Distribution of Transferrin Saturations in the African-American Population

By Victor R. Gordeuk, Christine E. McLaren, Anne C. Looker, Victor Hasselblad, and Gary M. Brittenham

To determine if transferrin saturations in African Americans may reflect the presence of a gene that influences iron metabolism, we analyzed the distribution of these values in 808 African Americans from the second National Health and Nutrition Survey. We tested for a mixture of three normal distributions consistent with population genetics for a major locus effect in which the proportion of normal homozygotes is $p^2$, of heterozygotes, $2pq$, of affected homozygotes, $q^2$; and in which $p+q=1$. Three subpopulations based on transferrin saturation were present ($P<.0001$) and the fit to a mixture of three normal distributions was good ($P=.2$). A proportion of .009 was included in a subpopulation with a mean $\pm$ standard deviation transferrin saturation of 63.4$\pm$5.7% (postulated homozygotes for a gene that influences iron metabolism), while a proportion of .136 had an intermediate saturation of 36.0$\pm$5.7% (postulated heterozygotes) and .856 a saturation of 24.6$\pm$5.7% (postulated normal homozygotes). These proportions were consistent with population genetics because the sum of the square roots of the proportions with the lowest mean transferrin saturation ($P=.925$) and the highest ($q=0.093$) was approximately 1 ($1.018$). The results are consistent with the presence in African Americans of a common locus that influences iron metabolism.

© 1998 by The American Society of Hematology.

IRON OVERLOAD IS COMMON in Africa, but this condition is not widely considered to be a problem among African Americans whose ancestors originated in Africa. Recognized for more than 60 years, iron overload in Africa is etiologically related to increased dietary iron. A recent study suggested that a non-HLA–linked iron-loading gene may also be implicated in the pathogenesis, with heterozygotes for the iron-loading locus developing iron overload only in the face of high dietary iron, but with homozygotes becoming iron-loaded even without increased dietary iron.

In the United States, primary iron overload is regarded as predominantly a problem among Caucasian Americans in the form of HLA-linked hemochromatosis. Based on screening the population for homozygotes, the estimated gene frequency for this recessive disorder is .067. Recently, we used a novel application of statistical mixture modeling to analyze the distribution of transferrin saturations in white Americans and to estimate the prevalence of hemochromatosis heterozygotes in that population. Our findings were consistent with a distinct distribution of transferrin saturations in hemochromatosis heterozygotes and with a gene frequency of .07 to .08.

The present study was prompted by a concern that a primary iron overload condition may be present in the African-American population, but be largely unrecognized and untreated. We postulated that, in an analogous situation to that of HLA-linked hemochromatosis in whites, an iron-loading locus in African Americans might be manifested by distinct, statistically discernible subpopulations in the distribution of population transferrin saturation data.

MATERIALS AND METHODS

Source of data. The second National Health and Nutrition Examination Survey, which studied a representative sample of the nonsituationalyzed United States population ages 6 months to 74 years from 1976 to 1980, was the source of data for the present study. In this survey, 20,322 persons were examined from 64 primary sampling units (counties or small groups of contiguous counties). Based on an interviewer’s observation, each person was classified as white, black, or other. Of those surveyed, 2,763 (13.6%) were African Americans. Serum iron and total iron binding capacity were measured by a modification of the Automated Technicon AAI-25 method, which is a colorometric method using ferrozine; the transferrin saturation was calculated from these values by dividing the serum iron by the total iron binding capacity and multiplying by 100.

Selection criteria. We selected transferrin saturations from African-American men and women aged 20 years to 74 years for whom the mean corpuscular volume was between 80 fL and 100 fL and the erythrocyte protoporphyrin was <70 mg/dL red blood cells. Additional selection criteria included hemoglobin concentration $\geq 13.5$ g/dL and hematocrit $\geq 40\%$ for men and hemoglobin $\geq 11.5$ g/dL and hematocrit $\geq 34\%$ for women. We excluded subjects with abnormally low hemoglobin or hematocrit values because anemias of various causes are associated with abnormally high or low transferrin saturations.

We excluded subjects with abnormal values for mean corpuscular volume because a low mean corpuscular volume can be associated with iron deficiency or inflammation and a high mean corpuscular volume can be associated with megaloblastic conditions and drug effects. We excluded subjects with elevated erythrocyte protoporphyrin levels because of the associations with iron deficiency and inflammation. After applying the selection criteria, there were 836 individuals in the data set. As described later, an additional 28 subjects were excluded as possible heterozygotes for HLA-linked hemochromatosis. Table 1 gives the numbers of subjects excluded using specific criteria.

