J Forensic Sci, Jul. 1978, Vol. 23, No. 3

B. W. Grunbaum,[1] *Ph.D., M.Crim.; S. Selvin,*[2] *Ph.D.; N. Pace,*[3] *Ph.D.; and D. M. Black,*[1] *B.A.*

Frequency Distribution and Discrimination Probability of Twelve Protein Genetic Variants in Human Blood as Functions of Race, Sex, and Age

The identification of large numbers of genetic variants for antigens, enzymes, and other proteins, combined with knowledge of their frequencies of occurrence in the population, enhances the ability to detect the individuality of a given blood specimen. Furthermore, the greater the number of variant protein systems considered, the greater is the likelihood of discriminating between individuals. As new techniques become available, forensic scientists will have increasing need for statistical data to use in attempting to establish the individuality of a blood specimen.

The probability always exists that two randomly selected blood samples are identical with respect to a series of genetic phenotypes by chance alone. This probability is often called the matching probability and will be denoted by P. The quantity $1 - P$ is often called the discrimination probability. The quantity P is typically calculated by multiplying together the probabilities based on the frequency of a series of genetic variants; that is, $P = [P_1P_2P_3 \cdots P_n]$, where P_i is the sum of the squared phenotypic frequencies of each variant [*1*]:

$$P_i = \sum_{j=1}^{k} p_j^2 \tag{1}$$

where k is the number of different phenotypes within the ith blood group system and p_j is the phenotypic frequency of the jth phenotype.

The magnitude of P is often of critical importance. Equally important is the validity of the process that produces the discrimination probability. Three assumptions intrinsic to the calculation of P were investigated for a set of twelve protein genetic variants for four ethnic groups (white, black, Chicano/Amerindian, and Asian). The twelve genetic variants

The preparation of these materials was financially assisted through a federal grant from the Law Enforcement Assistance Administration and the California Office of Criminal Justice Planning and under Title I of the Crime Control Act of 1973. The opinions, findings, and conclusions in this publication are those of the author and not necessarily those of OCJP or LEAA, and OCJP and LEAA reserve a royalty-free, nonexclusive, and irrevocable license to reproduce, publish, and use these materials, and to authorize others to do so. A copy of these materials may be obtained from OCJP or LEAA upon payment of the cost for reproducing the materials. Received for publication 16 Sept. 1977; revised manuscript received 19 Oct. 1977; accepted for publication 31 Oct. 1977.

[1] Research biochemist and research assistant, respectively, White Mountain Research Station, University of California, Berkeley, Calif. 94720.

[2] Associate professor, School of Public Health, University of California, Berkeley, Calif. 94720.

[3] Professor of physiology, Emeritus, University of California, Berkeley, Calif. 94720.

TABLE 1—*Genetic systems investigated.*

Variant	Abbreviation	Total Number	Substrate	Sample Size, μl	Completion Time, min	Method of Visualization	Method Reference
Autosomal, by blood typing							
ABO-system	ABO	11 678	...	...	...	...	...
Rhesus-system	Rh	11 676	...	...	...	...	...
Autosomal, by electrophoresis							
Phosphoglucomutase	PGM	11 626	cellulose acetate	0.50	60	formazan	*6*
Adenylate kinase	AK	10 312	cellulose acetate	0.25	30	formazan	*4*
Adenosine deaminase	ADA	9 891	cellulose acetate	0.25	30	formazan	*4*
Erythrocyte acid phosphatase	EAP	9 547	cellulose acetate	0.50–1.0	40	4 MUP[a]	*7*
Esterase D	EsD	10 959	cellulose acetate	1.0	30	4 MUA[b]	*8*
Hemoglobin	Hb	11 678	cellulose acetate	0.25	20	*O*-dianisidine	*3* and *4*
Haptoglobin	Hp	935	acrylamide gel	1.0–3.0	180	*O*-dianisidine	*9*
Gc system	Gc	9 775	cellulose acetate	0.25	40	IMF[c] + Ponceau	*10*
6-Phosphogluconate dehydrogenase	PGD	9 647	cellulose acetate	0.50	30	formazan	*4*
Sex-Linked							
Glucose-6-phosphate dehydrogenase	G-6-PD	8 650 male, 2 834 female	cellulose acetate	0.25	30	formazan	*11*

