Am J Hum Genet 36:904–915, 1984

Some Fallacies in the Computation of Paternity Probabilities

MIKEL AICKIN[1]

SUMMARY

Legal identification of fathers by means of a "paternity probability" has been used in European courts for decades, and has recently been introduced into American courts and accepted by some of them. The voluminous literature on this topic contains virtually no fundamental criticism of the logical basis for the probabilistic computations. Here I suggest that the "paternity probability" suffers from three basic fallacies: (1) contrary to claims, the figure is not, in fact, the probability that the alleged father is the true father, (2) the denominator of the likelihood ratio used in the computation is driven by (sometimes self-contradictory) assumptions and is not based on facts, and (3) post-inclusionary computations are based on speculation about genotypes that does not constitute scientific evidence. It is recommended that pending the resolution of these difficulties "paternity probabilities" should not be computed or introduced as positive evidence of paternity.

INTRODUCTION

The recent rise of illegitimacy rates in the United States, together with a concomitant expansion of public financial support for single-parent families, has made the establishment of paternity an issue of increasing concern to state and local governments. Federal pressure, in the form of threats to reduce funding, has forced the state and county attorneys into a central, aggressive role in paternity suits, frequently displacing the traditional notion that the mother has the greatest interest in such proceedings. The consequence has been that the courts are not equipped to deal with the potential flood of litigation that would result if a substantial fraction of men accused of paternity refused out-of-court settlements.

An ideal solution here would be an infallible method of detecting paternity, one that would convince correctly accused fathers of the futility of further legal

Received December 7, 1983; revised April 3, 1984.

[1] Statistical Consulting Services, 517 East Lodge Drive, Tempe, AZ 85283.

 0002-9297/84/3604-0017$02.00

action and immediately exonerate incorrectly accused men. It was with the hope of bringing to the court a sufficiently reliable, if not infallible, test for paternity that the American Association of Blood Banks (AABB) sponsored a conference to resolve outstanding differences among experts concerning the computation and interpretation of the genetically based "probability of paternity." The published results of that conference [1] contain guidelines that essentially encourage the admission as evidence of any mathematical or statistical computation concerning paternity, provided only that it is clearly explained. The purpose of this paper is to show that in the light of fundamental difficulties in the computation of paternity probabilities these guidelines are ill-advised.

PROOF AND PROBABILITY IN PATERNITY TESTING

A paternity dispute occurs when a mother and child are presented to the court, and the issue is to determine who is the father. In "one-man" cases, a single man is identified as the alleged father, while in "several-men" cases, there is a collection of men known to the court who are possible fathers. Until recently, the role of genetics has been to use serological evidence as a means of testing whether an alleged father could be excluded from paternity ([1], p. 21 ff.). Barring a mutation, laboratory accident, or break in the chain of evidence, a negative result from the "inclusion-exclusion test" exonerates the alleged father, and this traditional use of blood tests will not be questioned here.

However, as a direct result of progress in serological testing that permits the identification of several genetic systems, especially HLA, it is currently believed that when the alleged father is included the genetic evidence not only admits him as a possible father but also actually provides some positive support for the proposition that he is the father ([1], p. 63 ff.). Long before the modern advances, a technology for computing a "probability of paternity" was developed in a classic paper by Essen-Möller [2] ("The Power of Similarities to Prove Paternity"). This paper has become a seminal work for paternity testers, and has been quoted frequently in court and in the literature as the justification for various computations [3].

Throughout the literature on paternity probabilities, there has been a reluctance to examine the fundamental logic of Essen-Möller's formulation of the problem, and a consequent acceptance of paternity probabilities as scientific evidence. Here I suggest that in many of its applications this methodology embodies basic fallacies. This is not done to show that genetic evidence has no part in paternity disputes, but rather to argue that the current practice is overly optimistic, if not opportunistic, in its attitude about what biostatistical evidence can establish.

THE FRAMEWORK FOR COMPUTING PATERNITY PROBABILITIES

In ([1], p. 155 ff.), Kaye and I suggested that a welter of idiosyncratic systems of notation has arisen in dealing with paternity probability calculations. We argued that adherence to the standard notation of elementary probability theory was more likely to introduce clarity of thought into the computations. The notation of that paper will be used for the purposes of this one.

