Pamela L. Reading,[1] *Ph.D. and Emily G. Reisner,*[1] *Ph.D.*

The Effect of Differences in Gene Frequency on Probability of Paternity

REFERENCE: Reading, P. L. and Reisner, E. G., **"The Effect of Differences in Gene Frequency on Probability of Paternity,"** *Journal of Forensic Sciences*, JFSCA, Vol. 30, No. 4, Oct. 1985, pp. 1130–1139.

ABSTRACT: Knowledge of gene frequencies in populations is required for the calculation of probability of paternity. The question remains open as to the degree of accuracy of gene frequency estimates required to give accurate probability of paternity figures. This is of special concern in the HLA system, which has haplotype frequencies known to vary in populations. This paper presents computer simulation data comparing probability of paternity calculations using HLA data from California and North Carolina. Comparisons were made between geographic regions, and between blacks and whites within a geographic region. It was found that when the absolute probability of paternity is high, the average differences induced were small, but at lower probabilities the changes can be large. Differences were most pronounced between black and white populations. Examples of individual cases are given to illustrate the huge differences that can be induced in some cases by changing gene frequency.

KEYWORDS: forensic science, paternity, probability

Calculation of the probability of paternity demands a knowledge of the frequencies of genes in populations. The fact that these frequencies are known with some degree of accuracy allows a calculation of a probability of paternity. Without such data, we could only label men as "possible fathers" with no quantitative measure of the probability of their paternity. Unfortunately, our knowledge of human gene frequencies is less than complete. Although a great many large surveys have been done around the world, we cannot always be sure that the frequencies obtained apply to the population of interest. This is especially true in the United States, where wide differences in ethnic background can exist between adjacent areas. Race and ethnic group can also be problematic factors, since fewer studies have been done on black, Hispanic, and American Indian populations than on white populations. Frequency data for population mixtures and isolated groups can also be difficult to obtain.

Unless the trio (mother, child, and alleged father) in question is drawn from a well-studied population, a laboratory's gene frequency data will probably not be an extremely accurate estimate of the true frequencies in that population. The question then is how accurate must the gene frequencies be to give a reliable probability of paternity? Or, how large a difference in gene frequency is needed to change significantly the probability of paternity? A recent paper by Aiken [*1*] considers this question a major obstacle for meaningful calculations of probability of paternity.

Previous research [*2*] has shown that in simple, two allele systems, the probability of pater-

Received for publication 28 July 1984; revised manuscript received 29 Dec. 1984; accepted for publication 3 Jan. 1985.

[1]Statistical consultant and director of HLA Typing Laboratory, respectively, Department of Microbiology/Immunology, Duke University Medical Center, Durham, NC.

nity calculations are affected by simulated "errors" in gene frequency only when the frequency of one gene is very low. Hummel and Claussen [3] have explored the same problem in the red blood cell systems. They generated trios by computer and compared paternity probabilities as calculated using Southwest German frequencies (in the form of the Essen-Moller value) to those produced using the appropriate frequencies for various ethnic groups. They report that for most groups, the two Essen-Moller values fall within one "verbal predicate"[2] of each other. As expected, the degree of coincidence decreases as the distance from Germany, both geographically and ethnically, increases.

In the present paper, an attempt to explore this question is made via a computer simulation for the complex HLA antigenic system. HLA, when used for paternity diagnosis, can have very large effects on the cumulative probability of paternity. This is due to the large number of alleles, all of which are of very low frequency. These traits have induced some worries in the field [5] as to the accuracy of paternity calculations obtained from HLA testing. By examining the effect of gene errors on this case, we hope to obtain an estimate of the extent of change that errors in gene frequency in the HLA system can have on the final product.

Statistical Methods

To study paternity calculations with a complicated blood group system such as HLA, it is necessary to make some simplifying assumptions. While these assumptions will certainly not invalidate the conclusions made, they may force us to limit our application of the information obtained.

To make the HLA system manageable, consideration was limited to cases in which only one possible father is involved. Thus, theoretical "trios" of mother, child, and alleged father are the subjects. Further, the trios were limited to genotypes in which no homozygotes or "blank" phenotypes appear. Thus, all three subjects have four different and distinct antigens, two at the A locus and two at the B locus. Restricting the subjects to those having four clearly distinguishable antigens means that the true father's genetic contribution to the child is absolutely clear.

