Bivariate mixture modeling of transferrin saturation and serum ferritin concentration in Asians, African Americans, Hispanics, and whites in the Hemochromatosis and Iron Overload Screening (HEIRS) Study

CHRISTINE E. MCLAREN, VICTOR R. GORDEUK, WEN-PIN CHEN, JAMES C. BARTON, RONALD T. ACTON, MARK SPEECHLEY, OSWALDO CASTRO, PAUL C. ADAMS, BEVERLY M. SNIVELY, EMILY L. HARRIS, DAVID M. REBOUSSIN, GEOFFREY J. MCLACHLAN, and RICHARD BEAN, for the Hemochromatosis and Iron Overload Screening Study Research Investigators

IRVINE, CALIF; WASHINGTON, DC; ORANGE, CALIF; BIRMINGHAM, ALA; LONDON, ONTARIO, CANADA; WINSTON-SALEM, NC; PORTLAND, ORE; AND BRISBANE, AUSTRALIA

Bivariate mixture modeling was used to analyze joint population distributions of transferrin saturation (TS) and serum ferritin concentration (SF) measured in the Hemochromatosis and Iron Overload Screening (HEIRS) Study. Four components (C1, C2, C3, and C4) with successively age-adjusted increasing means for TS and SF were identified in data from 26,832 African Americans, 12,620 Asians, 12,264 Hispanics, and 43,254 whites. The largest component, C2, had normal mean TS (21% to 26% for women, 29% to 30% for men) and SF (43–82 μg/L for women, 165–242 μg/L for men), which consisted of component proportions greater than 0.59 for women and greater than 0.68 for men. C3 and C4 had progressively greater mean values for TS and SF with progressively lesser component proportions. C1 had mean TS values less than 16% for women (<20% for men) and SF values less than 28 μg/L for women (<47 μg/L for men). Compared with C2, adjusted odds of iron deficiency were significantly greater in C1 (14.9–47.5 for women, 60.6–3530 for men), adjusted odds of liver disease were significantly greater in C3 and C4 for African-American women and all men, and adjusted odds of any *HFE* mutation were increased in C3 (1.4–1.8 for women, 1.2–1.9 for men) and in C4 for Hispanic and white women (1.5 and 5.2, respectively) and men (2.8 and 4.7, respectively). Joint mixture modeling identifies a component with lesser SF and TS at risk for iron deficiency and 2 components with greater SF and TS at risk for liver disease or *HFE* mutations. This approach can identify populations in which hereditary or acquired factors influence metabolism measurement. (Translational Research 2007;xx:xxx)

Abbreviations: EM = expectation-maximization; *HFE* = hemochromatosis gene on chromosome 6p; HEIRS = Hemochromatosis and Iron Overload Screening; SF = serum ferritin concentration; TS = transferrin saturation; wt = wild type

From the Department of Epidemiology, University of California, Irvine, Calif; Department of Medicine, Howard University, Washington, DC; Chao Family Comprehensive Cancer Center, Orange, Calif; Southern Iron Disorders Center, Birmingham, Ala; Departments of Microbiology, Medicine, and Epidemiology and International Health, University of Alabama at Birmingham, Birmingham, Ala; Department of Epidemiology and Biostatistics, University of Western Ontario, London, Ontario, Canada; Department of Medicine, London Health Sciences Center, London, Ontario, Canada; Department of Public Health Sciences, Wake Forest University School of Medicine, Winston-Salem, NC; Kaiser Permanente Center for Health Research, Portland, Ore; and Department of Mathematics, The University of Queensland, Brisbane, Australia.

The HEIRS Study was initiated and funded by NHLBI in conjunction with NHGRI. Supported by contracts N01-HC-05185 (University of Minnesota), N01-HC-05186 (Howard University), N01-HC-05188 (University of Alabama at Birmingham), N01-HC-05189 (Kaiser Permanente Center for Health Research), N01-HC-05190 (University of California, Irvine), N01-HC-05191 (London Health Sciences Centre), and N01-HC-05192 (Wake Forest University). Also supported by Grant R01 HL083328-01A1 from the National Heart, Lung, and Blood Institute (to C.E.M.); Grant M01-RR00032 from the University of Alabama at Birmingham General Clinical Research Center (GCRC); the Southern Iron Disorders Center (to J.C.B.); Grant M01-RR10284 from Howard University GCRC; Howard University Research Scientist Award UH1-HL03679-05 from the National Heart, Lung, and Blood Institute and the Office of Research on Minority Health (to V.R.G.); and Grant UC Irvine M01RR 00827-29 from the General Clinical Research Centers Program of the National Center for Research Resources National Institutes of Health.

Submitted for publication July 27, 2007; revision submitted October 10, 2007; accepted for publication October 11, 2007.

Reprint requests: Christine E. McLaren, Epidemiology Division, Department of Medicine, University of California, Irvine, 224 Irvine Hall, Irvine, Calif 92697-7550; e-mail: cmclaren@uci.edu.

1931-5244/$ – see front matter

doi:10.1016/j.trsl.2007.10.002

AT A GLANCE COMMENTARY

Background

This study evaluated the association among levels of iron measures, transferrin saturation (TS) and serum ferritin concentration (SF), and the presence of *HFE* mutations, self-reported liver disease, and iron deficiency in multiethnic populations. Mixture modeling identified components with decreased SF and TS at risk for iron deficiency, a component with normal mean SF and TS, and 2 components with increased SF and TS at risk for liver disease or *HFE* mutations.

Translational Significance

This approach identifies populations in which hereditary or acquired factors influence metabolism measurement, potentially complementing and enhancing genetic testing for assessment of disease complications.

TS and SF are iron measures, the levels of which are influenced by iron stores, a variety of inflammatory and neoplastic disorders, and inheritance of alleles in iron-related genes. C282Y is a common missense mutation of the *HFE* gene,[1] which is detected typically in whites of northern European ancestry; however, the *HFE* H63D allele occurs in most race/ethnic groups worldwide.[2] In Caucasian participants enrolled in a screening study conducted at the Kaiser Permanente San Diego Health Appraisal Clinic, components of TS identified by mixture modeling corresponded to distributions of *HFE* genotypes.[3] The HEIRS Study is a large, multicenter screening study in which TS, SF, and *HFE* mutations were measured for each participant.[4,5] Analyses of phenotypic and genotypic data collected from 44,136 non-Hispanic Caucasian participants in the HEIRS Study enrolled in different geographic regions demonstrated a strong association between the *HFE* genotype and the TS subpopulations. The analyses also confirmed the validity of the mixture modeling approach when applied to a convenience sample of patients observed at primary care clinics and at blood drawing facilities.[6]

In contrast with previous investigations that modeled the univariate distribution of TS, we now report detailed analyses of the bivariate distribution of TS and SF values from African Americans, Asians, Hispanics, and non-Hispanic whites enrolled in the HEIRS Study. The following hypotheses were evaluated in these analyses: 1) The residual distribution of TS and natural logarithm of SF can be modeled as a mixture of normal distributions for each gender after removal of the effects of age and study site; 2) it is possible to identify components with progressively increasing means of TS and SF, standardized to the median age that reflects major locus and environmental effects; and 3) for some race/ethnicities, an association exists between TS and SF components and presence of *HFE* mutations, self-reported liver disease, and iron deficiency. The statistical issues that have been addressed include the need for appropriate adjustment for known sources of variation in biologic markers; fitting a bivariate normal model to the data; estimation of the mixing proportion, mean values for TS and SF, and 95% confidence ellipses within each subpopulation; and evaluation of the association between levels of TS and SF and the presence of *HFE* mutations, self-reported liver disease, and iron deficiency.

