*Am. J. Hum. Genet.* 42:135–142, 1988

# Bayesian Interval Estimation of Genetic Relationships: Application to Paternity Testing

David E. Goldgar* and Elizabeth A. Thompson†

*Division of Biostatistics, Department of Preventive Medicine, University of Mississippi School of Medicine, Jackson; and †Department of Statistics, University of Washington, Seattle

## Summary

**Using genetic marker data, we have developed a general methodology for estimating genetic relationships between a set of individuals. The purpose of this paper is to illustrate the practical utility of these methods as applied to the problem of paternity testing. Bayesian methods are used to compute the posterior probability distribution of the genetic relationship parameters. Use of an interval-estimation approach rather than a hypothesis-testing one avoids the problem of the specification of an appropriate null hypothesis in calculating the probability of paternity. Monte Carlo methods are used to evaluate the utility of two sets of genetic markers in obtaining suitably precise estimates of genetic relationship as well as the effect of the prior distribution chosen. Results indicate that with currently available markers a "true" father may be reliably distinguished from any other genetic relationship to the child and that with a reasonable number of markers one can often discriminate between an unrelated individual and one with a second-degree relationship to the child.**

## Introduction

Use of the paternity index or likelihood ratio in paternity determination has become a somewhat controversial subject. Aickin (1984) and Li and Chakravarti (1985) have discussed what they consider to be fallacies and inconsistencies inherent in the paternity index. These criticisms have been rebutted to varying degrees by, among others, Brenner (1985), Walker (1985), Elston (1986), Mickey et al. (1986), and Thompson (1986). Part of this debate centers on whether the alleged father's phenotype is relevant to the question of paternity in any way other than simple determination of inclusion or exclusion. Incorporation of a prior probability of paternity has also been debated, although (1) there now seems to be a general consensus that incorporation of prior information is desirable and (2) Elston (1986) has shown that for any probability of paternity to be valid it must incorporate an appropriate prior probability. The purpose of the present paper is to formulate the problem in such a way that many of the controversial issues become irrelevant.

The paternity-index approach requires an appropriate null hypothesis for construction of the likelihood ratio. As pointed out by Aicken (1984), Li and Chakravarti (1985), and others, the usual interpretation of the denominator of the likelihood ratio as being the genotype probability for a "random individual" suffers from serious deficiencies. On the other hand, ignoring the information contained in the putative father's marker phenotypes (except from the standpoint of exclusion/inclusion), as proposed by Li and Chakravarti (1985), is not an optimal solution either. In this paper we abandon the odds or hypothesis-testing approach in favor of an interval-estimation approach. In particular, a Bayesian approach is used to estimate genetic relationships between individuals. Bayesian analysis has several advantages over the standard-likelihood approach, largely because it incorporates prior knowledge or belief in the accused individual's relationship to the child. In addition, the Bayesian approach avoids certain statistical problems at parameter boundaries as well as reliance on large-sample distributional theory.

Received October 23, 1986; revision received May 8, 1987.

Address for correspondence and reprints: David E. Goldgar, Ph.D., Department of Genetic Epidemiology, University of Utah, 410 Chipeta Way #105, Salt Lake City, UT 84108.


0002-9297/88/4201-0019$02.00

Because the problem of paternity determination is, after all, a special case of determining the genetic relationship between individuals, methods based on the theory of genetic relationships and kinship coefficients are applicable to paternity testing. Cotterman's (1940) first description of the use of $k$-coefficients for specifying the relationship between individuals has been extended by numerous other investigators (Edwards 1967; Yasuda 1968; Thompson 1974, 1975). In particular, Thompson (1975) examined the question of estimation and testing of pairwise relationships from a likelihood point of view. A theoretical overview of identity by descent, kinship coefficients, estimating genetic relationships, and reconstructing genealogies is provided by Thompson (1985, pp. 16–71). In the past, these methods have been applied largely toward identification of population structure, estimation of levels of population inbreeding, and reconstruction of complex genealogies. Our aim is to estimate the genetic correlation (kinship coefficient) of a putative father to a child—or, more precisely, to the true biological father of the child. Construction of confidence intervals for the true genetic relationship provides a paternity-testing method that is not dependent on a particular null hypothesis. Thus, on the basis of confidence intervals, various relationships between the true father and the alleged father can be excluded.