A few abbreviations have been used to symbolize the genotypes and haplotypes and their frequencies. The true father of the child must have one haplotype which he passed on to the child. Under the assumptions made, this haplotype is clearly indicated by the child's phenotype. We will call this haplotype, which the true father *must* possess, the *obligatory haplotype*, or OH. The other haplotype which the true father possesses, and which is unknown to us, is the *complementary haplotype*, or CH. To reiterate, the father's genotype is made up of the OH and the CH. We will symbolize the frequencies of these haplotypes in the general population by f(OH) and f(CH). By this symbolism, the frequency of the true father's genotype is the population is $2 \times f(OH) \times f(CH)$. We can ignore the homozygous case since our assumptions preclude it.

In our tests for the HLA system, however, we cannot detect the linkage relationships of the antigens. Therefore, for a man accused of paternity, we must calculate the probability that he has the correct *haplotypes*, given that his phenotype shows the required antigens. We do this by computing the percentage of people in the population with the same four antigens and who have the OH and CH arrangement instead of the opposite arrangement. Thus, for a man who has the antigens A1, A3, B5 and B7, the probability that he has the haplotypes A1, B5 and A3, B7 is:

$$\frac{2 \times f(A1\,B5) \times f(A3\,B7)}{2 \times f(A1\,B5) \times f(A3\,B7) + 2 \times f(A1\,B7) \times f(A3\,B5)} \tag{1}$$

[2]"Verbal predicates" are categorical guidelines designed to be used for paternity decisions. Originated by Hummel et al [4], they are assigned on the basis of the probability of paternity calculation, but are not based on statistical principles.

or, if we use the OH and CH symbolism:

$$\frac{2 \times f(OH) \times f(CH)}{2 \times f(OH) \times f(CH) + 2 \times f(A1\,B7) \times f(A3\,B5)} \tag{2}$$

This equation produces a frequency between zero and one. We will call this frequency the probability that a person has the obligatory haplotype given his phenotype, D. We can use this probability to calculate the possibility that a man of this phenotype fathered a child possessing the OH. Since the chance is 50% that a given haplotype is passed on to a child, the probability is 0.5 (D). To calculate the paternity index, comparing the alleged father's probability of fathering the child to that of a random man, we need the frequency of the OH in the population, or f(OH). The paternity index is:

$$PI = 0.5 \times \frac{D}{f(OH)} \text{ or } \frac{D}{2f(OH)} \tag{3}$$

This is the usual paternity index (PI), a likelihood ratio with no expression of prior probability included. Its derivation and use have been extensively discussed (for example, see Chakraborty [*6*]).

The true father in a particular case must have given the child the OH. The complementary haplotype the true father possesses is immaterial, and could be any one of the many haplotypes possible in the HLA system. Since the arrangement of antigens into haplotypes is unknown to us, the probability of paternity must depend on the frequencies of both haplotypes, and on the frequencies of the "opposite haplotypes," those composed of the same four antigens in opposite linkage relationship to the OH and CH. Using an inaccurate frequency for any of these four haplotype frequencies will result in an erroneous paternity calculation. If we acknowledge that our gene frequency estimates are always somewhat inaccurate, then we will be concerned with the degree to which this affects the paternity calculations we make.

We can select a haplotype as the obligatory one, the OH, and calculate the paternity indices for all men carrying that haplotype. That is, we study all possible combinations of the OH with other haplotypes (as the CH). This can be thought of as calculating the paternity index for all the possible fathers of a child. We can then use the mean of these figures as a gauge of the effects of changing the haplotype frequencies used. If we compare two sets of HLA frequency data and obtain mean paternity indices for various OHs, we can observe the effect of differing gene frequencies on the paternity index.

What can this tell us? By comparing two different sets of gene frequencies, we can tell how important it is to distinguish between them for paternity testing purposes. For example, if gene frequencies from Iowa and Nebraska give very different results, we should be cautious about assuming that Iowa data are sufficient for calculations about Nebraska trios. If black and white data sets give very different results, then extra caution must be taken in determining the racial identity of the trio.

