Critical Tests of Hypotheses for Race Mixture Using Gm Data on American Caucasians and Negroes

T. EDWARD REED[1]

INTRODUCTION

Numerous estimates have been made of the proportion (M) of the genes of American Negroes which are of Caucasian origin. The early estimate by Glass and Li (1953) for Baltimore Negroes of 0.30565, using the R^o allele at the *Rh* locus, is probably the best known value. Other investigators, using this and other loci, including *Gm*, have obtained values for M ranging from about 0.20 to 0.40 for Negroes living outside the "Deep South" (Alabama, Florida, Georgia, Louisiana, Mississippi, South Carolina). All estimates were made using the formula $M = (q_n - q_a)/(q_c - q_a)$, where q is the gene frequency in question and subscripts n, a, and c refer to U.S. Negroes, African Negroes, and Caucasians, respectively. There is thus general agreement that M for these Negroes lies between 0.2 and 0.4, although the universal absence of standard errors hinders critical comparisons among the various estimates.

The purpose of this paper is not to review the estimates of M (this will be done elsewhere) but to review Gm frequency data in U.S. populations, present new data and a new estimate of M, and illustrate a method for critically testing hypotheses on race mixture.

The *Gm* locus is especially useful for formulating and testing various "strong" hypotheses about Caucasian-Negro mixture. The alleles Gm^1, $Gm^{1,2}$, and Gm^5 [determining, respectively, Gm factors (1), (1) and (2), and (5)] are stated to be the only alleles present in unmixed Caucasians, when testing for these three factors only. All are said to be absent in pure African Negroes, the latter having only the $Gm^{1,5}$ allele determining factors (1) and (5) (Steinberg, 1967). (The previous symbols for factors 1, 2, and 5 were a, x, and b.) If this is correct, we can find M as the sum of the frequencies of these three "Caucasian" alleles in American Negroes. In addition, we can make and test the reasonable assumption that the present frequencies of these introduced alleles in Negroes are proportional to their frequencies in the Caucasian "ancestor" population. Such proportionality will exist in the absence of selective differences among Negroes of different *Gm* genotypes, assuming the above distribution of alleles in the "parent" Caucasian and African populations and no large *Gm* frequency changes in Caucasians over the period when hybridization was occurring.

Received May 20, 1968.

The work reported here was begun while the author was in the Child Health and Development Studies, Oakland, California, on leave from the University of Toronto. The work was supported by U.S. Public Health Service research grants HD 00718 and HD 00720 (to Professor J. Yerushalmy) from the National Institutes of Health.

[1] Departments of Zoology, Paediatrics, and Anthropology, University of Toronto, Toronto, Ontario, Canada.

Estimates of M from Gm data by Steinberg *et al.* (1960; estimate of 0.31 in Cleveland and Baltimore Negroes) and by Blumberg *et al.* (1964; estimate of 0.073 in Georgia Negroes) were made by summing the frequencies of the "Caucasian" *Gm* alleles; the assumption of proportionality was not tested. This paper will review the available Gm(1,2,5) data, estimate M subject to the above assumptions of original gene distributions and proportionality, and test the goodness of fit of observed and expected values. Certain hypotheses alternative to the above will also be tested.

After all calculations had been completed it was learned (A. G. Steinberg, personal communication) that the Gm(5) testing in the studies of Steinberg *et al.* (1961) and Blumberg *et al.* (1964) may be unreliable by an unstated amount because of the Ragg (from rheumatoid arthritis patients) reagents used. The present Gm data came from testing with SNagg (normal serum) reagents and should be correct. For the purpose of testing hypotheses, all data, previously published or first given here, are assumed to be correct. Final conclusions, however, will be drawn only from the new Gm data presented in this paper.

THE DATA

Gm(1,2,5) data on both Caucasians and Negroes are given by Steinberg *et al.* (1961) for the Cleveland, Ohio, area and by Blumberg *et al.* (1964) for an area in Georgia. Unpublished data on the Child Health and Development Studies (CHDS) of Oakland, California, also give Gm (1,2,5) data for both races in this region. The CHDS is an ongoing study of pregnancy and child development in families which are members of a large medical care plan. Bloods from gravidae and their husbands were sent by air to the Blood Grouping Laboratory, Boston, for Gm and blood grouping. The persons in CHDS appear to be typical of the residents of the eastern San Francisco Bay area (Reed, 1968). The reagents used in testing the CHDS bloods were the following (SNagg agglutinator/anti-D): Gm(1)—Murdoch/Gagnon; Gm(2)—Greene/Harman; Gm(5)—Harrington/Harrison. Further details are on file.

The observed phenotype frequencies and estimated allele frequencies in the Caucasians and Negroes of these three studies are presented in Table 1. Estimation of gene frequencies, here and in later tables, and of other parameters such as M, is by maximum likelihood. A general computer program, MAXLIK (Reed and Schull, 1968), estimates the parameters by likelihood scores (Rao, 1952). Iteration continues until all corrections to parameters being estimated are less than 10^{-6}. Chi squares testing goodness of fit of observed phenotype frequencies with those expected from Hardy-Weinberg equilibrium, using estimated gene frequencies, all show a good fit ($P > .05$). The Caucasians of the three areas do not differ significantly among themselves in their phenotypes ($\chi^2_{[8]} = 12.12$, $P > .10$) and their gene frequencies are quite similar, although the most extreme comparison, Cleveland versus Oakland, *appears* to be significant. In contrast, the three Negro populations are very nearly significantly different in their phenotypes [$\chi^2_{[6]} = 12.37$ after pooling the $(1,-2,-5)$ and $(1,2,-5)$ groups; $\chi^2_{[6]} = 12.57$ for $P = .05$]. The Cleveland and Oakland Negroes do not differ in their phenotype distributions ($\chi^2_{[3]} = 1.11$, $P > .75$). The gene frequencies show similar relations, all three of the "Caucasian" alleles of the Georgia Negro group being less than the corresponding frequencies in either of the other Negro groups.