METHODS

Sources of data. The source of data was the HEIRS Study, the goals of which were to evaluate the prevalence; genetic and environmental determinants; and potential clinical, personal, and societal impact of hemochromatosis and iron overload in a multiethnic, primary care–based sample of 101,168 adults over a 7-year period. Participants, who were at least 25 years old, were recruited from 5 field centers, 4 in the United States and 1 in Canada. The research was conducted according to the principles of the Declaration of Helsinki. Informed consent was obtained, and the study protocol was approved by the Institutional Review Board of each field center.

Laboratory screening tests included spectrophotometric serum iron and unsaturated iron binding capacity, turbidometric immunoassay of SF (Roche Applied Science/Hitachi 911, Indianapolis, Ind) and calculated TS on nonfasting blood samples. The central laboratory, located at the University of Minnesota Medical Center (Fairview, Minneapolis, Minn), performed all laboratory screening tests, with the exception of TS testing of the Canadian participants. These tests were performed at MDS Laboratory Services (Buraby, British Columbia, Canada) using an identical method. *HFE* C282Y and H63D alleles were determined from spots of whole blood with the use of a modification of the Invader assay (Third Wave Technologies, Madison, Wis) that increases the allele-specific fluorescent signal by including 12 cycles of locus-specific polymerase chain reaction before the cleavase reaction. Absence of a detectable C282Y or H63D mutation was designated as wt. Race/ethnicity was self-reported on an initial screening form; participants indicated whether they were of Hispanic or Latino heritage, and they could

mark as many race/ethnicity categories as necessary to describe their background. We examined data from participants who identified themselves as white or Caucasian only; black or African-American only at US field centers or black, African, Haitian, Jamaican, or Somal in Ontario only; Spanish, Latino, or Hispanic heritage, irrespective of additional racial/ethnic identification; or Asian only. SF varies with age and sex,[7,8] and TS is affected by diurnal and day-to-day variability.[9,10] In the HEIRS Study, thresholds for increases in TS and SF were TS greater than 50% and SF greater than 300 μg/L for men and TS greater than 45% and SF greater than 200 μg/L for women. Other details of study design, laboratory testing, data management, and analysis are described elsewhere.[5]

Selection criteria. We analyzed data from African-American (n = 26,832), Asian (n = 12,620), Hispanic (n = 12,264), and non-Hispanic white (n = 43,254) men and women for whom complete data on TS, SF, *HFE* genotypes, and age were available and for whom the age in years was specified. The 6 *HFE* genotypes were designated by the mutations present, or wt if neither mutation was found, as follows: wt/wt, H63D/wt, C282Y/wt, C282Y/H63D, H63D/H63D, and C282Y/C282Y. Data from participants who reported that they had been told previously by a doctor that they had hemochromatosis, iron overload, or increased iron in the body were excluded because of the possible decreasing of TS or SF as a consequence of treatment. Data from participants who reported hearing about the study exclusively from a family member were excluded to control for possible selection bias. The detection thresholds of the laboratory instruments were 3% for TS and 15 μg/L for SF. Values below these detection thresholds were imputed as 1.5% for TS and 7.5 μg/L for SF. The final samples for TS modeling consisted of observations of 94,970 participants (59,692 women and 35,278 men).

Adjusted TS and SF concentration. To remove known sources of variation, TS and SF values were adjusted for age and field center using separate multiple linear regression analyses for each gender and racial/ethnic group. Without appropriate adjustment, parameter estimates (eg, mean and variance for mixture components) might be biased. Because a nonlinear relationship existed between TS and age, the method of restricted cubic splines was used to estimate terms that represent age as predictors for TS.[11] Multiple linear regression for the outcome variable TS was then applied with predictors, including spline terms for age and dummy variables created for field center, which are expressed relative to the field center where mean TS was least. For each individual, the value of the regression residual was calculated and the adjusted TS was computed as the sum of the regression residual and a constant. The constant was calculated as the predicted TS at the median age with equal weights applied to parameters for field center. Median ages were 47 years for African-American women, 50 years for Asian women, 42 years for Hispanic women, 52 years for non-Hispanic Caucasian women, 48 years for African-American men, 52 years for Asian men, 43 years for Hispanic men, and 55 years for non-Hispanic Caucasian men.

The distributions of SF values were skewed. Box–Cox transformations were applied, and the transformation that indicated normality most closely was selected. The transformation was applied for data from Asians, African-American men, and whites. The natural log transformation was applied to data from African-American women and Hispanics. Because SF tends to increase with age[7,8] and to vary by field center, gender-specific and racial/ethnic-specific multiple linear regression equations were formed by regressing the SF values on spline terms for age and field center. For each individual, the adjusted SF value was calculated as the sum of the regression residual and a constant. The constant was calculated as the predicted SF at the median age with equal weights applied to parameters for field center.

Statistical mixture modeling of the bivariate distribution of TS and SF. For analysis of the bivariate distribution of TS and SF, adjusted for age, and field center, the physiologic models we considered were a single, bivariate normal distribution and mixtures of bivariate normal distributions.[12,13] This analysis is represented by a set of statistical models for a random sample $Y_1, Y_2, \ldots, Y_n$ of TS and SF values from n subjects. Let g denote the number of mixture components in the model (g = 1,2,3 . . .). Under the model with g components, each component i has proportion π_i, and each Y_j has density such that

$$f_g(y_j;\pi,\mu,\Sigma) = \sum_{i=1}^{g} \pi_i \phi(y_j;\mu_i,\Sigma_i)$$

where

$$0 \le \pi_i \le 1 \ (i = 1, 2, \ldots, g)$$

and

$$\sum_{i=1}^{g} \pi_i = 1.$$

The function $\phi(y_j;\mu_i,\Sigma_i)$ denotes the multivariate normal density with mean μ_i and component-covariance matrix Σ_i. As noted by McLachlan,[13] a normal mixture model without restrictions on the component–covariance matrices may be viewed as too general for many situations in practice. We compared the results of fitting mixtures with and without the assumption of homoscedasticity

Table I. Summary of bivariate 4-component mixture models for TS and SF in 59,692 women, adjusted for age and field center*

Racial/ethnic group for women	Mixture component	Mixing proportion	Mean TS (%)	Mean SF (μg/L)
African-American	C4	0.006	79.3	266
N = 17,085	C3	0.097	41.5	113
	C2	0.731	23.3	76
	C1	0.166	12.9	17
Asian	C4	0.008	70.2	221
N = 7,546	C3	0.174	41.8	120
	C2	0.707	25.6	82
	C1	0.111	12.4	22
Hispanic	C4	0.015	57.4	83
N = 8,469	C3	0.219	35.1	64
	C2	0.594	21.7	43
	C1	0.172	11.7	10
White	C4	0.007	80.0	203
N = 26,592	C3	0.120	44.8	103
	C2	0.720	26.1	80
	C1	0.154	15.1	27

Abbreviation: CI, confidence interval.
*Weighted prevalence of *HFE* mutations and self-reported liver disease within components.
†Race/ethnic distribution for 57,251 of 59,692 women (95.5%) who gave self-reported answers to a question at screening regarding history of liver disease: African American (n = 15,867, 92.8%), Asian (n = 7409, 98.2%), Hispanic (n = 7890, 93.2%), and white (n = 26085, 98.1%).
‡Presence of iron deficiency is defined as SF <15 μg/L.

and found that whereas the general model has flexibility, in most cases, the assumption of homoscedasticity where $\Sigma_i = \Sigma$ ($i = 1, 2, \ldots, g$), produced a better fit in the lower and upper tails of distributions (ie, lesser and greater values of TS and SF). For results reported in this article, the component–covariance matrices were restricted to being the same; thus, $\Sigma_i = \Sigma$ ($i = 1, 2, \ldots, g$).

