Int J Legal Med (2005) 119: 251–259
DOI 10.1007/s00414-004-0493-9

ORIGINAL ARTICLE

Yue-Qing Hu · Wing K. Fung

Evaluation of DNA mixtures involving two pairs of relatives

Received: 3 June 2004 / Accepted: 20 September 2004 / Published online: 12 February 2005

Abstract This paper considers the statistical evaluation of DNA mixtures in the following situations: (1) two unknown contributors are related respectively to two typed persons, (2) two of the unknown or untyped contributors are related and the third unknown contributor is related to a typed person, or (3) there are two pairs of related unknown contributors to the DNA mixture. The corresponding formulas for evaluating the likelihood ratios on the strength of DNA evidence are derived and the kinship coefficients for the related persons are incorporated into the calculations. Two examples are analyzed for illustration.

Keywords Forensic science · Kinship coefficient · Likelihood ratio · Mixed stain · Relatives

Introduction

The effect of kinship on the match probability and the likelihood ratio (*LR*) of DNA evidence receives researchers' attention when there are related persons involved in the pool of possible perpetrators. Evett (1992) established a formula for the likelihood ratio in a case where the defense is "It was my brother". Brookfield (1994) evaluated the effect of the possibility that the suspect and the source of the crime scene DNA are relatives upon the likelihood ratio. Balding et al. (1994) and Roeder (1994) addressed the evaluation of the match probability when the culprit is assumed to be a relative of the suspect. Donnelly (1995) quantified the effect of close relatives on the match probability and Sjerps and Kloosterman (1999) discussed the assessment of DNA profile for close relatives of an excluded suspect.

Recently, Fukshansky and Bär (2000) derived a formula for evaluating a mixed stain from one untyped suspect with one typed relative, limited to three kinship relationships. Hu and Fung (2003) resolved the problem on the evaluation of DNA mixtures when there are two related unknown (i.e. untyped) contributors or one unknown contributor is related to a typed person. They have developed a general formula for the calculation of the likelihood ratio *LR*, incorporating the kinship coefficients (k_0,$2k_1$,k_2). In fact, the kinship coefficients k_0,$2k_1$, and k_2 of two individuals are the probabilities that these two individuals share zero, one, and two pairs of ibd (identical by descent) alleles (Fung and Hu 2004), respectively. All of these formulas are limited to one pair of related persons. However, we may encounter difficulties in more complex cases. For example, if the number of unknown contributors is two, we may encounter that each unknown contributor is related to a single typed person. Interpreted more precisely, an unknown contributor X_1 is related to a typed person T_1, and independently the other unknown contributor X_2 is related to another typed person T_2. If the number of unknown contributors is three, we may encounter that one unknown contributor is related to a typed person and the other two unknown contributors are related; i.e., an unknown contributor X_1 is related to a typed person T_1, and X_2 and X_3 are two related unknown contributors. If the number of unknown contributors is four, we may have the situation of two pairs of relatives, i.e. X_1 and X_2 are related unknown contributors and so are X_3 and X_4. These three cases will render calculations of likelihood ratios more difficult than those given in Hu and Fung (2003).

Generally, denote $x>1$ the number of unknown contributors, $X_1, X_2, \ldots$ the unknown contributors, and $T_1, T_2, \ldots$ the typed persons. In the case of $x\geq 2$, we consider the following three problems on the evaluation of the DNA mixture

Y.-Q. Hu · W. K. Fung (✉)
Department of Statistics and Actuarial Science,
The University of Hong Kong,
Pokfulam Road,
Hong Kong, People's Republic of China
e-mail: wingfung@hku.hk
Tel.: +852-28591988
Fax: +852-28589041

involving two pairs of relatives, under Hardy-Weinberg equilibrium:

- Problem 1: there are x unrelated unknown contributors, among which X_1 is related to T_1, and X_2 is related to T_2
- Problem 2: there are x unknown contributors, among which X_1 is related to T_1, and X_2 and X_3 are related, and X_1 and the other $x-3$ unknown contributors are unrelated
- Problem 3: there are x unknown contributors, among which X_1 and X_2 are related, X_3 and X_4 are related, and the other $x-4$ unknown contributors are unrelated.

We are going to tackle these problems, and give three general formulas for calculating the likelihood ratio of DNA mixture. The implementation of these three formulas by computer programs is straightforward, and two examples from the literature are analyzed for illustration.

Notations and formulas

Let M be the genetic profile (distinct alleles) of the mixed stain at an autosomal locus, $K=T_1, T_{2,...}$ be the collection of the genotypes of the typed persons, H be a hypothesis declaring specifically (i) who the known (typed) contributors are, (ii) the number (x) of the unknown contributors, and (iii) the relationship among the unknown contributors and typed persons. In most of the studies, e.g. Weir et al. (1997), Fukshansky and Bär (1998), Fung and Hu (2001), all the involved individuals, known or unknown, are assumed to be unrelated. Recently, Fukshansky and Bär (2000) as well as Hu and Fung (2003) addressed the evaluation of DNA mixtures when two unknown contributors are related or one unknown contributor is related to one typed person.