TABLE 1

Gm PHENOTYPE AND ALLELE FREQUENCIES IN CAUCASIANS (C) AND NEGROES (N) IN THREE AREAS OF THE UNITED STATES

RACE	AREA*	PHENOTYPES†						ALLELES‡				
		(1, −2, −5)	(−1, −2, 5)	(1, −2, 5)	(1, 2, −5)	(1, 2, 5)	Total	1	1, 2	5	1, 5	χ^2§
C	(1) Cleveland, O.	14	149	87	11	42	303	.205 ± .017	.091 ± .012	.705 ± .019	.....	0.973
	(2) Southern Georgia	21	134	90	12	38	295	.241 ± .018	.088 ± .012	.671 ± .019	.....	2.007
	(3) Oakland, Calif.	27	183	168	25	75	478	.254 ± .014	.109 ± .010	.637 ± .016	.....	4.949
	Pooled: (1)+(3)	41	332	255	36	117	781	.235 ± .011	.102 ± .008	.663 ± .012	.....	4.264
	Pooled: (1)+(2)+(3)	62	466	345	48	155	1,076	.236 ± .009	.098 ± .007	.665 ± .010	.....	4.404
N	(1) Cleveland, O.	8	7	338	1	10	364	.145 ± .025	.015 ± .005	.139 ± .026	.701 ± .036	0.352
	(2) Southern Georgia	1	0	187	0	1	189	.071 ± .036	.003 ± .003	10^{-6} ± .037	.926 ± .051	‖
	(3) Oakland, Calif.	3	8	239	2	8	260	.114 ± .030	.020 ± .006	.175 ± .031	.691 ± .042	<1#
	Pooled: (1)+(3)	11	15	577	3	18	624	.133 ± .019	.017 ± .004	.155 ± .020	.695 ± .027	2×10^{-5}
	Pooled: (1)+(2)+(3)	12	15	764	3	19	813	.122 ± .017	.014 ± .003	.136 ± .017	.729 ± .024	0.009

NOTE.—See Introduction concerning possible errors in Gm(5) testing in studies (1) and (2).

* Cleveland from Steinberg *et al.* (1961); Southern Georgia from Blumberg *et al.* (1964); Oakland from present study.

† Populations tested for Gm factors (1), (2), and (5) only.

‡ Maximum-likelihood estimates and their standard errors. Alleles (1), (1,2), and (5) in Caucasians and (1), (1,2), (5), and (1,5) in Negroes are here assumed to be the only ones present.

§ Chi square testing goodness of fit of observed numbers of phenotypes to expected numbers. Degrees of freedom: C,2; N,1.

‖ Not calculable but fit is extremely good.

Chi square is 0.554, but the smallest expected number is 1.26 and next smallest is 3.39. Calculated value therefore may be unreliable by Cochran's (1954) rule for minimum expected values.

Note that the frequency of 0.000 ± 0.037 for the Gm^5 allele of Georgia Negroes is compatible with an actual frequency of .05–.06. This is reasonable since the $Gm^5Gm^{1,5}$ individual cannot be distinguished from a $Gm^{1,5}Gm^{1,5}$ person. It is therefore not correct to conclude, as do Blumberg *et al.* (1964), that absence of Gm(−1,−2,5) phenotype in this population implies absence of the Gm^5 allele.

These results suggest that the Georgia Negroes have less Caucasian ancestry than non-southern Negroes, in agreement with the results of Blumberg *et al.* (1964) and Workman *et al.* (1963) for other loci of this Georgia population. It therefore seems appropriate, assuming correctness of the data, to pool the Cleveland and Oakland Negroes to obtain a sample of non-southern Negroes but to consider the Georgia Negroes separately.

TABLE 2

NULL HYPOTHESIS (H_0) AND CORRESPONDING EQUATIONS FOR ESTIMATING M^* (H_0: *Gm* GENES IN CAUCASIANS—Gm^1, $Gm^{1,2}$, Gm^5; *Gm* GENES IN AFRICAN NEGROES—$Gm^{1,5}$; NO SELECTIVE DIFFERENCES AMONG *Gm* GENOTYPES OF U.S. NEGROES)

Race	Phenotype	Genotypes	Proportion in Total (Caucasian +Negro) Sample
Caucasian....	1, −2, −5	Gm^1Gm^1	$w_c x_1^2$
	−1, −2, 5	Gm^5Gm^5	$w_c x_3^2$
	1, −2, 5	Gm^1Gm^5	$2w_c x_1 x_3$
	1, 2, −5	$Gm^1Gm^{1,2}$, $Gm^{1,2}Gm^{1,2}$	$w_c(2x_1x_2+x_2^2)$
	1, 2, 5	$Gm^{1,2}Gm^5$	$2w_c x_2 x_3$
U. S. Negro...	1, −2, −5	Gm^1Gm^1	$w_n y_1^2$
	−1, −2, 5	Gm^5Gm^5	$w_n y_3^2$
	1, −2, 5	Gm^1Gm^5, $Gm^1Gm^{1,5}$, $Gm^5Gm^{1,5}$ $Gm^{1,5}Gm^{1,5}$	$w_n(2y_1y_3+2y_1y_4+2y_3y_4+y_4^2)$
	1, 2, −5	$Gm^1Gm^{1,2}$, $Gm^{1,2}Gm^{1,2}$	$w_n(2y_1y_2+y_2^2)$
	1, 2, 5	$Gm^{1,2}Gm^5$, $Gm^{1,2}Gm^{1,5}$	$w_n(2y_2y_3+2y_2y_4)$

NOTE.—Relationships specified by hypothesis: $y_1 = Mx_1$, $y_2 = Mx_2$, $y_3 = Mx_3$, $y_4 = 1 - M$.

