Prevalence of Heterozygotes for Hemochromatosis in the White Population of the United States


In previous studies, the prevalence of HLA-linked hemochromatosis, thought to be the most common genetic illness in whites, has been estimated by identifying homozygotes in the population. Because not all homozygotes express the disease phenotypically, the accuracy of these estimates is uncertain. We analyzed the distribution of transferrin saturation values in the second National Health and Nutrition Examination Survey to estimate the prevalence of hemochromatosis heterozygotes in the US population. After removing values for possible homozygotes, two populations were present \( (P < .01 \text{ for each gender}) \). When weighted to reflect the US adult white male population as a whole, a proportion of 850 per 1,000 (95% confidence interval, 0.81 to 0.89) were included in a population with a lower mean saturation of 29.7% (29.1% to 30.3%), whereas 150 per 1,000 (0.11 to 0.19) comprised a population with a higher mean saturation of 47.0% (45.1% to 49.0%). Similar results were found for the female population. The gene frequencies were estimated to be 0.081 from the male population and 0.070 from the female population corresponding to prevalences of homozygotes of 6.6 and 4.8 per 1,000, respectively. Our results confirm that the gene for hemochromatosis is common. © 1995 by The American Society of Hematology.

O NCE CONSIDERED to be a rare disorder, hemochromatosis is now recognized to be one of the most common autosomal recessive disorders in white populations. Substantial morbidity and mortality may result from untreated hemochromatosis.\(^1\) Although it is unusual for heterozygotes for hemochromatosis without underlying hematologic disorders to develop iron overload sufficient to cause overt organ damage,\(^2\) marked iron overload has developed in patients with idiopathic refractory sideroblastic anemia, hereditary spherocytosis, pyruvate kinase deficiency,\(^3,4,6\) or sporadic porphyria cutanea tarda who are also heterozygotes for hemochromatosis.\(^7\)

Surveys to determine the gene frequency of hereditary hemochromatosis have been based predominantly on the principle of identifying homozygotes in the population. Such studies conducted in Australia, England, Finland, and Sweden over the past 10 years have estimated the prevalence of homozygosity for hemochromatosis at 0.5 per 1,000 to 11.5 per 1,000, corresponding to gene frequencies of 0.022 to 0.107.\(^8,9\) A recent large study of predominantly white blood donors in Utah estimated that the prevalence of homozygotes was 4.5 per 1,000, corresponding to the hemochromatosis gene frequency of 0.067 and to a proportion of 125 per 1,000 for heterozygotes.\(^10\) If the results from Utah are applicable nationwide, more than 1 million Americans are hemochromatosis homozygotes who are at risk for major iron overload and more than 25 million are heterozygotes. Such a common condition might have important implications for disease screening and for policies regarding the fortification of food with iron.

Since the discovery that the gene for hemochromatosis is linked to the HLA locus,\(^11,12\) it has been possible through HLA typing of affected families to assign the homozygous affected, heterozygous, or homozygous normal genotype to family members. Transferrin saturation is regarded as the best single screening test for hemochromatosis, with a sustained level of greater than 62%, strongly suggestive of the homozygous state.\(^13,14\) Examination of family members has shown that mean transferrin saturation values for heterozygotes are higher than among unaffected subjects but lower than homozygotes. For example, in four studies the mean values for transferrin saturation ranged from 64% to 95% in homozygotes, from 35% to 48% in heterozygotes, and from 29% to 36% in individuals unaffected by the hemochromatosis gene.\(^15-18\)

If 12.5% of the US white population is heterozygous for the hemochromatosis gene, as suggested by the Utah study,\(^19\) then the higher mean transferrin saturation of heterozygotes might affect the distribution of transferrin saturation in the population as a whole. We postulated that the distribution of transferrin saturation in the United States reflects several populations based on individual genotype for hemochromatosis and that a statistical technique to separate finite mixtures of distributions could be used to quantify these groups.\(^20\) We examined the distribution of transferrin saturation in the second National Health and Nutrition Examination Survey (NHANES II) to determine the proportion of hemochromatosis heterozygotes and the proportion of unaffected individuals in the population and to obtain an indepen-

From the Department of Mathematics, Moorhead State University, Moorhead, MN; the Division of Hematology and Oncology, The George Washington University Medical Center, Washington, DC; the Division of Health Examination Statistics, National Center for Health Statistics, Hyattsville, MD; the Center for Health Policy Research and Education, Duke University, Durham, NC; the Department of Internal Medicine, University of Utah College of Medicine and LDS Hospital, Salt Lake City, UT; and the Department of Medicine, MetroHealth Medical Center, Case Western Reserve University, Cleveland, OH.