In this paper the conditional likelihood of a child's phenotype given the phenotypes of two hypothetically related individuals is derived as a function of a set of parameters that represent the degree of genetic relationship of these individuals to the child and to each other. This formulation allows for estimation, formation of confidence intervals, and hypothesis tests of the relationship parameters. Results of a study of the distribution of this likelihood function within the framework of the usual paternity-testing situation are presented.

## Derivation of the Likelihood

It is assumed that marker-phenotype information is available on three individuals—a child and two individuals whose relationship to the child we wish to investigate. These two individuals are assumed to be unilineally related to the child, one through the maternal line and one through paternal descent. Extension of the method to more than two potentially related individuals is relatively simple; for clarity it will not be dealt with in this paper. The two individuals are assumed to be noninbred but may be related in either a unilineal or bilineal fashion. The basic idea, then, is to write the conditional probability of the child's phenotypes for a battery of genetic markers, given the phenotypes of the two individuals, as a function of the coefficients of relationship between these three individuals and the marker-allele frequencies. The coefficient of relationship between two individuals, which we denote by $R$, was first proposed by Wright (1922) and, if it is assumed that neither individual is inbred, is defined to be twice their kinship coefficient, $\Psi$. Here $\Psi$ denotes the probability that two homologous genes drawn at random, one from each of two individuals, are identical by descent. In terms of Cotterman (1940) $k$-coefficients, we have $k_0 = (1 - R)^2$, $2k_1 = 2R(1 - R)$, and $k_2 = R_2$ for the simplest bilineal relationships and $k_0 = (1 - 2R)$, $2k_1 = 2R$, and $k_2 = 0$ for unilineal relationships. Although it is true that not all pairwise genealogical relationships can be specified by the single parameter $R$, all unilineal relationships are allowed for, in addition to those bilineal relationships satisfying $4k_2k_0 = 4k_1^2$. Thus we allow for relationships such as self, sib, and double first cousin but cannot specify quadruple half-first cousins, for example. Formally, the parameters of the likelihood are as follows:

$R_{mp}$ is the coefficient of relationship between the maternal and paternal individuals. This is typically assumed to be known, and for many situations it will be zero.

$R_{mc}$ ($R_{pc}$) is the coefficient of relationship between the maternally (paternally) related individual and the biological mother (father), $.0 \leq R_{mc}, R_{pc}, R_{mp} \leq 1.0$

$p = \{p_{sn}\}$ is the set of population frequencies of allele $s$ at marker locus $n$.

To derive the likelihood function for a single locus, let $P_m$, $P_p$, and $P_c$ be the phenotypes of the maternal and paternal individuals and of the child, respectively, and let $G_j$ be the $j$th of $g$ genotypes at this locus; $j = 1, 2, \ldots, g$; $\delta_{hj}$ = prob[phenotype $h$|genotype $j$] = 0 or 1; and $j = 1, 2, \ldots, g$; $h = m, p, c$. By means of this notation the relationships between the individuals under study are shown in the path diagram in figure 1. Note again that in our model $R$ is the relationship between the tested individual and the biological parent of the child and that $R/2$ is the relationship between the individual and the child. Given the combined relationships $\mathcal{R} = (R_{pc},$

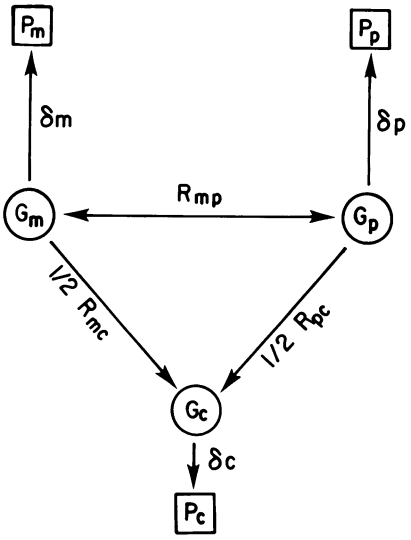


**Figure 1** Path diagram of the relationships between the tested individuals (m, p, and c). $P_m$, $P_p$, and $P_c$ refer to the phenotypes of these individuals, and $G_m$, $G_p$, and $G_c$ are their corresponding genotypes. The coefficients $\delta_m$, $\delta_p$, and $\delta_c$ represent the relationships between genotype and phenotype, and the $R_{ij}$ values are the coefficients of relationship between these individuals.

