Int J Legal Med (1999) 112 : 176–180

ORIGINAL ARTICLE

Marjan Sjerps · Ate D. Kloosterman

On the consequences of DNA profile mismatches for close relatives of an excluded suspect

Received: 8 December 1997 / Received in revised form: 16 July 1998

Abstract If the DNA profiles of a crime stain and the reference sample from the suspect do not match, the suspect is excluded as the donor of the crime stain. However, in some situations the DNA evidence can suggest that a close relative of the suspect might match the stain, in particular when the reference sample from the suspect and the crime stain share rare alleles. This finding can be important for the authorities. The forensic scientist has to decide whether or not to notify the authorities in these circumstances. To the best of our knowledge there is not yet an objective rule for making this decision. We propose such a decision rule for brothers of the suspect, investigate its performance and address some ethical, legal, and practical aspects. Our calculations can be simply adjusted for other relatives of the suspect.

Key words Forensic DNA typing · DNA profile · Exclusion · Relatives · Mismatch

Introduction

In forensic DNA typing the DNA profile of the crime stain is usually compared to that of the reference sample from the suspect. If these profiles do not match the suspect is excluded as the donor of the crime stain. However, since DNA evidence concerns inheritable features, there can be situations in which the two non-matching DNA profiles suggest that a close relative of the suspect might match the crime stain. This holds in particular for cases in which the two non-matching DNA profiles share several very rare alleles. This information may be important in a criminal investigation.

Evett (1992) has considered a situation in which a suspect matched the crime stain, and the defence was “it was my brother”. Others have investigated the effect of relatives on the strength of the evidence against a suspect which matches a crime stain (e.g., Balding and Nichols 1994; Brookfield 1994; Donnelly 1995; NRC 1996). As yet, however, the implications for relatives in case the suspect does not match the crime stain are rarely investigated and it is not clear if and when the forensic scientist should inform the authorities. Obviously, it is not desirable that the authorities are informed if in fact the crime stain was left by an unrelated person (false alarm). The forensic scientist thus needs an objective decision rule when to inform the authorities, with a low rate of false alarms.

In this paper, we develop such a decision rule for brothers of the suspect based on a pre-set low level of the false alarm rate. The rule can easily be extended to other relatives. We investigate the detection chance for brothers actually matching the crime stain by illustrating our method with a real case example. Furthermore, we discuss the effect of the population of the offender on the calculations. Ethical and legal issues and practical consequences are addressed in the discussion.

Definition of an objective decision rule

Suppose that the DNA profiles of a semen stain from the offender and the reference sample from the suspect do not match. Let X be the profile of the semen and Y that of the suspect ($X \neq Y$). Furthermore, suppose that the population in which the crime has occurred is not highly inbred. We can then calculate the likelihood ratio (LR) for the hypothesis that a brother of the suspect left the semen stain against the hypothesis that an unknown unrelated individual left the stain:

$$LR = \frac{\text{P[suspect has type Y, semen stain is type X} \mid \text{semen stain was left by a brother]}}{\text{P[suspect has type Y, semen stain is type X} \mid \text{semen stain was left by an unrelated individual]}} \quad (1)$$

M. Sjerps (✉) · A. D. Kloosterman
Gerechtelijk Laboratorium, Volmerlaan 17,
2288 GD Rijswijk, The Netherlands
e-mail: sjerps@bart.nl
Tel. +31 70 4135353/139; Fax +31 70 4135454

This can be rewritten as:

$$\frac{\text{P[semen stain is type X | semen stain was left by a brother]}}{\text{P[semen stain is type X | semen stain was left by an unrelated individual]}} \qquad (2)$$

multiplied by

$$\frac{\text{P[suspect has typeY | semen stain is type X and was left by a brother]}}{\text{P[suspect has typeY | semen stain is type X and was left by an unrelated individual]}} \qquad (3)$$

Under the assumption that in the denominator there is no correlation between the DNA profile of the donor of the stain and that of the suspect, which may not be true in case of informative search protocols (Balding and Donnelly 1995), and ignoring differences between populations for the moment, the LR can be approximated by

$$LR \approx \frac{\text{P[suspect has typeY | brother has type X]}}{\text{P[suspect has type Y]}} \qquad (4)$$

Under the assumption of independence of alleles within and between loci, this ratio can be estimated if an appropriate database is available (see e.g. the formulae in Weir 1996). The likelihood ratio can be interpreted as follows: the observation that the suspect's DNA type is Y and that of the semen stain is X is LR times as likely if the semen stain was left by a brother of the suspect than if the semen stain was left by an unrelated individual. Hence, a LR larger than 1 can be considered as evidence against a brother of the suspect. The magnitude of the LR can thus be used to define an objective decision rule for informing the investigating officer.