It should be obvious from an examination of Eq 3 that the frequency of the OH and CH will have a large effect on the paternity index. In fact, if D (the probability of the person possessing the OH and CH given the correct phenotype) changes, the paternity index will change by the amount $\Delta D/2\,f(OH)$. The absolute change in the PI is thus highly dependent on the frequency of the obligatory haplotype f(OH). If f(OH) is very small, say 0.0005, then a change in D of only 0.005 results in a five-point change in the PI. However, if f(OH) is 0.05, then a change in D of 0.5 is needed to produce the same effect. Thus, if errors in haplotype frequency occur, their effect on the paternity index depends on the frequency of the obligatory haplotype as well as on the size of the error itself. In other words, a small change in a low gene frequency produces a large change in the PI.

The choice of an OH for our simulation will be crucial to the results we get. By choosing an OH which is very different in frequency between the two data sets, we can get a "worst case" look at the change in the paternity indices. By choosing an OH which is identical in frequency in the two data sets, we can gauge how much effect the differences in the other gene frequencies have on the paternity index.

Procedure

For this research, two sets of HLA gene frequencies were compared. One set is that collected by the Terasaki Laboratory at UCLA,[3] commonly used as "American" frequencies. The other set is that compiled by Reisner et al [8]. Significant differences between the two data sets were found for several antigens and were reported by Reisner et al [8].

We compared the UCLA data to the North Carolina data by race, and significant differences in haplotype frequencies were noted. A significantly different haplotype frequency is defined as one whose frequency as given by Reisner et al was outside the 95% confidence interval for the North Carolina data. For this purpose, any haplotypes not found in the North Carolina data were not considered. Fourteen haplotypes were found to fit these criteria in the white population data and twenty in the data on blacks. Additionally, twelve haplotypes which are significantly different in frequency between whites and blacks in the UCLA data were studied. In this comparison, three haplotypes which are very close in frequency in the black and white groups were also selected for study.

For each haplotype chosen as a model OH, all possible phenotypes that included it were generated. For each such phenotype, two possible OH frequencies were used, one from the UCLA data and one from the North Carolina data. All other frequencies in the equation were taken from the UCLA data. The paternity index was computed twice on the basis of the two frequencies available, and a probability of paternity calculated from the paternity index. We chose to express our results as probability of paternity rather than paternity index for these reasons: (1) the probability of paternity is the more generally used figure; and (2) large differences in paternity indices engender very small shifts as the probability approaches 100%. Thus, using a probability rather than a paternity index is a more conservative estimation of differences observed. In the black versus white comparisons, calculations were done using only data for the particular race. That is, calculations were performed independently as one would do for a black trio and for a white trio. The average probability of paternity generated by the various data sets were computed over the 220 possible genotypes generated.

Table 1 gives a sample calculation of the difference in Paternity index between North Carolina and UCLA for a man of phenotype A2, w31; B5, 13, assuming that A2, B5 is passed on to a child. The frequencies of the A2, B5 haplotype in both North Carolina and California are used to calculate the population frequency of persons carrying the correct complement of haplotypes. Since the UCLA data is used as a standard, only the UCLA frequency of the complementary haplotype A31, B13 is used. This example is the case which produces the maximum difference in probability of paternity between the two gene frequency data sets.

Tables 2 and 3 list the haplotypes used, the frequencies observed in North Carolina and California, and the change in the probability of paternity. Data for white and black UCLA populations are given in Table 4.

Tables 5 and 6 give the maximum *difference* in probability of paternity between North Carolina and UCLA values. This was detected during the calculation of the averages listed in Tables 2 and 3. Table 7 lists representative values for the pairing of A2, B5 with A1, B12 through A36, B13. Table 8 lists the number of probabilities which fell into certain ranges when A2, B5 was paired with all possible haplotypes.

[3]We have used an early edition of data dated 2 Sept. 1980. A modified version of the same data is included in the paper by Dykes [7].