Adjustment of transferrin saturations for gender and diurnal variation. Because transferrin saturation has a diurnal variation, the inclusion of samples obtained at different times of the day, without appropriate adjustment, in an analysis of distribution might alter the

From the Department of Medicine, The George Washington University Medical Center, Washington D.C.; the Department of Mathematics, Moorhead State University, Moorhead MN; the Nutrition Statistics Branch, National Center for Health Statistics, Hyattsville MD; the Center for Health Policy Research and Education, Duke University, Durham, NC; and the Department of Medicine, MetroHealth Medical Center, Case Western Reserve University School of Medicine, Cleveland OH.

Submitted September 25, 1997; accepted October 28, 1997.

Supported in part by a grant from the Office of Minority Health to the Cell Biology and Metabolism Branch, National Institute of Child Health and Human Development (Bethesda, MD) and by a grant from the National Center for Health Statistics, Centers for Disease Control (Hyattsville, MD).


The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. section 1734 solely to indicate this fact.

© 1998 by The American Society of Hematology.

0006-4971/98/9106-00683.00/0

results. In addition, because of the relatively small size of the sample data from men (368) and women (468), analyzing the individual distributions could result in inappropriately large standard errors for parameter estimates. In this study, we adjusted transferrin saturation values for blood samples drawn in the afternoon or evening to reflect expected values had the blood samples been drawn in the morning. These expected values were determined using linear regression analysis in the following manner. First, for each gender, the transferrin saturation values were stratified by time of blood collection (morning, afternoon, or evening) and each stratum was divided into deciles. Second, a regression equation was determined using the average value of transferrin saturations for each decile from blood samples drawn in the morning as the dependent variable. Predictors included an indicator variable for gender and the average value of transferrin saturations by decile from blood samples drawn in the afternoon. Similarly, a separate regression equation was formed for use with transferrin saturation values from blood samples drawn in the evening. Third, the predicted average morning transferrin saturation value was then computed for samples drawn in the afternoon or evening using the appropriate regression equation. To form the frequency distributions of transferrin saturation values for each decile from blood samples drawn in the morning, the expected values had the blood samples been drawn in the morning. Similarly, the predicted morning transferrin saturation value was then computed for samples drawn in the afternoon or evening. The exclusions were applied sequentially, in the order shown.

Adjustment of the data set to account for a possible admixture of Caucasian HLA-linked hemochromatosis genes. The data were modified to take into account the possibility that the distribution of transferrin saturations from African Americans is affected by individuals who are heterozygotes or homozygotes for HLA-linked hemochromatosis. The gene frequency for the HLA-linked hemochromatosis locus in the Caucasian population is estimated to be .067. Assuming a 25% admixture of Caucasian genes in the African-American population, the gene frequency for the HLA-linked hemochromatosis locus in African Americans would be 25% of .067 or .017. Population genetics would then project for African Americans a proportion of homozygotes and heterozygotes for the HLA-linked hemochromatosis gene of 3 per 10,000 (.017 squared) and a proportion of heterozygotes of 33 per 1,000. We assumed that the transferrin saturation values from these projected African American heterozygotes for the HLA-linked hemochromatosis locus would be normally distributed with the same mean and SD as found in our previous study of Caucasian Americans in the second National Health and Nutrition Examination Survey (NHANES II). Under this assumption, we removed 3.3% of the 836 values (n = 28) from the data set that were closest to 28 random values generated from a normal distribution with a mean of 45.5% and a SD of 7%. In the case when several transferrin saturation values were equidistant to a random value, one of these values was selected randomly. Because the number of transferrin saturations arising from heterozygotes for HLA-linked hemochromatosis was projected to be less than one, no further adjustment was made to the data.

