B.S. Weir,[1] *C.M. Triggs,*[2] *L. Starling,*[3] *L.I. Stowell,*[3] *K.A.J. Walsh,*[3] *and J. Buckleton,*[3]

Interpreting DNA Mixtures*

REFERENCE: Weir BS, Triggs CM, Starling L, Stowell LI, Walsh KAJ, Buckleton J. Interpreting DNA mixtures. J Forensic Sci 1997;42(2):213–222.

ABSTRACT: The interpretation of mixed DNA stains is explained in the context of likelihood ratios. The probabilities for the mixed-stain profile are evaluated under alternative explanations that specify the numbers of contributors and the profiles of any known contributors. Interpretations based simply on the frequencies with which random members of a population would not be excluded from a mixed-stain profile do not make use of all the information, and may overstate the strength of the evidence against included people. The effects of the numbers of contributors depends on whether all the alleles at a locus are present in the mixed stain. A general equation is given to allow likelihood ratios to be calculated, and includes the "2*p*" modification suggested by the 1996 NRC report. This modification is not always conservative. A computer program to perform calculations is available.

KEYWORDS: forensic science, mixed stains, DNA profiles, likelihood ratios

For a large variety of crimes, the use of genetic markers has proved invaluable in identifying perpetrators or exonerating falsely accused suspects. However, the interpretation of genetic profiles from biological samples can be complicated when the samples contain material from more than one person. This is especially common in rape cases: The sample may contain material from the victim or consensual sexual partners as well as from the perpetrator, or there may be multiple perpetrators. Mixed-stain evidence was discussed by Evett et al. (1991), and more recently by Aitken (1995). Their method, which is the same as the one we use, has been endorsed by the National Research Council (1996).

This approach differs from previous treatments (National Research Council, 1992) that calculate only the proportion of the population that is either included or excluded in a mixed stain. Such calculations ignore the profiles of specific people associated with the mixture, and so are analogous to the "random man not excluded" probabilities in paternity cases (e.g., Walker et al., 1983). Our calculations do consider specific profiles, and so are more like the "paternity index." Moreover, the conventional calculations do not consider the number of contributors to the mixture, and therefore miss the interaction between the number of alleles and the number of contributors on the numerical strength of the evidence of a mixed stain. Only by comparing the probabilities of the evidentiary profile under alternative explanations is it possible to arrive at a complete analysis of mixtures. Indeed, we show by example that the conventional method can overstate the strength of evidence against a defendant not excluded from a mixed stain.

In this paper, we offer some rules to simplify the interpretation of mixed stains, and phrase these rules in terms of the probabilities of certain subsets of the alleles present in the mixed stain. The subsets are the alleles that come from people whose identity is not known with certainty. The rules we describe give an immediate way of deriving results such as those given as examples by the National Research Council (1996). Our results begin with the assumption that the number of contributors to the mixed stain is known, and then we point to methods that make conservative assumptions about this number. We also consider how to interpret mixed VNTR profiles, when there may be unseen alleles.

To keep the treatment as simple as possible, only discrete genetic systems will be treated. These include conventional blood group markers, binned RFLP markers, as well as the current PCR-based markers. We assume that a profile has been determined for an evidentiary sample, that there are indications of more than one contributor to the sample, and that the sample may contain the profiles of one or more known people. The task is to assign a numerical weight to this evidence, and this in turn requires that databases are available from which to estimate frequencies of components of the profile.

As a further simplification, we assume independence of alleles, within and between loci. We develop methods for handling single-locus profiles and then multiply results over loci. We also defer the issue of population structure and relatedness, both of which cause dependencies between the profiles of different people.

Likelihood Ratios

The convenient likelihood ratio framework for assigning weight to evidence will be used (e.g., Aitken, 1995; Balding and Nichols, 1994; Evett, 1983; Lindley, 1977; Weir, 1994). The likelihood ratio for comparing two alternative explanations, C and $\overline{C}$, for the evidentiary profile is formed as the ratio of the probabilities of the profile under the alternatives. The likelihood ratio is written as L and the profile as E, so

$$L = \frac{\Pr(E|C)}{\Pr(E|\overline{C})}$$

If a jury is being asked to make a choice between explanations C and $\overline{C}$, it can be told "The profile E is L times more likely to have arisen under explanation C than under explanation $\overline{C}$."

[1]Program in Statistical Genetics, Department of Statistics, North Carolina State University, Raleigh, NC.

[2]Department of Statistics, University of Auckland, Auckland, New Zealand.

[3]ESR: Forensic, Mt. Albert Science Center, Auckland, New Zealand.

*This work was supported in part by NIH grant GM45344, and by award 95-IJ-CX-0007 from the National Institute of Justice, Office of Justice Programs, U.S. Department of Justice. Points of view in this document are those of the authors and do not necessarily represent the official position of the U.S. Department of Justice.

Received 2 Jan. 1996; and in revised form 2 Aug. 1996; accepted for publication 5 Aug. 1996.

General Notation

We have found it very helpful to introduce notation that makes calculation of the likelihood ratio somewhat automatic. The key is to specify who are the known contributors to the profile, and therefore which alleles in the profile may be from unknown contributors, under each explanation. The likelihood ratio becomes the ratio of the probabilities of the sample profile, taking into account both the known and unknown contributors. For example, suppose an evidentiary sample has four alleles *abcd* at some locus and that it is known the sample contains DNA from the victim and the perpetrator. Suppose, further, the victim has type *ab* and a suspect has type *cd*. The prosecution explanation C would be that the known contributors are the victim and the suspect, whereas the defense explanation $\overline{C}$ may be that the victim is a known contributor, but that alleles *cd* came from an unknown contributor.

- The probability of the sample stain under C is one, because all four alleles can be explained with certainty. There are no unknown contributors, and the probability is written as $P_0(\phi|abcd) = 1$. The subscript 0 indicates the number of unknown contributors, the quantity ϕ indicates the "empty" set of alleles that were from unknown contributors, and *abcd* is the sample profile.
- The probability of the sample profile under explanation $\overline{C}$ is written as $P_1(cd|abcd)$. Now the subscript 1 indicates the one unknown contributor, the set *cd* is the alleles from this unknown person, and *abcd* is still the whole profile.
- The likelihood ratio is

$$L = \frac{P_0(\phi|abcd)}{P_1(cd|abcd)} \tag{1}$$

In general, the evidentiary profile contains a set E of alleles at some locus. There may be known contributors to the profile, having some or all of these alleles between them. Under at least one of the alternative explanations, there are unknown contributors. These people must carry the set U of alleles in E not carried by the known contributors, and they cannot carry any alleles not in E. If there are x unknown contributors to the profile, then what is needed is the probability $P_x(U|E)$ that these x people have the alleles in U between them, but do not have any alleles not in E. When there are no unknown contributors, the symbol P_0 will be used.

