

Probability and Paternity Testing

R. C. ELSTON¹

SUMMARY

A probability can be viewed as an estimate of a variable that is sometimes 1 and sometimes 0. To have validity, the probability must equal the expected value of that variable. To have utility, the average squared deviation of the probability from the value of that variable should be small. It is shown that probabilities of paternity calculated by the use of Bayes' theorem under appropriate assumptions are valid, but they can vary in utility. In particular, a recently proposed probability of paternity has less utility than the usual one based on the paternity index. Using an arbitrary prior probability in the calculation cannot lead to a valid probability unless, by chance, the chosen prior probability happens to be appropriate. Appropriate assumptions regarding both the prior probability and gene or genotypic frequencies can be estimated from prior experience.

INTRODUCTION

For the statement "there is a 10% probability of rain tomorrow" to have any practical meaning, two conditions must hold. First, when this particular statement is made on very many different occasions, on one-tenth of the succeeding days it must rain. This condition gives the probability *validity*. Second, statements of this nature must vary, with respect to the probability quoted, in such a way that it is more likely to rain on a day succeeding a day on which a high probability is quoted than on a day succeeding a day on which a low probability is quoted. Fulfillment of this second condition gives the probability *utility*: if the average probability of rain on any one day is 10%, there is little utility in always saying, regardless of cloud and wind conditions, "there is a 10% probability of rain tomorrow." A probability that is closer to 1 or 0, depending on whether it

Received December 4, 1985; revised March 3, 1986.

¹ Department of Biometry and Genetics, Louisiana State University Medical Center, New Orleans, LA 70112.

© 1986 by the American Society of Human Genetics. All rights reserved. 0002-9297/86/3901-0010\$02.00

does or does not principles, which genetic principle: ing recommendat which the probat ers in this area ev or "likelihood" f ment for which th illustrate these pr ties of paternity t conditions both c than the other. T means of Bayes' recently been crit in general, a valid probability of pat To avoid too n reference to a se mother, child, an the same, howev

It is helpful to quantity. In the alleged father is ti he truly is or is n some blood tests problem, thus, is unknown true par parameter is not 1 case for an overall on one of two diff be valid if its expe of alleged fathers analogy between a the values of a ge true parameter va the true probabilit a probability and t by the mean squa from the true val utility of our prob

We shall therefo to (1) their expecte of alleged fathers. are true fathers an

does or does not rain on the next day, clearly has more utility. These simple principles, which are surely self-evident, often appear to be forgotten when genetic principles are used to arrive at "probabilities." Most recently, conflicting recommendations have been published regarding the appropriate manner in which the probability of paternity should be calculated [1-3]. Some practitioners in this area even appear to believe that substituting the words "plausibility" or "likelihood" for "probability" somehow gives statistical meaning to a statement for which the first condition does not hold. The purpose of this paper is to illustrate these principles by comparing and contrasting two different probabilities of paternity that have been proposed. We shall see that under well-defined conditions both of these probabilities are valid, but that one has more utility than the other. The usual probability calculated from the paternity index by means of Bayes' theorem will be seen to have more utility, even though it has recently been criticized as being illogical and irrelevant [1]. It will be seen that, in general, a valid probability of paternity must include information on the prior probability of paternity.

To avoid too many statistical details, these principles will be illustrated by reference to a set of independent two-allele polymorphic loci for which the mother, child, and alleged father have been typed; the major principles remain the same, however, whatever the genetic systems used.

PROBABILITY CONSIDERED AS AN ESTIMATE

It is helpful to consider probability as an estimate of a true, but unknown, quantity. In the case of disputed paternity, this "true probability" that an alleged father is the father of a child is either 1 or 0, corresponding to whether he truly is or is not the father. The probability that we quote on the basis of some blood tests can be considered an estimate of this true quantity. The problem, thus, is similar to the general statistical problem of estimating an unknown true parameter, but with a slight difference: in this case, the unknown parameter is not the same for all members of the population (as would be the case for an overall mean, for example), but rather it is a variable that can take on one of two different values. We can, nevertheless, consider our estimate to be valid if its expected value, over the whole population, equals the proportion of alleged fathers who are actually true fathers. In this sense, we draw an analogy between a *valid* probability and an *unbiased* estimate. Similarly, just as the values of a good estimate will, on repeated sampling, vary little from the true parameter value, so will a good "probability of paternity" vary little from the true probabilities 1 or 0. We can thus draw an analogy between the utility of a probability and the efficiency of an estimate, which is conveniently measured by the mean squared error—the expected squared deviation of the estimate from the true value. The smaller the mean squared error, the greater is the utility of our probability estimate.