For each data set, we applied the EMMIX[13] program to fit models and to assess the number of normal components. This program evaluates mixtures of distributions via the EM algorithm. The advantages of this program are that several methods are available to provide starting values for parameters, restrictions may be placed on the component–covariance matrices, the range of local solutions can be viewed graphically, and a bootstrap assessment of the fit to components is provided. The EM algorithm is an approach to the iterative computation of maximum likelihood estimates for incomplete-data problems. For the current application, the observed data are viewed as being incomplete, as the associated mixture components are unknown. Two steps exist for each iteration of the EM algorithm. Starting from initial parameter estimates, the conditional expectation of the log-likelihood given the observed data is computed in the expectation step. New values for the mixture model parameter estimates are computed in the maximization step by global maximization of conditional expectation of the complete data log-likelihood. Iterations continue until convergence is reached according to prespecified criteria. Excellent descriptions and applications of this approach are given elsewhere.[12–15] For each data set, a random sample of 10% of the data was used for each of 6 random starts and equal diagonal covariance matrices were specified. The significance of the likelihood ratio test statistics at the 0.01 level was assessed.[13] For additional assessment of model fit, the AIC and BIC statistics and estimates of overall correct allocation rate were examined. Based on the final racial/ethnic-specific models for men and women, the proportions and the means and variances for adjusted TS and adjusted SF within components of the bivariate distribution were computed.

Prevalence of *HFE* mutations, self-reported liver disease, and iron deficiency, within TS and SF components. The prevalence within TS and SF components, adjusted for age and field center, was computed for any *HFE* mutation (H63D/wt, C282Y/wt, C282Y/H63D, H63D/H63D, or C282Y/C282Y), the C282Y *HFE* mutation (C282Y/wt, C282Y/H63D, or C282Y/C282Y), self-reported liver disease, and iron deficiency, which was defined as SF less than 15 μg/L. Differences in self-reported liver disease, iron deficiency, and presence of *HFE* mutations with respect to TS and SF components were assessed. Odds ratio estimates are reported with corresponding 95% bootstrap confidence intervals.

Table I. Continued

HFE mutations		Self-reported liver disease†		Iron deficiency‡	
Weighted prevalence (%)	Adjusted odds ratio (95% CI)	Weighted prevalence (%)	Adjusted odds ratio (95% CI)	Weighted prevalence (%)	Adjusted odds ratio (95% CI)
11.6	1.56 (0.87, 2.77)	14.6	11.7 (6.33, 20.7)	5.0	2.69 (0.90, 5.94)
12.7	1.74 (1.51, 1.94)	2.6	1.81 (1.35, 2.42)	2.0	1.05 (0.79, 1.36)
7.7	1	1.4	1	1.9	1
6.7	0.86 (0.77, 0.98)	0.9	0.61 (0.45, 0.80)	44.2	40.0 (37.4, 42.7)
6.8	0.81 (0.43, 1.27)	8.4	1.71 (0.74, 3.77)	2.0	0.72 (0.01, 3.80)
11.0	1.35 (1.19, 1.50)	6.0	1.19 (1.02, 1.37)	0.9	0.31 (0.23, 0.45)
8.3	1	5.1	1	2.8	1
7.8	0.93 (0.80, 1.07)	3.7	0.71 (0.60, 0.86)	29.8	14.9 (13.5, 16.11)
27.3	1.47 (1.06, 2.03)	7.6	2.84(1.37, 5.37)	2.0	0.64 (0.11, 1.45)
26.6	1.42 (1.31, 1.53)	2.6	0.91 (0.73, 1.56)	1.6	0.53 (0.43, 0.68)
20.4	1	2.8	1	3.0	1
19.4	0.94 (0.85, 1.05)	2.0	0.71 (0.53, 1.01)	59.6	47.5 (44.3, 50.6)
75.8	5.18 (3.74, 7.18)	6.8	3.13 (1.64, 5.37)	1.4	0.57 (0.71, 1.81)
52.1	1.79 (1.70, 1.87)	2.6	1.13 (0.96, 1.34)	1.4	0.57 (0.45, 0.71)
37.8	1	2.3	1	2.4	1
34.6	0.87 (0.84, 0.91)	1.8	0.79 (0.69, 0.91)	29.3	17.0 (16.4, 17.6)

RESULTS

Statistical mixture modeling of TS and SF. The final analytic sample consisted of TS and SF values from 26,832 African Americans (17,085 women, 9,747 men), 12,620 Asians (7,546 women, 5,074 men), 12,264 Hispanics (8,469 women, 3,795 men), and 43,254 non-Hispanic Caucasians (26,592 women, 16,662 men). Tables I and II display the results of statistical mixture modeling, which reflect 4 components with successively increasing means for TS and SF. Scatterplots of adjusted TS and SF are displayed in Figs 1–4. The 95% confidence ellipses are superimposed on each and reflect the component probability densities with clear separation between 95% confidence ellipses for the lowest and highest components. The estimated proportions for the 4-component models differ by race/ethnic group and gender. Mean TS and SF in corresponding mixture components are greater for men than women.

C4. For each race/ethnicity and gender, C4 contained the smallest estimated proportion of values. For men, the estimated means for TS and SF in C4 exceeded HEIRS Study thresholds for increased TS and SF: TS greater than 83.8% and 886 μg/L in African Americans, 83.5% and 686 μg/L in whites, 79.4% and 660 μg/L in Asians, and 80.4% and 503 μg/L in Hispanics (Table II). As shown in Table I, for women, the estimated means for TS and SF are 79.3% and 266 μg/L in African Americans, 70.2% and 221 μg/L in Asians, 80.0% and 203 μg/L in whites, and 57.4% and 83 μg/L in Hispanics. In most cases, these mean estimates exceed HEIRS Study thresholds for increased TS and SF.

C1. For all groups, C1 contained the second-to-the-smallest estimated proportion of values. For men, the estimated means for TS and SF in C1 are 15.3% and 38 μg/L in African Americans, 6.4% and 24 μg/L in Hispanics, 17.6% and 43 μg/L in whites, and 19.1% and 46 μg/L in Asians. For women, the estimated means are 11.7% and 10 μg/L in Hispanics, 12.9% and 17 μg/L in African Americans, 12.4% and 22 μg/L in Asians, and 15.1% and 27 μg/L in whites.