[a] 4-Methyl umbelliferyl phosphate. The zones are visualized as the fluorescence of 4-methyl umbelliferone.
[b] 4-Methyl umbelliferyl acetate. The zones are visualized as the fluorescence of 4-methyl umbelliferone.
[c] Immunofixation.

are listed in Table 1 along with the total number of individual determinations. The three assumptions are as follows:

1. The phenotypic frequencies are the same in both males and females.
2. The phenotypic frequencies do not vary with age of the donor.
3. The phenotypic frequencies are statistically independent, since the multiplication of probabilities used in calculating the discrimination probability P is only valid for independent frequencies.

Verification of these three assumptions permits accurate estimation of twelve phenotypic and gene frequencies for four ethnic groups when all possible data are used. These frequencies are then used to determine the discrimination probability.

Sample Collection

Within the past year, this laboratory has undertaken a large-scale study, funded by the California Office of Criminal Justice Planning, to establish statistical data on a number of genetic variants in human blood. Fresh blood specimens from donors were provided by blood banks in sealed segments from blood bags. The segments were identified by a code to protect the anonymity of the donors. Information concerning the age, sex, ethnic group, and ABO and Rh blood types of the donor as determined by the blood bank accompanied each sample.

All blood samples were to be collected within California; however, within a few months it was found that some of the ethnic subgroups were infrequent blood donors. It became apparent that a statistically valid sampling of blood from Asians, particularly females, could not be obtained from California blood banks within a reasonable time period. Consequently, samples from Asian donors were obtained from the Blood Bank of Hawaii in Honolulu to supplement those from California blood banks.

Method

The contributing blood banks typed each sample for the ABO system and for positive or negative Rh. This laboratory then analyzed each sample for up to ten additional genetic variants (Table 1), utilizing a standardized electrophoretic technology that has been under development here for several years [*2–4*].

All analytical methods, except for determination of haptoglobin (Hp), utilized cellulose acetate membranes as a substrate in the electrophoretic procedures. Each method required 20 to 60 min for completion.

Acrylamide gel was used as the substrate for Hp electrophoresis. Determination of Hp with acrylamide gel as a substrate required 180 min for completion. The method is also relatively expensive, and it was primarily for this reason that fewer Hp determinations were carried out.

Results

Basic to the estimation of phenotypic or gene frequencies is the assumption that these frequencies do not differ between males and females for autosomal genetic systems. The present study afforded the opportunity to test this assumption for the eleven autosomal genetic systems examined.

The usual χ^2 test for homogeneity was used [*5*]. Only the analysis of the ABO system among whites is given in detail in Table 2. The significance probability greater than 0.05 indicates that there is no difference in ABO frequencies between males and females. The other 43 sex analyses are summarized in Table 3 and show no association between sex

TABLE 2—*Analysis of ABO system among whites. The expected numbers are in parentheses. χ^2 = 6.97; degrees of freedom = 3; and significance probability = 0.073.*

	Phenotype				
Sex	A	B	AB	O	Total
Male	1439 (1400.0)	472 (471.3)	161 (159.4)	1836 (1881.8)	3908
Female	712 (750.9)	252 (252.7)	77 (83.1)	1055 (1009.3)	2096
Total	2151	724	238	2891	6004

TABLE 3—*Significance levels of χ^2 tests conducted to test for differences between phenotypic frequencies in males and females for four ethnic groups.*

Genetic Variant	White *P* Value	Black *P* Value	Chicano/ Amerindian *P* Value	Asian *P* Value
ABO	0.073	0.753	0.650	0.118
Rh	0.394	0.735	0.953	0.193
PGM	0.904	0.626	0.360	0.977
AK	0.624	0.963	0.094	0.970
ADA	0.332	0.693	0.081	0.146
EAP	0.904	0.893	0.387	0.753
EsD	0.567	0.779	0.813	0.546
G-6-PD[a]	...	...	...	...
Hb	0.995	0.733	0.529	0.999
Hp	0.456	0.476	0.876	0.393
Gc	0.653	0.554	0.421	0.068
PGD	0.497	0.570	0.717	0.778

[a] Sex-linked, therefore known to differ.

and phenotype. Since no systematic difference exists between males and females, the data may be combined to yield estimates with increased precision.