We have found that this approach and notation allows us to approach any situation. For example, suppose a victim reports having been raped by two men, and one suspect is identified. Suppose, also, it is known that there have been no consensual partners. The evidentiary sample is assumed to contain DNA from the victim and the perpetrators, and is found to have alleles *abcd*. (Although preferential extraction, Gill et al., 1985, may minimize the DNA from the victim in the male fraction, it may not eliminate her DNA.) The victim is of type *ab* and the suspect of type *cd*. Explanation C is that there is one unknown contributor to the sample (the second perpetrator), but this person is not required to have any specific alleles. He cannot have alleles other than *abcd*. Under explanation $\overline{C}$, there are two unknown contributors (both perpetrators) and these two must have contributed alleles *cd* between them. The likelihood ratio is

$$L = \frac{P_1(\phi|abcd)}{P_2(cd|abcd)} \tag{2}$$

Calculation of Probabilities—We now illustrate how the $P_x(U|E)$ probabilities can be calculated. We also give a general formula, suitable for computer programs, and list the cases most likely needed in casework in Appendix 1.

Profiles with One Allele

If the profile has only alleles of type a at a locus, then all contributors must be *aa* homozygotes. The possibility of unseen alleles will be addressed later. The profile at other loci, or other factors, indicate that the evidentiary stain is a mixture. Assuming independence of alleles within loci, the probability that one unknown contributor has allele a, and no allele other than a, is

$$P_1(a|a) = p_a^2$$

where p_a is the population frequency of allele a, and p_a^2 is the frequency of homozygotes in the population. If there are x unknown contributors, they must all be *aa* homozygotes and

$$P_x(a|a) = p_a^{2x}$$

If a known contributor to the profile is homozygous *aa*, then there are no unexplained alleles. The probability needed, for x unknowns, is $P_x(\phi|a)$, where ϕ is the empty set. Although these unknown people do not need to provide a specific allele, in fact they can only be of type *aa*, so

$$P_x(\phi|a) = p_a^{2x}$$

This is a good place to point out that allele frequencies like p_a refer to the "relevant" population: those people who might be considered potential contributors to the evidentiary sample. This population is defined by circumstances of the crime, rather than by attributes of the suspect. The circumstances may prescribe a particular ethnic group, in which case frequencies should be used for that group. Otherwise, frequencies from different ethnic groups should be considered. The ethnicity of the suspect is not material as we are not considering population structuring.

Profiles with Two Alleles

If the profile has alleles *ab*, then all contributors must be of genotype *aa*, *ab* or *bb*. If the source of only allele a is in question, the contributor of that allele must be of type *aa* or *ab*. People of type *ac*, where c is any other allele, are excluded because the sample profile does not contain c. Therefore the probability that one unknown contributor has allele a, and no allele other than a or b, is

$$\begin{aligned} P_1(a|ab) &= p_a^2 + 2p_ap_b \\ &= (p_a + p_b)^2 - p_b^2 \end{aligned}$$

The generalization to x contributors is

$$P_x(a|ab) = (p_a + p_b)^{2x} - p_b^{2x}$$

This expression gives the probability of all combinations of x people who have only alleles a and b, and it excludes only the situation where all x are *bb* homozygotes because that set of people could not contribute allele a. When $x = 2$, for example, the possible

sets of people are *aa*, *aa*; *aa*, *ab*; *aa*, *bb*; *ab*, *ab*; *ab*, *bb*. The probabilities of these sets are p_a^4, $4p_a^3p_b$, $6p_a^2p_b^2$, $4p_ap_b^3$ which add to $(p_a + p_b)^4 - p_b^4$. If known contributors already have alleles *ab*, then unknowns need not have any specific alleles, but they cannot have alleles other than *ab*, so

$$P_x(\phi|ab) = (p_a + p_b)^{2x}$$

When $x = 2$, the possible sets of people are *aa*, *aa*; *aa*, *ab*; *aa*, *bb*; *ab*, *ab*; *ab*, *bb*; *bb*, *bb*. The probabilities of these sets add to $(p_a + p_b)^4$. Finally, if there are no known contributors, then both alleles *ab* must be carried by the unknown contributors

$$P_x(ab|ab) = (p_a + p_b)^{2x} - p_a^{2x} - p_b^{2x}$$

This allows for all combinations of x people who have no alleles other than *a* or *b*, except the cases where all x people are *aa* or all are *bb*. These last two cases could not result in an *ab* profile for the mixed stain. When $x = 2$, the possible sets of people are *aa*, *ab*; *aa*, *bb*; *ab*, *ab*; *ab*, *bb*. These sets have probabilities adding to $(p_a + p_b)^4 - p_a^4 - p_b^4 = 2p_ap_b(2p_a^2 + 3p_ap_b + 2p_b^2)$.

Profiles with Three Alleles

If the profile has alleles *abc*, then all contributors must be of genotype *aa*, *ab*, *bb*, *ac*, *bc*, or *cc*. If known contributors already have alleles *abc*, then unknowns need not have any specific alleles but cannot have alleles other than *abc*, so

$$P_x(\phi|abc) = (p_a + p_b + p_c)^{2x}$$

The probability that x unknown contributors have allele *a*, and no allele other than *a*, *b* or *c*, is

$$P_x(a|abc) = (p_a + p_b + p_c)^{2x} - (p_b + p_c)^{2x}$$

When $x = 1$ this simplifies to $p_a^2 + 2p_ap_b + 2p_ap_c$, corresponding to *aa*, *ab*, or *ac* genotypes.

The probability that x unknown contributors have alleles *ab*, and no allele other than *ab* or *c*, is

$$P_x(ab|abc) = (p_a + p_b + p_c)^{2x} - (p_b + p_c)^{2x} - (p_a + p_c)^{2x} + p_c^{2x}$$

When $x = 1$, this simplifies to $2p_ap_b$ as only *ab* heterozygotes can contribute both *a* and *b*.

Finally, if there are no known contributors, then all alleles *abc* must be carried by the unknown contributors, which means that $x > 1$, and

$$P_x(abc|abc) = (p_a + p_b + p_c)^{2x} - (p_a + p_b)^{2x} - (p_b + p_c)^{2x} - (p_a + p_c)^{2x} + p_a^{2x} + p_b^{2x} + p_c^{2x}.$$

This expression is zero when $x = 1$, as it is not possible for one person to contribute all of alleles *a*, *b*, and *c*.