$R_{mc}$, $R_{mp}$), the probability of the child's phenotype conditional on the maternally and paternally related individuals' phenotypes can be written as

$$\Pr(P_c|P_m, P_p) = \sum_i^g \Pr(p_c|G_i) \cdot \Pr(G_i|P_m, P_p)$$

$$= \sum_i^g \sum_j^g \sum_k^g \Pr(P_c|G_i) \cdot \Pr(G_i|G_j, G_k)$$

$$\cdot \Pr(G_j, G_k|P_m, P_p) = \sum_i^g \sum_j^g \sum_k^g \delta_{ci} \Pr(G_i|G_j, G_k)$$

$$\cdot \left[ \frac{\delta_{jm}\, \delta_{kp}\, \Pr(G_j, G_k)}{\sum_{j'} \sum_{k'} \delta_{j'm} \delta_{k'p} \Pr(G_{j'}, G_{k'})} \right], \quad (1)$$

where all Pr (probabilities) are computed under the relationship $\mathcal{R}$.

Now we require the child's genotype probability given the maternal and paternal individuals' genotypes, $\Pr(G_i|G_j, G_k)$, as a function of the parameters $R_{mc}$ and $R_{pc}$, and the allele frequencies $\boldsymbol{p}$. Let the child's genotype, $G_i$, consist of alleles $g_1$ and $g_2$. Either $g_1$ is derived from the paternal relative and $g_2$ from the maternal or the opposite is true. We have

$$\Pr(G_i|G_j, G_k) = \begin{bmatrix} f(g_1, G_j; R_{mc}, \boldsymbol{p}) \cdot f(g_2, G_k; R_{pc}, \boldsymbol{p}) \\ \text{if } g_1 = g_2; \\ \text{and} \\ f(g_1, G_j; R_{mc}, \boldsymbol{p}) \cdot f(g_2, G_k; R_{pc}, \boldsymbol{p}) \\ + f(g_1, G_k; R_{pc}, \boldsymbol{p}) \cdot f(g_2, G_j; R_{mc}, \boldsymbol{p}) \\ \text{if } g_1 \neq g_2 \end{bmatrix}, \quad (2)$$

where the functions $f(g, G; R, \boldsymbol{p})$ depend on the allele $g$ and the genotype $G$ and are shown in table 1. The derivation of these functions is straightforward. For example, for the child's allele $A_i$ and for the maternally related individual with genotype $A_iA_i$, the $A_i$ allele in the child is either identical by descent to that contributed by the maternal relative (with probability $R = R_{mc}$) or is derived from the population at large with probability $(1 - R)p_i$. This results in the table entry of $R + (1 - R)p_i$. Other table entries are derived similarly.

In addition, we need the joint probability of the maternal and paternal individuals' genotypes, $\Pr(G_j, G_k)$, expressed as a function of $R_{mp}$ and $\boldsymbol{p}$. These functions, $f(G_j, G_k; R, \boldsymbol{p})$, which express the joint probability of pair genotypes as a function of the genetic correlation, $R$, and the gene frequencies, $\boldsymbol{p}$, may be obtained by transformating corresponding results for $k$-coefficients (Crow and Kimura 1970, p. 137; Kimberling and Goldgar 1980).

Substitution of these functions in equation (1) allows calculation of the probability of any triplet of maternal, paternal, and child marker phenotypes for a given locus at specific values of $R_{mc}$, $R_{pc}$, $R_{mp}$, and

**Table 1**

**Conditional Parent-Child Genotype Probabilities Expressed as a Function of $R$ and Gene Frequency**

| Parental Genotype | Child's Allele | $F(R, p)$ |
|---|---|---|
| $A_iA_i$ ..... | $A_i$ | $R + (1 + R)p_i$ |
| | $A_j$ | $(1 - R)p_j$ |
| $A_iA_j$ ..... | $A_i$ | $\frac{1}{2}R + (1 - R)p_i$ |
| | $A_j$ | $\frac{1}{2}R + (1 - R)p_j$ |
| | $A_k$ | $(1 - R)p_k$ |

*p;* that is, it allows calculation of the likelihood function for the parameters. For a battery of genetic markers, the corresponding likelihood is the product of the likelihoods for each locus, if free recombination and no linkage disequilibrium between loci are assumed. Standard likelihood theory is used to obtain point and interval estimates and to test hypotheses about the unknown relationship parameters.