We shall therefore compare different probabilities of paternity with respect to (1) their expected values and (2) their mean squared errors, over a population of alleged fathers. We shall assume that a proportion π of these alleged fathers are true fathers and that the remaining $1 - \pi$ are simply randomly chosen men

from the same population, which will be assumed to be randomly mating with regard to the polymorphic markers used. In other words, we shall assume the existence of a population that is randomly mating with respect to the polymorphisms and that within this population there is a population of alleged fathers made up as follows: a proportion π belong to random father-mother-child trios from this population and a proportion $1 - \pi$ are random men to whom mother-child pairs from the same population have been randomly appended.

TWO PROPOSED PROBABILITIES OF PATERNITY AND THEIR VALIDITY

In what follows, the subscript i will denote a constellation of phenotypes for the trio comprising the alleged father, the mother, and the child. Let X_i denote the conditional probability of such a constellation given that the alleged father is the true father, and Y_i , the conditional probability given that a random man from the same population is the true father. For the case of two codominant autosomal alleles A and a , with respective gene frequencies p and $q (= 1 - p)$, table 1 presents the values of X_i and Y_i for each of the 27 phenotype (also, in this case, genotype) trios. This table assumes Hardy-Weinberg equilibrium and absence of mutation and selection and that there is no doubt concerning the mother-child relationship. The values X_i and Y_i are likelihoods [4], and from them we can derive the likelihood ratios $L_i = X_i/Y_i$ which are also shown in the table (six of the possible genotype trios cannot occur, and for them L_i is undefined). These results are well known and agree, for example, with table 1 of Chakraborty and Ryman [5] and table 1 of Majumder and Nei [6].

Inspection of table 1 reveals that in this particular situation L_i can take on only one of six values: $1/p$, $1/2p$, 1 , $1/2q$, $1/q$, or 0 . Let us pool all phenotype trios that have the same value of L_i , and use the word constellation to mean such a set of trios. Then the information in table 1 can be compressed as shown in table 2. This table allows for every possibility. With probability π , the alleged father is the true father and the conditional probabilities X_i sum to unity. Similarly, with probability $1 - \pi$, a random man is the true father and the conditional probabilities Y_i sum to unity. For a particular constellation i , the probability of paternity is obtained, using Bayes' theorem, as

$$P_i = \frac{\pi X_i}{\pi X_i + (1 - \pi) Y_i} = \frac{\pi L_i}{\pi L_i + 1 - \pi} \quad (1)$$

This is the usual probability of paternity, and L_i is the usual paternity index.

Over the whole population, the expected value of P_i is

$$\begin{aligned} & \pi \sum_i X_i P_i + (1 - \pi) \sum_i Y_i P_i \\ &= \pi \sum_i X_i \frac{\pi L_i}{\pi L_i + 1 - \pi} + (1 - \pi) \sum_i Y_i \frac{\pi L_i}{\pi L_i + 1 - \pi} \end{aligned}$$

TABLE 1

CONDITIONAL PROBABILITIES OF TRIO PHENOTYPES, IN THE CASE OF A CODOMINANT AUTOSOMAL LOCUS WITH ALLELES *A* AND *a* AND GENE FREQUENCIES *p* AND *q*, GIVEN THE ALLEGED FATHER (X_i) OR A RANDOM MAN (Y_i) IS THE TRUE FATHER AND THE LIKELIHOOD RATIOS $L_i = X_i/Y_i$

