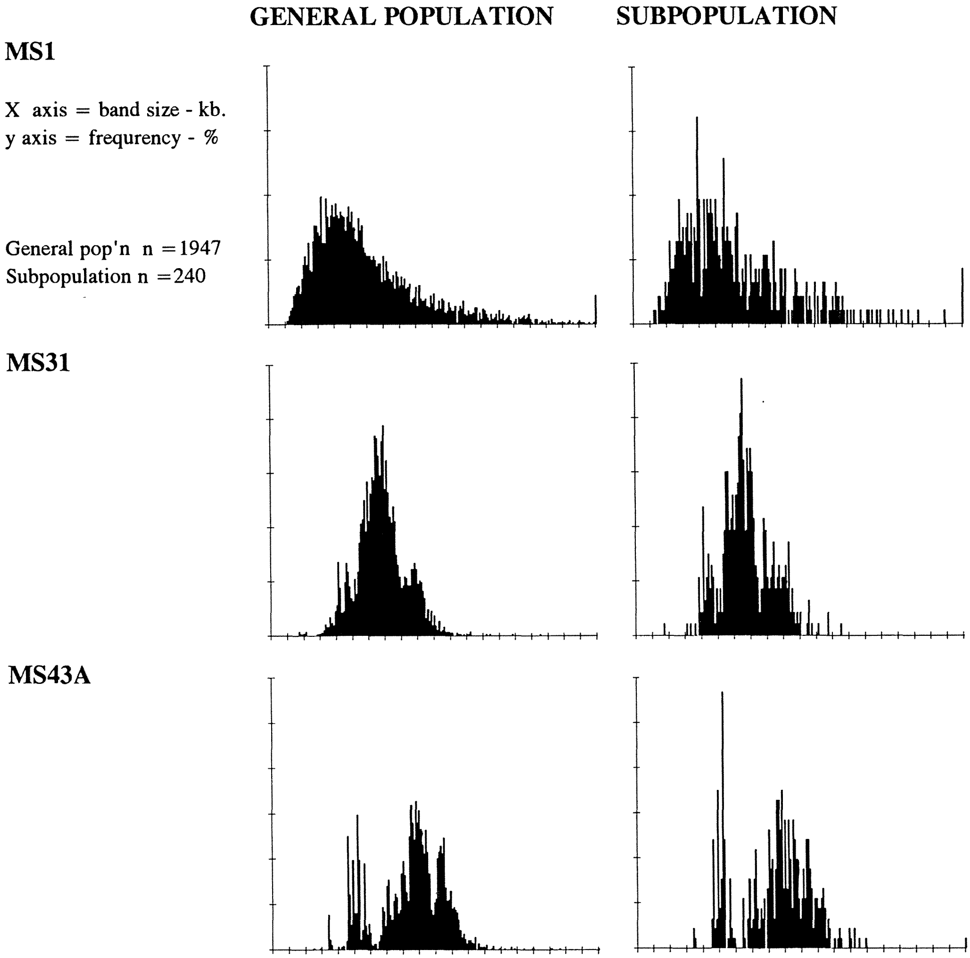
For each case the DNA profiles of the "subpopulation" individuals were plotted to generate the allele distributions of the four loci used at this lab. These might then be considered the "correct" databases for the two cases. Comparisons were made with plots of the remaining individuals who form the broad general databases.

For the two cases, we have taken the target crime profiles and made estimates of the occurrence of the profile using the "correct" database for that subpopulation and compared this with the value obtained using the general database. We have also included examples of the frequency calculation using F_{ST} coefficients as a means of modifying the probabilities from the general database..[Nichols R.A. & Balding D.J.(1991) *Heredity,* **66**, 297-302]

Finally on the assumption that greater relatedness amongst a subpopulation might lead to an increased sharing of bands, we have used the "between person" comparison method [Evett I.W. Scranage J.K. & Pinchin R.(1992) *Journal of the Forensic Science Society,* Vol 32,(4),307-326]. to see if it is possible to detect any differences between the subpopulation and the general population.

Results.

1). Plots of the allele distributions at three of the loci are here shown for the Caucasian case.



The results from the two populations appear very similar. The smaller size of the subpopulation databases (approximately 200 individuals for both cases) makes a less smooth distribution and some of the more defined low molecular weight alleles (see MS43A) appear to be relatively more common in the subpopulation. The results for the Afro Caribbean population are also very similar when general and subpopulations are compared to each other.

2). For both cases, estimates were made of the chance occurrence of the target crime profile. As can be seen in the table below, the differences that occur when the "correct" subpopulation database is used as opposed to the general database are relatively small. Examples of the values obtained when the F_{ST} coefficient is used to modify the general database results are given in the two right hand columns. These arbitrary levels of F_{ST} which are used in casework at the Metropolitan Police Laboratory can be seen to be extremely cautious. The true value of this parameter in urban populations is probably at least an order of magnitude less.

Comparison of calculations using the general and subpopulation databases:

	Subpopulation database	General database	1% F_{ST}	5% F_{ST}
Caucasian case	1 in 17,000,000	1 in 27,000,000	1 in 7,000,000	1 in 200,000
Afro Caribbean case	1 in 3,000,000,000	1 in 3,000,000,000	1 in 200,000,000	1 in 2,000,000

3). The "between person" comparisons method takes each profile in a database and compares it with every other. In each of our subpopulations this amounts to more than 20,000 comparisons. Likelihood ratios for each comparison are calculated and the results plotted graphically. It is important to emphasise that using four loci, none of the profiles matched any others within the databases. In order to detect any small effects of increased band sharing, the analysis was performed on each locus individually. Random samples of the same size as the subpopulations were selected from the general databases to act as controls.

For all the loci and populations examined the probability of obtaining a likelihood ratio of greater than 100 (medium strength evidence) was extremely low. There appear to be no detectable differences between the subpopulations and the general databases using this test.

Conclusion.

That human populations are substructured and do not interbreed randomly is self evident. The question that we as forensic scientists need to answer is whether this significantly affects the strength of the evidence we present.

The results from these two cases suggest that in British urban populations of Caucasian and Afro Caribbean people, the effects of subpopulations defined by place of residence and general appearance are insignificant. However in many forensic cases we are unable to accurately define the population from which the offender or innocent suspects originate, and it is for this reason that we at the Metropolitan Police Laboratory use a modification of the product rule. In all but the most unusual cases this is likely to be highly conservative.

Acknowledgment: We are very grateful to John Scranage of the Forensic Science Service for his assistance in analysing the data.