Application of this method to the specific problem of paternity determination allows for several simplifications of the likelihood function specified in equations (1) and (2). First, we will assume that the two ancestral individuals are not related (i.e., $R_{mp} = 0$). Inspection of equation (1) shows, however, that even if this assumption is violated, the likelihood does not depend on $R_{mp}$ if there is a one-to-one correspondence between marker phenotypes and marker genotypes. In this case, the likelihood for each marker locus reduces to equation (2); that is, $\Pr(P_c|P_m, P_p) = \Pr(G_c|G_m, G_p)$. In the paternity-determination formulation we naturally assume that the woman tested is the true mother of the child (i.e., $R_{mc} = 1$) and examine the likelihood as a function of a single unknown parameter, $R_{pc}$. Under this parameterization, the paternity index can be expressed as the ratio of the likelihood function given by equations (1) and (2) evaluated at $R_{pc} = 1$ to the same likelihood evaluated at $R_{pc} = 0$. However, we prefer to take a Bayesian interval-estimation approach to the problem.

## Bayesian Estimation of $R_{pc}$

In a paternity-testing context, we are concerned with estimation of the genetic relationship $R_{pc}$ between the child's true father and the accused father. In the following treatment, $R$ (without subscript) will denote $R_{pc}$. In a Bayesian analysis, one must first decide on an appropriate prior distribution for the parameter(s) of interest. For the prior distribution of the genetic relationship parameter, $R$, we chose to use the beta distribution—$g(R) = B(\alpha, \beta) \cdot R^{\alpha-1}(1 - R)^{\beta-1}$—owing to its ability to take on a variety of forms depending on the values of the parameters $\alpha$ and $\beta$. Two sets of values of $\alpha$ and $\beta$ were chosen for presentation in this paper: (1) $\alpha = 1.0$, $\beta = 1.0$ and (2) $\alpha = .1$, $\beta = .05$. The first of these corresponds to a uniform prior and was chosen for comparison with the standard likelihood approach since in this case the posterior density of $R$ is simply the likelihood function normalized to integrate to one. The second prior, which we will denote the empirical prior, was chosen to represent the typical situation in which we have a fairly strong prior belief that the accused individual is the true father, a somewhat lesser belief that he is unrelated to the child, and we allow for the possibility that he is related to the true father. The empirical posterior distribution of $R$ for a given set of marker phenotypes and a given prior, $g(R)$, is obtained in the following manner: We first evaluate the likelihood function—$L(P_m, P_p, P_c|R)$—for a given mother-child-man trio at a large number of points in the parameter space (in this case the interval [0, 1]); in our analyses we typically evaluated the likelihood at .01 intervals. Next we evaluate $g(R)$ and compute $L(R) \cdot g(R)$ at each point. This function is then normalized to a probability density function, $p(R)$, by integrating and dividing each point value by the resultant integral. A posterior Bayes estimate (PBE) of $R$ is the mean of $p(R)$ and is obtained by numerical integration. The cumulative posterior distribution of $R$, $Q(R)$, is given by $Q(R) = \int_0^R p(R')dR'$. A $100\gamma\%$ Bayesian interval estimate of $R$ is found by examining the cumulative posterior distribution for the smallest interval $(R_1, R_2)$ for which the difference $Q(R_2) - Q(R_1) \geq \gamma$. The procedure can be extended to obtain joint estimates and confidence regions for two parameters, e.g., $R_{mp}$ and $R_{pc}$.

## Simulation Methods

Rather than examining the likelihood under an artificial special case such as $N$ markers each with $n$ equally frequent alleles, we chose to look at situations likely to be encountered in practice. Two sets of markers were used. The first (STD) corresponds to the standard blood-group/enzyme/protein battery used by many paternity-testing laboratories. The 24 markers that constitute this battery had heterozygosity frequencies ranging from 6% to 69%, with a mean heterozygosity of 37.5%. The second marker battery (COM) was a group of 44 markers consisting of those in the first battery with the addition of 20 RFLPs. Although not currently used for paternity testing, these are expected to supplement or replace standard markers in the near future (Balazs et al. 1984; Smouse and Chakraborty 1986). RFLPs were selected from the 100 or so available markers on the basis of their informativeness and diversity of chromosomal location. Heterozygosity in these 20 markers ranged from 46% to 78%, with a mean of 56.8%. Although it was impossible to avoid some degree of linkage between markers, the overall effect on our

analyses is probably negligible. To maximize computational efficiency in the simulation study, availability of genotype data was assumed for all marker loci. Use of a single highly polymorphic system such as HLA was not considered.