MOTHER	ALLEGED FATHER	CHILD								
		AA			Aa			aa		
		X_i	Y_i	L_i	X_i	Y_i	L_i	X_i	Y_i	L_i
AA	AA	p^4	p^5	$\frac{1}{p}$	0	p^4q	0	0	0	...
	Aa	p^3q	$2p^4q$	$\frac{1}{2p}$	p^3q	$2p^3q^2$	$\frac{1}{2q}$	0	0	...
	aa	0	p^3q^2	0	p^2q^2	p^2q^3	$\frac{1}{4}$	0	0	...
	Subtotal	p^3	p^3	...	p^2q	p^2q	...	0	0	...
Aa	AA	p^3q	p^4q	$\frac{1}{p}$	p^3q	p^3q	1	0	p^3q^2	0
	Aa	p^2q^2	$2p^3q^2$	$\frac{1}{2p}$	$2p^2q^2$	$2p^2q^2$	1	p^2q^2	$2p^2q^3$	$\frac{1}{2q}$
	aa	0	p^2q^3	0	pq^3	pq^3	1	pq^3	pq^4	$\frac{1}{q}$
	Subtotal	p^2q	p^2q	...	pq	pq	...	pq^2	pq^2	
aa	AA	0	0	...	p^2q^2	p^3q^2	$\frac{1}{p}$	0	p^2q^3	0
	Aa	0	0	...	pq^3	$2p^2q^3$	$\frac{1}{2p}$	pq^3	$2pq^4$	$\frac{1}{2q}$
	aa	0	0	...	0	pq^4	0	q^4	q^5	$\frac{1}{4}$
	Subtotal	0	0	...	pq^2	pq^2	...	q^3	q^3	
Total	p^2	p^2		$2pq$	$2pq$...	q^2	q^2		

Substituting X_i for Y_iL_i , this becomes

$$\pi \sum_i \frac{\pi X_i L_i + (1 - \pi) X_i}{\pi L_i + 1 - \pi} = \pi \sum_i X_i = \pi$$

Two things should be noted about this result. First, it proves that the usual probability of paternity is valid. Second, the proof depends only on the conditions $\sum_i X_i = 1$ and $X_i = Y_i L_i$ for all *i*. Any other definitions of X_i , Y_i , and L_i , that

TABLE 2

CONDITIONAL PROBABILITIES X_i AND Y_i AND LIKELIHOOD RATIOS L_i , OBTAINED FROM TABLE 1 BY POOLING THE TRIO PHENOTYPES INTO SIX BASIC CONSTELLATIONS

Constellation, i	X_i	Y_i	L_i
1	$p^2(p + q^2)$	$p^3(p + q^2)$	$\frac{1}{p}$
2	$pq(p + q^2)$	$2p^2q(p + q^2)$	$\frac{1}{2p}$
3	pq	pq	1
4	$pq(p^2 + q)$	$2pq^2(p^2 + q)$	$\frac{1}{2q}$
5	$q^2(p^2 + q)$	$q^3(p^2 + q)$	$\frac{1}{4}$
6	0	$pq(1 - pq)$	0
Total	1	1	...

satisfy these conditions will also lead to a valid probability of paternity. In particular, L_i and Chakravarti's [1] proposal is equivalent to pooling the first five constellations shown in table 2, so that we take account only of whether an exclusion has occurred (paternity index = 0) or not (paternity index > 0). Doing this, we arrive at the two values of L_i shown in table 3. These are still likelihood ratios (but with respect to differently defined events), and in table 3, we still have $\sum_i X_i$ and $X_i = Y_i L_i$ for both $i = 1$ and $i = 2$. Thus, calculating P , from these values of L_i results in a probability of paternity that is also valid. From now on, paternity probabilities calculated using the paternity index will be denoted P_{1i} and those calculated using the likelihood ratio based on exclusion vs. nonexclusion will be denoted P_{2i} .

All these conditional probabilities and likelihood ratios can be easily generalized to allow for multiple allelism, dominance, or multiple systems. If the allele

TABLE 3

CONDITIONAL PROBABILITIES X_i AND Y_i AND LIKELIHOOD RATIOS L_i , OBTAINED FROM TABLE 2 BY POOLING INTO THE TWO CONSTELLATIONS DETERMINED BY PRESENCE OR ABSENCE OF AN EXCLUSION

Constellation, i	X_i	Y_i	L_i
1	1	$1 - pq(1 - pq)$	$\frac{1}{1 - pq(1 - pq)}$
2	0	$pq(1 - pq)$	$\frac{1}{1 - pq(1 - pq)}$
Total	1	1	...