Profiles with Four Alleles

Similar arguments lead to the four-allele results:

$$P_x(\phi|abcd) = (p_a + p_b + p_c + p_d)^{2x}$$

$$P_x(a|abcd) = (p_a + p_b + p_c + p_d)^{2x} - (p_b + p_c + p_d)^{2x}$$

$$P_x(ab|abcd) = (p_a + p_b + p_c + p_d)^{2x} - (p_b + p_c + p_d)^{2x} - (p_a + p_c + p_d)^{2x} + (p_c + p_d)^{2x}$$

$$P_x(abc|abcd) = (p_a + p_b + p_c + p_d)^{2x} - (p_b + p_c + p_d)^{2x} - (p_a + p_c + p_d)^{2x} - (p_a + p_b + p_d)^{2x} + (p_c + p_d)^{2x} + (p_b + p_d)^{2x} + (p_a + p_d)^{2x} - p_d^{2x},\ x > 1$$

$$P_x(abcd|abcd) = (p_a + p_b + p_c + p_d)^{2x} - (p_b + p_c + p_d)^{2x} - (p_a + p_c + p_d)^{2x} - (p_a + p_b + p_d)^{2x} - (p_a + p_b + p_c)^{2x} + (p_c + p_d)^{2x} + (p_b + p_d)^{2x} + (p_b + p_c)^{2x} + (p_a + p_d)^{2x} + (p_a + p_c)^{2x} + (p_a + p_b)^{2x} - p_a^{2x} - p_b^{2x} - p_c^{2x} - p_d^{2x},\ x > 1$$

Examples

These results allow us to return to the examples in Equations 1 and 2. Setting $x = 1$

$$P_1(\phi|abcd) = (p_a + p_b + p_c + p_d)^2$$

$$P_1(cd|abcd) = 2p_cp_d$$

and with $x = 2$

$$P_2(cd|abcd) = 2p_cp_d[6(p_a + p_b)^2 + 6(p_a + p_b)(p_c + p_d) + 2(p_c + p_d)^2 - p_cp_d]$$

For Equation 1, therefore,

$$L = \frac{P_0(\phi|abcd)}{P_1(cd|abcd)} = \frac{1}{2p_cp_d}$$

which is just the standard result for the case when all alleles are explained by C and there are two alleles, *cd*, from an unknown person under $\overline{C}$.

For Equation 2,

$$L = \frac{P_1(\phi|abcd)}{P_2(cd|abcd)} = \frac{(p_a + p_b + p_c + p_d)^2}{2p_cp_d[6(p_a + p_b)^2 + 6(p_a + p_b)(p_c + p_d) + 2(p_c + p_d)^2 - p_cp_d]}$$

An important feature of this result is that it is not the reciprocal of the frequency of the suspect's profile. The analysis of the mixed profile has required the use of likelihood ratios.

Profiles with m *Alleles*

The general expressions for x contributors when there are three or four alleles seem fairly cumbersome, but they follow a pattern.

We noticed that a single equation can be given for the probability that x unknown contributors have at least the alleles in some set of alleles U, but have no alleles not in the mixed-profile allele set E. A formal proof of the equation was supplied by C.H. Brenner and is contained as Appendix 2 to this paper. The equation is

$$P_x(U|E) = (T_0)^{2x} - \sum_j (T_{1_j})^{2x} + \sum_{j,k} (T_{2_{j,k}})^{2x} - \sum_{j,k,l} (T_{3_{j,k,l}})^{2x} + \cdots \quad (3)$$

where T_0 is the sum of probabilities of all alleles in E, T_{1_j} is the sum of probabilities of all alleles in E except for the jth allele in U, $T_{2_{j,k}}$ is the sum of probabilities of all alleles in E except for the jth and kth alleles of U, and so on. Such equations are useful for constructing computer programs, and we give instructions below for obtaining such a program. However, the explicit results we list in Appendix 1 will generally be sufficient.

As an illustration of the use of Appendix 1, suppose an evidentiary sample E has only three alleles *abc* and includes contributions from a victim and a perpetrator. If the victim has genotype *ab* and a suspect has profile *aa*, then the explanation C including the suspect still requires unknown contributors for allele *c*. The alternative explanation $\overline{C}$ may still include the victim but excludes the suspect, and so includes some unknown contributors. If the circumstances of the crime specify that there were three contributors,

$$L = \frac{P_1(c|abc)}{P_2(c|abc)} = \frac{p_c(2p_a + 2p_b + p_c)}{(p_a + p_b + p_c)^4 - (p_a + p_b)^4} = \frac{p_c}{(p_a + p_b + p_c)^2 + (p_a + p_b)^2} \quad (4)$$

Number of Contributors

The results give so far depend on the number of contributors to the mixed sample. Sometimes the circumstances of the crime will dictate that number, and the forensic scientist must use that number in the numerical analysis. Other times, the number may not be known but the presence of more than two alleles at a locus implies that there was more than one contributor. However, if all the loci in the profile had no more than four alleles, but these loci had more than four alleles in the population, an argument could be made that explanations with other than two contributors may safely be ignored. Other situations that may indicate the number of contributors are when a mixed stain occurs amongst a collection of single-contributor stains.

Whenever there is doubt as to the number of contributors, there can be considerable variation in the likelihood ratio. The issue has been discussed in some detail by C.H. Brenner, R. Fimmers and M. Baur (submitted, personal communication from C.H. Brenner). For loci with more alleles than are present in the sample profile, so that the sum of probabilities of those alleles is less than one, larger numbers of unknown people make $P_x(U|E)$ smaller—it becomes increasingly difficult for a large number of people to have only the sample alleles between them. This result also follows from Equation 3, since $T_0 < 1$ means that $P_x(U|E) \to 0$ as x increases. If the numerator of the likelihood ratio is fixed, the stronger is the evidence against the suspects included under $\overline{C}$ as the number of contributors increases. Conversely, smaller x values for the denominator will make the likelihood ratio smaller and will weaken the evidence against suspects included under C but not $\overline{C}$.

When all the alleles at a locus are present in the evidentiary sample, however, an increased number of contributors makes the profile more likely. It becomes easier for large numbers of people to have all the alleles at the locus between them. Equation 3 now shows that $P_x(U|E) \to 1$ as x increases since $T_0 = 1$. The likelihood ratio is now smaller for larger x values.

These results on the effect of x on $P_x(U|E)$ are strictly true only when x is large and may not hold for small x. Care is needed, and it is advisable to obtain numerical values for several values of x.