For each marker, genotypes of the four grandparents were randomly simulated according to allele frequencies. The genotype of the mother was obtained by randomly choosing one allele from each maternal grandparent. From the paternal grandparents, genotypes of the "true" father and his brother were similarly obtained. The child's genotype was then obtained by randomly selecting one allele from the mother and one from the true father. In addition, one random individual was generated. This procedure was repeated for each marker in the battery under consideration. Thus, for each paternity "case," there were three putative fathers: the true father, his brother, and an individual drawn at random from the population, providing us with the ability to look at true $R_{pc}$ values of 1.0, .5, and .0, respectively. Two hundred case sets were independently simulated for each marker battery.

For each of the 200 cases, the value of the likelihood function was obtained for each putative father for $R_{pc}$ values of .00, .01, .02, . . . , 1.00. From this empirical likelihood function, the cumulative posterior distribution, PBE of $R_{pc}$, and 95% Bayesian interval estimates were obtained as described above for each prior. All simulation and analysis were performed on a Compaq Deskpro microcomputer by means of a set of TURBO PASCAL programs that were specifically written for this purpose.

## Results

Results of the analyses of the simulated paternity cases are summarized in table 2 and figure 2. Table 2 describes the distribution of the point and interval estimates of $R_{pc}$ among the 200 simulated cases for each true value of $R_{pc}$ considered in the simulation. Data presented are the mean ± SD and range of the PBE of $R_{pc}$, the average upper and lower 95% confidence limits, and the percentage of cases in which the interval contained the values of .0, .5, and 1.0. The results are presented for both sets of markers and both prior distributions of $R_{pc}$ that were used. Figure 2 shows the average cumulative posterior distributions for the combined marker battery and the uniform prior for each simulated value of $R_{pc}$: 1.0, .5, and .0.

When the man tested is, in fact, the biological father of the child (i.e., $R_{pc} = 1.0$; table 2A), the average estimates of $R_{pc}$ are .903 and .950 for the combined set of markers and for uniform and empirical priors, respectively. From table 2A we also see that for the combined marker set, all 200 cases had lower 95% limits of >.5. This was true irrespective of the prior distribution and indicates *exclusion (with 95% probability) of all standard genealogical relationships other than true father!* For the standard marker battery, the proportion of cases with lower Bayesian probability limits >.5 was 55% for the uniform prior and 88% for the empirical prior. Under the empirical prior, the mean lower limits of these intervals were .85 and .65 for the combined and standard marker sets, respectively; corresponding figures for the uniform prior were .75 and .49.

When the putative father is the brother of the true father (i.e., $R_{pc} = .5$; table 2B), the average estimates of $R_{pc}$ are very close to the simulation values for both priors and both sets of markers. Moreover, 94% of the interval estimates using the uniform prior and combined marker set contained the true value of .5, and 76.5% of these intervals included neither $R = 0$ nor $R = 1$. Results for the empirical prior show a reduced frequency of confidence intervals containing the true value, and in only 47% of the intervals could both extremes be excluded, even when the larger marker battery was used, a result illustrative of the fact that the empirical prior gives low weight to values of $R_{pc}$ near .5.

Finally, we examined the distribution of estimates and confidence intervals when the putative father is genetically unrelated to the child (i.e., $R_{pc} = 0$; table 2C). The upper confidence limits for this case tell us to what degree other relationships can be ruled out when the putative father is unrelated to the child. For the combined marker battery using the empirical prior, relationships of $R_{pc} \geq .50$ could be excluded with 95% confidence in 87.5% of the cases, whereas in 39.0% of the cases relationships of $R \geq .25$ could be so excluded. Even with the standard battery, 62.5% of intervals had upper limits of <.5.

## Discussion

The methods described in this paper were developed as an application of general methodology for investigation, on the basis of marker phenotype data, of the genetic relationships between a set of individuals. In addition to the usual paternity problem, these