On the other hand, glucose-6-phosphate dehydrogenase (G-6-PD) is a sex-linked variant and is expected to have different phenotypic frequencies between males and females which contribute differently to the computation of the discrimination probability.

Like the sex of the individuals sampled, the relationship between age and phenotypic frequencies is also of potential importance in estimating valid proportions of genetic variants. If the age distribution of a population affects the distribution of phenotypic frequencies, then estimates made from combining all age groups will not have a clear-cut meaning. Although it is not expected that age would be highly associated with changes in phenotypic frequencies, it is possible that small effects could be detected with the large number of individuals in the present data set. The small effects might come, for example, from different degrees of racial admixture that depend on age within the four ethnic groups recorded, or, perhaps, the effects of differential migration by age could cause phenotypic frequencies to differ with age. In order to investigate the possible relationship between age and phenotypic frequency the mean age of the donors was computed for each phenotype within a system for each of the four ethnic groups. The mean values were compared by an analysis of variance (one-way classification) to test for possible differences in age associated with specific phenotypes. Again, only the ABO system for whites is given in detail (Tables 4 and 5). The frequency f value being equal to 1.85 implies that the differences in mean age are not statistically different. The significance probability is equal to

TABLE 4—*Analysis of ABO system among whites.*

Phenotypes	Mean Age	Standard Deviation	Sample Size[a]
A	36.67	12.25	2147
B	37.14	11.97	724
AB	35.34	12.38	238
O	36.25	12.35	2887

[a] There were eight cases without age recorded.

TABLE 5—*Analysis of variance for ABO system among whites.*

Source	Degrees of Freedom	Mean Square	*f*
Between phenotypes	3	278.91	1.85
Within phenotypes	5992	150.49	...
Total	5992	...	...

0.136. Inasmuch as it is higher than 0.05, it may be concluded that age is not a factor in determining phenotypic frequencies for the ABO system among whites.

The other 51 age analyses are summarized by reporting only the significance probabilities in Table 6. Even with the resolution power of this data set, no consistent relationship between age and phenotypic frequencies emerges.

Since the data collected showed little evidence of an effect from sex or age of the individual on the phenotypic frequencies, the data were combined to give a series of estimates of phenotypic frequencies, as shown in Table 7. Also included are the maximum likelihood estimates of the gene frequencies.

Some variants showed sizable differences in phenotypic frequencies among the four ethnic groups (for example, the Rh system) and other frequencies show less notable differences (such as phosphoglucomutase). Nevertheless, all differences were highly statistically significant ($P < 0.001$) for the twelve genetic variants. Even the small differences, as in

TABLE 6—*The significance probabilities associated with a one-way analysis of variance for mean age assorted into genetic variant classes for four ethnic groups.*

Genetic Variant	White *P* Value	Black *P* Value	Chicano/ Amerindian *P* Value	Asian *P* Value
ABO	0.136	0.614	0.088	0.176
Rh	0.544	0.103	0.131	0.560
PGM	0.418	0.083	0.228	0.118
AK	0.228	0.261	0.375	0.227
ADA	0.081	0.403	0.384	0.511
EAP	0.097	0.340	0.483	0.055
EsD	0.570	0.209	0.928	0.934
G-6-PD (male)	0.110	0.217	0.923	0.790
G-6-PD (female)	0.456	0.916	0.969	0.060
Hb	0.839	0.275	0.862	0.585
Hp	0.544	0.319	0.754	0.111
Gc	0.382	0.077	0.083	0.545
PGD	0.558	0.636	0.882	0.277

TABLE 7—*Phenotypic frequencies (percent)[a] and gene frequencies of twelve blood group variants for four ethnic categories.*