Writing $P(\text{Evidence}|H)=P(K,M|H)=P(K|H)P(M|K,H)$ and noting the probability of observing the genotypes K, is independent of the hypothesis H, i.e. $P(K|H)=P(K)$, the calculation of the likelihood ratio is then just a ratio of two conditional probabilities like the term $P(M|K,H)$. Given the hypothesis H, denote U the allele set that cannot be explained by the known contributors declared in H and χ the genetic profile of the x unknown contributors, then $P(M|K,H)=P(U\subset\chi\subset M|K)$, which is the conditional probability that the x unknown contributors explain allele set U and do not have any alleles not in set M, given the typed genotypes K (Hu and Fung 2003). It is convenient to denote i=1,2,... as the alleles and p_i the corresponding allele frequencies. By the principle of inclusion and exclusion, Hu and Fung (2003) derived

$$P(M|K,H) = W(M) - \sum_{i\in U} W(M\setminus\{i\}) + \sum_{i,j\in U} W(M\setminus\{i,j\}) - \cdots + (-1)^{|U|} W(M\setminus U), \quad (1)$$

where

$$W(D) = P(\chi \subset D|K) \quad (2)$$

is the conditional probability that every allele of the x unknown contributors is in set D given the genotypes of the typed persons, D is an arbitrary subset of M, and $|U|$ represents the cardinality of set U. As we do in this paper for simplicity, the hypothesis H mainly concerns the number and composition of the unknown contributors, although who the known contributors are, is an essential part of the hypothesis itself.

For a simple representation of our results, we define

$$Q(n,D) = s^n - \sum_{i\in D}(s-p_i)^n + \sum_{i,j\in D}(s-p_i-p_j)^n + \cdots \quad (3)$$

for any set $D\subset M$ and non-negative integer n, where $s=\sum_{i\in M} p_i$.

It is noted that $Q(0,\varphi)=1$ and $Q(n,D)=0$ for $n<|D|$. In fact, the quantity $Q(n,D)$ is just the probability of a set of n alleles including the allele set D and being a subset of M. It is observed from Eq. 3 that the implementation of $Q(n,D)$ by a computer program is easy. It is noted that the quantity $Q(n,D)$ depends not only on set D and integer n, but also on the frequencies of the alleles in set M. In calculating the likelihood ratio at a locus, the allele frequencies and the mixture set M remain the same within that locus and so for brevity $Q(n,D)$ can be treated as a function of n and D only.

Under Hardy-Weinberg equilibrium and all involved persons being unrelated, it is obvious that

$$W(D) = \left(\sum_{i\in D} p_i\right)^{2x}$$

and thus the formulas reported in Weir et al. (1997) and Fukshansky and Bär (1998) for evaluating the weight of DNA evidence can be represented as

$$P(M|K,H) = s^{2x} - \sum_{i\in U}(s-p_i)^{2x} + \sum_{i,j\in U}(s-p_i-p_j)^{2x} + \cdots = Q(2x,U).$$

In the following three subsections, we will derive, under Hardy-Weinberg equilibrium, the corresponding formulas for calculating $P(M|K,H)$ in three different scenarios of untyped contributors and typed persons, incorporating the kinship coefficients into the calculation. All individuals involved are assumed to be unrelated unless otherwise specified.

Two unknowns are related to two typed persons

Assume that, at least two persons $T_1=t_{11}t_{12}$ and $T_2=t_{21}t_{22}$ are typed, and the hypothesis about who the x (≥ 2) unknown unrelated contributors are is:

H : Two of the x unknowns, X_1 and X_2, are related to T_1 and T_2, respectively. (4)

Then we have:

$$P(M|K,H) = (k_{10}, 2k_{11}, k_{12}) \begin{pmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} \end{pmatrix} \begin{pmatrix} k_{20} \\ 2k_{21} \\ k_{22} \end{pmatrix} \tag{5}$$

where $(k_{10},2k_{11},k_{12})$ are the kinship coefficients of individuals X_1 and T_1, $(k_{20},2k_{21},k_{22})$ are the kinship coefficients of individuals X_2 and T_2, and

$$\begin{aligned}
\sigma_{11} &= Q(2x, U), \\
\sigma_{12} &= I_M(t_{21})Q(2x-1, U\setminus\{t_{21}\})/2 + I_M(t_{22})Q(2x-1, U\setminus\{t_{22}\})/2, \\
\sigma_{13} &= I_M(t_{21})I_M(t_{22})Q(2x-2, U\setminus\{t_{21}\}\cup\{t_{22}\}), \\
\sigma_{21} &= I_M(t_{11})Q(2x-1, U\setminus\{t_{11}\})/2 + I_M(t_{12})Q(2x-1, U\setminus\{t_{12}\})/2, \\
\sigma_{22} &= I_M(t_{11})I_M(t_{21})Q(2x-2, U\setminus\{t_{11}\}\cup\{t_{21}\})/4 \\
&\quad + I_M(t_{11})I_M(t_{22})Q(2x-2, U\setminus\{t_{11}\}\cup\{t_{22}\})/4 \\
&\quad + I_M(t_{12})I_M(t_{21})Q(2x-2, U\setminus\{t_{12}\}\cup\{t_{21}\})/4 \\
&\quad + I_M(t_{12})I_M(t_{22})Q(2x-2, U\setminus\{t_{12}\}\cup\{t_{22}\})/4, \\
\sigma_{23} &= I_M(t_{11})I_M(t_{21})I_M(t_{22})Q(2x-3, U\setminus\{t_{11}\}\cup\{t_{21}\}\cup\{t_{22}\})/2 \\
&\quad + I_M(t_{12})I_M(t_{21})I_M(t_{22})Q(2x-3, U\setminus\{t_{12}\}\cup\{t_{21}\}\cup\{t_{22}\})/2, \\
\sigma_{31} &= I_M(t_{11})I_M(t_{12})Q(2x-2, U\setminus\{t_{11}\}\cup\{t_{12}\}), \\
\sigma_{32} &= I_M(t_{11})I_M(t_{12})I_M(t_{21})Q(2x-3, U\setminus\{t_{11}\}\cup\{t_{12}\}\cup\{t_{21}\})/2 \\
&\quad + I_M(t_{11})I_M(t_{12})I_M(t_{22})Q(2x-3, U\setminus\{t_{11}\}\cup\{t_{12}\}\cup\{t_{22}\})/2, \\
\sigma_{33} &= I_M(t_{11})I_M(t_{12})I_M(t_{21})I_M(t_{22})Q(2x-4, U\setminus\{t_{11}\}\cup\{t_{12}\}\cup\{t_{21}\}\cup\{t_{22}\})
\end{aligned}$$

where, for example, $I_M(t_{11})$ is the indicator function defined by $I_M(t_{11})=1$ if $t_{11}\in M$ and 0 otherwise. The proof of Eq. 5 is given in the Appendix.

