Am J Hum Genet 33:617–628, 1981

Assigning a Probability for Paternity in Apparent Cases of Mutation

EDWARD D. ROTHMAN,[1] JAMES V. NEEL,[2] AND FRED M. HOPPE[1]

SUMMARY

Any direct estimator of mutation in a human population is subject to error due to nonpaternity. This paper deals with the quantification of this error by producing, under certain assumptions, the probability *for* paternity. In addition, a new direct estimator of the mutation rate is introduced.

INTRODUCTION

In studies of human mutation rates, a standard complication has been the possibility that some among the apparent mutations result from the discrepancy between legal and biological parentage customarily referred to as "nonpaternity." This possibility has been largely ignored in studies in which the apparent mutant is an individual with one of the dominantly inherited diseases on the dual grounds that affected individuals who might be a source of the allele in question are so rare (population frequency usually $< 10^{-4}$) and/or so stigmatized that, compared to mutation, this possibility was relatively unlikely. When, however, the mutant phenotype is a variant protein, the possibility assumes a larger dimension. Now the frequency of the phenotype on which studies of mutation will be based is usually approximately 20 per 10,000 persons, and there are usually no external stigmata.

Here, we will develop an appropriate statistical framework for treating this complication when the study of mutation is based on protein variants. There are a number of protocols that such a study can pursue. That which we are currently employing—and which undoubtedly is one of the most representative—is as follows [1]: Subjects are screened for rare variants of a defined series of proteins. In the event of a variant, the parents are also examined for presence of this variant. Periodically, as the study progresses, a subset of the parents and children are subjected to the

Received April 4, 1980; revised November 10, 1980.
This study was supported by contract EY-76-C-02-2828 from the Department of Energy.
[1] Departments of Human Genetics and Statistics, University of Michigan, Ann Arbor, MI 48109.
[2] Department of Human Genetics, University of Michigan.

genetic typings employed in questions of parentage to establish a baseline for "nonpaternity" in the series. Any child found to exhibit an apparent mutation and his or her parents will also be subjected to the aforementioned typings. Such an approach enables us, first, to develop an estimate of the mutation rate and, second, to rank apparent mutations on the basis of the odds that the legal father is the biological father.

THE GIVENS AND THE ASSUMPTIONS IN THIS TREATMENT

We now state in precise form exactly what are the "givens" and the "assumptions" in the treatment we will develop.

The Givens

(1) The prior probability of nonpaternity will have been established through the appropriate studies. This is an important distinction from the situation obtaining in medico-legal questions, where the prior probability is usually unknown. (2) The apparent mutation in question will be characterized by the gain of a rare attribute, rather than in the loss of an attribute that might under the usual circumstance be expected to be present in a child. Thus the study will not be obfuscated by the occurrence of inherited "null" variants, as illustrated by the situation in which a father who is phenotypically haptoglobin type 2 (but genetically 2/0) could, when married to a type 1 woman, legitimately father a type 1 child. (3) Paternity studies will be performed on every apparent mutation in the series, with a battery of tests whose combined probability of detecting nonpaternity can be calculated. (4) The frequency in the population under study of the types of "rare" variants being employed as possible indicators of mutation is known, both from the study itself and the literature. For these purposes, we will define a rare variant as one with an allele frequency not greater than .01. In fact, for most of the genetic systems to be employed in studies of this type on civilized populations, the combined gene frequency for the various types of rare variants that are detected approximates 10^{-3} (reviewed in [2]).

The Assumptions

(1) Nonpaternity, when it occurs, is at random with respect to those traits on which the detection of nonpaternity is based. Let N denote the size of the population of potential fathers and let X_i, $i = 1, 2, \ldots N$, represent a complete listing of the multilocus genotype for which observations are available. So $X_1, X_2, \ldots X_N$ then represents the genetic structure of this male population. Further, let F, C, M, and T represent the genotypes, respectively, of the legal father, the child, the mother, and the biological father. We have assumed that a male affine or consanguine of the mother of the child is no more likely to be the true father than any other male from the population of potential fathers. In our notation, we express the probability that F is the true father as

$$P(F|X_1, \ldots, X_N) = \begin{cases} \frac{1}{N} & F = X_i,\ 1 \leq i \leq N \\ 0 & \text{else} \end{cases}.$$