**Table 2**

**Characteristics of the PBE and 95% Bayesian Interval Estimates of $R_{pc}$**

| Parameter | Prior | | | |
|---|---|---|---|---|
| | Standard Marker Battery | | Combined Marker Battery | |
| | Uniform | Empirical | Uniform | Empirical |
| | A. $R_{pc} = 1.0$ | | | |
| PBE: | | | | |
| Mean ± SD | .81 ± .05 | .90 ± .04 | .90 ± .02 | .95 ± .01 |
| Range | .54–.89 | .59–.95 | .85–.93 | .93–.96 |
| 95% Bayesian interval estimates: | | | | |
| Average limits | .49–1.00 | .65–1.00 | .75–1.00 | .85–1.00 |
| % Of intervals containing: | | | | |
| .00 | 0.0 | 0.0 | 0.0 | 0.0 |
| .50 | 45.0 | 12.0 | 0.0 | 0.0 |
| 1.00 | 100.0 | 100.0 | 100.0 | 100.0 |
| | B. $R_{pc} = .5$ | | | |
| PBE: | | | | |
| Mean ± SD | .50 ± .16 | .49 ± .23 | .48 ± .15 | .47 ± .19 |
| Range | .20–.89 | .11–.88 | .15–.90 | .08–.95 |
| 95% Bayesian interval estimates: | | | | |
| Average limits | .09–.82 | .11–.83 | .12–.74 | .09–.75 |
| % Of intervals containing: | | | | |
| .00 | 43.0 | 63.5 | 20.5 | 46.5 |
| .50 | 98.0 | 87.5 | 94.0 | 86.5 |
| 1.00 | 15.5 | 30.5 | 3.0 | 6.0 |
| | C. $R_{pc} = .0$ | | | |
| PBE: | | | | |
| Mean ± SD | .25 ± .10 | .16 ± .11 | .18 ± .08 | .11 ± .07 |
| Range | .11–.68 | .06–.80 | .07–.51 | .04–.50 |
| 95% Bayesian interval estimates: | | | | |
| Average limits | .00–.55 | .00–.47 | .00–.39 | .00–.31 |
| % Of intervals containing: | | | | |
| .00 | 95.5 | 98.5 | 96.0 | 100.0 |
| .50 | 58.5 | 37.5 | 18.0 | 12.5 |
| 1.00 | 0.5 | 1.0 | 0.0 | 0.0 |

methods could be applied, for example, to the more general problem of estimating relationships for remote ancestries (a problem recently addressed from the point of view of exclusion by Darlu and Cavalli-Sforza [1985]) or could be used as a statistical test for consanguinity. In this paper, however, we have concentrated on the standard paternity-testing situation because of both its widespread application and the recent controversies surrounding its implementation. Parameterization of the likelihood in terms of the single genetic correlation $R_{pc}$ is attractive for several reasons. First, under this parameterization it is quite obvious that, contrary to the claim of Li and Chakravarti (1985), the standard paternity index is a true likelihood ratio. It also allows tests of more general hypotheses than do previous approaches, and, more important, it allows one to discard the hypothesis-testing approach altogether. By computing the likelihood of the observed marker phenotypes as a function of the parameter $R$ and by assuming a prior distribution for $R$, one can easily compute the posterior distribution of $R$. The posterior distribution can be used to derive certain probabilities relevant to the paternity issue and to obtain both point and interval estimates of $R$.

The advantage of the confidence-interval approach is partially one of semantics. Both the paternity index and odds of paternity compare the probability of the