It is noted that if $(k_{20},2k_{21},k_{22})=(0,0,1)$, i.e. $X_2=T_2$, then the typed person T_2 becomes a known contributor and the hypothesis H becomes "one of the $x-1$ unknown contributors, X_1, is related to T_1"; if $(k_{10},2k_{11},k_{12})=(k_{20},2k_{21},k_{22})=(0,0,1)$ then both typed persons T_1 and T_2 become the known contributors and the hypothesis H becomes "there are $x-2$ unknown contributors". It is thus concluded that the hypothesis H can cover a variety of hypotheses, which will be shown in the example given in the next section.

Remark 1 If $(k_{20},2k_{21},k_{22})=(1,0,0)$, i.e. X_2 and T_2 are unrelated, then the hypothesis H becomes "X_1, one of the x unknown contributors, is related to $T_1=t_{11}t_{12}$" and the conditional probability $P(M|K,H)$ in Eq. 5 can be simplified to

$$\begin{aligned}
P_{UN}(x, U) &= k_{10}Q(2x, U) \\
&\quad + k_{11}(I_M(t_{11})Q(2x-1, U\setminus\{t_{11}\}) \\
&\quad + I_M(t_{12})Q(2x-1, U\setminus\{t_{12}\})) \\
&\quad + k_{12}I_M(t_{11})I_M(t_{12})Q(2x-2, U \\
&\quad \setminus\{t_{11}\}\cup\{t_{12}\}),
\end{aligned}$$

which coincides with the result of Hu and Fung (2003), with a more concise form given here.

Remark 2 If $(k_{20},2k_{21},k_{22})=(0,0,1)$, i.e. $X_2=T_2$, then the conditional probability $P(M|K,H)$ in Eq. 5 can be simplified to

$$\begin{aligned}
P_{TW}(x, U) &= I_M(t_{21})I_M(t_{22})\{k_{10}Q(2x-2, U\setminus\{t_{21}\}\cup\{t_{22}\}) \\
&\quad + k_{11}I_M(t_{11})Q(2x-3, U\setminus\{t_{11}\}\cup\{t_{21}\}\cup\{t_{22}\}) \\
&\quad + k_{11}I_M(t_{12})Q(2x-3, U\setminus\{t_{12}\}\cup\{t_{21}\}\cup\{t_{22}\}) \\
&\quad + k_{12}I_M(t_{11})I_M(t_{12})Q(2x-4, U\setminus\{t_{11}\}\cup\{t_{12}\}\cup\{t_{21}\}\cup\{t_{22}\})\} \\
&= I_M(t_{21})I_M(t_{22})P_{UN}(x-1, U\setminus\{t_{21}\}\cup\{t_{22}\}),
\end{aligned}$$

which is equivalent to the expression given in Remark 1 where the set U is replaced by $U\setminus\{t_{21}\}\cup\{t_{22}\}$ and the number of unknown contributors x is replaced by $x-1$ if $t_{21}, t_{22}\in M$. It is evident that since alleles t_{21} and t_{22} of individual T_2 are contributed to the mixture, the set U should be changed to $U\setminus\{t_{21}\}\cup\{t_{22}\}$ and the unknown contributors number is decreased by 1 to $x-1$.

Remark 3 If $(k_{20},2k_{21},k_{22})=(0,1,0)$, i.e. X_2 and T_2 have a parent-child relationship, then the conditional probability $P(M|K,H)$ in Eq. 5 can be simplified to

$$\begin{aligned}
P_{PC}(x,U) &= I_M(t_{21})\{k_{10}Q(2x-1,U\setminus\{t_{21}\})\\
&\quad +k_{11}I_M(t_{11})Q(2x-2,U\setminus\{t_{11}\}\cup\{t_{21}\})\\
&\quad +k_{11}I_M(t_{12})Q(2x-2,U\setminus\{t_{12}\}\cup\{t_{21}\})\\
&\quad +k_{12}I_M(t_{11})I_M(t_{12})Q(2x-3,U\setminus\{t_{11}\}\cup\{t_{12}\}\cup\{t_{21}\})\}/2\\
&\quad +I_M(t_{22})\{k_{10}Q(2x-1,U\setminus\{t_{22}\})\\
&\quad +k_{11}I_M(t_{11})Q(2x-2,U\setminus\{t_{11}\}\cup\{t_{22}\})\\
&\quad +k_{11}I_M(t_{12})Q(2x-2,U\setminus\{t_{12}\}\cup\{t_{22}\})\\
&\quad +k_{12}I_M(t_{11})I_M(t_{12})Q(2x-3,U\setminus\{t_{11}\}\cup\{t_{12}\}\cup\{t_{22}\})\}/2\\
&= I_M(t_{21})P_{UN}(x-1/2,U\setminus\{t_{21}\})/2+I_M(t_{22})P_{UN}(x-1/2,U\setminus\{t_{22}\})/2.
\end{aligned}$$

This result can be explained by the following reasons. Without loss of generality, assume $T_2=t_{21}t_{22}$ is the father of X_2, then either t_{21} or t_{22} is transmitted to X_2 with equal chance. If $t_{21}\in M$ and is transmitted to X_2, the set U should be changed to $U\setminus\{t_{21}\}$ and the number of unknown contributors is $x-1/2$, where 1/2 can mean that one allele of X_2, t_{21}, is typed and the other allele of X_2 is unknown, i.e. one out of two alleles of X_2 is known. So we have the term $P_{UN}(x-1/2,U\setminus\{t_{21}\})$ obtained by replacing U with $U\setminus\{t_{21}\}$ and x with $x-1/2$ in the expression of Remark 1. Similarly, we have the term $P_{UN}(x-1/2,U\setminus\{t_{22}\})$. Finally we have equation $P_{PC}(x,U)$ as given above by the law of total probability.