We imagine that there is a parameter, b, that indicates what the true but unknown situation is. We use $b = 0$ to stand for the case in which none of the possible fathers before the court is the actual father, and $b = i$ to mean that of the n possible fathers, the ith is the true father. In one-man cases, the only possibilities are $b = 0$ or $b = 1$.

We also use the symbol T to stand for the genetic evidence that is to be introduced in court. This evidence may differ widely from case to case, and so we intend to use T in a flexible way. In some cases, T will stand for the inclusion-exclusion test, so it will represent the outcome that a particular alleged father is either included or excluded. In other cases, T may stand for an entire list of the phenotypes of the mother, child, and all possible fathers known to the court.

Regardless of how T is used, the probabilities $P(T|b = 0)$ (computed under the assumption that none of the known possible fathers is the actual father) and $P(T|b = i)$ (computed under the assumption that the ith possible father is the real father) are the relevant figures to compute for any statistical inference. According to the Bayesian philosophy, each individual judge or juror should assess these values in the following way. Before the introduction of the serological evidence, one should formulate prior probabilities $P(b = 0)$ and $P(b = i)$ that reflect the degree of belief in each of these situations justified by the previous evidence in the case. To express degrees of belief after the serological evidence is admitted, each judge or juror should use generally accepted rules of probability to compute $P(b = i|T) = P(T|b = i)P(b = i)/K$, where K is the sum of the expressions $P(T|b = i)P(b = i)$ for i running from 0 to n. The number on the left in this equation is to be interpreted as the degree of belief that $b = i$ after the introduction of T.

It should not be thought that the Bayesian method follows in some way inexorably from the laws of probability. There is substantial controversy whether the situations $b = i$ have the same logical status as real-world observations such as T, and whether the same laws ought to apply to both of them. An alternative theory of generating beliefs from evidence has been proposed by Shafer [4], and stands as a logical alternative to the Bayesian method. Nevertheless, most paternity testers accept the Bayesian framework, and even those who do not still regard the computation of $P(T|b = i)$ as the ingredient necessary for coming to a conclusion about paternity.

In one-man cases, the Bayesian computation can be cast in a particularly simple form:

$$\frac{P(b = 1|T)}{P(b = 0|T)} = LR \frac{P(b = 1)}{P(b = 0)},$$

where LR is the likelihood ratio, and equals $P(T|b = 1)/P(T|b = 0)$. Even non-Bayesians recognize the LR as a crucial summary of the evidence in one-man cases, and it has come to be called the "paternity index." The Essen-Möller method is the special case in which the prior odds ratio $P(b = 1)/P(b = 0)$ equals 1. Large values of LR or $P(b = 1/T)$ are taken as scientific evidence of paternity.

THE ALTERNATIVE HYPOTHESIS FALLACY

In one-man cases, the parameter b is defined so that $b = 1$ corresponds to the situation in which the alleged father is the biological father, and $b = 0$ corresponds to the case in which he is not. This seems appropriate, because it correctly models the central factual finding that the judge or jury needs to make. Regardless of whether one chooses to use Bayesian or other inferential schemes, the problem is to come to a conclusion whether $b = 1$ or $b = 0$ is, in fact, correct.

When the *LR* assumes a value greater than 100, which it frequently does in cases of nonexcluded men where a rich genetic system like HLA is used, then it is common practice to follow Essen-Möller in reporting that the "probability of paternity" exceeds .99, so that the evidence in favor of paternity is quite great. When *LR* exceeds 1,000, the Essen-Möller formula yields .999, which the average judge or juror might be excused for thinking provides not only strong support for paternity, but also nearly a proof that the alleged father is the biological father. Indeed, one source [5] even goes so far as the use of the phrase "practically proved" in referring to the interpretation of such a high *LR*, and the AABB guidelines evidently approve the use of this phrase ([1], p. xiv). But what is it that has been "practically proved" by an *LR* greater than 1,000? The formal mathematical statement of the problem suggests that it is the paternity of the alleged father, but is this really what the biological evidence says?

The answer is clearly no, for the simple reason that with current technology the genetic evidence produced by a laboratory cannot distinguish which of two nonexcluded men is the actual father. Supporters of the use of the Essen-Möller and other similar formulas agree that the biological evidence cannot draw such distinctions, but they go on to argue that the *LR* correctly reflects the chances of the observed biological evidence under the two hypotheses ($b = 0$ and $b = 1$), and that use of standard statistical techniques yields the formulas for the "probability of paternity" that they put forth. Although this argument appears sound, and has been almost universally accepted by experts in paternity testing, it does not seem to deal with the problem that no laboratory evidence can pick the correct father from a group of men biologically capable of being the father.