* Symbols: x_1, x_2, and x_3 are frequencies of Gm^1, $Gm^{1,2}$, and Gm^5, respectively, in Caucasians, and $x_1 + x_2 + x_3 = 1$; y_1, y_2, and y_3 are corresponding frequencies of these alleles in U.S. Negroes; y_4 is the frequency of $Gm^{1,5}$ in U.S. Negroes, and $y_1 + y_2 + y_3 + y_4 = 1$; $w_c = N_c/(N_c + N_n)$ and $w_n = N_n/(N_c + N_n)$, where N_c and N_n are total numbers of individuals in the Caucasian and Negro samples, respectively; and M = proportion of genes at *Gm* locus in U.S. Negroes which are of Caucasian origin.

ESTIMATION OF M, ASSUMING NO SELECTION

An estimate of M can be obtained from a given Negro population and a given Caucasian population when the gene frequencies in the African "parent" population and other relations are properly specified. Considering Gm factors (1), (2), and (5), there are 10 race-phenotype classes (five Caucasian Gm phenotypes and five Negro Gm phenotypes) as shown in Table 2. These classes have frequencies specified by certain parameters, including Caucasian gene frequencies and M. The estimation problem is to find the estimates of the parameters which produce the best fit between observed and expected phenotype frequencies. It is worth noting (T. E. Reed, unpublished) that when this procedure is applied to a single "Caucasian" allele, such as Fy^a, the estimate of M obtained is identical with that obtained from the usual gene frequency ratio (here $M = q_n/q_c$, since $q_a = 0$). This identity is a necessary consequence of using the equivalent relation $q_n = Mq_c$ in the maximum-likelihood estimation.

Table 2 presents the assumptions, genotype proportions, and specified parameter relationships for the first hypothesis (the null hypothesis, H_0), used in estimating M. The H_0 accepts the statement of Steinberg (1967), referred to above, on *Gm* alleles present in Caucasian and African Negro populations. It further assumes equal fitnesses in U.S. Negroes of *Gm* alleles and genotypes. This latter assumption implies that, within sampling error, the present frequencies of the "Caucasian" alleles in U.S. Negroes are proportional to their frequencies in Caucasians. This may be expressed as $y_i = Mx_i$, where y_i is the frequency of the ith "Caucasian" allele, say Gm^1, in U.S. Negroes and x_i is the corresponding allele frequency in Caucasians.

Table 3 presents estimates of $M \pm$ standard error and goodness of fit for 20 U.S. Negro–U.S. Caucasian pairs of populations under H_0. The results of this variety of Negro-Caucasian population pairs are presented to show the relatively small effect, other than the southern–non-southern Negro difference previously mentioned, that choice of particular Negro or Caucasian populations makes on the value of M. However, note the large differences in goodness of fit. The best estimate of M will clearly be based on the largest homogeneous populations; for the non-southern Negro this will involve the Cleveland plus Oakland Negroes and the total Caucasian population. The southern Negro estimate will use the Georgia Negroes and the total Caucasians. The corresponding M's for these two Negro groups are 0.265 ± 0.024 and 0.062 ± 0.032, respectively. The goodness of fit for the non-southern Negro (Cleveland and Oakland) is very poor, however ($\chi^2_{[5]} = 37.26$, $P \ll 0.001$). Note, however, that using Oakland Caucasians and Negroes only gives $M = 0.273 \pm 0.037$ with a *good* fit.

It should be noted that the Georgia Negroes give very little information on Caucasian admixture. Only two of these 189 Negroes have non-African Gm phenotypes; yet Gm^1 and Gm^5 could be present in heterozygotes with the African $Gm^{1,5}$ and not be identifiable. The estimate of M is therefore poor as shown by the relatively large standard error. The small numbers of non-African phenotypes in the other two Negro populations (26 and 21) make the effect of sampling error important here also. This error, plus the larger size of the Cleveland Negro sample, and perhaps the Gm(5) testing problem, may explain why the Cleveland Negro–total Caucasian estimate also has a very poor fit ($\chi^2_{[5]} = 39.99$, $P \ll 0.001$) while the Oakland Negro–total Caucasian estimate fits much better ($\chi^2_{[5]} = 10.23$, $P = 0.07$). In all goodness-of-fit comparisons with chi square, the recommendations of Cochran (1954) on minimum expectations were followed.

It may be noted that estimates of M in non-southern Negroes, derived from the sum of the three "Caucasian" alleles of Table 1, agree within sampling errors with the estimates of H_0 (Table 3). For example, the above-mentioned estimate for Cleveland plus Oakland Negroes was 0.265 ± 0.024, while the sum of Gm^1, $Gm^{1,2}$, and Gm^5 in this pooled group is 0.305 ± 0.027. Exact agreement is not expected, since the assumption of proportionality is not made in the latter estimate.