This definition must satisfy two conditions: firstly, the investigating officer should only be notified if there is fair or strong evidence against a brother of the suspect and secondly the rate of false alarms should be low. The first condition involves choosing a minimum level for the strength of the evidence, such that the investigating officer is only informed if this level is exceeded. Since the LR is a measure of the strength of the evidence, this involves choosing a minimum level, LR_{min}.

The second condition involves choosing an acceptable upper limit for the percentage of times that the authorities are informed when the suspect's brothers are not the donors of the crime stain (false alarm). This is equivalent to setting an upper limit to the type I error level in conventional statistical hypothesis tests. Let α be the chosen percentage, e.g. 1%, and let LR_{crit} satisfy

$$\text{P}[LR > LR_{crit} \mid \text{semen stain was left by an unrelated individual}] \le \alpha. \qquad (5)$$

The second condition is satisfied if the authorities are informed only if LR_{crit} is exceeded. LR_{crit} can be determined using the frequency distribution of the LRs in an appropriate population database. This distribution can be estimated by calculating for each possible genotype an estimate of its frequency and the LR. LR_{crit} is defined as the value above which a percentage α of LRs lies. Since the LRs depend on X, the crime stain profile, LR_{crit} will also depend on X.

Combining the two conditions above, we come to the following decision rule: notify the investigating officer if, and only if, LR exceeds both LR_{min} and LR_{crit}. We will now study the performance of this rule.

Performance of the method

The chosen levels of α and LR_{min} determine the detection probability β i.e. the probability that the authorities are informed if indeed a brother of the suspect left the semen stain. The probability 1-β is comparable to the type II error in conventional statistical hypothesis tests. The detection probability β can be determined using the frequency distribution of persons whose brothers have DNA type X. This distribution can be estimated by calculating an estimate of the frequency (given a brother with type X) and the LR for each possible genotype. Then β is defined as the percentage of LRs larger than LR_{crit} and LR_{min}. Since the LR_{crit} depends on the crime stain profile X and the chosen level of the false alarm rate α, β depends on X, α and LR_{min}. For a specific profile X, one can make a table for various choices of LR_{min} and α.

Example

The DNA profiles of samples in a particular case, submitted to the Dutch Forensic Science Laboratory are given in Table 1. For each locus the LR concerning a brother of the suspect is calculated by using the allele frequencies in a sample from the Dutch Caucasian population (Sjerps et al. 1995; Ovington et al. 1997; Kloosterman manuscript in preparation).The combined LR is obtained by multiplying the LRs for each locus which yields 9653. Figure 1 shows the estimated frequency distribution of the LR in the Dutch Caucasian population. This frequency distribution is obtained by calculating the LR and the frequency of occurrence (estimated by multiplying allele frequencies) for each possible genotype. It is estimated that 95% of the Dutch Caucasians have a LR smaller than 1, and 99.995% have a LR smaller than 9653. LR_{crit} is found using the fre-

Table 1 the DNA profiles in a rape case, and the LR for each locus

Locus	Semen stain	Suspect	LR
HUMTHO1	6/10	6/ 8	0.8
HUMFES/FPS	11/11	10/11	0.8
HUMVWA31/1	14/18	14/17	2.2
HUMF13A1	5/ 6	6/ 7	0.7
D21S11	63/67	63/67	9.6
D18S51	14/15	12/14	0.9
D8S1179	13/16	13/16	20.4
HUMFIBRA (FGA)	25/26	25/26	57.6