To put this in concrete terms, imagine a situation in which the alleged father A has the same serological type as another possible father B, who is not known to the court. Suppose that the mother had sexual relations with both men. No matter which man the mother chooses to accuse of paternity, the "probability of paternity" will be the same. Even if the mother herself doesn't know which man is the father, and flips a coin to decide which to accuse, we can arrive at the result that paternity is "practically proved" for the randomly accused man. There is obviously something wrong with the reasoning that leads to the high "probability of paternity" in this case, and it is equally clear that it has something to do with the fact that the laboratory cannot produce any evidence suggesting that one or the other of the men is more likely to be the father.

To uncover the source of the fallacy, we have to go back to the fundamental definitions from which all the computations flow. It is not difficult to see that if we change $b = 1$ to mean that the alleged father "has the same serological

phenotype as the true father," rather than saying that he "is the true father," then *none of the calculations producing* LR *are in any way changed.* It is also not difficult to see that the probabilistic behavior of *LR* over many cases does not change when we alter the meaning of $b = 1$. It follows that the use of *LR* in statistical arguments, in effect, tests the null hypothesis that the alleged father is genetically distinguishable from the true father against the alternative hypothesis that he is not. If we make the redefinition, then we are using the biological evidence to determine whether or not the alleged father has the same genetic characteristics as the actual father, and this is the only kind of information that the serological evidence can supply.

It might be argued that this redefinition of b is a minor point, not materially altering the overall conclusion. Whether this turns out to be correct or not in any particular case is an interesting question. But one cannot go so far as to say that, in general, there is no real difference between asserting that it is "practically proved" that a man is the father and asserting that it is "practically proved" that he has the same phenotype as the father. Depending on the other evidence in the case, the distinction between these two conclusions can be either trivial or enormous. In the case of the serologically identical men, the difference is crucial, and the proper definition of b eliminates the undesirable property of "paternity probabilities" by which they might guarantee conviction of a man chosen at random. Here, it is quite reasonable to conclude that both men have a high probability of being genetically indistinguishable from the father, but it is nonsense to say that they both have high probabilities of being the father.

This fallacy is similar to a problem that is well known in the theory of linear models [6], and it is somewhat remarkable that statisticians had not discovered it earlier.

THE NULL HYPOTHESIS FALLACY IN ONE-MAN CASES

The alternative hypothesis fallacy seems to suggest that the current method of interpreting the *LR* needs to be changed, but it does nothing to suggest that the use of *LR* involves any basic impropriety. On the other hand, the null hypothesis fallacy suggests that in "one-man" cases the *LR* should not be computed or entered into the proceedings in any fashion. This second fallacy is independent of the first.

In obtaining the numerator of *LR*, one assumes for the purpose of computation that $b = 1$. This assumption, together with accepted principles of genetics, makes it possible (although in some cases complex) to compute the numerator, and it is difficult to raise any objections to the computation. However, the same may not be said of the denominator, where the assumption is that $b = 0$.

Let it be emphasized that the cases under investigation here are "one-man" cases, in which the only evidence before the court involves a single accused father. Here, the hypothesis $b = 0$ suffers from a certain vagueness. Simply saying that the alleged father is not the biological father (or is distinguishable from the biological father) does not in and of itself dictate how $P(T|b = 0)$ must be computed. What is required is a more precise description of the mechanism by which T might have arisen, when the alleged father is falsely accused. It is

nearly universally accepted by paternity-testing experts that the way to proceed is as if the actual father were drawn at random (in terms of his genetic characteristics) from some population. In some formulations of the problem, the assumption is that the alleged father was drawn at random, but the same basic reasoning is present in both approaches. Once the genetic frequencies of characteristics of that population are assumed known, then standard probabilistic computations using the rules of genetics yield the denominator probability.