Examples

We now show the effects of both the numbers of contributors and the profiles of specific people, by considering an example of a rape case (R. Cotton, personal communication). The Polymarker™ profile from a vaginal swab, as well as profiles from the victim and a suspect, are shown in Table 1, along with some allele frequencies used for illustrative purposes. Both the victim and the suspect profiles are included in the evidence profile.

If the two explanations for this evidence are:

C: Contributors were the victim and the suspect,
$\overline{C}$: Contributors were the victim and an unknown person.

then the evidence is certain under C. The probabilities under $\overline{C}$ are shown in Table 2, and the likelihood ratio is 8.9.

Evidence from a rape case may be collected from somewhere other than victim's body, and explanation $\overline{C}$ may then be that both contributors are unknown. With C still that the evidence is from the victim and the suspect, the probabilities are as shown in Table 3. The likelihood ratio is now 46.6, and has increased because of the larger number of unknowns in the denominator. The evidence did not contain all the alleles at each locus.

Finally, consider the case where the evidence in Table 1 is known to be from two perpetrators, but only one (the "suspect" of Table 1) is known. The evidence is no longer regarded as being

TABLE 1—*Polymarker™ example.*

Profile	LDLR	GYPA	HBGG	D7S8	Gc
Evidence	B	AB	AB	AB	ABC
Victim	B	AB	AB	AB	AC
Suspect	B	A	A	A	B
p_A		0.538	0.566	0.543	0.253
p_B	0.567	0.462	0.429	0.457	0.195
p_C					0.552

TABLE 2—*Calculations for rape case, one unknown.*

	$Pr(E\|C)$	$Pr(E\|\overline{C})$	
LDLR	1	$P_1(B\|B) = p_B^2$	0.321
GYPA	1	$P_1(\phi\|AB) = (p_A + p_B)^2$	1.000
HBGG	1	$P_1(\phi\|AB) = (p_A + p_B)^2$	0.990
D7S8	1	$P_1(\phi\|AB) = (p_A + p_B)^2$	1.000
Gc	1	$P_1(B\|ABC) = p_B(p_B + 2p_A + 2p_C)$	0.352

TABLE 3—*Calculations for rape case, two unknowns.*

	$Pr(E\|C)$	$Pr(E\|\overline{C})$	
LDLR	1	$P_2(B\|B) = p_B^4$	0.103
GYPA	1	$P_2(AB\|AB) = (p_A + p_B)^4 - p_A^4 - p_B^4$	0.871
HBGG	1	$P_2(AB\|AB) = (p_A + p_B)^4 - p_A^4 - p_B^4$	0.843
D7S8	1	$P_2(AB\|AB) = (p_A + p_B)^4 - p_A^4 - p_B^4$	0.869
Gc	1	$P_2(ABC\|ABC) = 12p_Ap_Bp_C(p_A + p_B + p_C)$	0.327

from a rape case and the "victim" profile in Table 1 is ignored. The two explanations are now:

C: Contributors were the suspect and an unknown,
$\overline{C}$: Contributors were both unknown.

The evidence is not certain under either C or $\overline{C}$, and the probabilities are shown in Table 4. For four of the five loci, the evidence is more likely under $\overline{C}$ than under C. It is more likely that two people will carry alleles AB at GYPA between them, for example, than it is for a single person to carry only allele B. The likelihood ratio is less than one for these loci, although it is 1.1 for all five loci combined. Simply considering the frequency of included people, at loci other than LDLR, would give a frequency of 1.0 and would imply no probative value of the evidence, whereas the evidence at these four loci is actually favorable to the suspect. The presence of alleles from an unknown person has weakened the evidence against the suspect.

Profiles with Unseen Alleles

Adding Null Allele

An RFLP locus can give rise to a single band c on an electrophoretic gel if it is homozygous cc, or if it is heterozygous cn for c and some unseen allele n. Such an unseen allele may need to be added to profiles of people for whom only one band is seen, and then also added to mixture profiles to which that person is supposed to have contributed.

For example, suppose the evidentiary sample has profile abc, the victim has type ab and a suspect has type c. If it is appropriate to assume the existence of unseen bands, the suspect's profile could be cc or cn and the sample profile abc or $abcn$. In either case, the evidence is certain under explanation C that the contributors were the victim and suspect. However, if $\overline{C}$ includes the victim but excludes the suspect, the unknown contributor of allele c must possess allele c and not possess an allele, other than a or b, that would be "seen." The possible genotypes for this person are ac, bc, cc, or nc and

$$L = \frac{1}{P_1(c|abcn)} = \frac{1}{p_c(2p_a + 2p_b + 2p_n + p_c)}$$

Estimates of null frequencies have all been less than 0.05 (Chakraborty et al., 1994).

Although null alleles require care, the machinery previously established is adequate provided the null n is added to the evidentiary profile. It need not appear in the set U in the expression $P_x(U|E)$.

TABLE 4—*Calculations for one or two unknowns.*

	$Pr(E\|C)$	$Pr(E\|\overline{C})$
LDLR	$P_1(\phi\|B) = 0.321$	$P_2(B\|B) = 0.103$
GYPA	$P_1(B\|AB) = 0.711$	$P_2(AB\|AB) = 0.871$
HBGG	$P_1(B\|AB) = 0.670$	$P_2(AB\|AB) = 0.843$
D7S8	$P_1(B\|AB) = 0.705$	$P_2(AB\|AB) = 0.869$
Gc	$P_1(AC\|ABC) = 0.215$	$P_2(ABC\|ABC) = 0.327$

The "2p" Rule

For the interpretation of single-contributor stains, it is customary to estimate the frequency of single-band VNTR patterns as twice the frequency assigned to that band. If the frequency of allele a is p_a, then the total frequency of individuals showing only that allele is $p_a^2 + 2p_ap_n$ where n is the possible unseen allele. The customary rule is the conservative replacement of this expression by $2p_a$, and the National Research Council (1996) recommends the same procedure for mixed stains. The general expression for the probabilities $P_x(U|E)$ can be modified to accommodate this rule, and we begin with some examples. Once the possibility of single bands is allowed for alleles from known contributors, we allow it for all contributors (and all alleles).