TABLE 1—*Calculation of probability of paternity for alleged father A2, B5/A31, B13.*

Frequency of A2, B5 in NC	0.0115
Frequency of A2, B5 in UCLA data	0.0205
Frequency of A31, B13 in UCLA data	0.0001
Frequency of A2, B5/A31, B13 in population:	
NC	$2 \times (0.0115) \times (0.0001) = 0.0000023$
UCLA	$2 \times (0.0205) \times (0.0001) = 0.0000041$
But alleged father could also be A2, B13/A31, B5:	
Frequency in UCLA data	$2 \times (0.0061) \times (0.0047) = 0.00005734$
Probability of A2, B5/A31, B13 =	
NC	$0.0000023/(0.0000023 + 0.00005734) = 0.03856$
UCLA	$0.0000041/(0.0000041 + 0.00005734) = 0.06673$
Paternity index =	
NC	$0.03856/(2 \times 0.0205) = 0.9406$
UCLA	$0.06673/(2 \times 0.0205) = 1.6276$
Probability of paternity =	
NC	$0.9406/1.9406 = 48.4\%$
UCLA	$1.6276/2.6276 = 61.9\%$

TABLE 2—*Average change in the probability of paternity between North Carolina and UCLA white populations.*

Obligate Haplotype	Frequency of Obligate Haplotype		Resulting Average Probability of Paternity		
	NC	UCLA	NC	UCLA	Difference
A3, B15	0.0016	0.0060	96.9	98.1	3.9
A9, B40	0.0016	0.0086	93.7	97.0	3.3
A9, B35	0.0049	0.0142	92.5	95.1	2.6
A2, B35	0.0066	0.0139	91.4	93.8	2.4
A10, B12	0.0016	0.0061	95.1	97.5	2.4
A9, B7	0.0033	0.0143	91.3	95.2	2.1
A2, B15	0.0181	0.0291	91.6	93.2	1.6
A2, B5	0.0115	0.0205	91.5	93.0	1.5
A9, B12	0.0066	0.0147	92.2	94.5	1.3
A3, B40	0.0016	0.0051	96.8	98.1	1.3
A2, B40	0.0197	0.0316	90.2	91.3	1.1
A9, B21	0.0016	0.0056	97.2	98.2	1.0
A10, B27	0.0016	0.0051	97.7	98.4	0.7
A10, B16	0.0066	0.0129	96.3	96.8	0.5

Discussion

Examination of the data in Tables 2 and 3 shows that the differences observed between the North Carolina and California haplotype frequencies were quite sufficient to generate changes in the probability of paternity. As expected, the effect was most serious when the haplotype frequencies were small.

In the white population comparisons (Table 2), the probability changes by less than four percentage points, and in half of the cases by less than two percentage points. In the black pop-

TABLE 3—*Average change in the probability of paternity between North Carolina and UCLA black populations.*

Obligate Haplotype	Frequency of Obligate Haplotype		Resulting Average Probability of Paternity		
	NC	UCLA	NC	UCLA	Difference
A9, B35	0.0029	0.0199	81.9	92.7	10.8
A33, B35	0.0029	0.0182	89.1	94.6	5.5
A2, B35	0.0214	0.0359	87.5	89.4	1.9
A3, B35	0.0071	0.0161	93.0	94.8	1.8
A30, B35	0.0114	0.0210	91.1	92.9	1.8
A28, B35	0.0071	0.0155	93.4	95.0	1.6
A1, B21	0.0014	0.0049	97.4	98.4	1.0
A30, B42	0.0100	0.0253	93.6	94.4	0.8
A3, B7	0.0100	0.0185	94.2	95.0	0.8
A28, B7	0.0157	0.0018	99.5	98.7	0.8
A28, B12	0.0114	0.0030	98.9	98.3	0.6
A29, B12	0.0043	0.0100	96.7	97.3	0.6
A30, B5	0.0171	0.0062	98.3	97.7	0.6
A33, B5	0.0129	0.0028	99.3	98.8	0.5
A28, B5	0.0129	0.0041	98.9	98.5	0.4
A10, B17	0.0129	0.0041	98.8	98.4	0.4
A32, B7	0.0100	0.0016	99.6	99.2	0.4
A36, B5	0.0086	0.0016	99.6	99.4	0.2
A31, B5	0.0100	0.0009	99.8	99.6	0.2
A31, B14	0.0071	0.0003	99.9	99.8	0.1

TABLE 4—*Average change in the probability of paternity between black and white UCLA populations.*