It is observed from Remarks 1–3 that:

$$\begin{aligned}
P(M|K,H) &= k_{20}P_{UN}(x,U)+2k_{21}P_{PC}(x,U)\\
&\quad +k_{22}P_{TW}(x,U)\\
&= k_{20}P_{UN}(x,U)\\
&\quad +k_{21}I_M(t_{21})P_{UN}(x-1/2,U\setminus\{t_{21}\})\\
&\quad +k_{21}I_M(t_{22})P_{UN}(x-1/2,U\setminus\{t_{22}\})\\
&\quad +k_{22}I_M(t_{21})I_M(t_{22})\\
&\quad P_{UN}(x-1,U\setminus\{t_{21}\}\cup\{t_{22}\}).
\end{aligned}$$

So $P(M|K,H)$ can be found through $P_{UN}(\cdot,\cdot)$, which can facilitate the calculation of the conditional probability in designing the computer program.

One unknown is related to a typed person and two other unknowns are related

Assume that there are at least three unknown contributors involved in a criminal case, among them, X_1 is related to a typed person T_1, and X_2 and X_3 are related. The hypothesis about who the contributors are, is given as follows:

$$\begin{aligned}
H:\ &\text{One of the } x(\geq 3) \text{ unknowns, } X_1,\\
&\text{is related to a typed person } T_1 \text{ and two of the}\\
&x \text{ unknowns, } X_2 \text{ and } X_3, \text{ are related.}
\end{aligned} \tag{6}$$

In this situation, the corresponding formula for calculating $P(M|K,H)$ is

$$P(M|K,H)=(k_{10},2k_{11},k_{12})\begin{pmatrix}\sigma_{11}&\sigma_{12}&\sigma_{13}\\ \sigma_{21}&\sigma_{22}&\sigma_{23}\\ \sigma_{31}&\sigma_{32}&\sigma_{33}\end{pmatrix}\begin{pmatrix}k_{20}\\ 2k_{21}\\ k_{22}\end{pmatrix} \tag{7}$$

where $(k_{10},2k_{11},k_{12})$ are the kinship coefficients of individuals X_1 and T_1, $(k_{20},2k_{21},k_{22})$ are the kinship coefficients of individuals X_2 and X_3, and

$$\begin{aligned}
\sigma_{11} &= Q(2x,U),\\
\sigma_{12} &= Q(2x-1,U),\\
\sigma_{13} &= Q(2x-2,U),\\
\sigma_{21} &= I_M(t_{11})Q(2x-1,U\setminus\{t_{11}\})/2+I_M(t_{12})\\
&\quad Q(2x-1,U\setminus\{t_{12}\})/2,\\
\sigma_{22} &= I_M(t_{11})Q(2x-2,U\setminus\{t_{11}\})/2+I_M(t_{12})\\
&\quad Q(2x-2,U\setminus\{t_{12}\})/2,\\
\sigma_{23} &= I_M(t_{11})Q(2x-3,U\setminus\{t_{11}\})/2+I_M(t_{12})\\
&\quad Q(2x-3,U\setminus\{t_{12}\})/2,\\
\sigma_{31} &= I_M(t_{11})I_M(t_{12})Q(2x-2,U\setminus\{t_{11}\}\cup\{t_{12}\}),\\
\sigma_{32} &= I_M(t_{11})I_M(t_{12})Q(2x-3,U\setminus\{t_{11}\}\cup\{t_{12}\}),\\
\sigma_{33} &= I_M(t_{11})I_M(t_{12})Q(2x-4,U\setminus\{t_{11}\}\cup\{t_{12}\}).
\end{aligned}$$

The proof of Eq. 7 is outlined in the Appendix.

Again Eq. 7 displays $P(M|K,H)$ as a linear combination of $Q(\cdot,\cdot)$. $P(M|K,H)$ can also be expressed as

$$P(M|K,H) = k_{20}P_{UN}(x,U) + 2k_{21}P_{UN}(x-1/2,U) + k_{22}P_{UN}(x-1,U)$$

which is convenient for designing the computer program.

Two pairs of related unknowns

In this subsection, we consider the hypothesis

$$H: \text{Amongst the } x(\geq 4) \text{ unknowns contributors, } X_1 \text{ and } X_2 \text{ are related, and } X_3 \text{ and } X_4 \text{ are related.} \quad (8)$$

Under this hypothesis, we can obtain

$$P(M|K,H) = (k_{10}, 2k_{11}, k_{12}) \begin{pmatrix} Q(2x,U) & Q(2x-1,U) & Q(2x-2,U) \\ Q(2x-1,U) & Q(2x-2,U) & Q(2x-3,U) \\ Q(2x-2,U) & Q(2x-3,U) & Q(2x-4,U) \end{pmatrix} \begin{pmatrix} k_{20} \\ 2k_{21} \\ k_{22} \end{pmatrix} \quad (9)$$

where $(k_{10},2k_{11},k_{12})$ are the kinship coefficients of individuals X_1 and X_2, and $(k_{20},2k_{21},k_{22})$ are the kinship coefficients of individuals X_3 and X_4. See Appendix for the proof of Eq. 9.