C2 and C3. For all groups, C2 contained the largest proportion of values; C3 contained the second-to-the-largest estimated proportion. For men, the estimated means for TS in these components are between 21% and 41% with estimated means for SF between 165 μg/L and 347 μg/L. For women, the estimated means for TS in these components are between 21% and 45% with estimated means for SF between 43 μg/L and 120 μg/L.

Prevalence of *HFE* mutations and self-reported liver disease within TS and SF subpopulations. Tables I and II display weighted prevalence of *HFE* mutations within each mixture component and the adjusted odds of any *HFE* mutation, self-reported liver disease, and iron

Table II. Summary of bivariate 4-component mixture models for TS and SF in 35,278 men, adjusted for age and field center*

Racial/ethnic group for men	Mixture component	Mixing proportion	Mean TS (%)	Mean SF (μg/L)
African-American	C4	0.009	83.8	886
N=9,747	C3	0.105	48.0	310
	C2	0.856	28.6	208
	C1	0.029	15.3	38
Asian	C4	0.013	79.4	660
N=5,074	C3	0.245	45.2	347
	C2	0.690	29.3	242
	C1	0.052	19.1	46
Hispanic	C4	0.010	80.4	503
N=3,795	C3	0.130	48.8	281
	C2	0.839	29.9	188
	C1	0.022	16.4	24
White	C4	0.009	83.5	686
N=16,662	C3	0.138	48.0	234
	C2	0.802	28.9	165
	C1	0.050	17.6	43

Abbreviation: CI, confidence interval.
*Weighted prevalence of *HFE* mutations and self-reported liver disease within components.
†Race/ethnic distribution for 33,926 of 35,278 men (96.2%) who gave self-reported answers to a question at screening regarding history of liver disease: African American (n = 9077, 93.1%), Asian (n = 4972, 98.0%), Hispanic (n = 3603, 94.9%), and white (n = 16274, 97.7%).
‡Presence of iron deficiency is defined as SF <15 μg/L.

deficiency, relative to that of C2. On the basis of TS and SF, in C4, the odds of any *HFE* mutation were increased significantly in white women (5.18; 95% confidence interval 3.74–7.18) and in Hispanic women (1.47; 1.06–2.03). Odds of any *HFE* mutation were increased significantly in white men (4.74; 3.35–6.63), Hispanic men (2.82; 1.48–5.07), and African-American men (2.38; 1.34–4.03).

Similarly, for those with adjusted TS and SF in C4, odds of self-reported liver disease were greatest for African-American women (11.7; 6.33–20.7) and men (10.4; 5.76, 18.9) and were significantly increased for white women (3.13; 1.64–5.37), Hispanic women (2.84; 1.37–5.37), Asian men (7.26; 4.22–12.2), Hispanic men (10.17; 4.71–22.6), and white men (4.93; 3.11–8.62), compared with C2.

The adjusted odds of iron deficiency were greater for whites (women 17.0, 16.4–17.6; men 60.6, 53.4–68.3), Hispanics (women 47.5, 44.3–50.6; men 3530, 1888–5914), and African Americans (women 40.0, 37.4–42.7; men 332, 234–440) for those with age-adjusted TS and SF attributed to the lowest mixture C1, relative to C2, as shown in Tables I and II.

DISCUSSION

For data from 94,970 Asians, African Americans, Hispanics, and whites enrolled in the HEIRS Study, after removal of sources of variability, including age and field center, bivariate mixtures of 4 normal components were identified with successively increasing means for TS and SF (Tables I and II, Figs 1–4). On a population basis, race/ethnic-specific components were identified with increased mean TS and SF, increased odds of *HFE* mutations, and increased odds of self-reported liver disease that possibly reflect iron overload and/or substantial hepatic dysfunction, or the effect of genes that may be modifiers of iron overload. The mixture modeling analyses also identified a component, C1, with low mean TS and SF and greater prevalence of iron deficiency. For African Americans, Hispanics, and whites, this component also was associated with the least prevalence of C282Y *HFE* mutations compared with the largest mixture component, C2, with predominantly normal values for TS and SF, which possibly reflects the protective effect of the C282Y *HFE* mutation against iron deficiency.

Our previous univariate mixture modeling studies with various population data sets have indicated that TS follows 3 components in Caucasians and African Americans. These components are consistent with a major locus or loci that influence iron metabolism in that the largest component has the lowest mean TS and that 2 progressively smaller components have progressively greater mean TS values.[3,6,16–18] Furthermore, these components of increasing mean TS have increasing iron stores as reflected by SF.[3,6,16] In Caucasians,

Table II. Continued

HFE mutations		Self-reported liver disease†		Iron deficiency‡	
Weighted prevalence (%)	Adjusted odds ratio (95% CI)	Weighted prevalence (%)	Adjusted odds ratio (95% CI)	Weighted prevalence (%)	Adjusted odds ratio (95% CI)
16.2	2.38 (1.34, 4.03)	17.4	10.4 (5.76, 18.9)	0.00	0.00 (0.00, 0.00)
12.7	1.80 (1.57, 2.12)	3.6	1.82 (1.40, 2.51)	0.00	0.00 (0.00, 0.00)
7.5	1	2.0	1	0.09	1
6.2	0.81 (0.57, 1.17)	2.9	1.46 (0.71, 2.51)	23.5	332 (234.3, 440.2)
8.9	1.10 (0.45, 2.14)	31.5	7.26 (4.22, 12.2)	0.00	0.01 (0.00, 0.03)
9.8	1.22 (1.06, 1.30)	8.9	1.54 (1.32, 1.82)	0.04	1.37 (0.03, 4.54)
8.2	1	6.0	1	0.03	1
10	1.25 (0.95, 1.65)	3.9	0.64 (0.46, 0.91)	9.91	412 (248.2, 553.7)
43.8	2.82 (1.48, 5.07)	28.4	10.7 (4.71, 22.6)	0.00	0.00 (0.00, 0.00
32.5	1.75 (1.52, 2.03)	5.8	1.65 (1.22, 2.35)	0.00	0.05 (0.01, 0.12)
21.6	1	3.6	1	0.01	1
28.5	1.45 (0.97, 2.25)	4.3	1.19 (0.51, 2.91)	32.9	3530 (1888, 5914)
73.4	4.74 (3.35, 6.63)	12.5	4.93 (3.11, 8.62)	0.01	0.02 (0.00, 0.12)
52.6	1.90 (1.79, 2.01)	4.4	1.59 (1.37, 1.86)	0.09	0.33 (0.11, 1.13)
36.8	1	2.8	1	0.27	1
34.1	0.89 (0.81, 0.96)	2.6	0.92 (0.74, 1.13)	14.1	60.6 (53.4, 68.3)

the greatest TS component is enriched for *HFE* C282Y and H63D homozyotes or compound heterozygotes and the middle TS component is enriched for *HFE* C282Y heterozygotes.[3,6] We hypothesized that bivariate modeling of SF and TS would be superior to univariate modeling of TS because it might enable the identification of additional components and the ability to test for an association with disorders of iron metabolism. For this component, the odds of iron deficiency in women varied from 14.9 times greater in Asians to 40 times greater in African Americans when compared with the component with normal mean TS and SF. Iron deficiency is common among women in the population[19,20] and would be expected to affect substantially the distribution of measures of iron status in women. In the current study, the refinement of adding SF-enabled identification of a 4th component with lower SF and TS at risk for iron deficiency. Thus, the current bivariate analysis of SF and TS, which identifies a component with lower iron measures than the largest component, C2, seems to reflect more elegantly and accurately the distribution of measures of iron status in the population as compared with univariate TS modeling.