Obligate Haplotype	Frequency of Obligate Haplotype		Resulting Average Probability of Paternity		
	White	Black	White	Black	Difference
A1, B8	0.0747	0.0127	85.4	97.0	11.6
A3, B7	0.0524	0.0185	88.7	95.0	6.3
A2, B15	0.0291	0.0041	92.5	98.7	6.2
A3, B35	0.0139	0.0359	95.4	89.3	6.1
A30, B42	0.0001	0.0253	99.9	94.4	5.5
A30, B7	0.0007	0.0188	99.7	94.3	5.4
A33, B35	0.0011	0.0182	99.7	94.6	5.1
A30, B17	0.0010	0.0149	99.7	95.4	4.3
A2, B7	0.0332	0.0175	90.0	94.2	4.2
A1, B17	0.0206	0.0042	94.9	98.4	3.5
A36, B35	0.0001	0.0130	99.9	96.4	3.5
A1, B37	0.0021	0.0024	98.4	99.4	1.0
A2, B14	0.0036	0.0036	97.8	98.6	0.8
A2, B17	0.0098	0.0121	96.4	95.7	0.7
A3, B35	0.0221	0.0161	94.9	94.8	0.1

ulation (Table 3), 14 of 20 haplotypes show changes of 1 percentage point or less. However, several haplotypes show very large changes, up to eleven percentage points.

In the North Carolina versus UCLA average comparisons, no results shifted the probability of paternity from below 95 to above 95%. However, in the comparisons of blacks and whites within the UCLA data set (Table 4), some shifts from above to below 95% probability were observed. For example, for the haplotype A1, B8, assuming the trio was black produced an aver-

TABLE 5—*Obligate and complementary haplotypes producing the maximum change in white population.*

Obligate Haplotype	Frequency of Obligate Haplotype		Complementary Haplotype	UCLA Frequency	Resulting Probability of Paternity		
	NC	UCLA			NC	UCLA	Difference
A9, B40	0.0016	0.0086	A31, B13	0.0001	42.6	79.1	36.5
A9, B7	0.0033	0.0143	A3, B13	0.0005	36.3	70.1	33.8
A10, B12	0.0016	0.0061	A29, B22	0.0003	41.3	72.4	31.1
A9, B21	0.0016	0.0056	A30, B15	0.0001	49.5	77.0	27.5
A3, B40	0.0016	0.0051	A31, B7	0.0030	57.7	80.8	23.1
A9, B35	0.0049	0.0142	A11, B21	0.0008	58.6	79.2	20.6
A9, B12	0.0066	0.0147	A29, B18	0.0004	37.7	57.0	19.3
A3, B15	0.0016	0.0060	A2, B14	0.0036	70.5	89.3	18.8
A2, B35	0.0066	0.0139	A33, B40	0.0001	40.1	58.0	17.9
A2, B5	0.0115	0.0205	A31, B13	0.0001	48.4	61.9	13.5
A10, B27	0.0016	0.0051	A11, B16	0.0011	81.0	92.5	11.5
A2, B40	0.0197	0.0316	A31, B13	0.0001	42.8	52.9	10.1
A10, B16	0.0066	0.0129	A33, B18	0.0001	83.9	90.0	6.1
A2, B15	0.0181	0.0291	A3, B13	0.0005	77.3	83.0	5.7

TABLE 6—*Obligate and complementary haplotypes producing the maximum change in black population.*

Obligate Haplotype	Frequency of Obligate Haplotype		Complementary Haplotype	UCLA Frequency	Resulting Probability of Paternity		
	NC	UCLA			NC	UCLA	Difference
A9, B35	0.0029	0.0199	A2, B42	0.0008	25.0	68.0	43.0
A33, B35	0.0029	0.0182	A36, B35	0.0003	31.1	72.3	41.2
A30, B5	0.0171	0.0062	A2, B42	0.0008	75.5	53.4	22.1
A1, B21	0.0014	0.0049	A2, B37	0.0004	68.1	87.6	19.5
A3, B35	0.0071	0.0161	A11, B17	0.0004	52.2	70.3	18.1
A28, B35	0.0071	0.0155	A36, B14	0.0003	55.3	72.1	16.8
A32, B7	0.0100	0.0016	A3, B22	0.0003	95.7	79.2	16.5
A29, B12	0.0043	0.0100	A2, B42	0.0008	65.6	80.8	15.2
A30, B35	0.0114	0.0210	A11, B42	0.0004	41.4	55.9	14.5
A28, B12	0.0114	0.0035	A29, B14	0.0003	91.7	78.1	13.6
A28, B7	0.0157	0.0018	A31, B35	0.0007	97.7	84.7	13.0
A2, B35	0.0214	0.0359	A33, B12	0.0025	53.9	65.0	11.1
A30, B42	0.0100	0.0253	A29, B14	0.0003	75.5	86.3	10.8
A3, B7	0.0100	0.0185	A31, B17	0.0005	70.0	80.1	10.1
A10, B17	0.0129	0.0041	A3, B22	0.0003	95.5	88.3	7.2
A28, B5	0.0129	0.0041	A11, B17	0.00042	95.9	89.6	6.3
A33, B5	0.0129	0.0028	A1, B15	0.0003	97.9	92.5	5.4
A36, B5	0.0086	0.0016	A32, B35	0.0007	98.7	94.8	3.9
A31, B14	0.0071	0.0003	A30, B40	0.0009	99.8	96.2	3.6
A31, B5	0.0100	0.0009	A30, B40	0.0009	99.5	96.0	3.5