Taking $(k_{20},2k_{21},k_{22})=(1,0,0)$ in Eq. 9, i.e. there exists only one pair of related unknowns among the x unknowns, will simplify the conditional probability $P(M|K,H)$ as $k_{10}Q(2x,U)+2k_{11}Q(2x-1,U)+k_{12}Q(2x-2,U)$, which was reported in Hu and Fung (2003). Taking $(k_{20},2k_{21},k_{22})=(0,0,1)$, i.e. there exists only one pair of related unknowns among the $x-1$ unknowns, will simplify the conditional probability $P(M|K,H)$ as $k_{10}Q(2x-2,U)+2k_{11}Q(2x-3,U)+k_{12}Q(2x-4,U)$. Taking $(k_{20},2k_{21},k_{22})=(0,1,0,)$, e.g. X_4 is the father of X_3, will simplify the conditional probability $P(M|K,H)$ as $k_{10}Q(2x-1,U)+2k_{11}Q(2x-2,U)+k_{12}Q(2x-3,U)$. In this situation, there is one and only one allele of the father X_4 which is ibd to one allele of child X_3, so only three among those four alleles of X_3 and X_4 are not ibd and hence we regard the number of unknown contributors as $x-1/2$. Thus replacing x in the first case of this paragraph corresponding to $(k_{20},2k_{21},k_{22})=(1,0,0)$ by $x-1/2$ will lead to the result. It is convenient to calculate the conditional probability given in Eq. 9 by the computer program.

It is interesting to note that Eqs. 5 and 7 are equal when the two alleles of T_2 satisfy $t_{21}, t_{22} \in M \setminus U$, and Eqs. 7 and 9 are equal when the two alleles of T_1 satisfy $t_{11}, t_{12} \in M \setminus U$. We will meet the latter case in the following section.

Examples

Example 1

The first example is taken from Stockmarr (2000). The mixture was typed as {18,24,28,31,33,36} at locus D1S80 from a mixed stain recovered from a crime scene, and the victim V was typed as {24,33}. Stockmarr (2000) considered two hypotheses namely "the victim and $n-1$ unknowns are contributors" versus "n unknowns are contributors", with a range of values of n. In the following, for illustration of using Eqs. 7 and 9, we take $n=4$ and investigate two competing propositions:

- H'_p: the victim and three unknowns are contributors, one unknown contributor is related to the victim, and the other two unknowns are related
- H'_d: two unknown contributors are related, and the other two unknowns are also related.

Since there can be various sorts of relationships, for simplicity, we consider only the following two particular hypotheses:

- H_p: the victim, one untyped relative R of the victim V, and two untyped full siblings are the contributors
- H_d: two related unknowns X_1 and X_2, and two untyped full siblings are the contributors.

In order to calculate the likelihood ratio, i.e. $P(M|K,H_p)/P(M|K,H_d)$, we first employ Eq. 7 to find $P(M|K,H_p)$.

Under the hypothesis H_p, x=3, U={18,28,31,36}, M={18,24,28,31,33,36}, and the two alleles of typed victim V satisfy 24,33 $\in M \setminus U$, so:

$$P(M|K,H_p) = (k_{10}, 2k_{11}, k_{12}) \begin{pmatrix} Q(6,U) & Q(5,U) & Q(4,U) \\ Q(5,U) & Q(4,U) & Q(3,U) \\ Q(4,U) & Q(3,U) & Q(2,U) \end{pmatrix} \begin{pmatrix} 1/4 \\ 1/2 \\ 1/4 \end{pmatrix}$$

where $(k_{10}2k_{11}k_{12})$ are the kinship coefficients of individuals R and V declared in H_p. Under the hypothesis H_d, x=4 and U={18,24,28,31,33,36}, we can find using Eq. 9

$$P(M|K,H_d) = (k_{10}, 2k_{11}, k_{12}) \begin{pmatrix} Q(8,U) & Q(7,U) & Q(6,U) \\ Q(7,U) & Q(6,U) & Q(5,U) \\ Q(6,U) & Q(5,U) & Q(4,U) \end{pmatrix} \begin{pmatrix} 1/4 \\ 1/2 \\ 1/4 \end{pmatrix}$$

where $(k_{10},2k_{11},k_{12})$ are the kinship coefficients of individuals X_1 and X_2 declared in H_d. So we can find the likelihood ratios for different kinship relationships between R and V, and X_1 and X_2.

We consider the 7 most common relationships including unrelated and monozygotic (MZ) twins, and the corresponding 7×7 likelihood ratios are shown in Table 1. The relationship is ordered from the most to the least related, i.e. monozygotic twins, parent-child, full siblings, ..., unrelated. The following is observed:

1. The looser the relationship between R and V in H_p, the larger the likelihood ratio. Within each column, the largest LR (unrelated case) is about 10 times the lowest LR (MZ twins case) and 2 times the second lowest (parent-child case).
2. The looser the relationship between X_1 and X_2 in H_d, the smaller the likelihood ratio. Within each row, the smallest LR (unrelated case) is about one-tenth of the largest LR (MZ twins case) and one-half of the second largest LR (parent-child case).
3. Excluding the case of MZ twins, the effect of relationship of R and V, or X_1 and X_2 on the LR is not large; it only has at most double or half the LR value of the unrelated relationship.