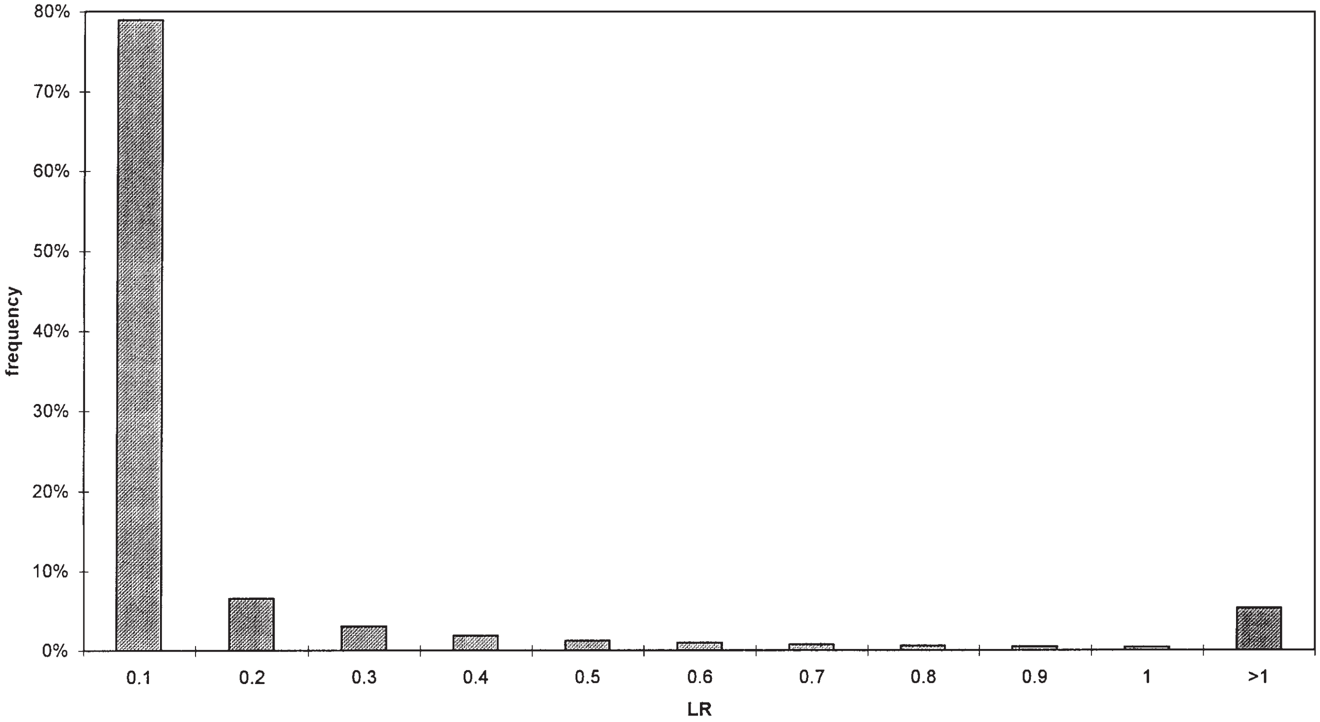

Fig. 1 Histogram of the frequency distribution of likelihood ratios of Dutch Caucasians, based on the DNA profile of the crime stain in Table 1

quency table corresponding to Fig. 1. A false alarm rate $\alpha = 1\%$ corresponds with $LR_{crit} = 8.7$ and $\alpha = 0.1\%$ corresponds with $LR_{crit} = 105.2$.

Figure 2 shows the estimated frequency distribution of the LR in Dutch Caucasians who have a brother with the same DNA profile as the crime stain. The detection probability β is found using the frequency table corresponding to Fig. 2. Table 2 gives the values of β for several values of α and LR_{min}. The high LR (9653) in this particular case is mainly due to the rare alleles shared by the DNA pro-

Fig. 2 Histogram of the frequency distribution of likelihood ratios of Dutch Caucasians, given that their brother matches the crime stain, based on the DNA profile of the crime stain in Table 1