Again, this computation is unobjectionable—under the assumptions that are being made. The question that needs to be addressed is whether the assumptions are reasonable. Several authors have pointed to an obvious weakness in the standard computation which results from the fact that the population, from which the true father is supposed to be drawn, may be poorly defined, and there may not be very good evidence about the distribution of genetic characteristics in that population. One feature of the argument is that in a rich system, like HLA, there are so many alleles and testing is so expensive that no proper statistical studies have been done on any population in order to estimate haplotype frequencies. The estimates that are available are evidently from "convenience samples," that is, samples consisting of whomever appeared at the blood bank or laboratory for testing. Since such a sampling plan has unknown characteristics, there is no theoretical way to establish how precise are the resulting estimates of frequencies. If the published frequencies misrepresent the population they are supposed to describe, then the subsequent computation of $P(T|b = 0)$ is incorrect, and worse, there is no way to estimate how much in error it may be.

On the other hand, supporters of the use of *LR* in one-man cases might point to the fact that with a system like HLA each possible phenotype occurs in the population with such a small frequency that it does not make much difference whether the frequencies are estimated precisely or not. For example, if *LR* is computed on the basis of the inclusion-exclusion test, then the denominator is the probability of including a randomly drawn man, and, again, assuming that the genetic systems investigated are sufficiently rich, this denominator will be very small. Whether the resulting *LR* turns out to be 1,000 or 2,000 doesn't seem like a very important issue: in either case, the evidence is strongly in favor of paternity, and we are only quibbling about exactly how strong it is.

However, the arguments against *LR* based on our ignorance of true gene frequencies miss the main point. The issue is not whether in a Los Angeles court case we should use the population of Los Angeles, of California, or of the entire country in computing the denominator. Neither these nor any other geographically determined populations are relevant to the case. What is pertinent is the population of "plausible fathers." This is the collection of men for whom one might have some reasonable belief, however small, that they might be the biological father. If, in fact, the alleged father is falsely accused, then any formulation of the hypothesis $b = 0$ must refer to the population of plausible fathers, because it is obvious that whatever selection process the mother used to choose whom to accuse, she used this population.

To see what effect might be expected from using plausible fathers rather than geographical populations, imagine a hypothetical case in which an observer in

the courtroom knows who the plausible fathers are and has the same genetic information on them that has been collected on the alleged father. This observer is capable of formulating the hypothesis $b = 0$ as a random draw from the collection of plausible fathers, and so he has no difficulty computing the denominator required for *LR*. For instance, if the inclusion-exclusion test is used, and if all of the other plausible fathers would be excluded as biological fathers, then his denominator would be zero, and it would be logically (not just probabilistically) proved that the alleged father were guilty. At the other extreme, if the other plausible fathers would all be included as possible biological fathers, then his denominator would be equal to one, and the genetic evidence in *LR* would say nothing about the alleged father. In intermediate cases, if the alleged father were included and if m of the n other plausible fathers were biologically capable of being the father, then the observer would compute m/n for the denominator.

There are three intriguing features of this. First, the computations for our observer are very simple and would be easily understood by anyone with a vague grasp of random draws and what it means to be excluded from paternity. This would be in contrast to the elaborate probabilistic computations that have occupied the attention of many paternity testers. Second, the population of other plausible fathers may be very difficult to identify; although one certainly would not want to admit the entire population of Los Angeles into this circle, it might be exceedingly difficult to determine reasonably whom this group ought to include. Third, the results obtained by our observer would usually differ materially from those based on a random draw from some geographical population (again, assuming a rich genetic system under test). This third fact is crucially important because it shows the degree to which the *assumptions* determine the conclusion.

Supporters of the use of *LR* will, of course, want to emphasize the second point above. They will argue that it would certainly be a good idea to discover the collection of plausible fathers, and they would have no qualms about permitting the denominator to be computed on this basis. But in practical situations, the difficulties standing in the way of discovering the plausible fathers are enormous. And even if they could be identified, it might be excessively time-consuming and costly to undertake genetic testing on all of them. Thus, the substitution of a known geographical population for the obviously more relevant population of plausible fathers is to be seen as an acceptable substitution, not only on the grounds of cost but also because we can imagine that the actual plausible fathers were drawn genetically at random from the wider population, and sampling at random from a sample drawn at random is mathematically equivalent to drawing at random from the original population (see [7]).