Example 2

This is a group rape analyzed by Fukshansky and Bär (1998), where three persons, the victim, the suspect S_1 and the suspect S_2, were typed at three loci DQa, FES, and F13A1. The mixed stain was assumed to be contributed by the victim and two assailants. The prosecution proposition H_p is taken as "S_1 and S_2 are both contributors to the

Table 1 Likelihood ratios for the hypotheses "oneuntyped relative R of the victim V, and two untyped full siblings arecontributors" versus "two related unknowns X_1 and X_2, and two untyped full siblings are contributors" about the criminal case reported in Stockmarr (2000)

(R,V)	(X_1, X_2)						
	MZ twins	Parent-child	Full sibs	Half-sibs	First cousins	Second cousins	Unrelated
MZ twins	20.92	4.48	3.94	2.69	2.24	1.99	1.92
Parent-child	100.87	21.61	19.02	12.97	10.81	9.61	9.27
Full sibs	115.88	24.82	21.85	14.90	12.42	11.04	10.65
Half-sibs	170.87	36.60	32.21	21.98	18.32	16.28	15.70
First cousins	205.87	44.09	38.81	26.48	22.07	19.62	18.92
Second cousins	232.12	49.72	43.76	29.86	24.88	22.12	21.33
Unrelated	240.86	51.59	45.41	30.98	25.82	22.96	22.14

Different relationships for R and V, and for X_1 and X_2 are considered.

Table 2 Overall likelihood ratios for the hypotheses" S_1 andS_2 are contributors" versus "R_1, one relative of S_1, and R_2, one relative of S_2, are contributors" about the case of agroup rape reported in Fukshansky and Bär (1998)

(R_1, S_1)	(R_2, S_2)						
	MZ twins	Parent-child	Full sibs	Half-sibs	First cousins	Second cousins	Unrelated
MZ twins	1	16	7	46	101	254	410
Parent-child	14	66	50	153	281	550	754
Full sibs	6	44	30	110	213	432	596
Half-sibs	35	135	112	297	521	949	1,245
First cousins	74	243	209	513	872	1,533	1,972
Second cousins	187	495	444	980	1,604	2,710	3,425
Unrelated	319	713	650	1,350	2,154	3,562	4,464

mixed stain." Fukshansky and Bär (1998) considered three different sets of defense propositions where all persons involved are assumed to be unrelated. When there is one pair of related persons involved, we can employ the result of Hu and Fung (2003) for weighting the DNA evidence. In the following, we consider a particular defense proposition involving two pairs of related persons:

– H_d: R_1, one relative of S_1, and R_2, and one relative of S_2 are both contributors to the mixed stain.

Using Eq. 5, we can find the likelihood ratio $1/P(M|K, H_d)$ for each of the three loci separately and then the overall one by multiplication. Table 2 shows the likelihood ratios when R_1 and S_1, and R_2 and S_2 take 49 possible combinations of commonly encountered relationships including unrelated and monozygotic twins. It is noted that the first row of Table 2 corresponds to the likelihood ratios about the prosecution hypothesis H_p versus the defense hypothesis "the victim, the suspect S_1, and one relative of the suspect S_2 are the contributors;" the first column of Table 2 corresponds to the likelihood ratios about the prosecution hypothesis H_p versus the defense hypothesis "the victim, the suspect S_2, and one relative of the suspect S_1 are the contributors." It is noted that Table 2 includes all three likelihood ratios reported in Fukshansky and Bär (1998). Specifically, the value of 410 at the upper right hand side of Table 2, is just the likelihood ratio value about H_1 versus H_2 reported in Fukshansky and Bär (1998); the lower left value of 319 is just the likelihood ratio value about H_1 versus H_3 in Fukshansky and Bär (1998) (N.B. there seems to be a typing error in the paper); the lower right value of 4,464, is just the likelihood ratio value about H_1 versus H_4 in Fukshansky and Bär (1998). Besides these values, Table 2 provides the likelihood ratios with respect to various combinations of relationships between R_1 and S_1 and between R_2 and S_2. For example, when R_1 and S_1 are full siblings and R_2 and S_2 are half-siblings, the likelihood ratio is 110, and when R_1 and S_1 are first cousins and R_2 and S_2 are second cousins, the likelihood ratio is 1,533.

It can be observed from Table 2 that the differences among the *LR*s of different kinship relationships can be very large. For example, the likelihood ratio for both full sibling relationships is 30, which is only 0.67% of the highest likelihood ratio 4,464, corresponding to both unrelated relationships.

Discussion

This paper extends our earlier work reported in Hu and Fung (2003) to the situation of two pairs of relatives and thus widens the scope of the application. Although the expressions, e.g. Eq. 5, for calculating likelihood ratios seem complex, they are in essence just linear combinations of $Q(\cdot,\cdot)$. Moreover, the implementation of $Q(\cdot,\cdot)$ by a computer program is not difficult.

The idea shown in this paper can be used to tackle more complex problems involving more than two pairs of relatives. For example, if there are three pairs of related unknowns among the x unknown contributors with corresponding kinship coefficients $(K_{i0},K_{i1},K_{i2})=(k_{i0},2k_{i1},k_{i2})$, i=1,2,3, then the conditional probability can be expressed as

$$P(M|K,H) = \sum_{i,j,k=0,1,2} K_{1i}K_{2j}K_{3k}Q(2x-i-j-k, U),$$

which is an extension of Eq. 9. The expressions of the conditional probabilities $P(M|K,H)$ for some other cases are obtained, but they are omitted for brevity.

Acknowledgements We thank the referees for helpful comments that improved the presentation of the paper. This work was partially supported by the Hong Kong RGC Competitive Earmarked Research Grant HKU 7022/04P, and the National Natural Science Foundation of China (10329102).