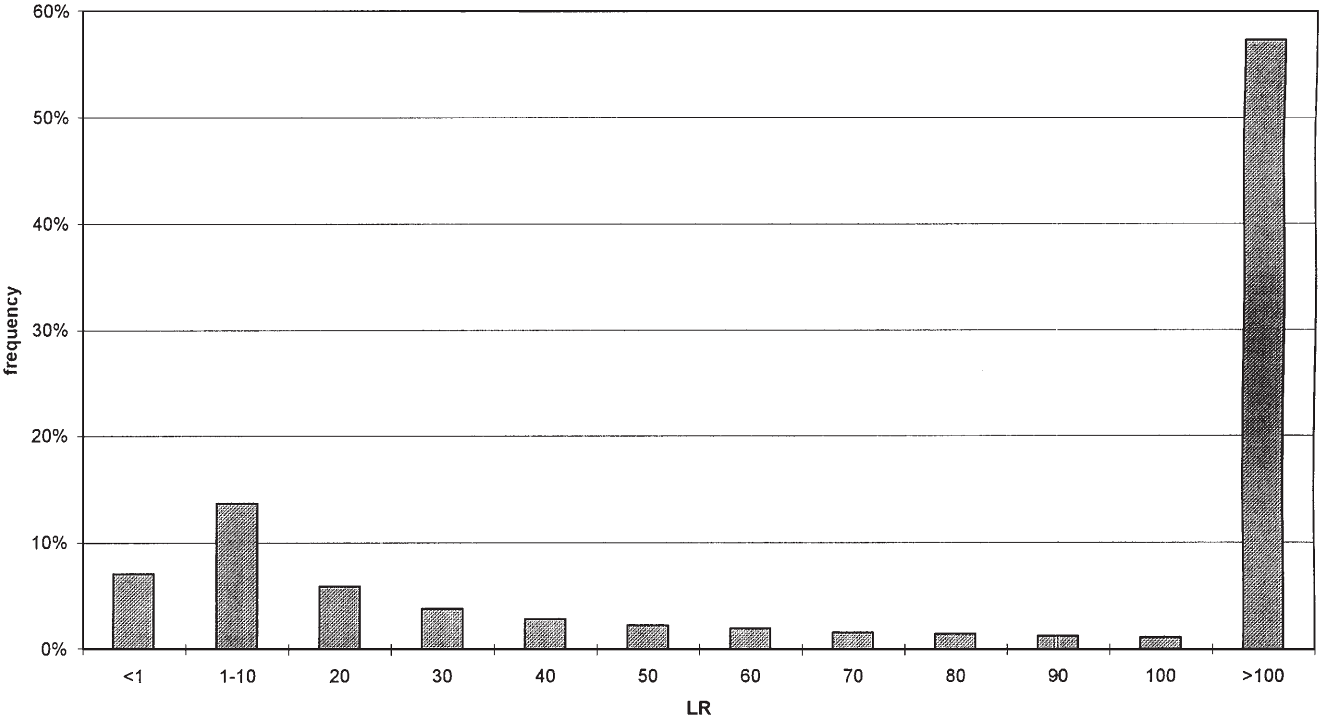

Table 2 The detection probability β as a function of α and LR_{min}

α	LR_{min}		
	10	33	100
0.1% (LR_{crit} = 105.2)	57%	57%	57%
1% (LR_{crit} = 8.7)	79%	69%	57%

files of the crime stain and the suspect at loci D8 and FGA. If in this case example the suspect's profile had the most common alleles at the loci D8 and FGA (13/14 and 21/22 respectively), the LR would have been 1.27. Hence the evidence becomes so weak that for reasonable choices of LR_{min} the authorities would not have been notified. If the suspect's profile had been the most common profile in the Dutch Caucasian population, the LR would have been only 0.15. This LR is in favour of the hypothesis that the suspect's brothers did not leave the crime stain, as expected.

Population issues

Since allele frequencies differ between populations, the LR will depend on which population data are used. Depending on the genetic differences between the populations, the LRs may differ considerably. To see which populations are relevant, we will consider two situations: (a) it is known to which population the donor of the crime stain belongs and (b) it is unknown to which population the donor of the crime stain belongs. In situation (a) the suspect must also be a member of the donor population. Hence, the ratio in Eq. (2) is 1 and the LR is, under some assumptions of independence (Balding and Donnelly 1995), given by Eq. (4). It is important to note, however, that Eq. (4) is not valid in case the crime has been commited in a highly inbred population.

In situation (b) we calculate the LR for several possible donor populations according to the NRC report II (1996). Let p_X be the frequency of type X in the suspect's population, and let q_X be the frequency in the donor population. Then Eq. (2) can be written as p_X/q_X and similarly Eq. (3) can be written as $p_{Y|\text{brother type }X}/p_Y$, so that Eq. (4) is replaced by the product of these two ratios. From this it can be seen that if the donor population is unknown, the LR depends on the allele frequencies in the suspect's population. Note that this differs from situations in which the suspect matches the crime stain, since in these cases one has to report the profile frequency in several populations (NRC report II 1996).

Discussion

In criminal cases where the suspect does not match the crime stain but where the DNA profiles indicate that a close relative of the suspect might be the actual donor of the crime stain, the forensic scientist needs an objective decision criterion whether to inform the authorities. In case the crime has not been commited in a highly inbred population group, we have shown that the likelihood ratio (LR) can be used as such for brothers of the suspect. Our method can easily be adjusted for other relatives. By doing so, the frequency of "false alarms" is below a pre-set low threshold, with a reasonable detection probability of brothers matching the crime stain. This decision rule may also be used as a guideline for deciding to perform additional tests (for example Y-chromosome specific STR loci).