ALTERNATIVE HYPOTHESES

The very poor fit of observed and expected phenotype frequencies under H_0, using Cleveland Negroes, is mainly due to an excess number of Gm(1,−2,−5) Negroes relative to the expected number. One possible explanation for this poor fit is the

TABLE 3

ESTIMATES OF M AND GOODNESS OF FIT UNDER HYPOTHESES H_0, H_1, AND H_2

Hypothesis	Caucasian Population	Negro Population											
		(1) Cleveland			(2) Southern Georgia			(3) Oakland			(1)+(3)		
		M	z	P	M	z	P	M	z	P	M	z	P
H_0	(1)	.258 ± .031	.....	< .001	.064 ± .033	.....	> .05	.274 ± .036	.....	> .05	.269 ± .024	.....	< .001
	(2)	.265 ± .032	.....	< .001	.066 ± .034	.....	> .05	.282 ± .037	.....	> .05	.276 ± .025	.....	< .001
	(3)	.255 ± .031	.....	< .001	.058 ± .031	.....	> .05	.273 ± .037	.....	> .05	.266 ± .024	.....	< .001
	(1)+(3)	.254 ± .031	.....	< .001	.060 ± .031	.....	> .05	.272 ± .036	.....	> .05	.264 ± .024	.....	< .001
	(1)+(2)+(3)	.255 ± .031	.....	< .001	.062 ± .032	.....	> .05	.274 ± .036	.....	> .05	.265 ± .024	.....	< .001
H_1	(1)	.188 ± .031	.131 ± .031	> .05	.026 ± .024	.068 ± .037	> .05	.240 ± .037	.085 ± .040	> .05	.212 ± .024	.113 ± .025	> .05
	(2)	.197 ± .032	.121 ± .032	> .05	.026 ± .025	.067 ± .037	> .05	.251 ± .039	.071 ± .041	> .05	.221 ± .025	.102 ± .026	> .05
	(3)	.188 ± .032	.119 ± .032	> .05	.022 ± .022	.067 ± .037	< .05*	.244 ± .039	.068 ± .040	> .05	.214 ± .025	.099 ± .025	> .05
	(1)+(3)	.188 ± .031	.124 ± .031	> .05	.024 ± .023	.067 ± .037	< .05*	.242 ± .038	.075 ± .040	> .05	.212 ± .024	.105 ± .025	> .05
	(1)+(2)+(3)	.190 ± .031	.123 ± .032	> .05	.024 ± .023	.067 ± .037	< .05*	.244 ± .038	.074 ± .040	> .05	.214 ± .024	.104 ± .025	> .05
H_2	(1)	.188 ± .031	3.75 ± .98	> .05	.026 ± .024	13.7 ± 14.8†	> .05	.240 ± .037	2.32 ± .74	> .05	.212 ± .024	3.06 ± .65	> .05
	(2)	.197 ± .032	3.05 ± .80	> .05	.026 ± .025	11.2 ± 12.1†	> .05	.251 ± .039	1.88 ± .60	> .05	.221 ± .025	2.48 ± .52	> .05
	(3)	.188 ± .032	3.02 ± .77	> .05	.022 ± .022	12.6 ± 13.8†	> .05	.244 ± .039	1.83 ± .58	> .05	.214 ± .025	2.43 ± .49	> .05
	(1)+(3)	.188 ± .031	3.28 ± .82	> .05	.024 ± .023	12.9 ± 14.0†	> .05	.242 ± .038	2.00 ± .62	> .05	.212 ± .024	2.66 ± .52	> .05
	(1)+(2)+(3)	.190 ± .031	3.22 ± .80	> .05	.024 ± .023	12.4 ± 13.5†	> .05	.244 ± .038	1.97 ± .61	> .05	.214 ± .024	2.62 ± .51	> .05

NOTE.—See Tables 2 and 4 for definitions of hypotheses. See Table 1 for definitions of Caucasian populations. [Note possible Gm(5) errors in studies (1) and (2).] The z is a parameter specified by H_1 or H_2. For H_1 it is the frequency of Gm^1 in African Negroes. For H_2 it is the selective advantage V in U.S. Negroes of the Gm^1 allele relative to that of other alleles (all have unit relative fitness). P = goodness-of-fit probabilities.

* $P < .05$ after pooling the four small Negro classes; unpooled χ^2 has $P > .05$.

† The last iteration correction for this estimate was in the range .003–.000005 and so is not as accurate as other estimates.

above-mentioned Gm(5) testing problem in data using Ragg reagents. False negatives in testing for Gm(5) could produce an apparent excess of Gm(1,−2,−5) in Negroes. Such an excess is not apparent in the Oakland Negroes, who were tested with SNagg reagents.

If there is no Gm testing problem, there are several other important possibilities, and these, like H_0, can be tested critically. One of these possibilities is that Gm^1 might be present in low frequency in African Negroes. This hypothesis, H_1, is described in Table 4 and the resulting estimates of M and goodness of fit are given in Table 3.

TABLE 4

HYPOTHESES H_1 AND H_2 AND THEIR CORRESPONDING EQUATIONS FOR ESTIMATING M*

H_1
Assumptions: *Gm* genes in Caucasians: Gm^1, $Gm^{1,2}$, Gm^5 *Gm* genes in African Negroes: Gm^1, $Gm^{1,5}$ No selective differences among *Gm* genotypes
Definitions: y_5 = frequency of Gm^1 in African Negroes
Relationships specified by hypothesis:
$y_1 = Mx_1 + (1 - M)y_5$
$y_2 = Mx_2$
$y_3 = Mx_3$
$y_4 = (1 - M)\ (1 - y_5)$
H_2
Assumptions: *Gm* genes in Caucasians: Gm^1, $Gm^{1,2}$, Gm^5 *Gm* genes in African Negroes: $Gm^{1,5}$ A selective advantage in U.S. Negroes of the Gm^1 allele, relative to other alleles, *may* exist.
Definitions: The fitness in U.S. Negroes of allele Gm^1, relative to that of other alleles, is V.
Relationships specified by hypothesis:
$y_1 = Mx_1V$
$y_2 = Mx_2$
$y_3 = Mx_3$
$y_4 = 1 - y_1 - y_2 - y_3$

* Symbols, phenotypes, genotypes, and proportions as in Table 2. Note special definitions and relationships for each hypothesis here.