Compared with C2 of the current study, C3 and C4 are enriched for *HFE* C282Y and H63D mutations. This enrichment is consistent with our previous studies with univariate TS modeling.[3,6] A new feature of the current analysis is that we determined how self-reported liver disease corresponds to the components identified by mixture modeling. Hepatocellular damage leads to increases in SF and TS independently of iron stores[21,22] and would be predicted to be represented in the components with increases in these variables. Compared with C2 of the current study, both C3 and C4 have increased odds of self-reported liver disease in most ethnic groups by sex, especially C4.

An interesting feature of the current study is that *HFE* mutations contributed most predominantly to C4, the component with the greatest TS and SF values, in Caucasians (weighted prevalence of 76% women and 73% men) and progressively less so in Hispanics (27% women and 44% men), African Americans (12% women and 16% men), and Asians (7% women and 9% men). Nevertheless, compared with Caucasians, C4 made up similar or greater parts of the overall distribution for Hispanics, African Americans, and Asians, and it had similar or greater mean TS and SF values. One possible explanation for these observations is that common, major loci that influence iron metabolism are yet to be discovered in these population groups. Another possibility is that liver disease is dramatically more common in these populations, but the historical information of the current study does not reflect this possibility.

McGrath et al[23] developed a predictive nomogram for the prediction of *HFE* C282Y homozygotes from TS and SF. For clinical use, the approach allowed

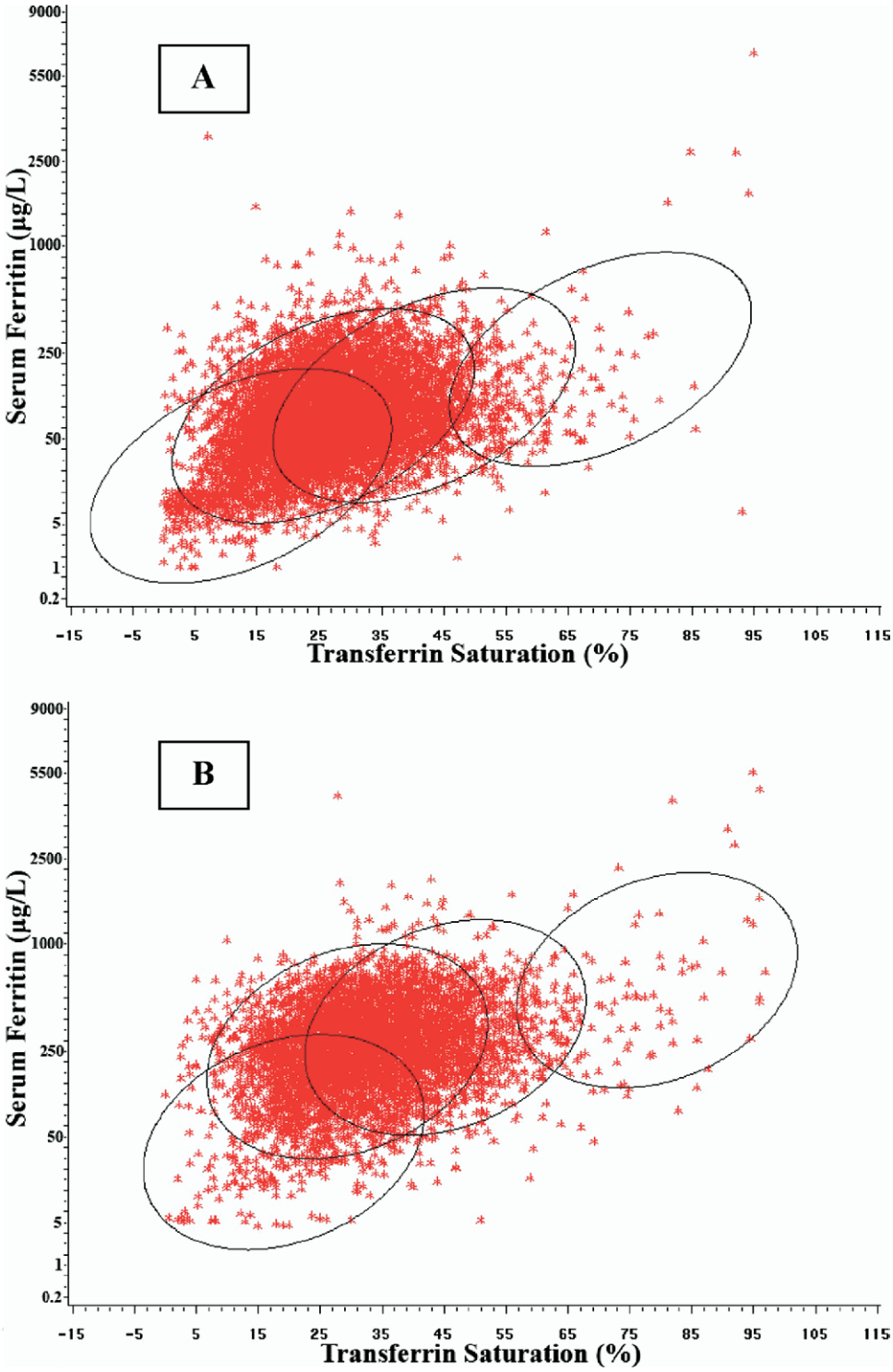

Fig 1. Scatterplot of age-adjusted TS and SF with 95% confidence ellipses based on the 4-component mixture model: (**A**) Asian women and (**B**) Asian men.

prediction of the probability that a patient was a C282Y homozygote over a wide range of SF and TS values. This prediction is an example of the use of the 2 biochemical tests, SF and TS, to improve the predictive ability of the single SF test.[23] In contrast, we have taken a population-based approach to examine the distribution of phenotypic data in 4 different race/ethnic groups and the association with prevalence of *HFE* gene mu-