(2) Within the population of potential, nonlegal fathers of the child, there is no bias increasing (or decreasing) the probability of paternity because of ethnic extraction, religion, etc. While this assumption is not strictly correct, it is necessary to the treatment. In our notation,

$$P(T|X_1, \ldots, X_N, F) = \begin{cases} \alpha & \text{if } F = T \text{ and } T = X_i \text{ for some } i \\ \dfrac{1-\alpha}{N-1} & \text{if } F = T \text{ and } T = X_i \text{ for some } i \\ 0 & \text{else} \end{cases},$$

where α is the prior probability of paternity. (3) The results of the genetic determinations are accurate, or, if there is error, its magnitude can be specified. (4) The identity of the mother is certain. (5) The genetic markers involved in the detection of nonpaternity occur in Hardy-Weinberg equilibrium and segregate independently of one another. As the number of markers that can be brought to bear on the question increases (see DISCUSSION), this assumption will certainly be violated, but for the present it is reasonable.

So, $P(C|X_1, \ldots, X_N, F, T)$ is determined by independent Mendelian segregation at each locus. This conditional probability thus depends only on T and the genotype of the mother M, which is, of course, implicit in the above probabilities. It will turn out that the precise value of N is relatively unimportant since it enters only through a multiplicative factor $[(N-1)/N]$, and we may therefore consider it to be fixed at some large unknown level.

STATISTICAL METHODOLOGY

Two procedures are involved in the extraction of information concerning mutation from the study. Step 1 simply estimates the mutation rate, whereas step 2 generates a probability statement for each presumptive mutant, from which one can rank apparent mutants as to the probability they are truly mutants.

Step 1

This method, originally suggested by Neel [3], will be extended here by a discussion of the statistical errors inherent in the methodology. Let I = frequency of nonpaternity, W = average frequency at the loci under consideration of alleles responsible for rare private variants, and D = probability of detecting nonpaternity with the available laboratory tests. Then the frequency with which undetected nonpaternity results in an apparent mutation is

$$IW(1-D) . \tag{1}$$

The corrected mutation rate (μ_{EST}) is derived from the apparent mutation rate (μ_{OBS}) simply by subtracting this term, that is,

$$\mu_{EST} = \mu_{OBS} - IW(1-D) . \tag{2}$$

D, computed from the allele frequencies of the genetic polymorphisms used for paternity exclusion, is customarily treated as a fixed probability. W, in any proper

study, will have been established on the basis of at least 2×10^5 determinations, and can also be treated as a fixed probability with small error. I, however, will be based on a relatively small sample, of several hundred determinations, and the error term must be taken into consideration.

An expression for the standard error of μ_{EST} can be obtained from the formula for the variance of the sum of two random variables. This yields

$$SD(\mu_{EST}) = \sqrt{\text{var}(\mu_{OBS}) + [W(1 - D)]^2 V(I) - 2X(1 - D)\text{cov}(\mu_{OBS}, I)} \quad (3)$$

To obtain an upper bound for the standard error (SE), we use EST of $\text{var}(\mu_{OBS}) = [\mu_{OBS}(1 - \mu_{OBS})]/N_{OBS}$ and EST of $\text{var}(I) = [I(1 - I)]/N_I$, and take the correlation to be zero. Here N_{OBS} and N_I refer to the number of observations available for the calculation of μ_{OBS} and I, respectively.

Let us assume that μ_{OBS} is 2×10^{-5} on the basis of 5×10^5 locus tests, that I based on 10 systems studied in 500 randomly selected trios is 2×10^{-2}, that on the basis of the study from which mutation rates are being estimated as well as an extensive literature, $W = 1 \times 10^{-3}$, and that the battery of tests available together will detect .8 of all instances of nonpaternity in the population. Then our estimate of mutation rate is $\mu_{EST} = 1.6 \times 10^{-5}$, and the upper bound of the SE of the estimate is found to be $.64 \times 10^{-5}$. Since the random variable μ_{OBS} is likely to be highly correlated with the random variable I, this SE may be quite conservative.