The fit of the non-southern Negro populations is seen to be uniformly good, while that of the Georgia Negroes is clearly good for two pairs and probably good for the other three (see note to Table 3). For the pooled non-southern Negro–total Caucasian estimate, M is 0.214 ± 0.024 and the frequency estimate of Gm^1 in African Negroes is 0.104 ± 0.025. The reasonableness of the latter estimate is discussed below.

Another alternative hypothesis to H_0, H_2, could also formally explain the observed excess of Gm(1,−2,−5) Negroes in Cleveland data by assuming that in U.S. Negroes the Gm^1 allele has a selective advantage over the $Gm^{1,2}$ and Gm^5 alleles. This hypothesis is described in Table 4, and the corresponding values of M and tests of fit are in Table 3. The H_2 differs fundamentally from H_1; yet the estimates of M and their

standard errors are the same. The fit is perhaps even better. The estimates of V, the cumulative relative fitness in U.S. Negroes of the Gm^1 allele relative to that of $Gm^{1,2}$ and Gm^5 alleles, have large standard errors and so are not very accurate. Those for the Georgia Negroes are so large as to make their estimates of V meaningless. Using total Caucasians, the estimate of V for the Cleveland plus Oakland Negroes is 2.62 ± 0.51. This large (significantly larger than unity) value is, by H_2, effectively the ratio of the present Gm^1 frequency in non-southern Negroes to the Gm^1 frequency which would have existed in these Negroes in the absence of a selective advantage. As such, it represents the cumulative selective advantage over the generations since Caucasian-Negro race crossing in North America first began. The Gm^1 alleles have existed in U.S. Negroes for perhaps an average of five to seven generations. With this range of time and assuming, as is probably incorrect, constant selective advantage over these periods, 1.15–1.21 is the range of relative fitness of Gm^1 on a generation basis, v ($v^t = V$, where t is time in generations). Note that the estimate of V for Oakland data only is not significantly greater than unity.

In addition to H_1 and H_2 as possible "explanations" for the present U.S. Negro–Caucasian Gm frequencies, a third hypothesis, H_3, may be briefly considered here. Ropartz *et al.* (1965) have suggested that, in addition to the three Gm alleles already considered to be present in Caucasians, a fourth allele, $Gm^{1,5}$, similar to the Negro allele but of Caucasian origin, is also present. This suggestion is not generally accepted (Steinberg, 1967, and personal communication). If one does estimate these four Gm allele frequencies in the present Caucasian populations, it is found that this possibility can neither be excluded nor supported; the frequency of $Gm^{1,5}$ in the pooled sample of 1,076 Caucasians, under H_3, is 0.012 ± 0.015, chi square for fit being 3.75 (1 df). For the 478 Oakland Caucasians, the frequency is estimated as 0.042 ± 0.022; chi square is 1.39. If one estimates M under H_3 (Gm^1, $Gm^{1,2}$, Gm^5, and $Gm^{1,5}$ present in Caucasians, $Gm^{1,5}$ only present in African Negroes, no selective differences in U.S. Negroes among Gm genotypes), one finds that the chi square for goodness of fit for Cleveland plus Oakland Negroes and total Caucasians is very high (37.50, df = 4, $P \ll 0.001$). Apparently, H_3 does not "explain" this observed Negro-Caucasian distribution. Using only Oakland Negroes and Caucasians, however, the fit is good ($\chi^2_{[4]} = 7.89$) and $M = 0.279 \pm 0.038$.

DISCUSSION

In the discussion to follow we will assume, initially, that all Gm testing is correct. We will further assume that, at the present three-factor level of testing used, the only alleles present in Caucasians are Gm^1, $Gm^{1,2}$, and Gm^5. The above results for H_3 give no convincing support for the presence of $Gm^{1,5}$. We will also assume that the genetic contribution of American Indians to American Negroes is negligible for the present problem. There appears to be no evidence for an appreciable contribution, and Glass (1955), in reviewing blood-group frequency data, particularly in the ABO and Rh systems, found evidence against this possibility. If, as seems likely, the Caucasians (usually the male slave owners) who introduced Caucasian genes into the U.S. Negro population were representative in their Gm genotypes, the initial frequencies of these "Caucasian" alleles in the Negroes must have been proportional to the Caucasian

frequencies. If there were no large changes in the frequencies of Gm phenotypes in Caucasians over the approximately 12-generation period since this race crossing began, which seems reasonable, and none of these alleles were initially present, and there were no selective differences among Negro *Gm* genotypes, we must then expect the frequencies of the three "Caucasian" alleles in U.S. Negroes to continue to be proportional to the Caucasian frequencies. This is the null hypothesis, H_0. This prediction of present proportionality is a "strong inference" in the sense of Platt (1964). It was tested and found not to agree with observation when data on Cleveland Negroes were used but did agree when Oakland Negroes were used.