Although this line of reasoning is attractive on practical grounds, it does not withstand a test of whether the final computation gives a number that is sufficiently pertinent to the particular case at hand. It asks, in effect, that we substitute probabilities in place of knowledge, a practice that is sometimes acceptable and sometimes not. For instance, in one case [8], the court desired to determine the proportion of a state's automobiles that were covered by insurance. A statistician was commissioned to design and execute a study and to report on its conclusions. In this case, the sample drawn by the statistician was, in effect, substituted for

the population from which it was drawn in order to permit the court to draw its conclusions. This is a legitimate substitution of probabilities for knowledge because the statistical characteristics of the sampling plan were well understood, and even though the reported percentage might not have been precisely correct, there was very little chance that it was substantially wrong. One could question whether the reported figure of 97% under- or overestimated the true value by one or two percentage points, but no one could reasonably maintain that the true figure were 50%.

But the substitution of probability for knowledge in paternity testing does not have these features. When an *LR* value of 1,000 is reported, based on the assumption of sampling from a geographical population, but, in fact, there are only two other plausible fathers, one of whom would be excluded, then the value of *LR* should be equal to 2 (based on the inclusion-exclusion test). In this case, the difference between the conclusion based on probabilities for an unknown pool of plausible fathers is substantially different from that based on knowledge of the pool, and the probability-based evidence is clearly misleading. When the customary *LR* value is not correct, it can be very wrong, and it is the *divergence* between the two conclusions in this case that makes the probability-based evidence unacceptable. In effect, by using the standard *LR*, we are employing a procedure that sometimes points us in the right direction, and sometimes points in exactly the wrong direction, and we have no means of telling in each particular case whether we are being informed or misinformed by the "paternity probability." Although paternity testers are fond of pointing out that the *LR* should make these disastrous errors relatively infrequently, this is only a statement about the long-run performance of the *LR*, and provides no comfort in individual cases.

Serious although it is, this difficulty is eclipsed by a further and even more fundamental objection to the standard calculations of *LR* that has gone unnoticed by paternity experts. Let us imagine ourselves as a juror who finds some elements of credibility in the stories of both sides, so that perhaps $P(b = 1) = .6$ and $P(b = 0) = .4$. Then the *LR* is computed in the standard fashion from the genetic evidence, and found to be 500. The posterior odds favoring paternity are $500(.6/.4) = 750$, and so the "probability of paternity" is .99867. Perhaps we may be forgiven for wondering how our .4 belief in the father's story dwindled to .00133 just on the basis of the fact that he is not excluded from biological fatherhood. The answer is that when the *LR* was computed, the hypothesis $b = 0$ was framed in a way that discounts our belief in the father's story. It is easy to see that this is true. We only have to approach the court observer (who, we recall, knows all about the plausible fathers) who has computed an entirely different *LR* and a rather smaller posterior "probability of paternity." Now, as a juror, we are not in the position of the court observer, because we do not know the plausible fathers, and so we do not, in fact, have access to his computation. But, by the same token, we are not in the position of someone who has no knowledge whatsoever about the plausible fathers, and who might be driven to use the draw from the geographical population as a device for coming up with the denominator of *LR*. We are, in fact, intermediate between these two extreme positions. To the extent that we believe the father's story (.4), we also believe that the collection of other

plausible fathers contains at least one man who would not be excluded, namely, the actual father. The standard computation of the denominator makes this possibility very unlikely as a consequence of its *assumptions*, in effect ignoring our .4 belief that it is true.

The core of the fallacy is this. The standard method carries out its computation of the denominator of *LR* by choosing to interpret the probabilistic effect of $b = 0$ in a certain way, but anyone who had some prior belief in $b = 0$ would almost certainly choose to represent its probabilistic effect in another way, one more favorable to the alleged father. In effect, jurors who have some prior belief in the alleged father's story are being swindled out of that belief.

To be fair, we ought to point out that if a juror had virtually no prior belief in the alleged father's story, and the standard *LR* turned out to be very small, then he, too, might have been swindled.

This is a fundamental objection to the use of *LR*. Neither the Bayesian approach nor any other approach takes into consideration how the proper probabilistic consequences of $b = 0$ are to be determined.

Supporters of the standard *LR* method might well ask, if their method is wrong, what is right? The answer here is painfully clear. Without any additional information about the collection of plausible fathers, the juror is unable to compute the denominator of *LR* in such a way that it avoids distorting his perception of the evidence. He may feel that the standard computation gives a value that is too low (or too high), but he has very little basis on which to substitute a more accurate figure, because his belief about the plausible fathers is so vague. Consequently, the current methods must be abandoned until a statistical technology is developed for incorporating into the computations the judge's or juror's vague belief about the class of plausible fathers.