The decision to inform the investigating officer has several ethical and legal aspects (Schneider et al.1997). Using this method, the consequence may be that innocent relatives become suspects, merely because of their relatedness to the primary suspect. This may have serious consequences for these relatives. On the other hand, if the forensic scientist has valuable information in a particular case, it is unjustified not to share this with the authorities. Although to the best of our knowledge there are no reliable data on the frequency of arrests of close relatives of an offender, such cases may not be unusual (Balding et al. 1994; Lempert 1994). Hence, there may be quite a number of cases in which this kind of information is valuable.

It is questionable whether a high LR is sufficient justification for the investigating officer to obtain a DNA reference sample from relatives of the suspect. In the Netherlands this can only be taken under very strict regulations, in particular there must be a strong reason for suspicion (Kloosterman and Janssen 1997). As yet, there are no guidelines whether a high LR is sufficient reason. In addition, if the DNA profiles cannot be obtained from the relatives of a suspect, it is unclear whether the DNA profile of the exonerated suspect can be used as evidence against a close relative.

The choice of the "false alarm" frequency α and the minimum evidence strength LR_{min} is also an ethical question. This choice is the responsibility of the legal authorities and not of the forensic scientist. However, no guidelines are available at this moment. The forensic scientist therefore needs to contact the proper authorities and agree with them on these thresholds. Once these thresholds have been established, the forensic scientist can use the resulting rule.

A practical point for the forensic scientist is the clear explanation of the impact of the results to the authorities. Suppose for example that in a rape case the profiles from the suspect and the semen stain differ by both alleles in several loci. We suggest to use the following expression: "We observed that the semen stain has type X and the suspect has type Y. Hence, the suspect did not leave the semen stain. The father and any sons of the suspect can also be excluded as the donor of the semen stain. However, our observation is 1000 times as likely if a brother of the suspect left the crime stain than if an unrelated man left the crime stain. We consider this as a strong indication that a brother of the suspect has left the crime stain. Care must be taken with the interpretation of the results if the crime was committed in a highly inbred population group."

In conclusion, we propose that this method should become a standard procedure for the forensic scientist in cases where the suspect does not match the crime stain. The legal authorities should set an acceptable upper limit to the frequency of "false alarms" and the minimum strength of the evidence. Furthermore, they must decide whether this type of indirect DNA evidence is sufficient grounds to start an investigation against the relatives of the suspect and to demand DNA samples from the relatives. If no DNA samples can be obtained they should decide whether this type of indirect evidence may be used.

References

Balding DJ, Donnelly P (1995) Inference in forensic identification. J R Statist Soc A 158 Part 1 : 21–53

Balding DJ, Nichols RA (1994) DNA profile match probability calculation: how to allow for population stratification, relatedness, database selection and single bands. Forensic Sci Int 64 : 125–140

Balding DJ, Donnelly P, Nichols RA (1994) Comment: some causes for concern about DNA profiles. Statist Sci 9 : 248–251

Brookfield JFY (1994) The effect of relatives on the likelihood ratio associated with DNA profile evidence in criminal cases. J Forensic Sci Soc 34 : 193–197

Donnelly P (1995) Nonindependence of matches at different loci in DNA profiles: quantifying the effect of close relatives on the match probability. Heredity 75 : 26–34

Evett IW (1992) Evaluating DNA profiles in a case where the defence is "It was my brother". J Forensic Sci Soc 32 : 5–14

Kloosterman AD, Janssen HJT (1997) Forensic DNA testing and its legislation in the Netherlands. Forensic Sci Int 88 : 55–58

Lempert R (1994) Comment: theory and practice in DNA fingerprinting. Statist Sci 9 : 255–258

National Research Council (1996) The evaluation of forensic DNA evidence. National Academy Press, Washington DC

Ovington A, Daselaar P, Sjerps M, Kloosterman AD (1997) A Dutch population study of the STR loci D21S11 and HUMFIBRA. Int J Legal Med 110 : 14–17

Schneider PM, Rittner C, Martin PD (eds) (1997) Proceedings of the European symposium: ethical and legal issues of DNA typing in forensic medicine. Forensic Sci Int 88 : 1–110

Sjerps M, van der Geest N, Pieron C, Gajadhar M, Kloosterman AD (1995) A Dutch population study of the STR loci HUMTHO1, HUMFES/FPS, HUMVWA31/1 and HUMF13A1, conducted for forensic purposes. Int J Legal Med 108 : 127–134

Weir BS (1996) Genetic data analysis II Sinauer Associates, Sunderland, pp 219