TABLE 4

CONDITIONAL PROBABILITIES X_i AND Y_i AND LIKELIHOOD RATIOS L_i , OBTAINED FROM TABLE 1 BY POOLING **AA** AND **Aa** INDIVIDUALS AND THEN POOLING THE TRIO PHENOTYPES INTO SIX BASIC CONSTELLATIONS

Constellation, i	X_i	Y_i	L_i
1	$p^2(1 + 2q)$	$p^2(1 + q)(1 + pq)$	$\frac{1 + 2q}{(1 + q)(1 + pq)}$
2	pq^2	$pq^2(1 + pq)$	$\frac{1}{1 + pq}$
3	pq^2	$p^2q^2(1 + q)$	$\frac{1}{p(1 + q)}$
4	pq^2	$pq^2(1 + q)$	$\frac{1}{1 + q}$
5	q^3	q^4	$\frac{1}{q}$
6	0	pq^4	0
Total	1	1	...

A is dominant to the allele a, for example, the entries in table 2 become as shown in table 4 and those in table 3 become as shown in table 5. If we have multiple independent systems, then each X_i , Y_i , and L_i becomes the product of the corresponding quantities for the individual systems. If the systems are not independent, it is still possible to define the conditional probabilities X_i and Y_i and analogous ratios L_i ; in one case, all the phenotypic information available is used, and in the other case, use is made only of the fact of whether or not there is an exclusion. In every case, we have $\sum_i X_i = 1$ and $X_i = Y_i L_i$ for all i , and so the resulting probabilities of paternity are valid.

TABLE 5

CONDITIONAL PROBABILITIES X_i AND Y_i AND LIKELIHOOD RATIOS L_i , OBTAINED FROM TABLE 4 BY POOLING INTO THE TWO CONSTELLATIONS DETERMINED BY PRESENCE OR ABSENCE OF AN EXCLUSION

Constellation, i	X_i	Y_i	L_i
1	1	$1 - pq^4$	$\frac{1}{1 - pq^4}$
2	0	pq^4	0
Total	1	1	...

COMPARISON OF UTILITIES

The mean squared error of any set of paternity probabilities P_i can be defined as

$$\sum_i [\pi X_i (P_i - 1)^2 + (1 - \pi) Y_i P_i^2] \quad (2)$$

This definition is based on the fact that over the whole population the true value of P_i is 1, and so the squared error is $(P_i - 1)^2$, with probability π ; and the true value of P_i is 0, and the corresponding squared error $(P_i - 0)^2$, with probability $1 - \pi$. The X_i and Y_i are the appropriate conditional probability distributions for obtaining the mean values in the two cases. Substituting $P_i = P_{1i}$ or $P_i = P_{2i}$ into expression (2), the utilities of the two sets of probabilities can be contrasted—the larger the mean squared error, the less utility the set of probabilities has. Since the P_{2i} are based on less information, we might expect them to have a larger mean squared error and so have less utility. We shall now show that this is so.

First note that when there is an exclusion the contribution to the mean squared error is zero, since in this case both the estimated and the true value of P_i is 0. Thus, to obtain the mean squared error of a set of P_i , we need only sum over the terms in equation (2) that correspond to a nonexclusion. For example, we would sum over the first five constellations in the situations represented in tables 2 and 4, in which the L_i are relevant for calculating probabilities P_{1i} , but there is only one term in the sum for situations represented by tables 3 and 5, in which the L_i are relevant for calculating probabilities P_{2i} . However, this one term can always be written as a sum analogous to the one used for the corresponding set of probabilities P_{1i} , the only difference being that the P_{2i} in this sum are all equal. If we were to calculate the mean squared error by summing over the first five constellations in table 2, for example, using the entries in that table for X_i and Y_i , but substituting $L_i = 1/[1 - pq(1 - pq)]$ for all i , the result would be the same as if we had calculated the mean squared error from the entries in table 3. Thus, in each case, we can express the mean squared error as the sum (2), provided we define P_i by equation (1) for a set of probabilities P_{1i} , but as follows for the corresponding set of probabilities P_{2i} :