Appendix

Lemma 1

If the kinship coefficients of two individuals X_1 and X_2 are $(k_0,2k_1,k_2)$, then for any set of alleles D

$$P(\chi_1 \subset D, \chi_2 \subset D) = k_0\left(\sum_{i\in D} p_i\right)^4 + 2k_1\left(\sum_{i\in D} p_i\right)^3 + k_2\left(\sum_{i\in D} p_i\right)^2, \quad (10)$$

$$P(\chi_1 \subset D | X_2 = x_{21}x_{22}) = k_0\left(\sum_{i\in D} p_i\right)^2 + k_1(I_D(x_{21}) + I_D(x_{22}))\sum_{i\in D} p_i + k_2 I_D(x_{21})I_D(x_{22}), \quad (11)$$

where χ_1 and χ_2 are the genetic profiles of X_1 and X_2 respectively.

Proof They are just Eqs. 4 and 5 in Hu and Fung (2003).

Lemma 2

For any U⊂M, and any pairwise distinct alleles $a,b,c,d,\in U$,

$$Q(n, U \setminus \{a\}) = \sum_{M\setminus U\subset D\subset M} I_D(a)(-1)^{|M\setminus D|}\left(\sum_{i\in D} p_i\right)^n, \quad (12)$$

$$Q(n, U \setminus \{a, b\}) = \sum_{M\setminus U\subset D\subset M} I_D(a)I_D(b)(-1)^{|M\setminus D|}\left(\sum_{i\in D} p_i\right)^n, \quad (13)$$

$$Q(n, U \setminus \{a, b, c\}) = \sum_{M\setminus U\subset D\subset M} I_D(a)I_D(b)I_D(c) \times (-1)^{|M\setminus D|}\left(\sum_{i\in D} p_i\right)^n, \quad (14)$$

$$Q(n, U \setminus \{a, b, c, d\}) = \sum_{M\setminus U\subset D\subset M} I_D(a)I_D(b)I_D(c)I_D(d) \times (-1)^{|M\setminus D|}\left(\sum_{i\in D} p_i\right)^n. \quad (15)$$

Proof It is sufficient to note that $I_D(a) = 1 \Leftrightarrow a \in D \Leftrightarrow (M \setminus U) \cup \{a\} \subset D \subset M \Leftrightarrow M \setminus (U \setminus \{a\}) \subset D \subset M$, so

$$\sum_{M\setminus U\subset D\subset M} I_D(a)(-1)^{|M\setminus D|}\left(\sum_{i\in D} p_i\right)^n = \sum_{M\setminus (U\setminus\{a\})\subset D\subset M} (-1)^{|M\setminus D|}\left(\sum_{i\in D} p_i\right)^n.$$

Thus Eq. 12 follows immediately from Eq. 3.

Similarly, $I_D(a)I_D(b) = 1 \Leftrightarrow a, b \in D \Leftrightarrow (M \setminus U) \cup \{a, b\} \subset D \subset M \Leftrightarrow M \setminus (U \setminus \{a, b\}) \subset D \subset M$. So Eq. 13 holds by $I_D(a)I_D(b)I_D(c) = 1 \Leftrightarrow a, b, c \in D \Leftrightarrow (M \setminus U) \cup \{a, b, c\} \subset D \subset M \Leftrightarrow M \setminus (U \setminus \{a, b, c\}) \subset D \subset M$, so Eq. 14 holds by $I_D(a)I_D(b)I_D(c)I_D(d) = 1 \Leftrightarrow a, b, c, d \in D \Leftrightarrow (M \setminus U) \cup \{a, b, c, d\} \subset D \subset M \Leftrightarrow M \setminus (U \setminus \{a, b, c, d\}) \subset D \subset M$, so Eq. 15 holds by Eq. 3.

Lemma 3

For any U⊂M, and any alleles $a,b,c,d,$ we have

$$\sum_{M\setminus U\subset D\subset M} I_D(a)(-1)^{|M\setminus D|}\left(\sum_{i\in D} p_i\right)^n = I_M(a)Q(n, U \setminus \{a\}), \quad (16)$$

$$\sum_{M\setminus U\subset D\subset M} I_D(a)I_D(b)(-1)^{|M\setminus D|}\left(\sum_{i\in D} p_i\right)^n = I_M(a)I_M(b)Q(n, U \setminus \{a\} \cup \{b\}) \quad (17)$$

$$\sum_{M\setminus U\subset D\subset M} I_D(a)I_D(b)I_D(c)(-1)^{|M\setminus D|}\left(\sum_{i\in D} p_i\right)^n = I_M(a)I_M(b)I_M(c)Q(n, U \setminus \{a\} \cup \{b\} \cup \{c\}) \quad (18)$$

$$\sum_{M\setminus U\subset D\subset M} I_D(a)I_D(b)I_D(c)I_D(d)(-1)^{|M\setminus D|}\left(\sum_{i\in D}p_i\right)^n = I_M(a)I_M(b)I_M(c)I_M(d)Q(n, U\setminus\{a\}\cup\{b\}\cup\{c\}\cup\{d\}). \quad (19)$$

Proof If $a\notin M$, then both sides of Eq. 16 are zero; if $a\in M\setminus U$, then Eq. 16 is just Eq. 3; if $a\in U$, then Eq. 16 is just Eq. 12. Thus, Eq. 16 holds.

If $a=b$, then Eq. 17 is just Eq. 16. Next, we consider $a\neq b$. If a or b is not in M, then both sides of Eq. 17 are zero; if $a,b, \in U$, then Eq. 17 is just Eq. 13; if $a\in U$ and $b\in M\setminus U$, then Eq. 17 is just Eq. 16; if $a,b\in M\setminus U$, then Eq. 17 is just Eq. 3. Thus, we have Eq. 17 by the symmetry of a and b.