For a profile with two alleles, ab, contributors must be a^*, b^* or ab where a^* or b^* means individuals showing only allele a or b. They may be homozygotes or heterozygotes for an unseen allele, and are assigned a frequency of $2p_a$ or $2p_b$. In the same notation as before

$$P_1(a|ab) = 2p_a + 2p_ap_b = 2[(p_a + p_b + p_ap_b) - p_b]$$

with a generalization to x contributors of

$$P_x(a|ab) = 2^x[(p_a + p_b + p_ap_b)^x - p_b^x]$$

For a profile with three alleles, abc, contributors must be a^*, b^*, c^*, ab, ac, or bc. For two contributors to have at least ab but no more than abc between them, the possibilities are a^* with b^*, ab or bc; and b^* with ab or ac; and ab with c^*, ab, ac or bc; and ac with bc. Therefore

$$\begin{aligned} P_2(ab|abc) &= 4p_a(2p_b + 2p_ap_b + 2p_bp_c) + 4p_b(2p_ap_b + 2p_ap_c) \\ &\quad + 4p_ap_b(2p_c + p_ap_b + 2p_ap_c + 2p_bp_c) + 8p_ap_bp_c^2 \\ &= 4[(p_a + p_b + p_c + p_ap_b + p_ap_c + p_bp_c)^2 \\ &\quad - (p_b + p_c + p_bp_c)^2 - (p_a + p_c + p_ap_c)^2 + p_c^2] \end{aligned}$$

with a generalization to x contributors of

TABLE 5—*Two-contributor examples from 1996 NRC Report. The contributors are the suspect and one unknown person under C, and two unknown people under $\overline{C}$.*

E	Suspect	General	Likelihood Ratio	
			p^2	$2p$
$abcd$	ab	$\frac{P_1(cd\|abcd)}{P_2(abcd\|abcd)}$	$\frac{1}{12p_ap_b}$	$\frac{1}{12p_ap_b}$
abc	bc	$\frac{P_1(a\|abc)}{P_2(abc\|abc)}$	$\frac{p_a + 2p_b + 2p_c}{12p_bp_c(p_a + p_b + p_c)}$	$\frac{1 + p_b + p_c}{4p_bp_c(3 + p_a + p_b + p_c)}$
abc	a	$\frac{P_1(bc\|abc)}{P_2(abc\|abc)}$	$\frac{1}{6p_a(p_a + p_b + p_c)}$	$\frac{1}{4p_a(3 + p_a + p_b + p_c)}$
ab	ab	$\frac{P_1(\phi\|ab)}{P_2(ab\|ab)}$	$\frac{(p_a + p_b)^2}{2p_ap_b(2p_a^2 + 3p_ap_b + 2p_b^2)}$	$\frac{p_a + p_b + p_ap_b}{2p_ap_b(2 + 2p_a + 2p_b + p_ap_b)}$
ab	a	$\frac{P_1(b\|ab)}{P_2(ab\|ab)}$	$\frac{2p_a + p_b}{2p_a(2p_a^2 + 3p_ap_b + 2p_b^2)}$	$\frac{1 + p_a}{2p_a(2 + 2p_a + p_b + p_ap_b)}$

TABLE 6—*Calculations for a three-allele VNTR mixed profile. Under C, the known contributors have all three alleles abc. Under $\overline{C}$ all contributors are unknown.*

No. unknowns under $\overline{C}$		No. of unknowns under C		
		0	1	2
2	p^2	$\frac{P_0(\phi\|abc)}{P_2(abc\|abc)} = 1{,}620$	$\frac{P_1(\phi\|abc)}{P_2(abc\|abc)} = 70$	$\frac{P_2(\phi\|abc)}{P_2(abc\|abc)} = 3$
	$2p$	$\frac{P_0(\phi\|abc)}{P_2(abc\|abc)} = 158$	$\frac{P_1(\phi\|abc)}{P_2(abc\|abc)} = 70$	$\frac{P_2(\phi\|abc)}{P_2(abc\|abc)} = 31$
	n	$\frac{P_0(\phi\|abcn)}{P_2(abc\|abcn)} = 3{,}380$	$\frac{P_1(\phi\|abcn)}{P_2(abc\|abcn)} = 226$	$\frac{P_2(\phi\|abcn)}{P_2(abc\|abcn)} = 15$
3	p^2	$\frac{P_0(\phi\|abc)}{P_3(abc\|abc)} = 21{,}600$	$\frac{P_1(\phi\|abc)}{P_3(abc\|abc)} = 938$	$\frac{P_2(\phi\|abc)}{P_3(abc\|abc)} = 41$
	$2p$	$\frac{P_0(\phi\|abc)}{P_3(abc\|abc)} = 54$	$\frac{P_1(\phi\|abc)}{P_3(abc\|abc)} = 24$	$\frac{P_2(\phi\|abc)}{P_3(abc\|abc)} = 11$
	n	$\frac{P_0(\phi\|abcn)}{P_3(abc\|abcn)} = 12{,}300$	$\frac{P_1(\phi\|abcn)}{P_3(abc\|abcn)} = 823$	$\frac{P_2(\phi\|abcn)}{P_3(abc\|abcn)} = 55$
4	p^2	$\frac{P_0(\phi\|abc)}{P_4(abc\|abc)} = 396{,}000$	$\frac{P_1(\phi\|abc)}{P_4(abc\|abc)} = 17{,}200$	$\frac{P_2(\phi\|abc)}{P_4(abc\|abc)} = 748$
	$2p$	$\frac{P_0(\phi\|abc)}{P_4(abc\|abc)} = 8$	$\frac{P_1(\phi\|abc)}{P_4(abc\|abc)} = 30$	$\frac{P_2(\phi\|abc)}{P_4(abc\|abc)} = 13$
	n	$\frac{P_0(\phi\|abcn)}{P_4(abc\|abcn)} = 108{,}000$	$\frac{P_1(\phi\|abcn)}{P_4(abc\|abcn)} = 7{,}250$	$\frac{P_2(\phi\|abcn)}{P_4(abc\|abcn)} = 483$

$$P_x(ab|abc) = 2^x[(p_a + p_b + p_c + p_ap_b + p_ap_c + p_bp_c)^x - (p_b + p_c + p_bp_c)^x - (p_a + p_c + p_ap_c)^x + p_c^x]$$

The "2p" modification of Equation 3, for x unknowns having at least the alleles in U but no alleles not in E, is

$$P_x(U|E) = 2^x\left[(S_0)^x - \sum_j (S_{1_j})^x + \sum_{j,k} (S_{2_{j,k}})^x - \sum_{j,k,l} (S_{3_{j,k,l}})^x \cdots\right]$$

where S_0 is the sum of all the probabilities of the alleles in E plus the sum of all the products of pairs of frequencies for alleles in E. Term S_{1_j} is S_0 with all the terms involving the jth allele of U removed, $S_{2_{j,k}}$ is S_0 with all the terms involving the jth and kth alleles in U removed, and so on. To see why this result holds, note that S_0 is the square of the sum of the frequencies of all the alleles in E, with each squared frequency (p^2) replaced by twice that frequency ($2p$). The most common examples are shown in Appendix 1, and in Table 5 we confirm the results given by the National Research Council (1996). We adopt the language of that report in writing "p^2" for calculations where unseen bands are not an issue, and "$2p$" where single band profiles are given twice the frequency of the band that is seen. It must be pointed out that the "$2p$" rule cannot be used for loci with few alleles when S_0 is greater than one.