Submitted December 1, 1994; accepted April 27, 1995.

Supported in part by Grants No. R01DK20630 and M01-RR00064. Computing equipment used in analyzing the data was provided by Instrumentation and Laboratory Improvement Grant No. USE-8851944 awarded by the National Science Foundation to the Department of Mathematics at Moorhead State University.

Address reprint requests to Christine E. McLaren, PhD, Department of Mathematics, Moorhead State University, Moorhead, MN 56563.

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.

© 1995 by The American Society of Hematology.

0006-4971/95/8605-0037$3.00/0
dent estimate for the frequency of the hemochromatosis gene in the United States based on these proportions.

MATERIALS AND METHODS

Sources of data. The primary source of data for this study was NHANES II, which surveyed a representative sample of the noninstitutionalized US population 6 months to 74 years of age in 1976 to 1980. Approximately 21,000 persons were examined from 64 primary sampling units (county or small group of contiguous counties). Based on an interviewer's observation, each person was classified by race as white, black, or other. Serum iron and total iron binding capacity were measured using a modification of the Automated Technicon AAI-25 method, which is a colorometric method using ferroin, and transferrin saturation was calculated from these values.

Selection criteria. For the analysis, we selected transferrin saturations from 4,748 white men and 5,151 nonpregnant white women 20 to 74 years of age. Of these, we then selected transferrin saturations computed from blood samples that had been drawn in the morning and for whom the mean red blood cell volume was between 80 and 100 fl and the erythrocyte protoporphyrin level was less than 70 mg/dL red blood cells. Additional selection criteria included transferrin saturation values from men for whom the hemoglobin level was $\geq 14$ g/dL and the hematocrit level was $\geq 42\%$ and from women for whom the hemoglobin level was $\geq 12$ g/dL and the hematocrit level was $\geq 36\%$. NHANES II was not designed to provide estimates for Hispanics; based on self-reported family ancestry, only 112 Hispanics were included in the sample used in this study. We excluded African-Americans from the analysis because HLA-linked hemochromatosis has only been described in populations derived from Europe and it is not clear that the hemochromatosis gene is found in populations of African origin. We excluded subjects whose blood samples were obtained in the afternoon and evening because transferrin saturation has a diurnal variation and including samples obtained at different times of the day might alter the distribution of transferrin saturation. 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 low mean corpuscular volume levels can be associated with iron deficiency and inflammation and high mean corpuscular volume levels can be associated with megaloblastic conditions and drug effects, all of which can lead to altered transferrin saturations. We excluded subjects with elevated erythrocyte protoporphyrin levels because of the association with iron deficiency and inflammation.

An objective of our study was to determine if data obtained for NHANES II support the theory that in whites the distribution of transferrin saturation values is affected by the presence of hemochromatosis heterozygotes. Given the limited sample size after applying the selection criteria, it would be difficult to detect a small subpopulation composed of hemochromatosis heterozygotes. We excluded 24 white men and 15 white women with transferrin saturation values $\geq 62\%$ to ensure that the final data set of 1,325 men and 1,547 women that would include mostly heterozygotes and normal homozygotes. Table 1 gives the number of subjects excluded using specific criteria.

<table>
<thead>
<tr>
<th>Exclusion</th>
<th>Number of Subjects</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total sample</td>
<td>4,748</td>
</tr>
<tr>
<td>Excluded for</td>
<td></td>
</tr>
<tr>
<td>Abnormal MCV</td>
<td>294</td>
</tr>
<tr>
<td>Abnormal protoporphyrin</td>
<td>24</td>
</tr>
<tr>
<td>Abnormal hemoglobin</td>
<td>451</td>
</tr>
<tr>
<td>Abnormal hematocrit</td>
<td>356</td>
</tr>
<tr>
<td>Blood drawn after noon</td>
<td>2,054</td>
</tr>
<tr>
<td>Probable homozygotes</td>
<td>24</td>
</tr>
<tr>
<td>Final analytic sample</td>
<td>1,325</td>
</tr>
</tbody>
</table>

Table 1. Exclusions From the Total Sample of White Men and Women 20 to 74 Years of Age

Abbreviation: MCV, mean corpuscular volume.