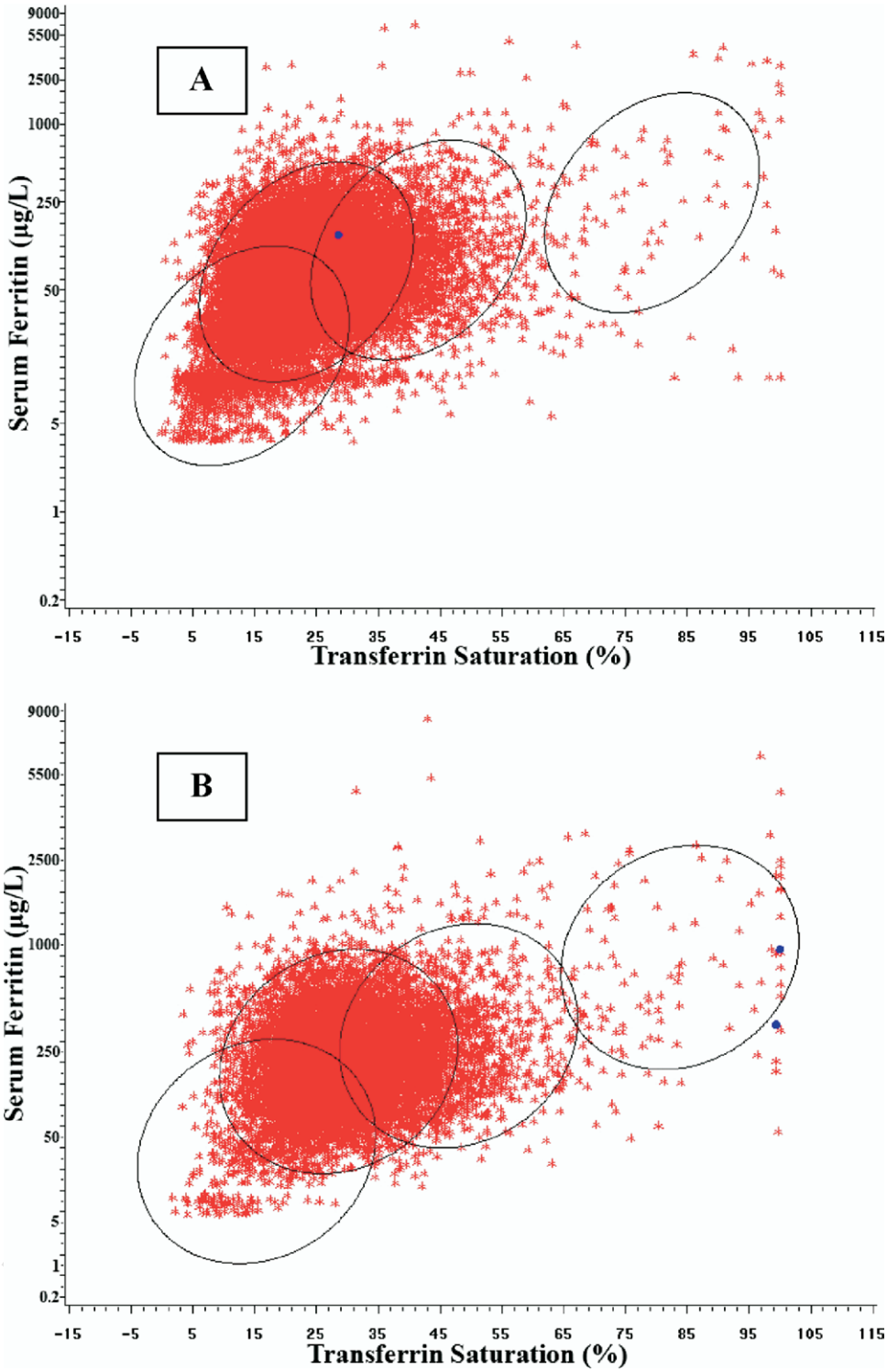

Fig 2. Scatterplot of age-adjusted TS and SF with 95% confidence ellipses based on the 4-component mixture model: (**A**) African-American women and (**B**) African-American men. Values from *HFE* C282Y homozygotes are shown as blue dots.

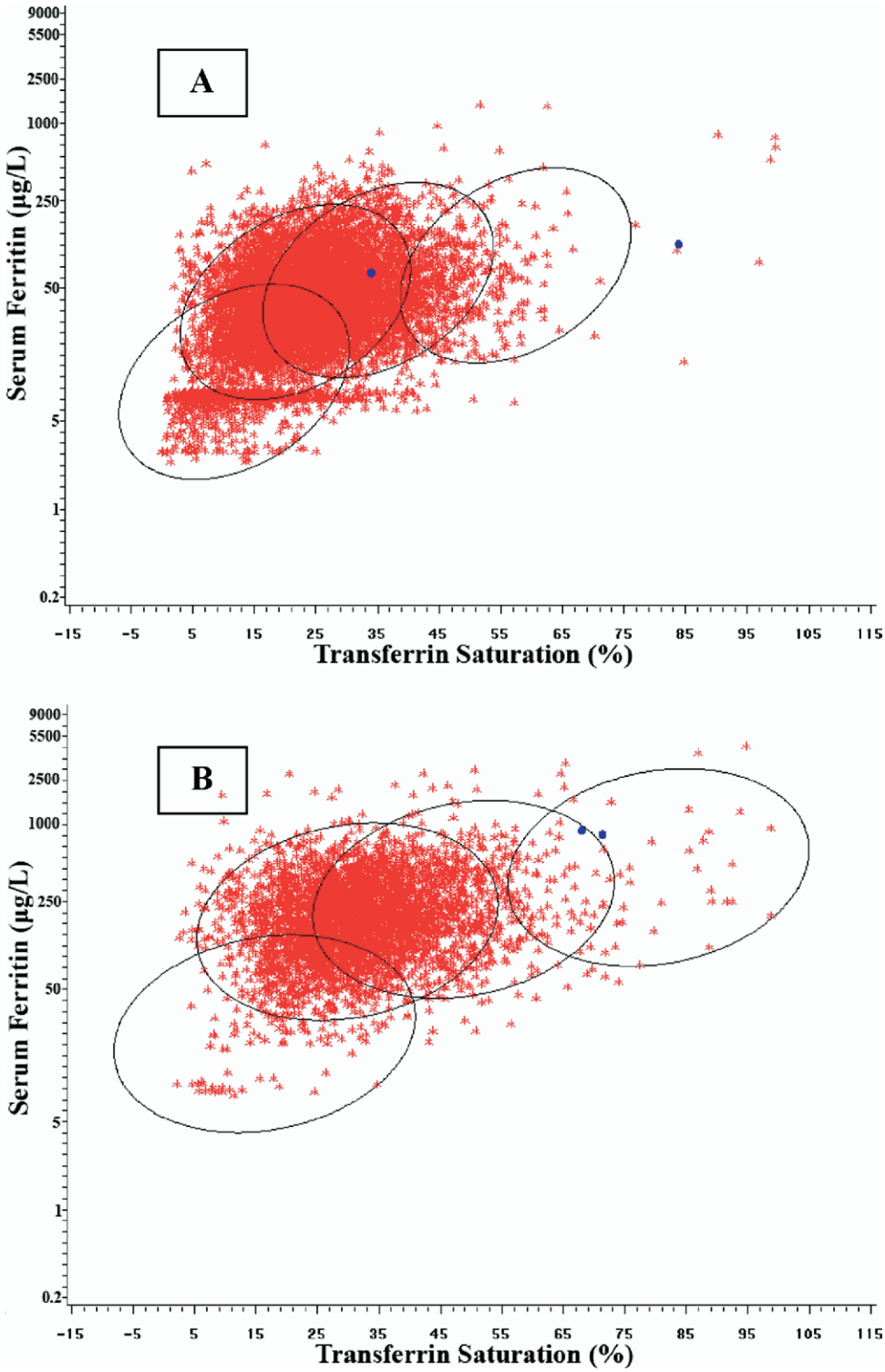

Fig 3. Scatterplot of age-adjusted TS and SF with 95% confidence ellipses based on the 4-component mixture model: (**A**) Hispanic women and (**B**) Hispanic men. Values from *HFE* C282Y homozygotes are shown in blue dots.