A comparison of the unseen-allele and the "$2p$" approaches is afforded by a three-band profile *abc* at D2S44 recovered from the center console of an automobile owned by the defendant in the case of People v. Simpson (Los Angeles County Case BA097211). This profile included the profile *ab* of the defendant OS and the profile *ac* of a victim RG (it did not include a second victim NB). Some frequencies for illustrative purposes are $p_a = 0.0316$, $p_b = 0.0842$, $p_c = 0.0926$. Three sets of calculations are laid out in Table 6: The possibility of unseen alleles could be ignored, an unseen allele (with frequency 0.05) could be allowed, or the "$2p$" rule could be used. In this instance, the court ordered that the number of contributors to the evidence profile be set to two, three or four. The "$2p$" rule is not always conservative, and we suggest caution in its use.

Discussion

The interpretation of mixed stains is straightforward in the likelihood ratio context. Alternative explanations for the mixed stain profiles need to be specified, and then compared on the basis of the probabilities of the profile under those explanations. Calculations that consider only single contributors are without a logical foundation. The 1992 report of the United States National Research Council recommended "If a suspect's pattern is found within the mixed pattern, the appropriate frequency to assign such a 'match' is the sum of frequencies of all the genotypes that are contained within (i.e., that are a subset of) the mixed pattern." We have shown by example that this method is not helpful, and may even be prejudicial. Indeed, the 1996 National Research Council report repudiates the 1992 recommendation: "This calculation is hard to justify, because it does not make use of some of the information available, namely, the genotype of the suspect. The correct procedure, we believe, was described by Evett et al. (1991)."

A program to calculate the likelihood ratio for mixed stains, for a range of numbers of unknown contributors under both C and $\overline{C}$, can be obtained from the senior author. (e-mail address: weir@ncsu.stat.edu or World Wide Web URL: http://www2.ncsu.edu/ncsu.CIL/stat__genetics/).

Acknowledgments

The wise counsel of Dr. Ian Evett, and the insights of Dr. Charles Brenner, are greatly appreciated.

References

1. Aitken CGG. Statistics and the evaluation of evidence for forensic scientists. New York: Wiley, 1995.
2. Balding DJ, Nichols RA. DNA profile match probability calculation: How to allow for population stratification, relatedness, database selection and single bands. Forensic Sci Int 1994;64:125–40.
3. Chakraborty R, Zhong Y, Jin L, Budowle B. Nondetectability of restriction fragment sizes within and between loci in RFLP typing of DNA. Am J Human Genet 1994;55:391–401.
4. Evett IW. What is the probability that this blood came from that person? A meaningful question? J Forensic Sci 1983;23:35.
5. Evett IW, Buffery C, Wilcott G, Stoney D. A guide to interpreting single locus profiles of DNA mixtures in forensic case. J Forensic Sci Soc 1991;31:41–7.
6. Gill P, Jeffreys AJ, Werrett DJ. Forensic application of DNA 'fingerprints.' Nature 1985;318:577–79.
7. Lindley DV. A problem in forensic science. Biometrika 1977; 64:207–13.
8. National Research Council. DNA technology in forensic science. Washington, DC: National Academy Press, 1992.
9. National Research Council. The evaluation of forensic DNA evidence. Washington, DC; National Academy Press, 1996.
10. Walker RH, Duquesnoy RJ, Jennings ER, Krause HD, Lee CL, Polesky H, editors. Inclusion probabilities in parentage testing. Arlington, VA; American Association of Blood Banks, 1983.
11. Weir BS. The effects of inbreeding on forensic calculations. Ann Rev Genet 1994;28:597–621.

Additional information and reprint requests:
Dr. B.S. Weir
Program in Statistical Genetics
Department of Statistics
North Carolina State University
Raleigh, NC 27695-8203

APPENDIX 1—SPECIAL CASES

One allele

	p^2 rule	$2p$ rule
$P_x(\phi\|a) = P_x(a\|a)$	p_a^{2x}	$2^x p_a^x$

Two alleles

	p^2 *rule*	$2p$ rule
$P_x(\phi\|ab)$	$(p_a + p_b)^{2x}$	$2^x(p_a + p_b + p_a p_b)^x$
$P_1(a\|ab)$	$p_a(p_a + 2p_b)$	$2p_a(1 + p_b)$
$P_2(a\|ab)$	$p_a(p_a^3 + 4p_a^2 p_b + 6p_a p_b^2 + p_b^3)$	$4p_a(1 + p_b)(p_a + p_a p_b + 2p_b)$
$P_x(a\|ab)$	$(p_a + p_b)^{2x} - p_b^{2x}$	$2^x[(p_a + p_b + p_a p_b)^x - p_b^x]$
$P_1(ab\|ab)$	$2p_a p_b$	$2p_a p_b$
$P_2(ab\|ab)$	$2p_a p_b(2p_a^2 + 3p_a p_b + 2p_b^2)$	$4p_a p_b(2 + 2p_a + 2p_b + p_a p_b)$
$P_x(ab\|ab)$	$(p_a + p_b)^{2x} - p_a^{2x} - \mathrm{p}_b^{2x}$	$2^x[(p_a + p_b + p_a p_b)^x - p_a^x - p_b^x]$