THE USE OF POST-INCLUSION GENETIC INFORMATION

Another class of fallacies involves the use of genetic information above and beyond what is used for the inclusion-exclusion test. Although these new fallacies are related to the null hypothesis fallacy, they have some special features that turn on the difference between a phenotype (directly observable inherited traits) and a genotype (often unobservable, microscopic chemical properties of cells). Only one such fallacy will be presented here.

The situation involves a genetic system with two codominant alleles, called *1* and *2*, and a relatively rare silent allele, called *s*. So, for instance, an individual with phenotype 1 could have genotype either (*1*,*1*) or (*1*,*s*)—that is, he or she had to have inherited the *1* gene from at least one parent, and the other gene could be anything but a *2*, for in that case, the phenotype would be 1-2. The silent allele is never expressed, so one cannot tell in this case whether or not it is present.

In the case to be considered, the mother, child, and alleged father all have phenotype 1. The collection of plausible fathers is known to consist of the alleged father and one other man, whose phenotype is 2.

In the hypothesis-testing framework, there are only two alternatives here, $b = 1$ or $b = 2$, and so we can continue to use the *LR*, in the form $P(T|b = 1)/$

$P(T|b = 2)$. Thus, large values of LR point in the direction of the alleged father.

If we take T to be the inclusion-exclusion test, then both the numerator and denominator of LR are 1, since the phenotypes of the mother and child do not exclude any man from paternity. Then $LR = 1$, and we see that no information about paternity is yielded by the inclusion-exclusion test.

The feeling among nearly all paternity testers is that we should now go further to speculate about the genotypes of the two men, based on our knowledge of their phenotypes. This is done by referring to table 1, which lists all the possible triples of father-mother-child genotypes for the cases consistent with the phenotypes observed. The probabilities of observing these triples can be computed from standard genetic principles, and are listed in the table, where p, q, and r stand for the known relative frequencies of the *1*, *2*, and *s* genes, respectively, in the population.

Since T now stands for the phenotype quadruple (mother, child, alleged father, other man), the numerator of LR is the probability of drawing the phenotypes we actually observed, assuming the alleged father is the true father, while the denominator is the same probability assuming the other man is. This gives

$$LR = \frac{p^2(p^2 + 4pr + 3r^2)(q^2 + 2qr)}{pqr(p + r)(p^2 + 2pr)} .$$

It is convenient to make the innocuous assumption that $p = q$, so we have the simple expression $LR = 3 + (p/r)$.

Let us now suppose that the silent allele is quite rare, so that r is very small. The paternity testers now compute that LR is very large, and so there is considerable evidence that the alleged father is guilty. Indeed, one paternity tester has suggested that the conclusion is obvious just from looking at table 1 and seeing that for the "other man" to be the father he must be the rare type (*2,s*) whereas the alleged father could be the common type (*1,1*). Thus, the LR is taken to be the proper way of quantifying this obviously informative evidence.

TABLE 1

GENOTYPE TRIPLES AND THEIR PROBABILITIES CONSISTENT WITH THE EVIDENCE PRESENTED, IN A HYPOTHETICAL *1-2-s* SYSTEM

Father	Mother	Child	Probability
1,1	*1,1*	*1,1*	p^4
1,1	*1,s*	*1,1*	p^3r
1,1	*1,s*	*1,s*	p^3r
1,s	*1,1*	*1,1*	p^3r
1,s	*1,1*	*1,s*	p^3r
1,s	*1,s*	*1,1*	p^2r^2
1,s	*1,s*	*1,s*	$2p^2r^2$
2,s	*1,1*	*1,s*	p^2qr
2,s	*1,s*	*1,s*	pqr^2

Let us try to see now what is obvious and of what we are being informed. The first possibility is that the other man is $(2,2)$. He is then excluded from being the father, and it is proved (logically) that the alleged father is guilty. The second possibility is that the other man is $(2,s)$, and then if we assume the alleged father is $(1,1)$, we can recompute $LR = 2(p + 2r)/(p + r)$, or if he is $(1,s)$, we can recompute $LR = 2(2p + 3r)/(p + r)$. Since r is very small, the first of these values will be nearly 2, the second nearly 4.

Thus, we see that in the three possible cases, one is definitive in convicting the alleged father, and the other two provide some modest evidence against him. What seems obvious here is that the value of $3 + (p/r)$, computed by the paternity tester, is in some sense an intermediate value between an enormous *LR* [if the other man is $(2,2)$] and two very modest *LR*s [if he is $(2,s)$]. It is like a weighted average with the enormous *LR* weighted more heavily than the modest ones.