As stated, the mutation rate is the average across all the loci under investigation. Eventually, data may accumulate to the point where specific locus rates can be obtained, although, with average rates as low as they appear to be [2], this situation will not be obtained for some years. When, however, such data are available, W now applies to the frequency of variants of specific proteins. Observed variant frequencies for specific proteins (excluding polymorphisms) thus far are usually in the range 1 to 50×10^{-4} (e.g., [2, 4–6]). It is apparent from equation (1) that the probability that an apparent mutation is not due to nonpaternity is directly proportional to the frequency of variants at that locus. At first thought, this would suggest that a study could to some extent avoid the complication of nonpaternity by concentrating on proteins for which variants are rare, but this practice would probably bias the study toward the choice of loci where mutation is less common than average.

Step 2

This method assigns a probability that the legal father is the biological father for each case of suspected mutation, based on extensive genetic typings of child, mother, and putative (nonexcluded) father. The rationale for this approach is made clear by a simulated example in which a child with a rare protein variant not present in either (nonexcluded) parent also possesses two other rare variants that are present in the father but not in the mother. This is too much for coincidence—we are intuitively persuaded of the validity of the legal relationship (and of the occurrence of the mutation).

The approach we have developed quantifies this intuition by obtaining a value for P(legal father is the biological father|data). It is more convenient to consider odds

rather than probabilities, so we define a parameter

$$\lambda = \frac{P(T = f | C = c, F = f, M = m)}{P(T \neq f | C = c, F = f, M = m)} . \tag{4}$$

This is computed by treating C as random and applying Bayes' theorem. λ is therefore the *posterior odds* in favor of the event that the genotype of the biological father is identical with that of the legal father.

Under conditions made precise below, λ may be factored into a product of terms having distinct probabilistic import,

$$\lambda = \frac{\lambda_0 I \cdot 2^{H+J+1} \cdot \mu}{K(m, c)} , \tag{5}$$

where λ_0 is the prior paternity odds; H, an integer measuring the amount of homozygosity; I and J, related integers; and $K(m, c)$, a "rarity factor" determined by the genotype frequencies in the population. We begin with the following proposition: Suppose $f \neq g$ are two genotypes. Then

$$P(T = f | F = f) = \alpha , \tag{6}$$

and $P(T = g | F = f) = [(1 - \alpha)/(N - 1)][P(g \text{ exists in population} | f \text{ exists in population})]$, where $P(F = f)$ is taken to be different from 0.

Before giving the proof, we observe that the second equality depends on the population being finite. If the population is large (effectively infinite), then the conditioning (f exists) is irrelevant. On the other hand, in the more realistic case of a finite population, the mere fact that a genotype f is the putative father implies that this genotype exists in the population, and thus the probabilities of existence of other genotypes must be changed. According to our assumptions, we see that the information ($F = f$) is equivalent to the information (f exists) as regards the updated (conditional) probabilities of existence of other genotypes. Although this may appear obvious, we note that such a result depends upon our assumptions. Other assumptions would not necessarily lead to expressions so appealing.

Proof of Proposition

$P(T = f | F = f) = \Sigma P(T = f | X_1, \ldots, X_N, F = f) \cdot P(X_1, \ldots, X_N | F = f)$, where Σ extends over all genetic structures $(X_1, \ldots, X_N) = \alpha \Sigma P(X_1, \ldots, X_N | F = f) = \alpha$ [by assumption (2)] and $P(T = g | F = f) = \Sigma P(T = g | X_1, \ldots, X_N, F = f) \cdot P(X_1, \ldots, X_N | F = f) = [(1 - \alpha)/(N - 1)] \Sigma P(X_1, \ldots, X_N | F = f)$. The foregoing is equivalent to $[(1 - \alpha)/(N - 1)] P(f \text{ and } g \text{ exist} | F = f)$ [by assumption (2)]. But because the event ($F = f$) is contained in the event (f exists), this reduces to

$$P(T = g | F = f) = \frac{1 - \alpha}{N - 1} P(g \text{ exists} | F = f) . \tag{7}$$