Three alleles

$P_x\phi\|abc)$	p^2	$(p_a + p_b + p_c)^{2x}$
	$2p$	$2^x(p_a + p_b + p_c + p_a p_b + p_a p_c)^x$
$P_1(a\|abc)$	p^2	$p_a(p_a + 2p_b + 2p_c)$
	$2p$	$2p_a(1 + p_b + p_c)$
$P_x(a\|abc)$	p^2	$(p_a + p_b + p_c)^{2x} - (p_b + p_c)^{2x}$
	$2p$	$2^x[(p_a + p_b + p_c + p_a p_b + p_a pc + p_b p_c)^x - (p_b + p_c + p_b p_c)^x]$
$P_1(ab\|abc)$	p^2	$2p_a p_b$
	$2p$	$2p_a p_b$
$P_2(ab\|abc)$	p^2	$2p_a p_b[2(p_a + p_b)^2 + 6p_c(p_a + p_b) + 6p_c^2 - p_a p_b]$
	$2p$	$4p_a p_b(p_a p_b + 2 + p_a + p_b) + 8p_a p_b p_c(3 + p_a + p_b + p_c)$
$P_x(ab\|abc)$	p^2	$(p_a + p_b + p_c)^{2x} - (p_b + p_c)^{2x} - (p_a + p_c)^{2x} + p_c^{2x}$
	$2p$	$2^x[(p_a + p_b + p_c + p_a p_b + p_a p_c + p_b p_c)^x$ $- (p_b + p_c + p_b p_c)^x - (p_a + p_c + p_a p_c)^x + p_c^x]$
$P_1(abc\|abc)$	p^2	0
	$2p$	0
$P_2(abc\|abc)$	p^2	$12p_a p_b p_c(p_a + p_b + p_c)$
	$2p$	$8p_a p_b p_c(3 + p_a + p_b + p_c)$
$P_x(abc\|abc)$	p^2	$(p_a + p_b + p_c)^{2x} - (p_b + p_c)^{2x} - (p_a + p_c)^{2x}$ $- (p_a + p_b)^{2x} + p_a^{2x} + \mathrm{p}_b^{2x} + \mathrm{p}_c^{2x}$
	$2p$	$2^x[(p_a + p_b + p_c + p_a p_b + p_a p_c + p_b p_c)^x$

Four alleles

$P_x(\phi\|abcd)$	p^2	$(p_a + p_b + p_c + p_d)^{2x}$
	$2p$	$2^x(p_a + p_b + p_c + p_d + p_a p_b + p_a p_c + p_a p_d + p_b p_c + p_b p_d + p_c p_d)^{2x}$
$P_1(a\|abcd)$	p^2	$p_a(p_a + 2p_b + 2p_c + 2p_d)$
	$2p$	$2p_a(1 + p_b + p_c + p_d)$
$P_x(a\|abcd)$	p^2	$(p_a + p_b + p_c + p_d)^{2x} - (p_b + p_c + p_d)^{2x}$
	$2p$	$2^x[(p_a + p_b + p_c + p_d + p_a p_b + p_a p_c + p_a p_d + p_b p_c + p_b p_d + p_c p_d)^x$ $- (p_b + p_c + p_d + p_b p_b + p_b p_d + p_c p_d)^x]$

$P_1(abc\|abcd)$	p^2	$2p_ap_b$
	$2p$	$2p_ap_b$
$P_x(ab\|abcd)$	p^2	$(p_a + p_b + p_c + p_d)^{2x} - (p_b + p_c + p_d)^{2x} - (p_a + p_c + p_d)^{2x} + (p_c + p_d)^{2x}$
$P_x(ab\|abcd)$	$2p$	$2^x[(p_a + p_b + p_c + p_d + p_ap_b + p_ap_c + p_ap_d + p_bp_c + p_bp_d + p_cp_d)^x$ $- (p_b + p_c + p_d + p_bp_c + p_bp_d + p_cp_d)^x$ $- (p_a + p_c + p_d + p_ap_c + p_ap_d + p_cp_d)^x + (p_c + p_d + p_cp_d)^x]$
$P_1abc\|abcd)$	p^2	0
	$2p$	0
$P_2(abc\|abcd)$	p^2	$12p_ap_bp_c(p_a + p_b + p_c + 2p_d)$
	$2p$	$8p_ap_bp_c(3 + p_a + p_b + p_c + 3p_d)$
$P_x(abc\|abcd)$	p^2	$(p_a + p_b + p_c + p_d)^{2x} - (p_b + p_c + p_d)^{2x} - (p_a + p_c + p_d)^{2x}$ $- (p_a + p_b + p_d)^{2x} + (p_c + p_d)^{2x} + (p_b + p_d)^{2x} + (p_a + p_d)^{2x} - p_d^{2x}$
	$2p$	$2^x[(p_a + p_b + p_c + p_d + p_ap_b + p_ap_c + p_ap_d + p_bp_c + p_bp_d + p_cp_d)^x$ $- (p_b + p_c + p_d + p_bp_c + p_bp_d + p_cp_d)^x$ $- (p_a + p_c + p_d + p_ap_c + p_ap_d + p_cp_d)^x$ $- (p_a + p_b + p_d + p_ap_b + p_ap_d + p_bp_d)^x$ $+ (p_c + p_d + p_cp_d)^x + (p_a + p_d + p_ap_d)^x + (p_b + p_d + p_bp_d)^x - p_d^x$
$P_1(abcd/abcd)$	p^2	0
	$2p$	0
$P_2(abcd/abcd)$	p^2	$24p_ap_bp_cp_d$
	$2p$	$24p_ap_bp_cp_d$
$P_x(abcd/abcd)$	p^2	$(p_a + p_b + p_c + p_d)^{2x}$ $- (p_b + p_c + p_d)^{2x} - (p_a + p_c + p_d)^{2x} - (p_a + p_b + p_d)^{2x}$ $- (p_a + p_b + p_c)^{2x} + (p_c + p_d)^{2x} + (p_b + p_d)^{2x} + (p_b + p_c)^{2x}$ $+ (p_a + p_d)^{2x} + (p_a + p_c)^{2x} + (p_a + p_b)^{2x}$ $- p_a^{2x} - p_b^{2x} - p_c^{2x} - p_d^{2x}$
	$2p$	$2^x[(p_a + p_b + p_c + p_d + p_ap_b + p_ap_c + p_ap_d + p_bp_c + p_bp_d + p_cp_d)^x$ $- (p_b + p_c + p_d + p_bp_c + p_bp_d + p_cp_d)^x$ $- (p_a + p_c + p_d + p_ap_c + p_ap_d + p_cp_d)^x$ $- (p_a + p_b + p_d + p_ap_b + p_ap_d + p_bp_d)^x$ $- (p_a + p_b + p_c + p_ap_b + p_ap_c + p_bp_c)^x + (p_c + p_d + p_cp_d)^x$ $+ (p_a + p_d + p_ap_d)^x + (p_b + p_d + p_bp_d)^x$ $+ (p_a + p_b + p_ap_b)^x + (p_a + p_c + p_ap_c)^x + (p_b + p_c + p_bp_c)^x$ $- p_a^x - p_b^x - p_c^x - p_d^x$

APPENDIX 2—PROOF OF A MIXED STAIN FORMULA OF WEIR[1]

The genetic markers ("alleles") of an evidence stain may be identical to the alleles of a reference sample (such as a suspect for example). The likelihood ratio for the evidentiary strength favoring association is then simply the inverse of the profile frequency. However, the evidence stain is often complicated by the presence of additional alleles, variously from additional known or unknown suspects or victims. The likelihood ratio is clearly more complicated in such cases, but Weir et al. (1997) present in the foregoing paper without proof a general and elegant formula for the probabilities that occur as numerator or denominator. In this appendix, we give a proof.