$$P_i = \frac{\pi \sum_i X_i}{\pi \sum_i X_i + (1 - \pi) \sum_i Y_i} = \frac{\pi}{\pi + (1 - \pi) \sum_i Y_i / \sum_i X_i}$$

where the summation is over all the constellations corresponding to nonexclusion. The two mean squared errors we wish to compare are thus equal to

$$\sum_i \left\{ \pi X_i \left[\frac{(1 - \pi) Y_i}{\pi X_i + (1 - \pi) Y_i} \right]^2 + (1 - \pi) Y_i \left[\frac{\pi X_i}{\pi X_i + (1 - \pi) Y_i} \right]^2 \right\} \quad (3)$$

and

$$\sum_i \left\{ \pi X_i \left[\frac{(1 - \pi)K}{\pi + (1 - \pi)K} \right]^2 + (1 - \pi) Y_i \left[\frac{\pi}{\pi + (1 - \pi)K} \right]^2 \right\}, \quad (4)$$

respectively, where $K = \sum_i Y_i$ and all summations are over the terms corresponding to nonexclusion. Simplifying, we find that the i th terms in expressions (3) and (4) are, respectively, equal to

$$\frac{\pi(1 - \pi)X_i Y_i}{\pi X_i + (1 - \pi)Y_i} \text{ and } \frac{\pi(1 - \pi) [(1 - \pi)X_i K^2 + \pi Y_i]}{[\pi + (1 - \pi)K]^2}$$

Thus, to show that expression (3) is smaller than expression (4) it is sufficient (when $0 < \pi < 1$) to prove the inequality

$$\frac{X_i Y_i}{\pi X_i + (1 - \pi)Y_i} < \frac{(1 - \pi)X_i K^2 + \pi Y_i}{[\pi + (1 - \pi)K]^2}. \quad (5)$$

Multiplying both sides of inequality (5) by $[\pi X_i + (1 - \pi)Y_i][\pi + (1 - \pi)K]^2$ and simplifying, we find that inequality (5) is equivalent to $0 < \pi(1 - \pi)(KX_i - Y_i)^2$, which is clearly true for all situations of interest.

In order to quantify the difference between the two types of paternity probability, the mean squared errors were calculated for probabilities based on one through five independent autosomal codominant loci and also for probabilities based on one through five independent dominant loci. In each case, two alleles in Hardy-Weinberg equilibrium were assumed and equal gene frequencies at all loci. Eleven prior probabilities and 11 gene frequencies were considered: π , $P = .01, .1, .2, .3, \dots, .9, \text{ and } .99$. Thus, in all $2(P_{1i} \text{ or } P_{2i}) \times 2$ (codominant or dominant loci) $\times 5$ (number of loci) $\times 11$ (number of values of π) $\times 11$ (number of values of P) mean squared errors were calculated. A summary of the results follows.

As expected, the error for P_{2i} was always found to be larger than the error for P_{1i} . In the case of codominant loci, the error was always largest for gene frequencies equal to .01 and .99 and smallest for a gene frequency of .5. For one locus, the errors were found to be largest when π is .5. With increasing number of loci, the larger errors tended to occur, when $P \neq .01, .99$, for π equal to .4. Expressed as a percentage of the square root of the mean squared error of P_{1i} , the largest error in P_{2i} occurred for π about .4 and P about .2 or .8, being just over 2% in the case of one locus and rising to just over 10% in the case of five loci. Thus, the largest percentage increase in error in P_{2i} tended to occur near those values of π and P that also gave rise to the largest error in P_{1i} . In the case of dominant loci, the error was always largest when the frequency of the dominant gene was large, but it was almost equally large when the gene frequency was small. The error was usually at a minimum for P about .2. Unlike the case of codominant loci, the errors were always found to be largest for π equal to .5. Expressed as a percentage of the root mean squared error of P_{1i} , the

largest error in P_{2i} occurred for π about .4 or .5 and P about 8.2. As before, the largest percentage error increased from a little over 2% for one locus to over 10% for five loci and tended to occur not too far from those values of π and P that maximized the error in P_{1i} . As might be expected, in the case of dominance, the error was always larger than in the case of codominance—although trivially so for extreme values of π and P . Finally, it was noted that the largest percentage error in P_{2i} was always slightly larger in the case of dominance than in the case of codominance.