Next we claim that

$$P(F = f) = \frac{1}{N} P(f \text{ exists}) . \tag{8}$$

This follows immediately from $P(F = f) = \Sigma P(F = f|X_1, \ldots, X_N) \cdot P(X_1, \ldots, X_N)$ and invoking assumption (1). Finally, we compute $P(g \text{ exists}|F = f) = \Sigma P(X_1, \ldots, X_N|F = f)$, where the sum is taken over all $(X_1, \ldots, X_N)$ containing both f and g. But $P(g \text{ exists}|F = f) = \Sigma P(X_1, \ldots, F = f)/P(F = f) = \Sigma P(F = f|X_1, \ldots, X_N) \cdot P(X_1, \ldots, X_N)/P(F = f) = \Sigma\{[P(X_1, \ldots, X_N)]/[N\, P(F = f)]\}$ [by Assumption (1)] $= \Sigma P(X_1, \ldots, X_N)/P(f \text{ exists}) = P(f \text{ and } g \text{ exist})/P(f \text{ exists}) = P(g \text{ exists}|f \text{ exists})$ [by equation (8)]. Hence, $P(g \text{ exists}|f \text{ exists}) = P(g \text{ exists}|F = f)$ and substitution into equation (7) gives the second assertion of display (6).

It is possible to begin the foregoing derivation with equation (6). The advantage in deriving them from more primitive assumptions is in mathematical hygiene. A probabilistic model requires the specification of the joint distribution of $(X_1, \ldots, X_N, F, T, C)$. Assumptions (1) through (5) give precisely such data.

Returning to the expression for λ, $P(T = f|C = c, F = f, M = m) = P(T = f, C = c, F = f, M = m)/P(C = c, F = f, M = m) = [P(C = c|T = f, F = f, M = m)P(T = f|F = f, M = m)]/[P(C = c|F = f, M = m)]$. Similarly,

$$P(T \neq f|C = c, F = f, M = m) =$$

$$\sum_{g \neq f} \frac{P(C = c|T = g, F = f, M = m)P(T = g|F = f, M = m)}{P(C = c|F = f, M = m)} .$$

Using assumption (5) and equation (6)

$$\lambda = \left(\frac{\alpha}{1 - \alpha}\right) \frac{(N - 1)\, P(C = c|T = f, M = m)}{\Sigma P(C = c|T = g, M = m)P(g \text{ exists}|f \text{ exists})} ,$$

where the sum in the denominator is taken over all possible genotypes $g \neq f$. As we are interested only in a suspected mutation, we restrict attention to those cases in which the child possesses at least one rare electromorph not present in either parent. Moreover, because the mutation rate is so small and because of the rarity of existing but as yet undetected genes, we may further delimit considerations to those cases in which the child possesses such a gene or genes at a single locus (see below). Anything more corresponds to definite nonpaternity within the accuracy of our assumptions and calculations.

Our next goal is to transform λ into a suitable computational form on the basis of our model. First we take into consideration Hardy-Weinberg equilibrium and independent segregation of the genetic markers at different loci. Consider, therefore, the possible combinations at a single locus. Suppose the mother to be given by A_iA_j and the biological father given by A_kA_l. We denote a mutant gene by A_0. The mutation probability (μ) for electromorphs will have been estimated in step 1; assume for now an order of magnitude of 10^{-5}. Therefore, double mutations at

homologous loci occur with a negligible frequency and this possibility will be ignored. This is equivalent to ignoring terms of the order of μ^2 and smaller.

The probabilities for the genotype of the child are summarized as: $A_iA_l, A_iA_k, A_jA_l, A_jA_k$, each with probability $(1 - 2\mu)/4$; $A_iA_0, A_jA_0, A_kA_0, A_lA_0$, each with probability $\mu/2$.

When we examine what happens with a totality of L loci, and invoke independent segregation, it becomes clear that the probability of a child having r mutations (at all L loci) is $K\mu^r$, where K is a constant such that $K\mu$ is still much smaller than 1. Consequently, as we have already agreed to ignore terms of order μ^2 and smaller, it is consistent to ignore multiple mutations at the L loci taken as a whole. Thus, only those males in the population who are at most one allele from consistency with M and C can be considered as potential biological fathers. Two or more inconsistencies constitute absolute nonpaternity. These considerations lead to the following cases:

Case A

If C is completely consistent with M and T (in the sense of Mendelian segregation), then

$$P[C|T, M] = (1 - 2L\mu) \prod_{i=1}^{L} \frac{\nu_i}{4} + O(\mu^2) .$$

Case B

If C is one allele short of consistency, then

$$P(C|T, M) = 2\mu \prod_{i=1}^{L} \frac{\nu_i}{4} + O(\mu^2) .$$

Here $O(\mu^2)$ represents terms the order of μ^2, and the ν_i are combinatorial factors resulting from independent Mendelian segregation and are defined as follows: At a locus i where no mutation occurs, ν_i is the number of distinct combinations of the mother's alleles and the father's alleles that could produce the child's genotype. At a locus i where a presumptive mutant allele appears, ν_i is the number of times the other allele at this locus appears in both parents. A check of all possible cases shows that the product

$$\prod_{i=1}^{L} \nu_i$$

can be very simply represented. In case A,

$$\prod_{i=1}^{L} \nu_i = 2^{H+J} .$$

In case B,

$$\prod_{i=1}^{L} \nu_i = I\,2^{H+J} .$$

Here, H = the combined number of homozygous loci in the union of parental sets at those loci where no mutation appears; J = the number of heterozygous loci in the child at which both parents are identical with the child; I = the number of alleles that the mother and father have in common with the child at a mutant locus. The following example displays this formula in operation. Example: $L = 4$

$$M = \begin{bmatrix} A_1A_2 \\ B_1B_2 \\ C_1C_1 \\ D_1D_2 \end{bmatrix} \qquad T = \begin{bmatrix} A_2A_2 \\ B_3B_3 \\ C_1C_2 \\ D_3D_4 \end{bmatrix} \qquad C = \begin{bmatrix} A_1A_2 \\ B_0B_3 \\ C_1C_1 \\ D_1D_4 \end{bmatrix}$$

$$P(C|T, M) = \left(\frac{1-2\mu}{4} + \frac{1-2\mu}{4}\right)\left(\frac{\mu}{2} + \frac{\mu}{2}\right)\left(\frac{1-2\mu}{4} + \frac{1-2\mu}{4}\right)$$

$$\left(\frac{1-2\mu}{4}\right) = \frac{\mu}{16} + \frac{3\mu^2}{8} + \frac{3\mu^3}{4} + \frac{3\mu^4}{2} = \frac{\mu}{16} + \text{lower order terms} .$$

The combinatorial factors are as follows: $\nu_1 = 2$ because the mother's A_1 can segregate with either of the father's two A_2's; $\nu_2 = 2$ because the allele B_3 appears twice (in the father); $\nu_3 = 2$ because the father's C_1 can segregate with either of the mother's two C_1's; and $\nu_4 = 1$ because only the mother's single D_1 can segregate with the father's single D_4. Thus

$$2\mu \prod_{i=1}^{4} \frac{\nu_i}{4} = 2\mu\left(\frac{2}{4}\right)\left(\frac{2}{4}\right)\left(\frac{2}{4}\right)\left(\frac{1}{4}\right) = \frac{\mu}{16} ,$$

which of course is consistent with the above. Finally, we find that $I = 2, H = 2$, $J = 0$, giving $(2\mu\, I\, 2^{H+J})/4^4 = [2\mu(2)\,(2^2)]/4^4 = \mu/16$ as before.

In the numerator of λ, the term $P(C = c|T = f, M = m)$ corresponds to case B and equals $\mu\, I\, 2^{H+J+1}$, which accounts for the presence of these factors in the numerator of equation (2). In the denominator, the offspring probabilities will depend on whether or not C is a mutation. It is therefore necessary to divide the set of potential biological fathers into two subsets reflecting the two cases above. To this end, let R denote the subset of genotypes g satisfying case A. For such g, the triplet $(C = c, T = g, M = m)$ is completely consistent, and since C has a suspected mutation, R represents the rare potential biological fathers. If $g \in R$ then

$$P(C = c|T = g, M = m) = (1 - 2L\mu) \prod_{i=1}^{L} \frac{\nu_i}{4} = (1 - 2L\mu)\frac{\nu(g)}{4^L} .$$

Let NR denote those genotypes satisfying case B. For such g, the triplet $(C = c, T = g, M = m)$ is one allele inconsistent and thus

$$P(C = c|T = g, M = m) = 2\mu \prod_{i=1}^{L} \frac{\nu_i}{4} = 2\mu \frac{\nu(g)}{4^L} .$$