ABO	A	B	AB	O	Rare	Gene A	Gene B	Gene O	*n*
White	35.8	12.1	4.0	48.2	0.0	0.224	0.083	0.692	6004
Black	25.9	21.5	5.0	47.7	0.0	0.168	0.142	0.690	1025
Chicano/Amerindian	30.9	10.8	1.9	56.5	0.0	0.180	0.065	0.755	1596
Asian	37.8	21.0	8.8	32.5	0.0	0.338	0.068	0.594	3053
Rhesus	Rh^+	Rh^-	Rare	Gene Rh^+	Gene Rh^-	*n*			
White	79.2	20.8	0.0	0.543	0.457	6004			
Black	92.6	7.4	0.0	0.728	0.272	1024			
Chicano/Amerindian	91.9	8.1	0.0	0.716	0.284	1596			
Asian	99.0	1.0	0.0	0.903	0.097	3052			
PGM	1-1	1-2	2-2	Rare	PGM^1	PGM^2	*n*		
White	58.9	35.6	5.4	0.1	0.768	0.232	5972		
Black	66.2	29.5	4.0	0.3	0.812	0.188	1024		
Chicano/Amerindian	58.7	34.7	6.2	0.4	0.764	0.236	1586		
Asian	59.0	35.0	5.6	0.4	0.769	0.231	3044		
AK	1-1	1-2	2-2	Rare	AK^1	AK^2	*n*		
White	92.7	7.1	0.1	0.1	0.963	0.037	5969		
Black	98.4	1.6	0	0	0.992	0.008	965		
Chicano/Amerindian	95.6	4.3	0.1	0	0.978	0.022	1344		
Asian	99.8	0.2	0	0	0.999	0.001	2304		
ADA	1-1	1-2	2-2	Rare	ADA^1	ADA^2	*n*		
White	90.0	9.8	0.2	0	0.949	0.051	5883		
Black	97.8	2.2	0	0	0.898	0.011	927		
Chicano/Amerindian	93.8	5.9	0.3	0	0.964	0.036	1260		
Asian	95.2	4.6	0.2	0	0.975	0.025	1821		
EAP	AA	BA	BB	CA	CB	CC	Rare		
White	10.8	42.1	39.3	3.3	4.3	0.2	0.0		
Black	5.6	31.4	60.2	0.2	1.3	0.1	1.1		
Chicano/Amerindian	6.7	35.8	53.5	1.6	2.2	0	0.2		
Asian	5.2	35.6	59.2	0.0	0	0	0		
EAP		Gene *A*	Gene *B*	Gene *C*	*n*				
White		0.332	0.630	0.038	4850				
Black		0.217	0.776	0.008	875				
Chicano/Amerindian		0.254	0.726	0.019	1360				
Asian		0.230	0.770	0.000	2462				
EsD	1-1	1-2	2-2	Rare	EsD^1	EsD^2	*n*		
White	79.5	19.3	1.2	0.0	0.892	0.108	5377		
Black	83.6	16.0	0.4	0	0.916	0.084	973		
Chicano/Amerindian	73.9	23.8	2.3	0	0.858	0.142	1580		
Asian	41.6	44.2	14.2	0	0.637	0.363	3029		
G-6-PD (male)	B	A	A^-	Rare	Gene *B*	Gene *A*	*n*		
White	99.5	0.4	0	0.1	0.995	0.005	3845		
Black	73.0	19.4	7.3	0.3	0.730	0.270	896		
Chicano/Amerindian	97.6	1.7	0.3	0.4	0.976	0.024	1229		
Asian	98.9	0.1	0.0	0.9	0.989	0.011	2680		
G-6-PD (female)	B	A	AB	A^-	Rare	Gene *B*	Gene *A*	*n*	
White	99.6	0.1	0.3	0	0	0.997	0.003	2071	
Black	63.1	9.9	23.4	0	1.8	0.776	0.224	111	
Chicano/Amerindian	97.1	0.3	0.3	1.8	2.3	0.995	0.005	310	
Asian	99.4	0	0.3	0	0.3	0.998	0.002	342	
Hb	A	AS	AC	Rare	Gene *A*	Gene *S*	*n*		
White	99.8	0.2	0	0.0	0.999	0.001	6004		
Black	89.3	8.6	1.8	0.4	0.956	0.044	1025		
Chicano/Amerindian	99.6	0.1	0.2	0.1	0.999	0.001	1596		
Asian	99.9	0.0	0.0	0.1	0.999	0.001	3053		