In a separate study, data were collected on 232 index subjects with hemochromatosis from well-characterized Utah hemochromatosis pedigrees. Approximately 99% of the individuals studied were of European ancestry. Three different genotypes, determined by a one-locus, two-allele model, were proposed, ie, homozygous normal, heterozygous for the hemochromatosis allele, and homozygous for the hemochromatosis allele. Pedigree members were assigned genotypes on the basis of none, one, or two HLA haplotypes in common with the index subject. Four hundred thirty-three members of these 232 pedigrees did not share an HLA haplotype with the index subjects and were assigned the normal genotype. For these 433 individuals, it was not possible to apply the same inclusion criteria that we applied to the NHANES II data because a complete blood count was not determined for each subject. Because premenopausal women may have altered transferrin saturation values because of mild iron deficiency, women were not considered in this analysis. Data from the remaining 182 men were analyzed to determine if transferrin saturation values followed a normal or log normal distribution in a homogeneous population. Serum iron concentration and total iron binding capacity were measured using atomic absorption spectroscopy and transferrin saturation values were calculated from the results.

Using the computer program DISFIT, the observed, grouped frequency distribution from 182 men was fit to a single normal distribution and to a single lognormal distribution. The expectation-maximization algorithm, an iterative distribution fitting procedure, was used for parameter estimation. The maximized log-likelihood function under the null hypothesis of a single normal distribution was evaluated and compared with the maximized log-likelihood function under the alternative hypothesis of a single lognormal distribution using the likelihood ratio statistic. A resampling technique was used to examine the distribution of the statistic under the null hypothesis. The likelihood ratio statistic for the observed data (12,15) was near the median value (10.71) of the distribution of resampled likelihood ratio statistics. Thus, analysis of the observed data was consistent with the null hypothesis of a normal model for transferrin saturation ($P = .30$).

Distribution of transferrin saturation values in NHANES II. We examined the distributions of transferrin saturation values for the remaining 1,325 white men and 1,547 women from the unweighted NHANES II sample data using techniques developed for analysis of distributions in grouped, truncated data. Because the mean values of transferrin saturations differ significantly between men and women, we performed separate distribution analyses for each sex. We made no adjustments for age differences because there was no...
significant correlation between age and transferrin saturation for either sex (Pearson's $r = -0.19$, $P = .49$ for men; $r = -0.14$, $P = .58$ for women). To determine each grouped frequency distribution, transferrin saturation values were sorted into intervals of 1% and the frequency of values within each interval was computed. Because we had found that transferrin saturation is normally distributed in normal homozygotes, the physiologic models we considered were a single normal distribution and a mixture of two normal distributions. The expectation-maximization algorithm was applied to the distributions of transferrin saturation values for parameter estimation.23,26,34,43

The statistical test used to determine the best fitting model was based on the likelihood ratio statistic. For mathematical tractability of the distribution fitting algorithm, equal variances were assumed for fit to a mixture of two normal distributions. For each observed distribution, the maximized log-likelihood function for a mixture of normal distributions was evaluated ($Log L_0$) and compared with the maximized log-likelihood function for a single normal distribution ($Log L_1$). Significance of the likelihood ratio statistic ($-2 \log (L_0/L_1)$) was assessed using a resampling technique.44 A $P$ value less than .05 indicated a better fit to a mixture 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.

We examined the sensitivity of results to exclusion of transferrin saturation values $\geq 62\%$. The same statistical methods were used to fit a single normal distribution and a mixture of two normal to truncated data from 1,349 men and 1,552 women. Separate data sets were analyzed after systematic exclusion of transferrin saturation values from $\geq 60\%$ to $\geq 65\%$ in increments of 1%.

Weighting the results to reflect the US population. A multistage estimation procedure was used to calculate subsample sample weights so the subsample data could be adjusted to reflect the US population.23,45 To derive parameter estimates that reflect the US population as a whole, secondary analyses were performed on the weighted distributions of transferrin saturation from NHANES II. These weighted distributions reflected the transferrin saturation values from white men and white women in the US population fitting our inclusion criteria. The methods described above were used to compute parameter estimates and form confidence intervals for distribution parameters using the sample weights.