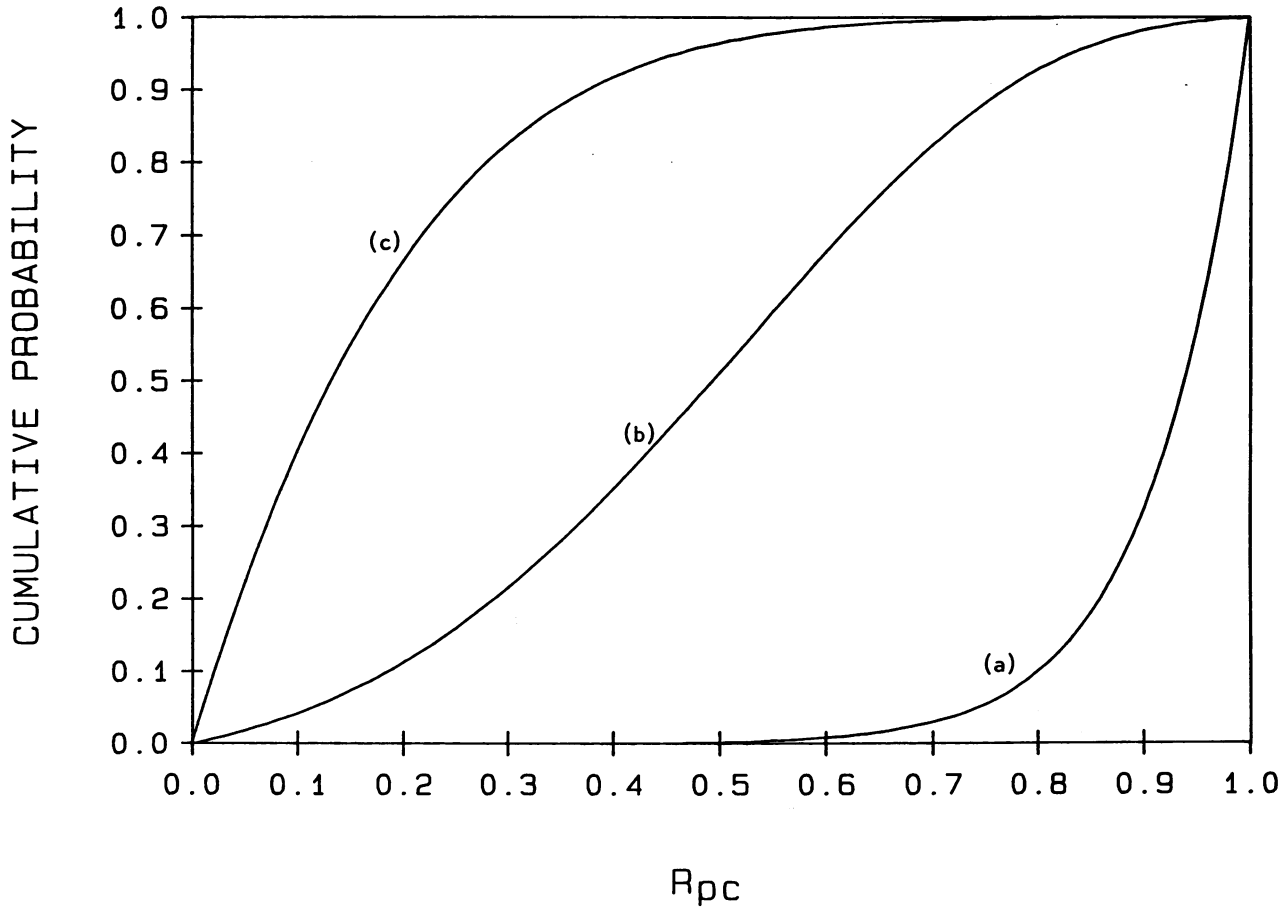


**Figure 2** Mean estimated cumulative posterior distribution of $R_{pc}$ for the combined marker battery when a uniform prior is used. (a), $R_{pc}$ = 1.0; (b), $R_{pc}$ = .5; and (c) $R_{pc}$ = .0.

child's phenotype, given the mother and putative father phenotypes, compared with the analogous probability for a "random" male. This interpretation of what is in fact a likelihood ratio for *two alternative relationships* has caused some confusion in the literature. By phrasing the paternity issue as one of *estimating* the genetic relationship of the accused man to the child, this issue is avoided. The confidence interval for the genetic relationship provides for the possibility of reliable exclusion of any biological relationships other than father—or, at the other end of the scale, could show that the accused individual has no close biological relationship to the child in question. Conversely, if the confidence interval was (.2, .8), for example, this could be indicative of the putative father being related to the true father of the child. In addition, an interval of this type may point to the presence of laboratory error, chimerism, null alleles, or sample mislabeling.

Use of the methods outlined in this paper require an assumption regarding the prior distribution of the parameter to be estimated. On the basis of our study several general observations can be made. First, as the amount of data (in this case number of markers) increases, the effect of the prior distribution becomes diluted. The effect of the prior distribution was also largest when the true value of $R_{pc}$ was small (e.g., .0). The reason for this is evident in figure 2, which shows the average likelihood function for the three simulated putative fathers. The likelihood function is considerably steeper in the vicinity of the true value of $R_{pc}$ for the case $R_{pc}$ = 1.0 than it is for the case $R_{pc}$ = 0. Thus, changes in the prior distribution will have a greater effect on the posterior distribution for the unrelated case. The choice of prior distribution in estimating genetic relationships depends on the particular application in which they are used. For the problem of genealogy reconstruction, for example,

one might choose a prior on the basis of general knowledge regarding the biology of the population or species, whereas for the typical paternity-testing situation, one would perhaps use a U-shaped prior similar to that analyzed in our study. If one wanted to consider solely the genetic evidence at hand for a particular case, then the uniform prior might be appropriate. The actual parameter values will often be chosen to fit some empirical criteria that may differ across populations.