Finally, setting $\lambda_0 = \alpha/(1 - \alpha)$ and $P(g|f) = P(g \text{ exists}|f \text{ exists})$, we get

$$\lambda = \frac{\mu\lambda_0 I 2^{H+J+1}}{(1 - 2L\mu) \sum_{g\epsilon R} P(g|f)\nu(g) + 2\mu \sum_{g\neq f\epsilon NR} P(g|f)\nu(g)} . \tag{9}$$

For computational purposes, equation (9) may be adapted for use with a numerical algorithm. However, it is worthwhile to make some approximations.

First, we replace $P(g \text{ exists}|f \text{ exists})$ by $P(g \text{ exists})$. This is not always a legitimate approximation, but it will do to demonstrate our methodology. Typically, the two sums occurring have on the order of 2^L terms, so that unless the functions that appear have a form that can be exploited, λ can be computed only numerically. Our approximation gives us an analytical grasp on the problem and permits a simplification of the computation.

Let $\pi(g)$ denote the probability that a single individual selected at random is type g, that is $\pi(g) = P(X_1 = g)$. However, $P(g \text{ exists}) = 1 - P(g \text{ doesn't exist}) = 1 - P(X_1 \neq g, X_2 \neq g, \ldots, X_N \neq g) = 1 - [1 - \pi(g)]^N$ [by assumption (5)]. If $\pi(g)$ is sufficiently small, then this expression can be replaced with $N\pi(g)$. We do this but note that each individual case must be checked.

Next we consider the sum on $g \neq f\epsilon NR$ in the expression for λ. Since the sum is premultiplied by μ and μ is small, little error is introduced if we permit $g = f$ in this sum. Again, this must be checked in each case. Thus the dependence on f, the genotype of the legal father, is eliminated from the denominator of λ.

Assumption (5) induces a factorization of $\pi(g)$ into a product of terms $\pi_i(g_i)$, g_i being the genotype at locus i, and π_i, the corresponding sampling probability at locus i. Bearing this in mind, together with the recollection that $\nu(g)$ itself is a product, we observe that both the sum on R and on NR can be factored into a product

$$\prod_{i=1}^{L} K_i = K ,$$

where K_i is the expectation of ν_i but only on the part of the sample space that satisfies R or NR, respectively, $K(R)$ and $K(NR)$.

This gives an expression for λ penultimate to assumption (5)

$$\lambda = \lambda_0 \left(\frac{N - 1}{N}\right) \frac{\mu I 2^{H+J+1}}{(1 - 2L\mu)K(R) + 2\mu K(NR)} .$$

Quite generally, $2\mu K(NR) << (1 - 2L\mu)K(R)$; and if, therefore, we ignore $2\mu K(NR)$ and replace $(N - 1)/N$ and also $1 - 2L\mu$ by 1, we obtain the very simple expression (2), where we have replaced $K(R)$ by $K(m, c)$ to indicate that it depends only on the genotype of the mother and child. The term $K(m, c)$ is a "rarity factor" in the following sense: it is made up of products of gene frequencies and combinatorial factors for permissible potential biological fathers. If there are several loci at which

the child has an uncommon gene not present in the mother, the corresponding factors in K will be small, boosting the odds, as expected.

Example. We give an example with 20 loci, two alleles at each locus except for three alleles at locus 1, the extra allele to account for the possibility that the child has inherited this rare variant from someone other than the legal father. The assumed gene frequency distribution is given in table 1. These values have been selected to correspond in general to the distribution of allele frequencies in the various genetic systems commonly used in studies of nonpaternity. The example thus approximates reality. Let the 20 loci be represented by $L_1, L_2, \ldots, L_{19}, L_{20}$ and let the subscripts 0, 1, and 2 refer to the alleles having frequencies given by the same subscripts in table 1. The genotypes of the mother and father at these 20 loci were obtained using 40 pseudorandom numbers, the gene frequencies given in table 1, and assumption (5). After endowing the child with a rare variant allele at locus 1, 39 pseudorandom numbers were used in conjunction with Mendelian laws [assumption (5)] to determine the genotype of the child. The results of this straightforward Monte Carlo simulation are presented in table 2.