Estimation of gene frequency. In previous population studies, the Hardy-Weinberg equilibrium equation ($q^2 + 2pq + p^2 = 1$) has been used to calculate the gene frequency and prevalence of hemochromatosis, assuming that the measured proportion of affected homozygotes in the population ($q^2$), and deriving the abnormal gene frequency ($q$) mathematically by taking the square root of $q^2$.46 In the present study, we excluded 39 subjects who had transferrin saturation values $\geq 62\%$ to remove the probable hemochromatosis homozygotes; thus we derived a scaled value for $p^2$ to approximate the proportion of normal homozygotes in the full population. This scaled value was derived by letting $p^2/(p^2 + 2pq)$ be equal to the estimated proportion of normal homozygotes in the truncated population and solving the resulting equation for $p^2$. An adjusted estimate of the gene frequency was then found from the equation $q = 1 - p$, and $q^2$ gave an adjusted estimate of the prevalence of homozygotes. Separate scaled estimates of the proportion of normal homozygotes and corresponding adjusted estimates of the gene frequency and the prevalence of homozygotes were made for men and women. We examined the effect of excluding individuals from the sample on gene frequency estimation. The weighted sample data sets were analyzed after systematic exclusion of transferrin saturation values from $\geq 60\%$ to $\geq 65\%$ in increments of 1% and the adjusted estimates of gene frequency were calculated for men and women. An alternative method would be to keep transferrin saturations $\geq 62\%$ in the data set and fit a mixture of three normal distributions, constraining the proportions is each component to satisfy the Hardy-Weinberg equilibrium equation. The accuracy of this alternative method would depend in part on the sparseness of data occurring at high transferrin saturation values.

RESULTS

Distribution of transferrin saturation values in NHANES II. The final analyses were performed on transferrin saturations for 1,325 men and 1,547 women. For unweighted sample data from NHANES II, analysis of the distribution of transferrin saturation showed a significantly better fit to two normal populations than to a single normal population for both men (likelihood ratio statistic, 85.3; $P < .01$) and women (likelihood ratio statistic, 108.3; $P < .01$). The goodness of fit to a mixture of two normal populations was excellent for men ($P = .813$) and good for women ($P = .177$). Figure 1 shows the observed and fitted distributions for transferrin saturation values according to sex. Table 2 gives estimates for the proportion of subjects included in each subpopulation and the mean transferrin saturation of each subpopulation. When proportions of subjects in corresponding subpopulations were compared for men and women they were similar ($P = .24$ for both subpopulations), but the mean transferrin saturations were significantly different ($P < .0001$).

The sensitivity of the results to exclusion of transferrin saturation values $\geq 62\%$ was examined. For men, data sets were analyzed after systematic exclusion of transferrin saturation values from $\geq 60\%$ to $\geq 65\%$ in increments of 1%. Among men, the estimated proportion of subjects in a subpopulation with the lower mean saturation increased from 0.82, when values $\geq 60\%$ were excluded, to 0.86, when values $\geq 65\%$ were excluded. The mean $(\pm SD)$ transferrin saturation values remained stable with estimates from a minimum of 29.2% $\pm$ 7% to a maximum of 29.9% $\pm$ 7%. The corresponding estimated proportion of subjects in a subpopulation with the higher mean decreased from 0.18 to 0.14 as fewer transferrin saturation values were excluded, and the mean $(\pm SD)$ transferrin saturation values increased slightly, with estimates from a minimum of 45.1% $\pm$ 7% to a maximum of 48.5% $\pm$ 7%.

Among women, the estimated proportion of subjects in a subpopulation with the lower mean transferrin saturation increased from 0.84, when values $\geq 60\%$ were excluded, to 0.87, when values $\geq 65\%$ were excluded. The mean $(\pm SD)$ transferrin saturation values remained stable, with estimates from a minimum of 26.5% $\pm$ 7% to a maximum of 27.0% $\pm$ 7%. The corresponding estimated proportion of subjects in a subpopulation with the higher mean decreased from 0.16 to 0.13 as fewer transferrin saturation values were excluded, and the mean $(\pm SD)$ transferrin saturation values increased slightly, with estimates from a minimum of 42.3% $\pm$ 7% to a maximum of 44.8% $\pm$ 7%.

Weighted results for the US population. The results after the data were weighted to reflect the US white population are also given in Table 2. Among men, the estimated proportion of subjects in a subpopulation with the lower mean $(\pm SD)$ transferrin saturation of 29.7% $\pm$ 7% was 0.85 (95%
percent confidence interval of .814 to .894) and the estimated proportion of subjects in a subpopulation with the higher mean transferrin saturation of 47% ± 7% was 0.15 (.106 to .186). Among women, the estimated proportion of subjects in a subpopulation with the lower mean transferrin saturation of 27% ± 7% was 0.87 (.837 to .901) and the estimated proportion of subjects in a subpopulation with the higher mean transferrin saturation of 44.7% ± 7% was 0.13 (.100 to .162).

