Am J Hum Genet 28:615–618, 1976

Brief Communication

A Note on the Distribution of the Number of Exclusions to be Expected in Paternity Testing

RANAJIT CHAKRABORTY[1] AND WILLIAM J. SCHULL[1]

As we have emphasized in two recent publications [1, 2], the immunogenetic and biochemical tests, more or less routinely available now, should offer solace to a falsely charged putative father in the legal defense of his innocence. To make evaluation of the probability of paternity exclusion intelligible to the nonspecialist in these publications, we ignored the complications resulting from mutations, suppressors (e.g., the Rh-null chromosome or the Bombay type), or misclassifications—at least in the basic equation developed in one of these articles [1]. Our recent experience in dealing with individual cases of such evaluation, however, has prompted us to probe the situation with more rigor.

In this communication, therefore, we set out in general form the distribution of the number of exclusions which may be obtained given a specific but arbitrary number of genetic determinations. Once such probabilities are obtained, we may then compare them with the odds of obtaining a similar number of exclusions on some basis other than nonpaternity (e.g., mutation).

DISTRIBUTION OF THE NUMBER OF EXCLUSIONS

Let n denote the number of tests performed (immunogenetic or biochemical) and X the number which excludes a falsely accused male from paternity. Clearly, X may assume any integral value from 0 to n. The cumulative probability of excluding a random male for the above set of n systems may be written as

$$P_1^{(n)} = 1 - \prod_{i=1}^{n} (1 - p_i), \qquad (1)$$

where p_i is the probability of exclusion on the basis of the ith system ($i = 1, 2, \ldots, n$). In terms of the random variable X, expression (1) is also Prob($X \geq 1$).

Received April 8, 1976; revised June 23, 1976.

This work was supported by a grant from the National Institute of General Medical Sciences (GM 19513).

[1] Center for Demographic and Population Genetics, University of Texas Health Science Center, Houston, Texas.

The probability of exclusion on the basis of exactly $k(= 0, 1, \ldots, n)$ of these n systems, is then given by

$$\text{Prob}(X = k) = \Sigma[p_{i_1} p_{i_2} \ldots p_{i_k} \prod_j (1 - p_j)], \tag{2}$$

where the product in equation (2) is formed over all $j \neq i_1, i_2, \ldots, i_k$ and the summation Σ extends over all possible combinations of $i_1, \ldots, i_k$ taken from 1, 2, $\ldots, n$. (In fact, Σ includes ${}^nC_k$ individual terms.) Expression (2) thus specifies the distribution of X, the number of possible exclusions. Simple as it might appear, expression (2) is truly a very tedious expression to compute in practice, particularly so if n is large (of the order of 20 or 30) or the p's are small, say, between 0–0.33 (only a few systems have p_i's above 0.20 level [1, 2]).

However, expression (1) lends itself to some algebraic manipulations which facilitate computation of expression (2) with precision and ease. When one additional system is put into our battery of n tests, the probability of one or more exclusions based on $n + 1$ systems, $P_1^{(n+1)}$, becomes

$$\begin{aligned} P_1^{(n+1)} &= 1 - \prod_{i=1}^{n+1} (1 - p_i) \\ &= P_1^{(n)} + p_{n \pm 1}(1 - P_1^{(n)}). \end{aligned} \tag{3}$$

Similarly, with $P_2^{(n)} = \text{Prob}(X \geqq 2$, when n systems are analyzed) we have the recurrence equation

$$P_2^{(n+1)} = P_2^{(n)} + p_{n+1}(P_1^{(n)} - P_2^{(n)}). \tag{4}$$

In general,

$$P_k^{(n+1)} = P_k^{(n)} + p_{n+1}(P_{k+1}^{(n)} - P_k^{(n)}) \tag{5}$$

holds for any $n \geqq 1$, and $k = 1, \ldots, n$.

Furthermore, noting that $p_k^{(n)} = 0$ for all $n < k$, and $P_k^{(k)} = p_1\ p_2 \ldots p_k$, equations (3)–(5) can be used successively to generate the distribution of X by the equation

$$\text{Prob}(X = k) = Pr(X \geqq k) - Pr(X \geqq k + 1) \tag{6}$$

for any specified n $(k = 1, 2, \ldots, n)$. The probability of obtaining no exclusion, obviously, being obtained as $\text{Prob}(X = 0) = 1 - \text{Prob}(X \geqq 1)$.

NUMERICAL EVALUATIONS

We previously enumerated 57 immunological and biochemical genetic systems where testing could easily become routine or is virtually routine now in many laboratories [1]. The information with respect to nonpaternity inherent in these systems under the test procedures usually available are also indicated in the literature [1, 3]. Some practitioners, however, argue that the sex-linked traits (e.g., Xg, Xm, and G6PD) are of only limited use in paternity exclusions (only if the child is a female) [4, 5]. For the present calculations, therefore, we shall exclude

these three systems. However, the nature of the distribution, and therefore the general confidence in the inferences drawn from such computations, does not alter substantially even with inclusion of these systems. We have also not included the HLA leukocyte types, in spite of their great potential power for paternity determinations, for different laboratories use different sets of antisera, and furthermore, the exclusion probability for this system depends upon whether the gene frequencies or the haplotype frequencies are used in the computations.