In this example, the value of H is 26 since both F and M have 13 homozygous loci (excepting the "mutant" locus). The value of J is seen to be 1 (B locus), while the value of I is found to be 4. Using a value for μ of 1.6×10^{-5} and taking $\lambda_0 = 49(.98/.02)$, the approximate posterior odds based on formula (5) are 246:1.

The only somewhat tedious part of this calculation involves the denominator of the odds ratio. However, the calculation of this term is facilitated by our earlier comments on the use of assumption (5). To assist the reader with this calculation, the following explicit illustration based on our example may be helpful. First, we note that

$$\sum_{g \epsilon R} \pi(g)\nu(g) = \prod_{L=1}^{20} \left[\sum_{g_i} \pi_i(g_i)\nu_i(g_i) \right],$$

where Σ on g_i is taken over all g_i consistent with M and C at locus i. Now for each locus, the possible genotypes for a father are listed along with the relative frequencies and the $\nu(g_i)$ factor. For example, at the D locus, we have

Potential father	$\pi(g)$	ν
D_1D_2	.42	2
D_2D_2	.49	4

This gives a value for $\sum_{gi} \pi_i(g_i)\nu_i(g_i)$ of 2.80 for this locus. A similar calculation based on the remaining 19 loci is then obtained and all factors multiplied together. The result of this computation in our example is 594.45.

With this factor at hand, the calculation of the posterior odds for any other potential father of this child, given this mother, is quite straightforward. The contribution to $H + J$ from the B, K, and L loci is always 3 for any potential father and this mother and child. Therefore, the posterior odds for any potential father is a function of only I and H^1, where H^1 is the number of homozygous loci in the father at loci other than A, B, K, and L. The posterior odds are ($I \times 2^{H^1} \times 2^{17} \times$

TABLE 1

ILLUSTRATIVE ALLELE FREQUENCIES FOR THE CALCULATION OF A PATERNITY PROBABILITY BY THE METHODOLOGY OF THIS PAPER

No. loci with indicated allele frequencies	1	1	4	6	8
P_1*	.4	.4	.3	.2	.05
P_2	.599	.6	.7	.8	.95
P_0	.001	0	0	0	0

* P_0 is the r frequency of a potential mutant allele, whereas P_1 and P_2 are the recognized alleles in a two-allele system.

$\mu)/594.45$ for potential fathers who are one allele short of consistency. A father with $H^1 = 0$ and $I = 2$ would yield the minimum odds ratio in our example of 7.05×10^{-3}. On the other hand, the maximum odds ratio results from an $H^1 = 16$ and $I = 4$, giving us an odds ratio of 9.47×10^5:1.

Although the numerator in the odds ratio does not depend on gene frequencies, the denominator does. It is apparent that loci N and S contribute to a smaller denominator in our example because of the presence of a relatively uncommon allele in the child. The presence of several uncommon alleles in the child greatly enhances our ability to make an inference about a potential father.