Distribution of transferrin saturation values in African Americans. We examined the distributions of adjusted transferrin saturation values for the remaining 808 men and women African Americans in the unweighted data using techniques developed for the analysis of distributions in grouped, truncated data. Transferrin saturation values were sorted into intervals and the frequency of values within each interval was computed. The physiologic models we considered were a single normal distribution and a mixture of two or three normal distributions. We have previously established that transferrin saturations in a homogeneous population follow a normal distribution. The expectation-maximization algorithm was applied to the distributions of transferrin saturation values for parameter estimation. Equal variances were assumed for fit to mixtures of normal distributions because models with unequal variances resulted in increased variances for the subpopulation with the highest transferrin saturation that were biologically implausible. The statistical test used to determine the best fitting model was based on the likelihood ratio statistic. For each observed distribution, the maximized log-likelihood function for a mixture of three normal distributions was evaluated (Log L1) and compared with the maximized log-likelihood function (Log L0) for either a single normal distribution or a mixture of two normal distributions. Significance of the likelihood ratio statistic, 2(Log L1-Log L0), was assessed by referring to the x^2 distribution with degrees of freedom based on the difference between the number of parameters being estimated under each model. A significance level of .05 was used. Similarly, the maximized log-likelihood function for a mixture of three normal distributions was compared with that of a single normal distribution and to that of two normal distributions. The x^2 statistic was then used to test goodness of fit of each observed distribution to the best fitting model. For the three-population model, the methods of Crump and Howe were used to compute confidence intervals for the proportion with the highest mean.

Weighting the results to reflect the African-American population as a whole. The assumptions underlying our analysis were that transferrin saturation values are independent and identically distributed, i.e., each observation has an equal chance of being selected, and that all observations come from the same distribution. However, because individuals in the NHANES II sample did not have an equal probability of selection, sample weights must be used to calculate parameter estimates that reflect the United States population. For NHANES II, a multistage estimation procedure was used to calculate sample weights so that point estimates would reflect the United States population. The methods described above were used to compute parameter estimates from the weighted transferrin saturation distribution to reflect results for African-American men and women in the United States population fitting our exclusion criteria. It was not possible to adjust the variance estimates to account for the complex design of NHANES II.

Estimation of gene frequency for a possible locus that influences iron status in African Americans. Accepting the possibility that our findings might represent the presence of a locus that influences iron metabolism among African Americans, we assumed that the three normal distributions of transferrin saturation in our analyses would represent predominantly a subpopulation of normal homozygotes, a subpopulation of heterozygotes, and a subpopulation of affected homozygotes. We estimated the proportions of normal homozygotes, heterozygotes, and affected homozygotes as the proportions in the populations with the lowest, intermediate, and highest mean transferrin saturations, respectively. According to the Hardy-Weinberg equilibrium equation, p^2 (the proportion of normal homozygotes) + 2pq (the proportion of heterozygotes) + q^2 (the proportion of abnormal homozygotes) = 1. We estimated the gene frequency of the abnormal allele (q) as the square root of the proportion from the population with the highest mean transferrin saturation. We estimated the gene frequency of the normal homozygotes (p) as the square root of the proportion from the population with the lowest mean transferrin saturation. We then
examined whether the modeled distributions were consistent with population genetics for a major locus effect in which \( p + q = 1 \).

**RESULTS**

**Analysis of unweighted data.** The primary analysis was performed on transferrin saturations for 808 individuals. The transferrin saturations had been adjusted for sex and diurnal variation and the data set had been adjusted for the possible presence of an HLA-linked hemochromatosis allele as described above. The unweighted data showed a significantly better fit to three normal populations than to two normal populations (likelihood ratio statistic, 32.1 with 3 degrees of freedom; \( P < .0001 \)) or to a single normal population (likelihood ratio statistic 123.1 with 7 degrees of freedom, \( P < .0001 \)). The fit of the data to a mixture of three normal populations was good (\( P = .20 \); in a goodness-of-fit analysis, a \( P \) value above .1 indicates an acceptable fit). Figure 1A shows the observed and fitted distributions for unweighted transferrin saturation values. An estimated proportion of .856 of the African Americans studied were included in a subpopulation with a mean saturation of 24.6%, while .135 comprised a subpopulation with an intermediate mean saturation of 38.0% and .009 (95% confidence interval of .004 to .017) formed a subpopulation with a mean saturation of 63.4% (Table 2). These proportions are consistent with population genetics for a single major locus affecting the distribution of transferrin saturations: the sum of the square roots of the proportion with the lowest transferrin saturation (\( p = .925 \)) and of the proportion with the highest saturation (\( q = .093 \)) is approximately 1 (1.018).