age probability of paternity of 97.0%; the assumption of a white trio produced an average of 85.4%.

Examining the three cases (Table 4) in which the frequency of the OH was similar in both populations (A1, B37; A2, B14; A2, B17), we find, as we would expect, that the degree of change decreases. While differences still exist in the two probabilities, they are all one percentage point or less.

These results produce an interesting comparison to the work of Hummel and Claussen [*2*]. Using actual trio data for the red blood cell systems, they found very little difference in prob-

TABLE 7—*Probability of paternity when obligate haplotype is A2, B5 (white population).*

Complementary Haplotype	UCLA Frequency	Resulting Probability of Paternity		
		NC	UCLA	Difference
A31, B13	0.0001	48.4696	61.9425	13.4729
A31, B21	0.0003	64.3142	75.22954	10.9153
A11, B17	0.0008	65.0576	75.7978	10.7402
A31, B12	0.0023	65.0868	75.8200	10.7332
A28, B15	0.0008	65.1975	75.9042	10.7067
A11, B15	0.0020	65.8658	76.4105	10.5447
A33, B42	0.0001	67.0402	77.2917	10.2515
A31, B15	0.0011	67.3725	77.5391	10.1666
A11, B12	0.0051	67.6437	77.7404	10.0967
A11, B42	0.0001	67.8948	77.9261	10.0313
A30, B15	0.0001	68.6817	78.5053	09.8236
A11, B21	0.0008	70.5119	79.8340	09.3221
A36, B12	0.0002	71.1920	80.3213	09.1293
A36, B7	0.0001	71.6283	80.6322	09.0039
A36, B42	0.0001	72.5172	81.2612	08.7440
A11, B8	0.0010	72.6977	81.3882	08.6905
A31, B13	0.0001	73.3747	81.8625	08.4878
A28, B5	0.0038	73.9317	82.2504	08.3187
A36, B15	0.0001	73.9512	82.2639	08.3127
A11, B7	0.0037	74.8888	82.9117	08.0237
A31, B8	0.0006	75.4227	83.2778	07.8551
A32, B21	0.0003	75.4565	83.3009	07.8444
A3, B12	0.0065	75.9300	83.6239	07.6939
A33, B7	0.0002	78.2638	85.1934	06.9296
A28, B17	0.0007	78.4201	85.2972	06.8771
A11, B13	0.0009	79.1571	85.7846	06.6275
A1, B42	0.0001	79.7289	86.1602	06.4313

TABLE 8—*Breakdown for probability of paternity when A2, B5 is obligatory haplotype (white population).*

Probability Range	Number of Calculations for NC Data	Number of Calculations for UCLA Data
95.0 +	27	51
90.0-94.9	75	118
80.0-89.9	91	39
70.0-79.9	16	11
Less than 69.9	11	1
Total	220	220

ability of paternity when different population gene frequencies were used. Our results for the average probability of paternity agree with their observations, but we produced some very large differences in individual cases. This could be explained by the fact that we have chosen to work with the HLA system, a system most likley to show differences because of the large number of genes of very low frequency. Also, we chose to study cases in which the difference between populations was known to be large.

Perhaps the most important results of this research can be seen in Tables 5 and 6, listing the maximum difference in the probability of paternity which was observed in the simulation for each haplotype. Even though for the obligate haplotypes we chose the average probability of paternity is fairly high and the differences induced small, in individual cases the probability of

paternity can be quite low and the difference very great. For example, although the average probability of paternity for the A2, B5 haplotype in whites is 91.5% using the North Carolina frequency and 93.0% using the California frequency, the respective probabilities are only 48.4 and 61.9% when A2, B5 is paired with A31, B13. While examination of Tables 7 and 8 makes clear that this is an isolated case, it is of concern to workers in the field of paternity testing that such large differences can occur.