DISCUSSION

Whereas under the assumptions made both P_{1i} and P_{2i} are valid probabilities of paternity, P_{1i} has more utility and is therefore preferable. The error of P_{2i} relative to that of P_{1i} increases (although, of course, both errors decrease) as the number of loci used for paternity testing increases. It is possible that P_{2i} is more robust than P_{1i} , that is, it may be less affected by deviations from the underlying assumptions. Provided appropriate assumptions are made in each case, however, a probability based on all available information will automatically have more utility than one that uses only the fact that there has been nonexclusion at each of a set of loci. This general principle does not depend upon the assumption of Hardy-Weinberg equilibrium. Nor does it depend upon the assumption that the only alternative to the true father being the alleged father is that he is a man chosen at random from the same genetic population as the mother. The usual paternity index can be extended in a straightforward manner to allow for many alternative situations. But in each case the underlying assumptions must be correct if the calculated probability is to be valid.

From the proof that both P_{1i} and P_{2i} are valid, it is obvious that a probability that does not incorporate the appropriate prior value of π cannot possibly (except by a chance fluke) be valid. Walker [3] illogically defends the arbitrary choice $\pi = .5$ as being a "neutral prior probability." Bayes [7] himself realized the logical difficulty in assuming that, for lack of knowledge, all possible events have equal prior probabilities; for that very reason, he never tried to have his famous essay published. If the true father is either the alleged father or a random man and we do not know which, the argument runs, the prior probabilities of these two events must be equal. Analogously, if I throw a die and know it will come up either a six or not a six, should I, in the absence of other knowledge, assume that the probabilities of these two events must be equal? Just as knowledge about dice in general should influence the priors chosen, so should knowledge about paternity tests in general. If in a particular case an appropriate prior probability could be obtained on the basis of other information, such as access to the mother, it should be used. But since it is usually impossible to quantify such probabilities, it makes sense to base the prior probability on experience of marker testing alone. Several statistical methods have been proposed for doing this [8-11], all based on realistic theoretical probability and genetic models. (Although it does not seem to have been pointed out, standard statistical methods can also be used to test the goodness-of-fit of past data to the actual models used for this purpose.)

An appropriate probability π is not the only essential for P, to be valid. The gene (or genotypic) frequencies that are used must also be correct and should be similarly estimated on the basis of past experience. It is not necessary to assume that the true father comes from the same genetic population as the mother, but it is necessary to know the relevant frequencies for the genetic population from which he does come. It is not necessary to assume that the true father, if he is not the alleged father, is "randomly chosen" from this genetic population. It is sufficient to assume that his relationship to both the mother and the alleged father is accurately known. If both these relationships are "unrelated," and there is no association between genotype and being the father of a child whose paternity is contested, the assumption that the father is randomly chosen is appropriate. The methods of calculating P_i in the more general situations, together with an investigation of the effects of misspecifying the gene frequencies or the relationships of the true father, are given by Reading [10].

Finally, it is interesting to consider one of the reasons why Li and Chakravarti [1] were misled into thinking that P_{1i} is not a probability of paternity. The paternity index (L, in tables 1, 2, and 4) can sometimes take on a value strictly between 0 and 1. If, for example, there is codominance, the mother is *aa*, the alleged father is *Aa*, and the child is *aa*, L, is $1/2q$ (table 1); this lies between 0 and 1 for q greater than .5. In such a situation, the alleged father is not excluded, and yet he is less likely to be the true father than a random man. Because of this, it is possible for P_{1i} to decrease with the additional information that the alleged father is not excluded on further genetic systems. Li and Chakravarti consider such a conclusion illogical: "By this method, we reach the illogical conclusion that the alleged father who failed to be excluded by the first three genetic systems is more likely to be the true father than one who failed to be excluded by four or five genetic systems!" Such a finding, when it occurs, is only illogical if we do not take account of the actual phenotypes observed, as opposed to basing conclusions solely on whether or not there has been an exclusion for each system. It is intuitively clear that presence of the same rare allele in both the alleged father and child increases the likelihood of paternity. By the same token, absence in the child of a rare allele that the father possesses decreases the likelihood of paternity. When the alleged father is *Aa* and the child is *aa*, the fact that *A* is rare ($q > .5$) makes the alleged father less likely to be the true father than someone else in the population. By taking account of this information, we arrive at a probability of paternity that has greater utility.