TABLE 2

GENOTYPES ASSIGNED AT 20 LOCI IN MOTHER, FATHER, AND CHILD FOR PURPOSES OF A WORKED EXAMPLE

Father	Mother	Child
$L_{1.2}L_{1.2}$	$L_{1.2}L_{1.2}$	$L_{1.0}L_{1.2}$
$L_{2.1}L_{2.2}$	$L_{2.1}L_{2.2}$	$L_{2.1}L_{2.2}$
$L_{3.1}L_{3.2}$	$L_{3.1}L_{3.2}$	$L_{3.1}L_{3.1}$
$L_{4.2}L_{4.2}$	$L_{4.2}L_{4.2}$	$L_{4.2}L_{4.2}$
$L_{5.2}L_{5.2}$	$L_{5.1}L_{5.2}$	$L_{5.2}L_{5.2}$
$L_{6.1}L_{6.2}$	$L_{6.2}L_{6.2}$	$L_{6.1}L_{6.2}$
$L_{7.2}L_{7.2}$	$L_{7.2}L_{7.2}$	$L_{7.2}L_{7.2}$
$L_{8.2}L_{8.2}$	$L_{8.2}L_{8.2}$	$L_{8.2}L_{8.2}$
$L_{9.1}L_{9.2}$	$L_{9.1}L_{9.2}$	$L_{9.2}L_{9.2}$
$L_{10.2}L_{10.2}$	$L_{10.2}L_{10.2}$	$L_{10.2}L_{10.2}$
$L_{11.2}L_{11.2}$	$L_{11.1}L_{11.2}$	$L_{11.1}L_{11.2}$
$L_{12.2}L_{12.2}$	$L_{12.1}L_{12.2}$	$L_{12.1}L_{12.2}$
$L_{13.2}L_{13.2}$	$L_{13.2}L_{13.2}$	$L_{13.2}L_{13.2}$
$L_{14.1}L_{14.2}$	$L_{14.2}L_{14.2}$	$L_{14.1}L_{14.2}$
$L_{15.2}L_{15.2}$	$L_{15.2}L_{15.2}$	$L_{15.2}L_{15.2}$
$L_{16.2}L_{16.2}$	$L_{16.2}L_{16.2}$	$L_{16.2}L_{16.2}$
$L_{17.2}L_{17.2}$	$L_{17.2}L_{17.2}$	$L_{17.2}L_{17.2}$
$L_{18.2}L_{18.2}$	$L_{18.2}L_{18.2}$	$L_{18.2}L_{18.2}$
$L_{19.1}L_{19.2}$	$L_{19.2}L_{19.2}$	$L_{19.1}L_{19.2}$
$L_{20.2}L_{20.2}$	$L_{20.2}L_{20.2}$	$L_{20.2}L_{20.2}$

NOTE: For further explanation of table, see text.

DISCUSSION

We have attempted here to provide a more rigorous basis than currently exists for evaluating the role of "nonpaternity" in apparent examples of mutation in our species. To this end, formulas have been provided for subtracting the contribution of nonpaternity from an apparent mutation rate and for ordering specific examples of possible mutation on the basis of the odds for paternity. The calculation of the odds ratio for paternity rests on five assumptions that are certainly not all literally true. Although the effects of departure from these assumptions would be difficult to determine in any generality, we believe that for our purpose (the ranking of potential mutations) they provide us with a suitable robust framework, since substantial changes in an odds ratio produce small changes in a probability scale. On the other hand, these assumptions could justifiably be questioned in the legal arena and are thus limited to our context.

Finally, it is worthwhile to note how the odds ratio formula may be used to obtain an independent direct estimator of the mutation rate. If $P_i(\mu)$ is the probability for paternity for fathers i, of mutant children, $i = 1, 2, \ldots, Z$, and each father is "one short of consistency" with the respective genotypes of mother and child, then an estimator of mutation rate is found by solving for $\hat{\mu}$, where

$$\hat{\mu} = \frac{\sum_{i=1}^{Z} P_i(\hat{\mu})}{\text{No. locus determinations}} .$$

This calculation would provide a check on the crude value obtained in step 1.

REFERENCES

1. NEEL JV, MOHRENWEISER HW, SATOH C, HAMILTON HB: A consideration of two biochemical approaches to monitoring populations for a change in germ cell mutation rates, in *Genetic Damage in Man Caused by Environmental Agents*, edited by BERG K, New York, Academic Press, 1979, pp 29–47
2. NEEL JV, MOHRENWEISER HW, MEISLER MH: Rate of spontaneous mutation at human loci encoding protein structure. *Proc Natl Acad Sci USA* 77:6037–6041, 1980
3. NEEL JV: The detection of increased mutation rates in human populations. *Perspect Biol Med* 14:522–537, 1971
4. HARRIS H, HOPKINSON DA, ROBSON EB: The incidence of rare alleles determining electrophoretic variants: data on 43 enzyme loci in man. *Ann Hum Genet* 37:237–253, 1974
5. NEEL JV, UEDA N, SATOH C, FERRELL RE, TANIS RJ, HAMILTON HB: The frequency in Japanese of genetic variants of 22 proteins. V. Summary and comparison with data on Caucasians from the British Isles. *Ann Hum Genet* 41:429–441, 1977
6. NEEL JV, SATOH C, HAMILTON HB, ET AL.: Search for mutations affecting protein structure in children of atomic bomb survivors: preliminary report. *Proc Natl Acad Sci USA* 77:4221–4225, 1980