The above rejection of H_0 raises the question of the estimate of M by Steinberg *et al.* (1960). They equated M to the sum of Gm^1 and Gm^5 allele frequencies [Gm(2) was not tested; therefore, their Gm^1 frequency includes Gm^1 and $Gm^{1,2}$]. For Negroes from the Cleveland and Baltimore areas and Caucasians from the same two areas, Steinberg *et al.* (1960) obtained an estimate for M of 0.310. However, if one estimates M from their data according to the present H_0 (Gm^1 and Gm^5 in Caucasians, $Gm^{1,5}$ in African Negroes, and Gm^1 and Gm^5 frequencies in present U.S. Negroes proportional to Caucasian frequencies), the goodness of fit is again poor ($\chi^2_{[2]} = 11.38$, $P < 0.005$). The major reason for this high chi square is the excess number of Gm(1,−5) Negroes relative to the number expected by H_0. The Gm(5) reagent used here was also a Ragg type, so the possibility of Gm(5) typing error is also raised here as it is in the data of Steinberg *et al.* (1961). (See also the present Introduction.) The estimate of M by Blumberg *et al.* (1964) is not informative here. Their Georgia data have already been considered in this paper and the little information provided by the Negroes (two non-African phenotypes in 189) was mentioned. Since blood-group data on Negroes from Charleston, South Carolina (Pollitzer, 1958), indicate an even lower value of M, it seems probable that southern Negroes have appreciably less Caucasian ancestry than do non-southern Negroes. For this reason we confine our attention to the latter group.

Hypothesis H_1 considered an alternative possibility that Gm^1 might be present in African Negroes. This assumption, unlike H_0, gave a good fit in all Negro-Caucasian pairs. The frequency of Gm^1 in African Negroes was estimated as 0.104 ± 0.025 in the pooled non-southern Negro–total Caucasian estimate. This frequency at first seems high considering that unmixed African Negroes are regarded as being homozygous for $Gm^{1,5}$ (Ropartz *et al.*, 1963; Steinberg, 1967). However, one must note that a Gm^1 frequency of 0.1 and $Gm^{1,5}$ frequency of 0.9 produce a Gm(1,−5) phenotype frequency of 0.01. The $Gm^1Gm^{1,5}$ genotype, which harbors most of the Gm^1 alleles, can be distinguished from the common $Gm^{1,5}Gm^{1,5}$ genotype only by progeny testing. A Gm^1 frequency of 0.05 would produce a Gm(1,−5) phenotype frequency of 0.0025. This Gm^1 frequency is about the lower limit required to make H_1 fit, since it produces a goodness of fit chi square of 11.95 (df = 5, $P < 0.05$).

The available Gm surveys in west African Negroes show the following Gm(1) and Gm(5) distributions: *Senegal area*, 8 Gm(1,−5), 3 Gm(−1,5), 2,340 Gm(1,5) in a total of 2,351 Negroes from various tribes (Ropartz *et al.*, 1960; Kane and Ruffié, 1963; Ropartz *et al.*, 1963; Gessain, Ruffié, *et al.*, 1965; Gessain *et al.*, 1965); *Nigeria*, 479 Gm(1,5) in a total of 479 (Boyer and Watson-Williams, 1961; Steinberg *et al.*,

1961). The frequency of the Gm(1,−5) phenotype in the total Senegal data is 0.0034 ± 0.0012, corresponding to a Gm^1 frequency of 0.058. The frequency of this phenotype in the Nigeria data should be less than 0.0062 (95% confidence), corresponding to a maximum Gm^1 frequency of 0.079.

Since the Senegal region was at the northern limit of the slave trade (Herskovits, 1941), the extensive data from there are not as relevant as would be desired. Also, it is the opinion of some authors [Ropartz *et al.*, 1960, who found two Gm(1,2,5) Negroes in 399] that there has been some white admixture. This possibility is not yet established. Nigeria was near the central area of the slave trade, and the upper limit of 0.079 for Gm^1 there is very pertinent. It is noteworthy that two of 35 Bashi from the eastern Congo, presumably east of the slave area, were Gm(1,−5) (Steinberg *et al.*, 1961). There is thus some evidence for Gm^1 around the margins of the slave-trading area but no evidence for its presence in the central part. The relatively high upper limit of its possible frequency, however, means that we cannot exclude H_1. It does seem somewhat unlikely.

Hypothesis H_2 tested the possibility that the poor fit of H_0 might be due to a selective advantage of the Gm^1 allele in U.S. Negroes. Like H_1, this hypothesis gave a good fit for all pairs and so is a second formal explanation of the observed frequencies. The estimate of 2.62 ± 0.51 for V in non-southern Negroes (Cleveland and Oakland), estimating the increase in Gm^1 frequency over what it would have been without selective advantage, seems very large, but the increase per generation, 15% to 20%, is not. Most Gm^1 alleles were introduced into U.S. Negroes before emancipation, about four generations ago. The estimate of five to seven generations, roughly 125 to 175 years, as the average time elapsed since the "ancestor" of any Gm^1 allele now in an American Negro was contributed by a Caucasian seems reasonable in the absence of useful data. It obviously is very desirable to know if Gm^1 really has this advantage. In the absence of direct observation on the relative fitness of Gm phenotypes of Negroes (which would be complicated by the inability to distinguish between $Gm^1Gm^{1,5}$ and $Gm^{1,5}Gm^{1,5}$) and the inability to exclude other explanations, such as Gm(5) typing problems or H_1, no decision can be made at present.

We may ask whether there are other hypotheses which could fit *all* the observed Negro-Caucasian data. The difficulty to be accommodated is the high frequency of the Gm^1 allele observed in pooled non-southern American Negroes relative to that expected under H_0. This could formally be explained by: (1) an excess, relative to the general Caucasian population, of Gm^1 in the Caucasian ancestors of U.S. Negroes, (2) a change in *Gm* allele frequencies in Caucasians in the last five to 10 generations such that Gm^1 is now much less frequent than it was, or (3) combinations of (1), (2), H_1, and H_2. Explanations (1) and (2) seem quite unlikely but cannot, of course, be disproved. On the other hand, genetic drift does seem excluded by the large numbers involved.