REFERENCES

1. LI CC, CHAKRAVARTI A: Basic fallacies in the formulation of the paternity index. *Am J Hum Genet* 37:809-818, 1985
2. AICKIN M: Some fallacies in the computation of paternity probabilities. *Am J Hum Genet* 36:904-915, 1984
3. WALKER RH: Guidelines for reporting estimates of probability of paternity. *Am J Hum Genet* 37:819-823, 1985

4. BAUR MP, GÜRTLER H, HENNINGSEN K, ET AL.: No fallacies in the formulation of the paternity index. Submitted for publication
5. CHAKRABORTY R, RYMAN N: Use of odds of paternity computations in determining the reliability of single exclusions in paternity testing. *Hum Hered* 31:363-369, 1981
6. MAJUMDER PP, NEI M: A note on positive identification of paternity by using genetic markers. *Hum Hered* 33:29-35, 1983
7. BAYES T: An essay towards solving a problem in the doctrine of chances. Reprinted in *Biometrika* 45:296-315, 1958
8. POTTHOFF RF, WHITTINGHILL M: Maximum-likelihood estimation of the proportion of nonpaternity. *Anz J Hum Genet* 17:480-494, 1965
9. BAUR MP, RITTNER C, WEHNER HD: The prior probability parameter in paternity testing. Its relevance and estimation by maximum likelihood, in *Lectures of the Ninth International Congress of the Society for Forensic Hemogenetics*, Bern, 1981, pp 389-392
10. READING PL: Problems in the attribution of paternity. Ph.D. dissertation, University of North Carolina at Chapel Hill, 1981
11. CHAKRAVARTI A, LI CC: Estimating the prior probability of paternity from the results of exclusion tests. *Forensic Sci Int* 24:143-147, 1984

Erratum

In the paper "Analysis of Three Restriction Fragment Length Polymorphisms in the Human Type II Procollagen Gene," by C. E. L. Eng and C. M. Strom (*Am J Hum Genet* 37:719-732, 1985), figure 3A was improperly assembled. Lanes 7 and 8 actually represent a shorter exposure of lanes 4 and 5, respectively, which were inadvertently added to the figure instead of the actual lanes 7 and 8. The correct figure, representing the original, uncut autoradiogram appears below. In addition, the figure legend neglected to mention that the individuals represented in figures 3A and B were patients with achondroplasia. The corrected figure legend appears below. The authors sincerely regret this error.

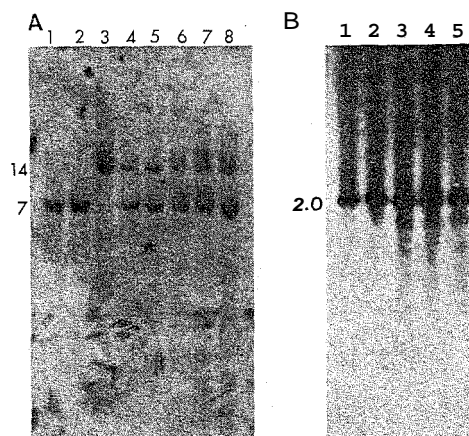


FIG. 3.—Autoradiograms of Southern filters of genomic DNA digested with *Hind*III or *Eco*RI/*Hind*III and hybridized to pgHCol(II)A. A, *Hind*III; B, sequential *Eco*RI/*Hind*III sequential digestions. Lanes 1, 2, 3, 4, and 5 in panel B correspond to lanes 1, 4, 5, 7, and 3, respectively, in panel A.