TABLE 7—*Continued*

Hp	1-1	1-2	2-2	Rare	Hp^1	Hp^2	*n*
White	14.6	49.3	36.1	0	0.392	0.608	274
Black	33.9	41.9	21.0	3.2	0.567	0.433	124
Chicano/Amerindian	21.7	55.9	22.4	0	0.497	0.503	161
Asian	7.7	37.8	53.5	1.1	0.269	0.731	376
Gc system	1-1	1-2	2-2	Rare	Gc^1	Gc^2	*n*
White	50.2	41.0	8.5	0.4	0.710	0.290	4488
Black	74.5	21.0	2.0	2.4	0.871	0.129	832
Chicano/Amerindian	59.1	35.3	5.2	0.5	0.771	0.229	1417
Asian	50.4	39.2	6.4	4.0	0.729	0.271	3043
				Gene	Gene		
PGD	A	AC	Rare	*A*	*C*	*n*	
White	96.2	3.7	0.0	0.981	0.019	4472	
Black	92.6	7.2	0.4	0.964	0.036	787	
Chicano/Amerindian	94.6	5.2	0.1	0.974	0.026	1494	
Asian	86.8	12.6	0.6	0.937	0.063	2894	

[a] 0 = no cases and 0.0 = less than 0.1%.

the case of the adenosine deaminase (ADA) system (maximum difference less than 7.1%), were statistically significant because of the high resolution power of the data set employed. The difference in phenotypic frequencies means that the computation of the discrimination probabilities must be carried out separately for each ethnic group.

As mentioned before, the valid computation of the discrimination probability assumes that the genetic systems are statistically independent. All possible comparisons of pairs of the 13 genetic systems (including the sex-linked G-6-PD for males and females) were tested for independence. This process produced 77 contingency tables for each of the four ethnic groups. Each of the 308 tables was evaluated with a χ^2 test for independence and a contingency coefficient. A contingency coefficient is a measure parallel to a correlation coefficient that summarizes the degree of association in a table ($C = 0$ implies perfect independence and $C = 1$ implies perfect dependence). Once again the ABO system among whites has been used to demonstrate the analysis employed (Table 8).

The other 307 analyses are summarized in Table 9, where only the contingency coefficients and the significant probabilities are reported. No consistent associations were demonstrated. These few showed some degree of association, but these differences were not consistent for all four ethnic groups. In light of the large amount of data available and the lack of consistency of the few significant associations found, it may be concluded that these 13 genetic variants behave statistically independently. The group specific component (Gc) showed significant association in five situations. The Gc system is associated with

TABLE 8—*Analysis of the EsD-ABO pair among whites. Expected numbers* [5] *are in parentheses. Contingency coefficient = 0.041; degrees of freedom = 6;* χ^2 *= 9.20; and significance probability = 0.162.*

	Phenotype				
EsD	A	B	AB	O	Total
EsD 1-2	1453 (1547.5)	496 (517.7)	181 (171.0)	2055 (2038.9)	4275
EsD 1-2	382 (375.7)	148 (125.7)	32 (41.5)	476 (495.1)	1038
EsD 2-2	21 (22.8)	7 (7.6)	2 (2.5)	33 (30.0)	63
Total	1946	651	215	2564	5376

TABLE 9—*Contingency coefficient and χ^2 significance probability reflecting degree of association (disequilibrium) among all possible pairs of 13 genetic variants (* = 0.01 < P < 0.05; ** = 0.01 < P < 0.001; and *** = P < 0.001).*