If two of $a,b,c,$ are identical, then Eq. 18 is just Eq. 17. So we consider that a,b and c are pairwise distinct. If one of a,b, and c is not in M, then both sides of Eq. 18 are zero; if $a,b,c\in U$, then Eq. 18 is just Eq. 14; if $a,b,\in U$ and $c\in M\setminus U$, then Eq. 18 is just Eq. 17; if $a\in U$ and $b,c\in M\setminus U$, then Eq. 18 is just Eq. 16; if $a,b,c\in M\setminus U$, then Eq. 18 is just Eq. 3. By the symmetry of a,b and c, Eq. 18 is hence proved.

If two of a,b,c and d are identical, then Eq. 19 is just Eq. 18. So we assume a,b,c and d are pairwise distinct. If one of a,b,c and d is not in M, then both sides of Eq. 19 are zero; if $a,b,c,d\in U$, then Eq. 19 is just Eq. 15; if $a,b,c,\in U$ and $d\in M\setminus U$, then Eq. 19 is just Eq. 18; if $a,b,\in U$ and $c,d\in M\setminus U$, then Eq. 19 is just Eq. 17; if $a\in U$ and $b,c,d\in M\setminus U$, then Eq. 19 is just Eq. 16; if $a,b,c,d\in M\setminus U$, then Eq. 19 is just Eq. 3. Thus we prove Eq. 19 by the symmetry of a, b, c and d.

Proof of Eq. 5

Under the hypothesis declared in Eq. 4, the only related individuals are X_1 and T_1, and X_2 and T_2. Let χ_1, χ_2, and χ_0 be the genetic profiles of X_1, X_2, and the other $x-2$ unknown contributors, respectively, then we have $\chi=\chi_1\cup\chi_2\cup\chi_0$ and further by Eq. 2

$$W(D) = P(\chi_1\subset D, \chi_2\subset D, \chi_0\subset D|K) = P(\chi_1\subset D|T_1)P(\chi_2\subset D|T_2)P(\chi_0\subset D). \quad (20)$$

Using Eq. 11 for the first two items in Eq. 20, $P(\chi_0\subset D) = \left(\sum_{i\in D}p_i\right)^{2(x-2)}$, and the results of Lemma 3, we can have Eq. 5 from Eq. 1 after some matrix manipulation.

Proof of Eq. 7

Under hypothesis declared in Eq. 6, the only related individuals are X_1 and T_1, and X_2 and X_3. Let χ_1, χ_2, χ_3, and χ_0 be the genetic profiles of X_1, X_2, X_3, and the other $x-3$ unknown contributors, respectively, then we have $\chi=\chi_1\cup\chi_2\cup\chi_3\cup\chi_0$ and further by Eq. 2

$$W(D) = P(\chi_1\subset D|T_1)P(\chi_2\subset D, \chi_3\subset D)P(\chi_0\subset D), \quad (21)$$

Substituting Eqs. 10, 11, and $P(\chi_0\subset D) = \left(\sum_{i\in D}p_i\right)^{2(x-3)}$ into the expression of $W(D)$ in Eq. 21 and then using Lemma 3, we can have Eq. 7 from Eq. 1 after some matrix manipulation.

Proof of Eq. 9

Under the hypothesis declared in Eq. 8, the only related individuals are X_1 and X_2, and X_3 and X_4. Let χ_1, χ_2, χ_3, χ_4, and χ_0 be the genetic profiles of X_1, X_2, X_3, X_4, and the other $x-4$ unknown contributors, respectively, then we have $\chi=\chi_1\cup\chi_2\cup\chi_3\cup\chi_4\cup\chi_0$ and further by Eq. 2

$$W(D) = P(\chi_1\subset D, \chi_2\subset D)P(\chi_3\subset D, \chi_4\subset D) P(\chi_0\subset D), \quad (22)$$

Substituting Eq. 10 and $P(\chi_0\subset D) = \left(\sum_{i\in D}p_i\right)^{2(x-4)}$ into Eq. 22 and then using Lemma 3, we have Eq. 9 from Eq. 1 after some matrix manipulation.

References

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

Brookfield FFY (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 case where the defense is "It is my brother". J Forensic Sci Soc 32:5–14

Fukshansky N, Bär W (1998) Interpreting forensic DNA evidence on the basis of hypotheses testing. Int J Legal Med 111:62–66

Fukshansky N, Bär W (2000) Biostatistics for mixed stain: the case of tested relatives of a non-tested suspect. Int J Legal Med 114:78–82

Fung WK, Hu YQ (2001) The evaluation of mixed stains from different ethnic origins: general result and common cases. Int J Legal Med 115:48–53

Fung WK, Hu YQ (2004). Interpreting DNA mixtures with related contributors in subdivided populations. Scand J Stat 31:115–130

Hu YQ, Fung WK (2003) Interpreting DNA mixtures with the presence of relatives. Int J Legal Med 117:39–45

Roeder K (1994) DNA fingerprinting: a review of the controversy. Stat Sci 9:222–247

Sjerps M, Kloosterman AD (1999) On the consequences of DNA profile mismatches for close relatives of an excluded suspect. Int J Legal Med 112:176–180

Stockmarr A (2000) The choice of hypotheses in the evaluation of DNA profile evidence. In: Gastwirth JL (ed) Statistical science in the courtroom. Springer, New York, pp 143–160

Weir BS, Triggs CM, Starling L, Stowell LI, Walsh KAJ, Buckleton J (1997) Interpreting DNA mixtures. J Forensic Sci 42:213–222