**Weighted results for the African American population.** The results after the data (adjusted for sex, diurnal variation, and the potential presence of HLA-linked hemochromatosis) were weighted to reflect the United States population of African Americans as a whole are given in Table 2. The weighted results are similar to our primary results using unweighted data, which suggests that the unequal probability of selection in NHANES II did not have a major input on the transferrin saturation distribution. An estimated proportion of .815 was included in a subpopulation with a mean saturation of 24.3%, while .172 comprised a subpopulation with an intermediate mean saturation of 39.1% and .012 formed a subpopulation with a mean saturation of 61.0%. The weighted findings are also consistent with population genetics for a single major locus affecting the distribution of transferrin saturations: the sum of the square roots of the proportion with the lowest transferrin saturation (\( P = .925 \)) and of the proportion with the highest saturation (\( q = .093 \)) is approximately 1 (1.018).

**DISCUSSION**

Our analysis of transferrin saturations from African Americans studied in the second National Health and Nutrition Examination Survey showed that three subpopulations of individuals could be detected. Using data that were adjusted for the time of day of collection of the blood sample, for gender, and for the possible presence of an HLA-linked hemochromatosis allele, we found that one subpopulation comprised an...
estimated proportion of .856 of the African Americans studied and had a mean transferrin saturation of 24.6%. A second subpopulation was made up of an estimated proportion of .136 and had a mean transferrin saturation of 38.0%. A third subpopulation comprised a proportion of .009 and had a mean transferrin saturation of 63.4%. Limitations to this analysis are that transferrin saturations were single rather than repeated determinations, that subjects were not screened for liver diseases that may be associated with elevations of transferrin saturation, and that the sample size was small for a population study. In addition, we were not able to fully account for the complex survey design when estimating variances.

The three subpopulations that we identified based on this mixture-modeling statistical analysis of transferrin saturation data from African Americans are consistent with the presence of a major locus that influences iron status. If we assume that the members of the small subpopulation with the highest mean transferrin saturation may be abnormal homozygotes for a locus that influences iron metabolism, then the abnormal gene frequency in this data set would be estimated to be .093. Similarly, if we assume that the members of the large subpopulation with the lowest mean transferrin saturation may be normal homozygotes for a locus that influences iron metabolism, then the normal gene frequency would be .925. Consistent with population genetics for a single major locus, the sum of the estimated normal and abnormal gene frequencies is approximately 1 (1.018).

If our present findings do reflect the presence in the African-American population of an abnormality of a major locus that influences iron metabolism, then this locus may be either (1) the HLA-linked hemochromatosis gene that has heretofore been described exclusively in people of European ancestry\textsuperscript{15,24} or (2) some different gene that influences iron metabolism. The fact that HLA-linked hemochromatosis has only been reported in Caucasians and that a genetic iron-loading disorder that is not linked to the HLA-locus may be common in Africa\textsuperscript{4} are supportive of the possibility that we are observing the effects of a genetic alteration that is unique to people of African ancestry. There is an admixture of Caucasian genes in the African-American population, estimated to be about 25\%,\textsuperscript{19} but such an admixture would not be sufficient to explain the findings of the present study. The gene frequency of HLA-linked hemochromatosis in the United States white population is estimated to be .067\textsuperscript{5} and an estimated 25% admixture of Caucasian genes could lead to an estimated gene frequency of only .017 for HLA-linked hemochromatosis in the African-American population. Furthermore, we made an adjustment to account for the probable presence of HLA-linked hemochromatosis heterozygotes in the present study.

Until recently, only rare mention has been made in the literature of African Americans with “hemochromatosis”\textsuperscript{25-29} Two recent reports underscore the fact that primary iron overload does occur among African Americans.\textsuperscript{30,31} Furthermore, one study raises that possibility that the condition may not be rare.\textsuperscript{31} The present statistical study of the distribution of transferrin saturation values is compatible with the possibility that an alteration in a gene that influences iron metabolism is present among African Americans. Clinicians should consider the diagnosis of primary iron overload in African-American patients and both treat the condition and screen family members of affected subjects. Investigations to define the prevalence, clinical consequences, and causes of primary iron overload in African Americans are needed.

REFERENCES

18. Hamilton LD, Gubler CJ, Cartwright GE, Wintrobe MM:


Distribution of Transferrin Saturations in the African-American Population

Victor R. Gordeuk, Christine E. McLaren, Anne C. Looker, Victor Hasselblad and Gary M. Brittenham