Identification of biological stains: probability of identity or of kinship

K. Hummel*), W. Bär**), and N. Fukshansky*)

*) Institute for Blood-Group Serology and Genetics. D-79008 Freiburg. Germany
**) Institute of Legal Medicine. CH-8057 Zürich. Switzerland

In cases of disputed identity, one should use probability values instead of only phenotypical frequencies in the population: $W_{id} = 1/[1+f(Ph)]$; W_{id} = probability of identity; f(Ph) = frequencies of the stain's phenotype in the adequate population. It is sufficient to report to the court probabilities because from these the judge like anyone else can appreciate the weight connected with a decision for or against identity in the actual case.

If $W_{id} \geq 99$ % the identity is "highly probable", if $W_{id} \geq 99{,}73$ % identity would be "practically proven". It must be mentioned, that *Bayes' probabilities*, as W_{id} values are, can only be valid if they include a *prior probability*. According to *Bayes' Postulate*, introduced 200 years ago, an expert should hold a neutral position with respect to the pros and cons and should therefore use a **neutral prior** of 0.5 which says that the suspect "*a priori*" - that is, for the first time - has an **equal** chance of being the producer of the stain or not. In reality real producers of stains are more frequently found among the suspected people than non-producers; therefore W_{id} values, based on a prior value of 0.5, are underrated. In court this may be considered as a "factor of caution".

The formula $W_{id} = 1/[1+f(Ph)]$ is only valid for cases with only **one** suspect and only if the stain is not a "mixed" one; i.e.: deriving from only **one** individual. Most forensic cases belong to this "standard type". In the other ones the simple formula can **not** be applied.

The **special** cases can be divided into different groups, depending on which mathematical method one has to apply.

One group might include cases, in which f(X) and f(Y) in the formula $W_{id} = f(X)/[f(X) + f(Y)]$ are frequencies of **pedigrees**, for example in cases in which the suspect claimed that not he but his brother produced the stain (= cases of "brother objection"). In other cases it is asked, if a certain individual is the sibling of the stain's producer. In this cases f(X) means the frequency of the pedigree in which the two persons are siblings, f(Y) the frequency of the two persons as not related with one another. W_K is the probability of kinship. In cases of "reverse paternity" it may be asked, if a couple will be the parents of a stain's producer. Then f(X) will be the frequency of the parent-child combination, and f(Y) the frequency of the 3 individuals, not related with one another.

The next group of special cases are "mixed-stain" cases. These have to be divided into two subgroups: The one without participation of the victim (= the stain does not show any band of the victim), the other one **with** findings of the victim (= the stain shows DNA bands deriving from the victim).

Let us assume that in a certain case of mixed stain two individuals are the producers of the stain, but only one suspect was arrested. The judge's question is: What is the probability that the suspect is one of the producers? The frequencies approach will never enable to give a reasonable

answer because there are two *hypotheses* to consider: "The suspected and an unknown produced the stain (=X)" and "Two unknown individuals are the producers of the stain (=Y)".

In another case arised the question, if two suspects (A;B) could be the producers of a stain. Then there exist four possible *hypotheses*:

1. A and B are the producers of the stain;
2. A and an unknown are the producers of the stain;
3. B and an unknown are the producers of the stain;
4. Two unknowns are the producers of the stain.

One has to calculate frequencies for all the four *hypotheses* - system for system. The final frequencies are reached by addition of the partial frequencies. Plausibilities for each *hypothesis* are found by comparison: f(1) + f(2) + f(3) + f(4) = 1. Calculations which have to be made by head are time consuming. A computer program does not exist as yet.

In the group of **mixed stains in which bands from the victim were found** one has to distinguish 4 different possibilities (also in combinations)

cases with one suspect;
cases with more then one suspect;
cases with one victim;
cases with more than one victim.

Table 1 shows the formulas of W_{id} for 5 different categories of band constellations found in mixed stains, this for cases with **one** suspect and **one** victim. Cases which are more complex (**more** than one suspect or/and **more** than one victim) need proper formulas, which are to be constructed depending on the special situation of the actual case.

Table 1: Formulas for calculate W_{id} in cases with mixed stains in which the victim participates

constell.	bands	frequencies	mixed stain	victim	suspect	formula
No.1	A	a	+	+	-	$W_{id} = 1/(1+2cd)$
	B	b	+	+	-	
	C	c	+	-	+	
	D	d	+	-	+	
No.2	A	a	+	+	-	$W_{id} = 1/(1+2bc)$
	B	b	+	-	+	
	C	c	+	-	+	
					or or	
No.3	A	a	+	+	- - +	$W_{id} = 1/\{1+2c[c+2(a+b)]\}$
	B	b	+	+	- + -	
	C	c	+	-	+ + +	
					or or	
No.4	A	a	+	+	+ + -	$W_{id} = 1/[1+(a+b)^2]$
	B	b	+	+	- + +	
					or	
No.5	A	a	+	+	- +	$W_{id} = 1/[1+(b^2+2ab)]$
	B	b	+	-	+ +	

It has to be emphasized, that W_{id} values were always **lower** when taking into account the victim's bands found in the stain. It is therefore indispensable to include the victim's bands in any biostatistical evaluation of stain results with regard to statements about their identity with possible producers.

Summary: In usual cases of disputed identity, one should use probability values instead of only phenotypical frequencies in the population.

If the case is more complicated, such as one with blood relationship between the involved persons or with mixed stains (with or without bands from the victim), then the probability of the *null hypothesis* will be defined by f(X)/[f(X)+f(Y)] and the probability of the *counter hypothesis* by f(Y)/[f(X)+f(Y)].

The frequencies f(X) and f(Y) can be calculated by head or by computer using an appropriate program (like the *kinship program* of Ihm and Hummel 1975). A further possibility is to allow our Institute in Freiburg to conduct the biostatistical evaluation of the findings in such an unusual case.

Reference:

Ihm P, Hummel K (1975) Ein Verfahren zur Ermittlung der Vaterschaftswahrscheinlichkeit aus Blutgruppenbefunden unter beliebiger Einbeziehung von Verwandten. Z ImmunForsch 149: 406-416