tations and disease manifestations. Another example of how this approach can be useful is given by Namboodiri et al[24] and Elston et al.[25] They analyzed the age-adjusted bivariate distribution of cholesterol and triglycerides in data from 247 individuals in 33 families where the probands had a type IIb lipoprotein phenotype. Results showed that the joint distribution had only 1 local maximum, which suggests the action of a single

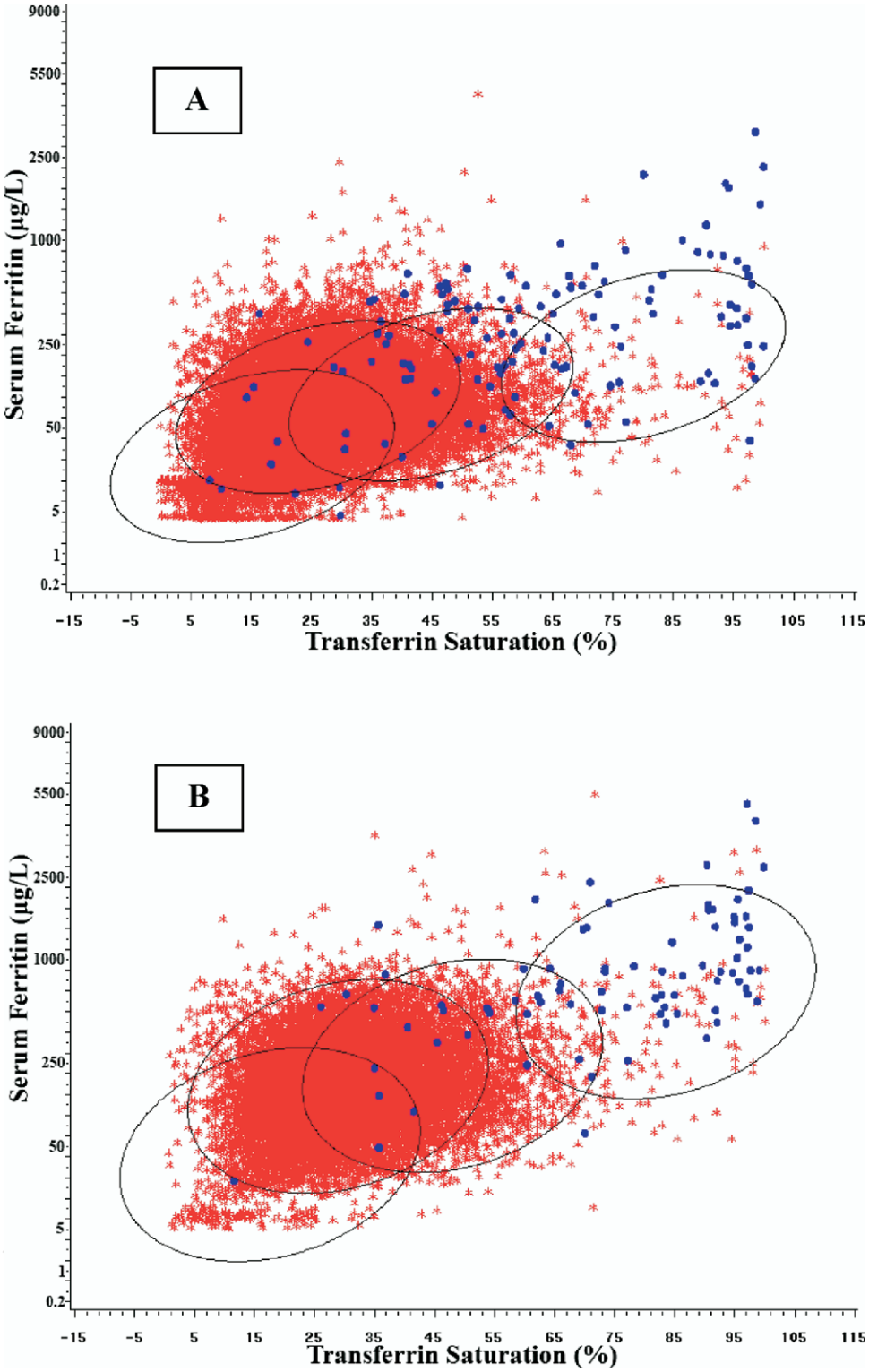

Fig 4. Scatterplot of age-adjusted TS and SF with 95% confidence ellipses based on the 4-component mixture model: (**A**) white women and (**B**) white men. Values from *HFE* C282Y homozygotes are shown as blue dots.

genetic determinant in the sample. They discuss the power of bivariate analyses of multigenerational data.[24,25]

Limitations of the current study are that TS and SF were based on single determinations of blood samples collected at various times during the day. Considerable day-to-day variation exists for TS,[9] and this measure is also affected by a substantial diurnal variation.[10] Both SF and TS are influenced by inflammation, with SF being increased and TS being decreased,[26,27] and we could not account for these potential changes in the current study. Despite these limitations, in the case of iron metabolism measures, bivariate mixture modeling seems to reflect the effects of a major genetic locus (*HFE*) and the effects of acquired factors (iron deficiency and self-reported liver disease).

In conclusion, bivariate mixture modeling can improve on univariate modeling in terms of reflecting the magnitude and complexity of health problems in the population as reflected in commonly available clinical tests. Such methodology has the potential to complement and to enhance genetic testing for assessment of disease complications.

PARTICIPATING INSTITUTIONS AND PRINCIPAL STAFF

Field centers. *Birmingham, Ala*—University of Alabama at Birmingham:

Dr. Ronald T. Acton (Principal Investigator), Dr. James C. Barton (Co-Principal Investigator), Ms. Deborah Dixon, Dr. Susan Ferguson, Dr. Richard Jones, Dr. Jerry McKnight, Dr. Charles A. Rivers, Dr. Diane Tucker, and Ms. Janice C. Ware.

Irvine, Calif—University of California, Irvine:

Dr. Christine E. McLaren (Principal Investigator), Dr. Gordon D. McLaren (Co-Principal Investigator), Dr. Hoda Anton-Culver, Ms. Jo Ann A. Baca, Dr. Thomas C. Bent, Dr. Lance C. Brunner, Dr. Michael M. Dao, Dr. Korey S. Jorgensen, Dr. Julie Kuniyoshi, Dr. Huan D. Le, Dr. Miles K. Masatsugu, Dr. Frank L. Meyskens, Dr. David Morohashi, Dr. Huan P. Nguyen, Dr. Sophocles N. Panagon, Dr. Chi Phung, Dr. Virgil Raymundo, Dr. Thomas Ton, Professor Ann P. Walker, Dr. Lari B. Wenzel, and Dr. Argyrios Ziogas.

London, Ontario, Canada—London Health Sciences Center:

Dr. Paul C. Adams (Principal Investigator), Ms. Erin Bloch, Dr. Subrata Chakrabarti, Ms. Arlene Fleischhauer, Ms. Helen Harrison, Ms. Kelly Jia, Ms. Sheila Larson, Dr. Edward Lin, Ms. Melissa Lopez, Ms. Lien Nguyen, Ms. Corry Pepper, Dr. Tara Power, Dr. Mark Speechley, Dr. Donald Sun, and Ms. Diane Woelfle.