A large family of additional hypotheses can be generated if one makes the reasonable assumption that a selective advantage of the Gm^1 allele in Negroes, if present, is actually due to an advantage of one or more genotypes containing this allele. Selection, if present, very likely acts primarily on genotypes and not on gametes. Since, by far, most Gm^1 alleles in U.S. Negroes are in $Gm^1Gm^{1,5}$ genotypes, one should test the hypothesis that the cumulative relative fitness of this genotype may be greater than

unity (all other genotypes having relative-fitness unity and allele distributions as in H_0). When this is done, the fit for total Caucasians–non-southern Negroes is found to be very poor ($V = 1.149 \pm 0.525$, $\chi^2_{[4]} = 36.85$, $P < .001$). A similarly poor fit is obtained if, instead, it is assumed that the fitnesses of all Negro genotypes containing a Gm^1 allele are equal (V) and may be greater than unity ($V = 1.785 \pm 0.390$, $\chi^2_{[4]} = 22.61$, $P < .001$). If, however, only Negro genotype Gm^1Gm^1 is allowed to have a fitness greater than unity, a good fit is obtained ($\chi^2_{[4]} = 7.600$, $P < .05$), but the estimate of V is 6.439 ± 2.484. This point estimate is unreasonably high, but the large standard error produces a confidence interval whose lower end is biologically possible. The above three hypothesis tests strongly indicate that it will not be possible, from the present data, to identify particular genotypes having selective advantages.

Throughout this study the suspected Gm(5) unreliability (A. G. Steinberg, personal communication) in the Cleveland and Georgia data has been referred to when necessary, but no decision concerning its existence or effect, if real, on estimation of M was taken. This deferral followed from uncertainty as to the extent of the unreliability and also from the desire to test a number of plausible alternative hypotheses. It is now time to attempt an over-all assessment. If Gm testing in Cleveland was accurate, then rejection of H_0 is meaningful and requires some alternate explanation such as H_1 or H_2. If Gm testing in Cleveland was subject to appreciable error, then the rejection of H_0 may simply reflect this error and have no further meaning. On the other hand, the Oakland data, tested with SNagg reagents, are believed to be reliable, and it is noteworthy that, using only these data, all three main hypotheses, H_0, H_1, and H_2, fit well. The parameters estimated in H_1 and H_2, however, do not differ significantly from those in H_0, and, consequently, for the Oakland data, H_1 and H_2 merge into H_0. Clearly the Oakland data agree well with the null hypothesis, and no other hypothesis seems to fit better.

Faced with these conflicting interpretations, it seems only reasonable to place confidence in the results from the Oakland data because the data appear to be reliable and no special hypotheses are required for a good fit. Results from the Cleveland data are suspect when judged by the same standards. We therefore further conclude that a critical test (H_0) reveals no evidence for selective differences among the three Caucasian alleles following their introduction into the American Negro.

The problem of deciding which data are more reliable should not obscure the value of the maximum-likelihood approach used for critically testing specific hypotheses on race mixture. Such tests apparently have not been made before. It should be noted that multiple alleles are not required to make these tests; two alleles, such as K and k, can be used (T. E. Reed, unpublished). We are now able to make various critical statements. For example, from the size of the standard errors in Table 3, we can say that a very considerable selective advantage of a given allele might be present but remain undetected, even in large samples.

It should not be thought that the problems, including misclassification, in estimating M from *Gm* alleles are unique to this locus. Rather, it is likely that they and perhaps others, too, exist in many of the estimates made with other loci (e.g., Workman *et al.*, 1963; Workman, 1968). (See Livingstone *et al.*, 1960, for an informative discussion of blood-grouping problems in certain studies.) What is unique to the

Gm locus, so far, is the opportunity to compare the proportions of two or three alleles in Negro hybrids, given the critical and reasonable assumption that these alleles came entirely from Caucasians and therefore should be proportional to the Caucasian allele frequencies. Comparable critical tests on other loci may well reveal similar problems.

SUMMARY

The *Gm* alleles Gm^1, $Gm^{1,2}$, and Gm^5 are believed to characterize unmixed Caucasians, while $Gm^{1,5}$ alone is believed present in unmixed west African Negroes [Steinberg, 1967, testing for Gm factors (1), (2), and (5) only]. Given these original gene distributions and the assumptions of no selective differences among the *Gm* genotypes in American Negroes, no *Gm* gene frequency changes in Caucasians, and negligible non-Caucasian contribution of genes, one can estimate the proportion M of Caucasian ancestry in American Negroes from the sum of the frequencies of the three "Caucasian" alleles.

The above assumptions, however, also permit a further, strong inference: the frequencies of these three "Caucasian" alleles in U.S. Negroes should be proportional to the corresponding frequencies of the appropriate Caucasian "parent" population. When this prediction is tested in maximum-likelihood estimation of M, it is found to be incorrect for one published body of data—the fit of observed to expected phenotype frequencies is very poor. New Gm data reported here, on California Caucasians and Negroes, however, agree well with this inference. Recent information indicates that there may have been testing errors in the published data.

The technique of using maximum-likelihood estimation for critical testing of hypotheses on race mixture is further illustrated by applying it to several alternative hypotheses. Two of these hypotheses fit the published Gm data, but, when tested with the new Gm data, they reduce to the original hypothesis described above.

Using the new Gm data, the estimate of M is 0.273 ± 0.037. There is no evidence for a selective advantage or disadvantage of any of the three "Caucasian" alleles (Gm^1, $Gm^{1,2}$, or Gm^5) in the American Negro, but the sensitivity for detection of such selection is rather low.