Variant Pair	Race	Contingency Coefficient	Significance Probability	Major Association
ABO by Rh	white	0.045	0.006**	excess of B and Rh^-
ABO by EAP	white	0.080	0.002**	excess of B and AA
ABO by ADA	Chicano/ Amerindian	0.082	0.037*	excess of B and ADA (1-2)
Rh by EsD	Chicano/ Amerindian	0.079	0.007**	excess of Rh^- and EsD (1-1)
Rh by EsD	Asian	0.048	0.031*	excess of Rh^- and EsD (1-1)
PGM by Hb	black	0.079	0.043*	excess of PGM (1-1) and Hb (AS)
PGM by Hp	black	0.292	0.024*	excess of PGM (1-2) and Hp (1-2)
EAP by Gc	Chicano/ Amerindian	0.120*	0.028*	lack of Gc (1-1) and EAP (CA)
EAP by Gc	white	0.061	0.049*	lack of CB and Gc (1-1)
AK by Gc	Chicano/ Amerindian	0.108	0.001***	excess of AK (1-2) and Gc (2-2)
6-P-GD by Gc	black	0.136	0.002**	excess of AC and Gc (2-2)
6-P-GD by Gc	Asian	0.050	0.034*	excess of AC and Gc (1-2)

erythrocyte acid phosphatase (EAP) systems for both whites and Chicano/Amerindians, the ABO system in Chicano/Amerindians, and the PGD system in both blacks and Asians. Nevertheless, the contingency coefficients are rather small and no substantial error would be incurred by assuming the Gc system to be statistically independent from the other eleven systems investigated.

Discussion

Since each of 308 pairs of genetic variants has been shown to be statistically independent, the phenotypic frequencies in Table 7 may be multiplied together to give an estimate of the discrimination probability.

Table 10 shows the probabilities that two randomly selected males have identical phenotypes for the four ethnic groups. Also included is the overall probability that two random males have the same phenotypes for all twelve genetic variants. The value of P varies slightly for the four ethnic categories. The same calculation for females yields virtually the same results (Table 11), the only difference being the use of the sex-linked G-6-PD system.

Clearly, the probability that two randomly selected individuals have the same phenotype for a given genetic system is a function of the frequency of that phenotype in the general population. For example, if all individuals in the general population have the same phenotype, the probability that any two are phenotypically identical is 1.0. On the other hand, there exists a theoretical minimum probability that two randomly selected individuals will be identical for a specific phenotype. This probability occurs when all phenotypes for a specific genetic system are equally frequent in the population. Mathematically, this probability is expressed as $1/k$, where k represents the number of variants within a genetic system. These theoretical minimum probabilities are given in Table 10.

The theoretical minimum is not achieved in reality but does give some idea of the efficacy of various genetic systems in discriminating between two randomly selected individuals. That is, a probability of 1.0 represents the worst possible situation for the purposes of

TABLE 10—*The probabilities that two randomly selected male individuals have the same phenotypes for four ethnic groups.*

System	White	Black	Chicano/ Amerindian	Asian	Theoretical Minimum
ABO	0.377	0.343	0.426	0.330	0.250
Rh	0.670	0.863	0.851	0.980	0.500
PGM	0.477	0.527	0.469	0.474	0.333
AK	0.864	0.969	0.916	0.996	0.333
ADA	0.820	0.957	0.883	0.908	0.333
EAP	0.346	0.464	0.419	0.480	0.167
EsD	0.632	0.725	0.603	0.388	0.333
G-6-PD (male)	0.990	0.576	0.953	0.978	0.500
Hb	0.996	0.805	0.992	0.998	0.333
Hp	0.395	0.335	0.410	0.435	0.333
Gc	0.427	0.600	0.477	0.413	0.333
PGD	0.927	0.862	0.898	0.793	0.333
P (identical) = P	0.0029	0.0039	0.0058	0.0036	...
P (not identical) = 1 − P	0.9971	0.9961	0.9942	0.9964	...

TABLE 11—*The probabilities that two randomly selected female individuals have the same phenotypes for four ethnic groups (that is,* P_i).

	White	Black	Chicano/ Amerindian	Asian
G-6-PD (female)	0.992	0.463	0.943	0.988
P (identical) = P	0.0029	0.0068	0.0061	0.0036
P (not identical) = 1 − P	0.9971	0.9932	0.9939	0.9964

discrimination. As the observed phenotypic frequency approaches the theoretical minimum the efficacy of the system for discrimination approaches maximum utility; that is, a genetic system with variant frequencies $1/k$ is the best possible discriminator. For example, the Gc system is fairly effective (P_{10} = 0.431 compared to the theoretical minimum of 0.333 for whites), whereas the hemoglobin (Hb) system is not effective in discrimination (P_8 = 0.996 versus the theoretical minimum of 0.333 for whites).