Conclusion

We can draw several conclusions from the simulations performed. The most important haplotype frequency to consider when frequency data are of questionable accuracy is the one required to be passed to the child from the father. If the frequency of the obligatory haplotype is changed significantly, the probability of paternity must change. When the average probability of paternity is high, a change in the obligatory haplotype frequency does not significantly change that probability but at lower levels of the average probability of paternity, especially below the 90% level, a change in the frequency of the obligatory haplotype can change the probability of paternity by up to twelve percentage points in our study. Thus if the obligatory haplotype in a particular case is known to vary significantly in different populations, the probability of paternity calculation may be less dependable. Even if the obligatory haplotype frequency is stable in different populations, small variations in the probability of paternity may occur due to the different frequencies in the "background" population data. Given the small size of the variations observed in this study, this is not likely to be of major concern.

It *must* be remembered that these conclusions are based on averages over many data points. Examination of the maximum differences observed shows that tremendous changes in the probability of paternity can be brought about by gene frequency errors. It is always advisable if there is reason to suspect that other gene frequencies may be applicable to calculate the paternity indices, giving all probable frequencies. For example, if questions exist on the race of the father, calculations should be made assuming both white and black gene frequencies.

It is also important to note the simplifying assumptions which were made for this work. These conclusions, while fairly reassuring, may not hold in other cases. While including blank antigens and homozygotes does not present any inherent reasons for altering the results we have obtained, it does present the possibility of further complication such as inbreeding.

It can be concluded that erroneous estimates of gene frequency can have severe effects on the calculation of a probability of paternity when using the HLA system. While this research has examined what might be thought of as a worst case situation, the UCLA population data are occasionally used as "representative" American values, and we have shown that North Carolina populations are quite different. The results also advise us to use extreme caution in the assignment of race/ethnic group to subjects being tested, since the difference between black and white probabilities can be very large for certain haplotypes.

References

[1] Aiken, M., "Some Fallacies in the Computation of Paternity Probabilities," *American Journal of Human Genetics*, Vol. 36, 1984, pp. 904-915.

[2] Reading, P. L., *Problems in the Attribution of Paternity*. doctoral dissertation, University of North Carolina, Chapel Hill, NC, 1982.

[3] Hummel, K. and Claussen, M., "Exclusion Efficiency and Biostatistical Value of Conventional Blood Group Systems in European and Non-European Populations; Suitability of Central European Tables for Non-German Speaking Populations," in *Biomathematical Evidence in Paternity*, K. Hummel and J. Gerchow, Eds., Springer-Verlag, Berlin, 1981, pp. 97-108.

[4] Hummel, K., Ihm, P., and Schmidt, V., *Biostatistical Opinion of Parentage*, Table Part 1, Fischer, Stuttgart, 1971.

[5] Selvin, S., "Some Statistical Properties of the Paternity Ratio," in *Inclusion Probabilities in Paternity Testing*, R. H. Walker, Ed., American Association of Blood Banks, Arlington, VA, 1983, pp. 77-85.
[6] Chakraborty, R. and Ferrell, R. E., "Paternity Index and Its Use in Paternity Diagnosis," in *Inclusion Probabilities in Paternity Testing*, R. H. Walker, Ed., American Association of Blood Banks, Arlington, VA, 1983, pp. 115-122.
[7] Dykes, D. D., "The Use of Frequency Tables in Parentage Testing, Appendix A," in *Probability of Inclusion in Paternity Testing*, H. Silver, Ed., American Association of Blood Banks, Arlington, VA, 1982, pp. 15-44.
[8] Reisner, E., Dawson, D. V., Ciftan, E. A., and Heise, E., "Studies of Random Black and White Populations in North Carolina," *Tissue Antigens*, Vol. 21, No. 1, Jan., 1983, pp. 14-23.

Address requests for reprints and additional information to
Emily G. Reisner, Ph.D.
Director, HLA Typing Laboratory
Duke University Medical Center
Box 3934
Durham, NC 27710