ACKNOWLEDGMENTS

I would like to thank Professor Curt Stern for pointing out to me the problem of Gm^1 and Gm^5 frequencies in one group of American Negroes and also for discussions of this problem. I would also like to thank Dr. Jack L. King for valuable comments on several drafts of this paper. Professors James V. Neel, C. C. Li, Hugh Fudenberg, Arthur Steinberg, Baruch Blumberg, and Louis Diamond kindly commented on certain aspects of this paper. I am grateful to Dr. Arthur E. Mourant and Mrs. K. Domaniewska Sobczak for supplying recent references on Gm distributions in west African populations. I thank Professor Jacob Yerushalmy for the opportunity to analyze the data from this study.

REFERENCES

BLUMBERG, B. S., WORKMAN, P. L., and HIRSCHFELD, J. 1964. Gamma-globulin, group specific, and lipoprotein groups in a U.S. white and Negro population. *Nature* **202**:561–563.

BOYER, S. H., and WATSON-WILLIAMS, E. J. 1961. The γ-globulin, Gm^{ab}, in Nigerians. *Nature* **190**:456.

COCHRAN, W. G. 1954. Some methods for strengthening the common χ^2 tests. *Biometrics* **10**:417–451.

GESSAIN, R., MOULLEC, J., and GOMILA, J. 1965. Groupes d'haptoglobine et de transferrine et groupes Gm des Coniagui et des Bassari. *Bull. Soc. Anthropol.* (Paris) Ser. 11. **8**:19–22.

GESSAIN, R., RUFFIÉ, J., KANE, Y., KANE, O., CABANNES, R., and GOMILA, J. 1965. Note sur la sero-anthropologie de trois populations de Guinée et du Sénégal: Coniagui, Bassari, et Bedik. *Bull. Mem. Soc. Anthropol.* (Paris) Ser. 11. **8**:5–18.

GLASS, B. 1955. On the unlikelihood of significant admixture of genes from the North American Indians in the present composition of the Negroes of the United States. *Amer. J. Hum. Genet.* **7**:368–385.

GLASS, B., and LI, C. C. 1953. The dynamics of racial intermixture—an analysis based on the American Negro. *Amer. J. Hum. Genet.* **5**:1–20.

HERSKOVITS, M. J. 1941. *The myth of the Negro past.* Harper, New York.

KANE, Y., and RUFFIÉ, J. 1963. Étude hémotypologique de quelques groupes Peul, Toucouleur, Ouoloff, et Sérère du Sénégal occidental. *Bull. Soc. Anthropol.* (Paris) Ser. 11. **4**:545–553.

LIVINGSTONE, F. B., GERSHOWITZ, H., NEEL, J. V., ZUELZER, W. W., and SOLOMON, M.D. 1960. The distribution of several blood group genes in Liberia, the Ivory Coast and Upper Volta. *Amer. J. Phys. Anthrop.* **18**:161–178.

PLATT, J. R. 1964. Strong inference. *Science* **146**:347–353.

POLLITZER, W. S. 1958. The Negroes of Charleston (S.C.); a study of hemoglobin types, serology, and morphology. *Amer. J. Phys. Anthrop.* **16**:241–263.

RAO, C. R. 1952. *Advanced statistical methods in biometric research.* Wiley, New York.

REED, T. E. 1968. Research on blood groups and selection from the Child Health and Development Studies, Oakland, California. III. Couple mating type and reproductive performance. *Amer. J. Hum. Genet.* **20**:129–141.

REED, T. E., and SCHULL, W. J. 1968. A general maximum likelihood estimation program. *Amer. J. Hum. Genet.* **20**:579–580.

ROPARTZ, C., RIVAT, L., and LENOIR, J. 1960. Fréquence des facteurs Gm(a), Gm(b), Gm(x), Gm-like et Inv chez 400 noirs africains. *Rev. Franç. Étud. Clin. Biol.* **5**:814–816.

ROPARTZ, C., ROUSSEAU, P.-Y., and RIVAT, L. 1963. Intérêt des groupes de γ-globuline Gm et Inv dans l'appréciation du métissage des populations: étude de ces groupes sérique dans l'ouest africain et l'extrême-orient. *Rev. Franç. Étud. Clin. Biol.* **8**:465–472.

ROPARTZ, C., ROUSSEAU, P.-Y., and RIVAT, L. 1965. Hypothèses sur la génétique formelle du système Gm chez les Caucasiens. *Humangenetik* **1**:483–496.

STEINBERG, A. G. 1967. Genetic variations in human immunoglobulins: the Gm and Inv types. Chap. 3 in T. J. GREENWALT (ed.), *Symposium on immunogenetics.* Lippincott, Philadelphia.

STEINBERG, A. G., STAUFFER, R., BLUMBERG, B. S., and FUDENBERG, H. 1961. Gm phenotypes and genotypes in U.S. whites and Negroes; in American Indians and Eskimos; in Africans; and in Micronesians. *Amer. J. Hum. Genet.* **13**:205–213.

STEINBERG, A. G., STAUFFER, R., and BOYER, S. H. 1960. Evidence for a Gm^{ab} allele in the Gm system of American Negroes. *Nature* **188**:169–170.

WORKMAN, P. L. 1968. Gene flow and the search for natural selection in man. *Hum. Biol.* **40**:260–279.

WORKMAN, P. L., BLUMBERG, B. S., and COOPER, A. J. 1963. Selection, gene migration and polymorphic stability in a U.S. white and Negro population. *Amer. J. Hum. Genet.* **15**: 429–437.