In figure 1 we present the probability distribution of the number of ex-

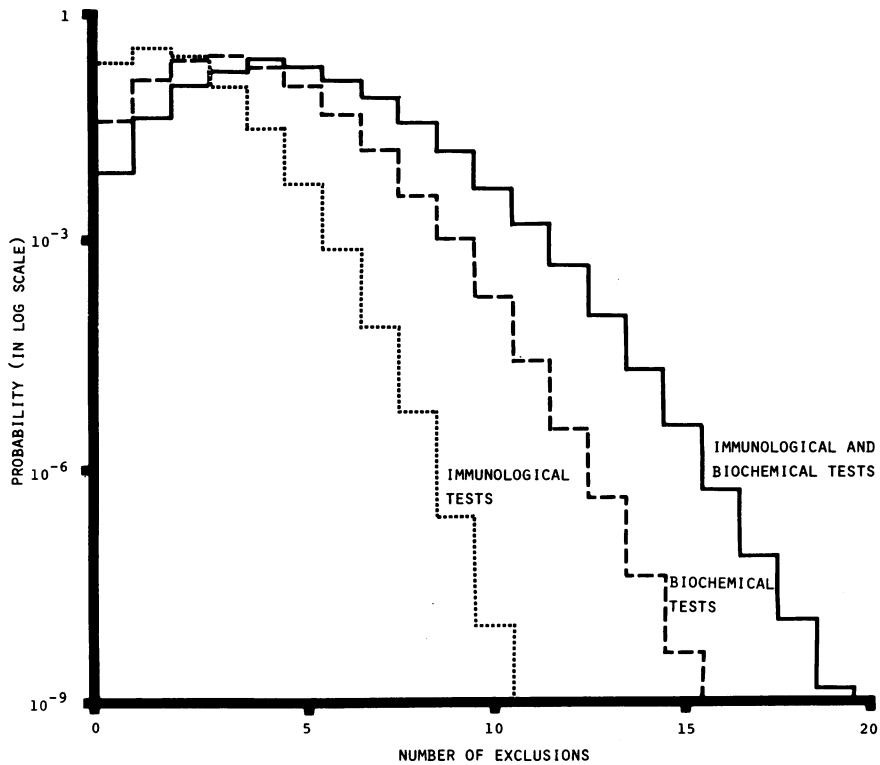

FIG. 1.—The probability distribution of the number of exclusions based on 24 immunological and 30 biochemical tests for United States white population.

clusions based on the exclusion probabilities for the individual systems in the United States white populations taken from reference [1]. The distributions are shown for 24 immunological, 30 biochemical, and 54 combined (biochemical and immunological) tests separately. The nature of the distributions for the other two major populations (black and Japanese) are also similar and, therefore, not presented here. It may be seen from the figure that the probability of obtaining a single exclusion for a random white male is .3629 for the 24 immunological tests, .1372 for the 30 biochemical tests, and .0416 for the 54 combined tests. Furthermore, if it is known that there is at least one exclusion observed from these tests, the above figures are increased by a certain extent since $\text{Prob}(X = k | X \geqq 1) =$

$\text{Prob}(X = k)/\text{Prob}(X \geqq 1)$. Use of this equation changes the above mentioned probabilities to .4546, .1426, and .0420, respectively.

One might wonder why the usual tests result generally in only one or two exclusions. This may be explained by computing the mean and variance of the number of exclusions, X. To do this let us define n random variables, $X_1, X_2, \ldots, X_n$ such that $\text{Prob}(X_i = 1) = p_i$, and $\text{Prob}(X_i = 0) = 1 - p_i$ for $i = 1, 2, \ldots, n$ (i.e., X_i takes value 1 if an exclusion is obtained on the basis of the ith system tested). Thus $X = \Sigma X_i$, and hence the mean and variance of X are given by

$$E(X) = \sum_i E(X_i) = \Sigma p_i,$$

and

$$V(X) = \sum_i V(X_i) = \Sigma p_i(1 - p_i),$$

where the summation Σ extends over all $i = 1, 2, \ldots, n$. The variance is obtained, of course, on the supposition that the genetic systems are all independent.

Using the same set of 24 immunological and 30 biochemical tests for the white population, the mean and variances are obtained as shown in table 1, which implies that it is reasonable to expect only a few exclusions in practice if the number of tests exceeds 50 as in these computations.

TABLE 1

MEAN AND VARIANCE OF THE NUMBER OF EXCLUSIONS

System	Mean	Variance
Immunological (24 systems)	1.390	1.222
Biochemical (30 systems)	2.862	2.399
Combined (54 systems)	4.348	3.538

SUMMARY

The distribution of the number of exclusions to be expected in paternity testing is derived, and the calculation of the mean and variance of this distribution is indicated.

REFERENCES

1. CHAKRABORTY R, SHAW M, SCHULL WJ: Exclusion of paternity: the current state of the art. *Am J Hum Genet* 26:477–488, 1974
2. CHAKRABORTY R, ROYCHOUDHURY AK: Paternity exclusion by genetic markers in Indian populations. *Indian J Med Res* 63:162–169, 1975
3. BUCHER KD, ELSTON RC: Estimation of nonpaternity for X-linked trait. *Am J Hum Genet* 27:689–690, 1975
4. WEINER AS: Chances of proving nonpaternity with tests for a X-linked trait. *Am J Hum Genet* 27:243–244, 1975
5. BIAS WB: Exclusion of paternity. *Am J Hum Genet* 27:245–246, 1975