Of what, then, does the paternity tester's *LR* inform us? It gives us an overall figure, highly incriminating to the alleged father, which obscures the fact that in one set of circumstances there is rather little evidence against him. It achieves this effect by substituting speculation about random draws of genotypes for our (admittedly noninformative) knowledge about phenotypes. It is therefore an example of those cases in which substitution of probabilities for knowledge is unacceptable, because of the divergence between the conclusions in particular cases.

It is also clear that the paternity tester's *LR* is capable of swindling us out of our prior beliefs in this case. Any prior belief we have that the other man is the true father entails a belief that he is $(2,s)$ rather than $(2,2)$, but by its *assumptions*, the *LR* computation makes the $(2,s)$ outcome unlikely.

This specific example was used here because it is on its way to becoming a touchstone for separating paternity-testing enthusiasts from those who have doubts about the method. Another example can be given in terms of the morphologically identical ABO system, in which the silent allele is common. For instance, using a table of *LR* for this system ([1], p. 79), we find that if the mother is B and the child O, then only AB men are excluded from paternity and the *LR*s for the other types are .63 for A, .71 for B, and 1.51 for O. The inclusion-exclusion test tells us only that the true father must carry the O allele. The post-inclusionary *LR* uses information about the "other" allele, the one not connected with paternity. Whether a man additionally carries A or B or another O is irrelevant to paternity. The type A father is, in a sense, carrying automatic protection against being found to be the father of any of his type O children. With enough systems under test, there could be men with rare-enough combinations that they would have almost zero paternity probability for all but a small fraction of their own children.

These are representative of a very large family of fallacious computations based on putatively precise genetic reasoning. In many cases, some of the speculative aspects could be removed from the calculations by obtaining more information on the individuals involved (by family studies, for example). In the silent gene example, knowledge of the genotype of the other man seems far more useful and informative than the paternity tester's *LR*.

CONCLUSION

The history of the use of probabilities in paternity testing is worthy of study. The methods were developed and codified primarily by researchers who did not pretend to be statistical experts, and as professional statisticians have moved into the field, they have seemed unwilling to raise fundamental questions about the assumptions underlying the earlier work. Perhaps this is because the whole problem appears so easy from the statistical standpoint that no rethinking seems to be called for.

The purpose of this paper has been to press the argument that considerable thought will be necessary before paternity probabilities become a useful tool for legal proceedings. Some courts may have gone too far already in accepting the expert testimony of paternity testers, and the new AABB guidelines unfortunately threaten to strengthen this trend. The immediate challenge is to develop methods of statistical reasoning that deal with fallacies like those presented here, before the current methods establish such a foothold in the legal arena that they become impossible to dislodge.

ACKNOWLEDGMENT

I have benefited in this work from stimulating conversations with my colleague, David Kaye.

REFERENCES

1. WALKER RH, ED: *Inclusion Probabilities in Parentage Testing*. Arlington, Va., American Association of Blood Banks, 1983
2. ESSEN-MÖLLER E: Die Beweiskraft der Ähnlichkeit im Vaterschaftsnachweis. Theoretische Grundlagen. *Anthropol Ges Wien Mitt* 68:2–53, 1938
3. ELLMAN IM, KAYE D: Probabilities and proof: Can HLA and blood group testing prove paternity? *NY Univ Law Rev* 54:1131–1162, 1979
4. SHAFER G: *A Mathematical Theory of Evidence*. Princeton, N.J., Princeton Univ. Press, 1976
5. HUMMEL K, CLAUSSEN M: Exclusion efficiency and biostatistical value of conventional blood group systems in European and non-European populations; suitability of Central European tables for non-German speaking populations, in *Biomathematical Evidence of Paternity,* edited by HUMMEL K, GERCHOW J, Berlin, Springer-Verlag, 1981, pp 97–108
6. SEARLE SR: *Linear Models*. New York, John Wiley, 1971, p 193 ff.
7. FINNEY DJ: Probabilities based on circumstantial evidence. *J Am Stat Assoc* 72:316–318, 1977
8. KATZ L: Presentation of a confidence interval estimate as evidence in a legal proceeding. *Am Stat* 29:138–142, 1975