Results of the simulation show that, for a wide variety of priors, if the alleged father is the true father, the standard markers do a reasonably good job of excluding other relationships with $\geq 95\%$ Bayesian posterior probability. For unrelated individuals, it appears that only with the empirical prior and combined marker battery do we get a large proportion of intervals of desirable width (i.e., $0 \leq R < .5$). The results also indicate that with a reasonable number of markers, it is often possible to discriminate between the two cases $R = .5$ and $R = .0$. To estimate other relationships reliably, however, a larger or more polymorphic marker battery would be required. Recent articles (Olson et al. 1986; Smouse and Chakraborty 1986) have described the use of Q-band chromosomal heteromorphisms and RFLPs in paternity testing, and the approach presented in this paper could also be used to analyze data obtained from Jeffreys et al.'s (1986) probes.

We do not expect that this approach will find immediate general use in paternity cases, since it has taken considerable time and effort on the part of geneticists to gain acceptance for the use of probability statements (i.e., odds of paternity, paternity index, etc.) as admissible evidence in paternity disputes. We do feel, however, that the approach that we have taken in this paper allows for more informative decisions to be made regarding the paternity issue. Clearly, as the number of markers routinely available for such analyses increases, an interval-estimation approach will shed important light on questions of disputed parentage.

## References


Aickin, M. 1984. Some fallacies in the computation of paternity probabilities. Am. J. Hum. Genet. 36:904–915.

Balazs, I., K. Wexler, L. Nicholas, L. Miyazaki, A. Giusti, M. Baird, P. Rubinstein, F. H. Allen, and L. Sussman. 1984. The use of restriction fragment length polymorphisms for the determination of paternity. Am. J. Hum. Genet. 36:161S.

Brenner, C. H. 1985. Evidence, probability, and paternity (letter). Am. J. Hum. Genet. 37:826–827.

Cotterman, C. W. 1974. A calculus for statistico-genetics. *In* P. A. Balonoff, ed. Genetics and social structure. Academic Press, New York. (Originally published as a Ph.D. thesis, Ohio State University, 1940.)

Crow, J. F., and M. Kimura. 1970. An introduction to population genetics theory. Harper & Row, New York.

Darlu, P., and L. L. Cavalli-Sforza. 1985. The probability of exclusion of ancestries based on genetic observations. Am. J. Hum. Genet. 37:581–590.

Edwards, A. W. F. 1967. Automatic reconstruction of genealogies from phenotypic information-AUTOKIN. Bull. Eur. Soc. Hum. Genet. 1:42–43.

Elston, R. C. 1986. Probability and paternity testing. Am. J. Hum. Genet. 39:112–122.

Jeffreys, A. J., V. Wilson, S. L. Thein, D. J. Weatherall, and B. A. J. Ponder. 1986. DNA "fingerprints" and segregation analysis of multiple markers in human pedigrees. Am. J. Hum. Genet. 39:11–24.

Kimberling, W. J., and D. E. Goldgar. 1980. The use of genetic markers in the study of human resemblance. Acta Genet. Med. Gemollol. 29:255–262.

Li, C. C., and A. Chakravarti. 1985. Basic fallacies in the formulation of the paternity index. Am. J. Hum. Genet. 37:809–818.

Mickey, M. R., D. W. Gjertson, and P. I. Terasaki. 1986. Empirical validation of the Essen-Möller probability of paternity. Am. J. Hum. Genet. 39:123–132.

Olson, S. B., R. W. Magenis, and E. W. Lovrien. 1986. Human chromosome variation: the discriminatory power of Q-band heteromorphism (variant) analysis in distinguishing between individuals, with specific application to cases of questionable paternity. Am. J. Hum. Genet. 38:235–252.

Smouse, P. E., and R. Chakraborty. 1986. The use of restriction fragment length polymorphisms in paternity analysis. Am. J. Hum. Genet. 38:918–939.

Thompson, E. A. 1974. Gene identities and multiple relationships. Biometrics 30:667–680.

———. 1975. The estimation of pairwise relationships. Ann. Hum. Genet. 39:173–188.

———. 1985. Pedigree analysis in human genetics. Johns Hopkins University Press, Baltimore.

———. 1986. Likelihood inference of paternity (letter). Am. J. Hum. Genet. 39:285–287.

Walker, R. H. 1985. Guidelines for reporting estimates of probability of paternity. Am. J. Hum. Genet. 37:819–825.

Wright, S. 1922. Coefficients of inbreeding and relationship. Am. Nat. 56:330–338.

Yasuda, N. 1968. Estimation of the inbreeding coefficient from phenotype frequencies by a method of maximum likelihood scoring. Biometrics 24:915–935.