Received 24 May 1996; and in revised form 7 Aug. 1996; accepted 9 Aug. 1996.

[1]Charles H. Brenner, Ph.D., Forensic Consultant in Mathematics, Berkeley, CA 94709.

Notation

Let E be the set of alleles observed in the evidence for some discrete-allele system. Of these some may be attributable to known parties; the remainder $U \subset E$ are to be explained by x people with two alleles (not necessarily distinct) each. *Explained* means that $U \subset X \subset E$, where X is the set of all alleles in the x people. More generally for a subset $S \subset U$, we will say that people with alleles X *exactly explain* S if $X \subset E$ and $X \cap U = S$, or equivalently, putting $W = U \backslash S$, if $U \backslash W \subset X \subset E \backslash W$.

The set-notation symbols are to be understood according to the standard conventions: $U \subset E$ means that the alleles U are among those of E, including the possibility that $U = E$. $X \cap U$ is the intersection—the set of alleles that are both in X and in U. $U \backslash S$ is the set difference—the set of alleles that are in U excluding those that are also in S. The *cardinality* (size, in number of alleles) of a set J is written $|J|$. The symbol $\in$ denotes membership; $j \in U$ means that j is an allele of the set U.

Following Weir et al. we write $P_x(U|E)$ for the probability that x random people explain U. Let J, $J \subset U$ be a set of alleles. We will be interested in sets of people who omit the set J. Let

$$T_{m_J} = \text{the total of the frequencies of the alleles in } E \backslash J. \quad (1)$$

Theorem

Weir discovered that

$$P_x(U|E) = T_0^{2x} - \sum_{j\in U} T_{1_j}^{2x} + \sum_{j,J\in U} T_{2_{j,k}}^{2x} - \sum_{j,k,J\in U} T_{3_{j,k,l}}^{2x} + - \cdots \quad (2)$$

but did not supply a proof.

Proof

A proof seems worthwhile. The general idea is clear enough—(2) is an instance of the principle of inclusion and exclusion (Hall, 1967). From the definition (1) and the assumption of a discrete allele system, $T_{m_J}^{2x}$ is the probability that x people's alleles are all in $E\backslash J$. As the basis for the inclusion-exclusion analysis, we note that

$$T_{m_J}^{2x} = \sum_{J\subset W\subset U} P_x(U\backslash W|E\backslash W) \quad (3)$$

because any set of people whose alleles are among $E\backslash J$ exactly explains some one and only one subset, $U\backslash W$, of $U\backslash J$. The summation is taken over all sets of alleles W that satisfy $J \subset W \subset U$. Introduction of the sets W, as a means of effectively classifying the various positive and negative contributions to the sum in (2), is the key idea in the proof.

Define

$$Q_m = \sum_{\substack{J\subset U\\ |J|=m}} T_{m_J}^{2x} \quad (4)$$

In this notation, (2) is expressed as

$$P_x(U|E) = Q_0 - Q_1 + Q_2 - + \cdots. \quad (5)$$

Summing (3) for fixed m over all sets $J \subset U$ of cardinality m we obtain from (4)

$$Q_m = \sum_{\substack{J\subset U\\ |J|=m}} \sum_{J\subset W\subset U} P_x(U\backslash W|E\backslash W) \quad (6)$$

$$= \sum_{k=m}^{n} \sum_{\substack{W\subset U\\ |W|=k}} \sum_{\substack{J\subset W\\ |J|=m}} P_x(U\backslash W|E\backslash W) \quad (7)$$

$$= \sum_{k=m}^{n} \sum_{\substack{W\subset U\\ |W|=k}} P_x(U\backslash W|E\backslash W) \sum_{\substack{J\subset W\\ |J|=m}} 1 \quad (8)$$

$$= \sum_{k=m}^{n} \sum_{\substack{W\subset U\\ |W|=k}} P_x(U\backslash W|E\backslash W)\binom{k}{m}. \quad (9)$$

On the right hand side of line (6), each W occurs many times, once for each J of which it is a superset. The object is to count how many times. Classifying the W's according to their size k on line (7), we see on (8) that it is the same as the number of m allele subsets of a k-set, which is exactly the definition of the binomial symbol $\binom{k}{m}$. Hence line (9).

To verify (5) form now the alternating sum over m, where the sum runs to $n = |U|$,

$$Q_0 - Q_1 + Q_2 - + \cdots$$

$$= \sum_{m=0}^{n} (-1)^m Q_m$$

$$= \sum_{m=0}^{n} (-1)^m \sum_{k=m}^{n} \sum_{\substack{W\subset U\\ |W|=k}} P_x(U\backslash W|E\backslash W)\binom{k}{m} \quad (10)$$

$$= \sum_{k=0}^{n} \sum_{\substack{W\subset U\\ |W|=k}} P_x(U\backslash W|E\backslash W) \sum_{m=0}^{k} (-1)^m\binom{k}{m}. \quad (11)$$

As line (10) shows the same set W may occur in several Q_m terms. To compute the net contribution due to each W, it is natural to reverse the order of summation so that the classification is on W first and then on m, which is formula (11). To verify the transition from (10) to (11) note that the index sets of the double summations $\sum_{m=0}^{n}\sum_{k=m}^{n}$ and $\sum_{k=0}^{n}\sum_{m=0}^{k}$ range over the same pairs (k, m)—namely the triangular array where $0 \le m \le k \le n$.

Hence the net number of times that a contribution from each set W is included and excluded is given by the last factor in (11). That factor is simply unity when $k = 0$, and when $k > 0$ is it even simpler, for by the binomial theorem $\sum_{m=0}^{k}(-1)^m\binom{k}{m} = (1 - 1)^k = 0$. So

$$Q_0 - Q_1 + Q_2 - + \cdots = P_x(U|E) + \sum_{k=1}^{n} \sum_{\substack{W\subset U\\ |W|=k}} P_x(U\backslash W|E\backslash W) \cdot 0$$

$$= P_x(U|E), \text{ Q.E.D.}$$

References

1. Hall M. Combinatorial Theory. New York: John Wiley & Sons, 1967.
2. Weir BS, Triggs CM, Starling L, Stowell LI, Walsh KAJ, Buckleton J. Interpreting DNA mixtures. J Forensic Sci 1997;42(2):213–222.

Additional information and reprint requests:
Charles H. Brenner, Ph.D.
2486 Hilgard Ave
Berkeley CA 94709