It should be noted that the overall ability to discriminate always increases with each additional genetic system employed. The twelve genetic systems used here yield the specific discrimination probabilities given in Tables 10 and 11. The addition of other genetic systems will improve the overall discrimination probability, particularly if genetic systems with variant frequencies close to the theoretical minimum of $1/k$ are chosen.

The discrimination probabilities derived here strictly apply only in California for whites, blacks, and Chicano/Amerindians, and in Hawaii for Asians, since these were the sampled populations. However, it is likely that these results generalize to most of the United States population with minor variation. This minor variation stems from the fact that samples from other parts of the United States may have phenotypic frequencies that are slightly different for the four major California-Hawaii groups. However, the broad demonstration that phenotypic frequencies do not significantly vary with age or sex but do vary with ethnicity, as well as the demonstration of statistical independence among different genetic systems, will undoubtedly apply in a great variety of situations.

Up to this point the discussion has focused on the discrimination probability 1 − P,

which is a relative measure of how well a set of genetic systems discriminates between two randomly chosen individuals. An equally important quantity is the probability that a randomly chosen individual of a given ethnic group possesses the same blood phenotypes as found in a predetermined sample of blood. This probability is simply the product of the phenotypic frequencies of the phenotypes considered in the predetermined sample and will be called the probability of coincidence (denoted by C). Specifically, if p_1 is the frequency of phenotype 1, p_2 is the frequency of phenotype 2, and p_3 is the frequency of phenotype 3, then the probability that a randomly chosen person possesses these three phenotypes is $C = p_1p_2p_3$. In general, the probability of matching a set of predetermined phenotypes is $C = p_1p_2 \cdots p_n$, where p_i is the phenotypic frequency of each variant for a set of n systems (see Eq 1).

The quantity $1 - C$ is the probability that an individual does not match the given set of genetic variants and is sometimes called the probability of exclusion.

The data in Table 7 can be used to calculate the probability C for any given set of the twelve genetic systems considered for any given ethnic group. For example, the probability that a randomly chosen white male from California matches the most common set of phenotypes among whites is computed as follows: for type O (ABO system), positive (Rh system), 1-1 (PGM system), and so forth, p_1 is then the frequency of type O (0.482), p_2 is the frequency of Rh^+ (0.792), p_3 is the frequency of PGM 1-1 (0.589), and so forth, and $C = p_1p_2p_3 \cdots p_{12} = (0.482)(0.792)(0.589) \cdots (0.962) = 0.015$.

In an investigation, when it is desired to know the probability that a randomly chosen person will or will not match a blood sample obtained in connection with a civil or criminal case, the process is the same. The probability that a random individual matches all the genetic variants found in a specific blood sample is the product of the phenotypic frequencies associated with each genetic variant in the given blood sample. It should be noted that this probability of matching applies only to the population for which the phenotypic frequencies p_i were estimated.

The fact that there is considerable difference among the phenotypic frequencies for the four ethnic groups studied here raises the possibility that the race of an individual may be predicted from a sample of blood. This question is under investigation and will be reported later.

It is not generally possible to apply the probability of coincidence to reflect on the guilt or innocence of a person in a courtroom setting. A defendant is not usually chosen at random with respect to phenotype since a trial is, in many cases, conditional on the outcome of a blood test. A trial will be conducted usually when the accused does match the given blood sample. A person would rarely stand trial if a genetic exclusion can be demonstrated. The process of excluding those individuals who do not match, and sending to trial only those who do match with respect to a series of genetic variants, means that most courtroom situations represent a selected (and therefore not random) population and the probability of coincidence no longer applies.