**Gene frequency estimation.** From scaled estimates of the proportions of normal homozygotes in the truncated distribution (0.85 for men and 0.87 for women), the gene frequency

![Fig 1. Distribution of transferrin saturation values in (A) white men and (B) white women. Observed data are indicated with an asterisk. Fitted distributions are indicated with a solid line. The relative frequency indicates the proportion of subjects having transferrin saturation values within a given interval, relative to that of the interval containing the maximum frequency. Each interval the relative frequency is calculated by (frequency/maximum frequency) × 100.](image)

<table>
<thead>
<tr>
<th>Table 2. Analysis of the Distribution of Transferrin Saturation for White Men and Women</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Postulated Population Unaffected by the Gene for Hemochromatosis</strong></td>
</tr>
<tr>
<td>---------------------------------------------</td>
</tr>
<tr>
<td><strong>Men</strong></td>
</tr>
<tr>
<td><strong>A. Unweighted sample data</strong></td>
</tr>
<tr>
<td>Proportion</td>
</tr>
<tr>
<td>95% confidence interval</td>
</tr>
<tr>
<td>Mean transferrin saturation (%)</td>
</tr>
<tr>
<td>Standard deviation</td>
</tr>
<tr>
<td><strong>B. Weighted population data</strong></td>
</tr>
<tr>
<td>Proportion</td>
</tr>
<tr>
<td>95% confidence interval</td>
</tr>
<tr>
<td>Mean transferrin saturation (%)</td>
</tr>
<tr>
<td>Standard deviation</td>
</tr>
</tbody>
</table>

Values are parameter estimates (±SE) for the best fit to two normal populations.
was estimated to be 0.081 from the male population and 0.070 from the female population, corresponding to a prevalence of homozygotes of 4.8 per thousand for women and 6.6 per thousand for men. Figure 2 shows that gene frequency is fairly robust to the transferrin saturation exclusion value used. For both men and women, the estimated gene frequency decreases slightly as fewer transferrin saturation values are excluded and appears to approach an asymptote of approximately 0.070 for men and 0.058 for women.

**DISCUSSION**

Our analysis of NHANES II transferrin saturation data from white men and women, truncated to remove the relatively small proportion of potential homozygotes, demonstrates that two subpopulations of individuals could be detected. One subpopulation, comprising an estimated 85% of men and 87% of women in the US population, with mean ±SD saturations of 29.7% ± 7% and 27.0% ± 7%, respectively, may predominately include individuals who are unaffected by the gene for hemochromatosis. In contrast, the second subpopulation, comprising an estimated 15% of men and 13% of women, with abnormally high mean transferrin saturations of 47% ± 7% and 44.7% ± 7%, respectively, may be composed predominantly of individuals who are heterozygous for the hemochromatosis gene. One limitation to this analysis is that transferrin saturations were single rather than repeated determinations. Another limitation is that NHANES II did not routinely include measurements of liver enzymes. We were unable to make a valid exclusion of subjects with abnormal liver function tests and therefore our findings conceivably could be influenced by increased transferrin saturations associated with minor hepatocellular damage due to concurrent viral disease, alcohol, or drugs.

Previous studies designed to estimate the frequency of the hemochromatosis gene in the population have been based on the identification of affected homozygotes and the calculation of their prevalence in the population. In contrast, we used a novel approach to approximate the gene frequency of hemochromatosis starting with a scaled estimate of the proportion of normal homozygotes in the population followed by calculation of the gene frequency. By this method, the gene frequency is estimated to be 0.081 from the male population and 0.070 from the female population, corresponding to a prevalence of homozygotes of 4.8 per thousand for females and 6.6 per thousand for males. The estimates of gene frequency are remarkably similar to that of 0.067 calculated on the basis of the large Utah study.13

If hemochromatosis is untreated, the condition is associated with substantial morbidity and mortality, whereas with early diagnosis and treatment, morbidity is prevented and life expectancy is normal.1 Although heterozygotes for hemochromatosis do not usually develop iron overload sufficient to cause overt organ damage,2 heterozygotes for HLA-linked hemochromatosis who also have idiopathic refractory sideroblastic anemia, hereditary spherocytosis, or pyruvate kinase deficiency may develop marked iron overload3-4 and the heterozygous state has been implicated in the pathogenesis of sporadic porphyria cutanea tarda.5 Our present study provides an independent estimate indicating that hemochromatosis is common in the US white population. We conclude that the homozygous condition is sufficiently prevalent to warrant increased efforts for early diagnosis.13,17

**ACKNOWLEDGMENT**

We thank Jim E. Askin and Wanxi Chen for technical assistance with statistical and graphical analyses.

**REFERENCES**

5. Mohler DN, Wheby MS: Hemochromatosis heterozygotes may have significant iron overload when they also have hereditary spherocytosis. Am J Med 292:320, 1986
Prevalence of heterozygotes for hemochromatosis in the white population of the United States