Summary

Fresh blood samples were obtained from 6004 whites, 1025 blacks, 1596 Chicano/Amerindians, and 3053 Asians of California and Hawaii. The samples were typed for ABO and Rh groups and were analyzed electrophoretically for ten genetically determined protein variant systems. The effects of race, age, and sex on phenotypic frequencies within each of the twelve genetic systems were investigated. Large frequency differences were found between races but not between different age and sex subgroups within races. It was also demonstrated that the twelve genetic systems behaved statistically independently. Discrimination probabilities were computed for each of the four ethnic groups. These serve as a measure of the effectiveness of the twelve genetic systems examined in individualizing

blood samples. The method is discussed for computing the probability that a randomly chosen individual of a given ethnic group possesses the same blood phenotypes as found in a predetermined sample of blood. The results presented here should prove useful in the investigation of civil and criminal cases involving blood samples.

Acknowledgments

We acknowledge with thanks the collaboration of the directors and staff of the following blood banks:

(1) Blood Bank of the Alameda-Contra Costa Medical Association, Oakland, Calif.;
(2) Blood Bank of San Bernardino-Riverside Counties, San Bernardino, Calif.;
(3) San Diego Blood Bank, San Diego, Calif.;
(4) Doctor's Blood Bank, Terminal Annex, Los Angeles, Calif.;
(5) The American National Red Cross, Central California Red Cross Blood Center, San Jose, Calif.; and
(6) Blood Bank of Hawaii, Honolulu, Hawaii.

References

[1] Jones, D. A., "Blood Samples: Probability of Discrimination," *Journal of the Forensic Science Society,* Vol. 12, 1972, pp. 355-359.
[2] Grunbaum, B. W., Zec, J., and Durrum, E. L., "Application of an Improved Microelectrophoresis Technique and Immunoelectrophoresis of the Serum Proteins on Cellulose Acetate," *Microchemical Journal,* Vol. 7, 1963, pp. 41-53.
[3] Grunbaum, B. W., "An Automatic One- to Eight-Sample Applicator for Fast Qualitative and Quantitative Microelectrophoresis of Plasma Proteins, Hemoglobins, Isoenzymes and Cross-Over Electrophoresis," *Microchemical Journal,* Vol. 20, 1975, pp. 495-510.
[4] Grunbaum, B. W., "A Microanalytical Electrophoresis Technique for the Determination of Polymorphic Blood Proteins for Medical and Forensic Applications," *Mikrochimica Acta,* Vol. 2, 1977, pp. 339-352.
[5] Conover, W. J., *Practical Nonparametric Statistics,* New York, Wiley, 1971.
[6] Grunbaum, B. W., "A Micro Procedure for Fast Typing of the Genetic Variants of Phosphoglucomutase," *Journal of the Forensic Science Society,* Vol. 14, 1974, pp. 151-157.
[7] Grunbaum, B. W. and Zajac, P. L., "Phenotyping of Erythrocyte Acid Phosphatase in Fresh Blood and in Bloodstains on Cellulose Acetate," *Journal of Forensic Sciences,* Vol. 23, No. 1, Jan. 1978, pp. 84-88.
[8] Grunbaum, B. W., Harmor, G. C., Del Re, B., and Zajac, P., "Electrophoresis of Esterase D in Fresh Blood and Bloodstains on Cellulose Acetate," *Journal of Forensic Sciences,* Vol. 23, No. 1, Jan. 1978, pp. 89-93.
[9] Grunbaum, B. W., "Phenotyping of Haptoglobin on Gradient Acrylamide Gel Slabs Using the Beckman Microzone System," *Journal of the Forensic Science Society,* Vol. 15, 1975, pp. 229-234.
[10] Grunbaum, B. W. and Zajac, P. L., "Rapid Phenotyping of the Group Specific Component by Immunofixation on Cellulose Acetate," *Journal of Forensic Sciences,* Vol. 22, No. 3, July 1977, pp. 586-589.
[11] Grunbaum, B. W. and Zajac, P. L., "Differentiation of the Genetic Variants of Glucose-6-Phosphate Dehydrogenase in Bloodstains by Electrophoresis on Cellulose Acetate," *Journal of the Forensic Science Society,* Vol. 16, 1976, pp. 319-324.

Address requests for reprints or additional information to
Dr. Benjamin W. Grunbaum
White Mountain Research Station
Bldg. T-2251
University of California
Berkeley, Calif. 94720