Portland, Ore and Honolulu, Hawaii—Kaiser Permanente Center for Health Research, Northwest and Hawaii, and Oregon Health and Science University:

Dr. Emily L. Harris (Principal Investigator), Dr. Mikel Aickin, Dr. Elaine Baker, Ms. Marjorie Erwin, Ms. Joan Holup, Ms. Carol Lloyd, Dr. Nancy Press, Dr. Richard D. Press, Dr. Jacob Reiss, Dr. Cheryl Ritenbaugh, Ms. Aileen Uchida, Dr. Thomas Vogt, and Dr. Dwight Yim.

Washington, DC—Howard University:

Dr. Victor R. Gordeuk (Principal Investigator), Dr. Fitzroy W. Dawkins (Co-Principal Investigator), Ms. Margaret Fadojutimi-Akinsiku, Dr. Oswaldo Castro, Dr. Debra White-Coleman, Dr. Melvin Gerald, Ms. Barbara W. Harrison, Dr. Ometha Lewis-Jack, Dr. Robert F. Murray, Dr. Shelley McDonald-Pinkett, Ms. Angela Rock, Dr. Juan Romagoza, and Dr. Robert Williams.

Central laboratory. *Minneapolis, Minn*—University of Minnesota and University of Minnesota Medical Center, Fairview:

Dr. John H. Eckfeldt (Principal Investigator and Steering Committee Chair), Ms. Susie DelRio-LaFreniere, Ms. Catherine Leiendecker-Foster, Dr. Ronald C. McGlennen, Mr. Greg Rynders, Dr. Michael Y. Tsai, and Dr. Xinjing Wang.

Coordinating center. *Winston-Salem, NC*—Wake Forest University:

Dr. David M. Reboussin (Principal Investigator), Dr. Beverly M. Snively (Co-Principal Investigator), Dr. Roger Anderson, Ms. Aarthi Balasubramanyam, Ms. Elease Bostic, Ms. Brenda L. Craven, Ms. Shellie Ellis, Dr. Curt Furberg, Mr. Jason Griffin, Dr. Mark Hall, Mr. Darrin Harris, Ms. Leora Henkin, Dr. Sharon Jackson, Dr. Tamison Jewett, Mr. Mark D. King, Mr. Kurt Lohman, Ms. Laura Lovato, Dr. Joe Michaleckyj, Ms. Shana Palla, Ms. Tina Parks, Ms. Leah Passmore, Dr. Pradyumna D. Phatak, Dr. Stephen Rich, Ms. Andrea Ruggiero, Dr. Mara Vitolins, Mr. Gary Wolgast, and Mr. Daniel Zaccaro.

NHLBI project office. *Bethesda, Md*—Ms. Phyliss Sholinsky (Project Officer), Dr. Ebony Bookman, Dr. Henry Chang, Ms. Kristianne Cooper, Dr. Richard Fabsitz, Dr. Cashell Jaquish, Dr. Teri Manolio, and Ms. Lisa O'Neill.

NHGRI project office. *Bethesda, Md*—Dr. Elizabeth Thomson.

Dr. Jean MacCluer, Southwest Foundation for Biomedical Research, also contributed to the design of this study.

REFERENCES

1. Feder JN, Gnirke A, Thomas W, Tsuchihashi Z, Ruddy DA, Basava A, et al. A novel MHC class I-like gene is mutated in patients with hereditary haemochromatosis. Nat Genet 1996;13: 399–408.

2. Merryweather-Clarke AT, Pointon JJ, Jouanolle AM, Rochette J, Robson KJ. Geography of HFE C282Y and H63D mutations. Genet Test 2000;4:183–98.
3. McLaren CE, Li KT, Garner CP, Beutler E, Gordeuk VR. Mixture distribution analysis of phenotypic markers reflecting *HFE* gene mutations. Blood 2003;102:4563–6.
4. Adams PC, Reboussin DM, Barton JC, McLaren CE, Eckfeldt JH, McLaren GD, et al. Hemochromatosis and iron overload screening (HEIRS) study: screening of a primary care population. N Engl J Med 2005;352:1769–78.
5. McLaren CE, Barton JC, Adams PC, Harris EL, Acton RT, Press N, et al. Hemochromatosis and Iron Overload Screening (HEIRS) study design for an evaluation of 100,000 primary care-based adults. Am J Med Sci 2003;325:53–62.
6. McLaren CE, Li KT, McLaren GD, Gordeuk VR, Snively BM, Reboussin D, et al. Mixture models of serum iron measures in populations screening for hemochromatosis and iron overload. Transl Res 2006;148:196–206.
7. Expert Scientific Working Group. Summary of a report on assessment of the iron nutritional status of the United States population. Am J Clin Nutr 1985;42:1318–30.
8. Zacharski LR, Ornstein DL, Woloshin S, Schwartz LM. Association of age, sex, and race with body iron stores in adults: analysis of NHANES III data. Am Heart J 2000;140:98–104.
9. Edwards CQ, Griffen LM, Kaplan J, Kushner JP. Twenty-four hour variation of transferrin saturation in treated and untreated haemochromatosis homozygotes. J Intern Med 1989;226:373–9.
10. Hamilton LD, Gubler CJ, Cartwright GE, Wintrobe MM. Diurnal variation in the plasma iron level of man. Proc Soc Exp Biol Med 1950;75:65–8.
11. Harrell FE. General aspects of fitting regression models. New York, NY: Springer Verlag; 2001.
12. Everitt BS, Hand DJ. Finite mixture distributions. London: Chapman and Hall; 1981.
13. McLachlan GJ, Peel D. Finite mixture models. New York: Wiley; 2000.
14. Titterington DM, Smith AFM, Makov UE. Statistical analysis of finite mixture distributions. Chichester, UK: Wiley; 1985.
15. McLachlan GJ, Krishnan T. The EM algorithm and extensions. New York: Wiley; 1997.
16. Gordeuk VR, McLaren CE, Looker AC, Hasselblad V, Brittenham GM. Distribution of transferrin saturations in the African-American population. Blood 1998;91:2175–9.
17. McLaren CE, McLachlan GJ, Halliday JW, Webb SI, Leggett BA, Jazwinska EC, et al. Distribution of transferrin saturation in an Australian population: relevance to the early diagnosis of hemochromatosis. Gastroenterology 1998;114:543–9.
18. McLaren CE, Li KT, Gordeuk VR, Hasselblad V, McLaren GD. Relationship between transferrin saturation and iron stores in the African American and US Caucasian populations: analysis of data from the third National Health and Nutrition Examination Survey. Blood 2001;98:2345–51.
19. Frith-Terhune AL, Cogswell ME, Khan LK, Will JC, Ramakrishnan U. Iron deficiency anemia: higher prevalence in Mexican American than in non-Hispanic white females in the third National Health and Nutrition Examination Survey, 1988–1994. Am J Clin Nutr 2000;72:963–8.
20. Looker AD, Cogswell M, Gunter E. Iron deficiency—United States, 1999–2000. MMWR Morb Mortal Wkly Rep 2002;51:897–9.
21. Di Bisceglie AM, Axiotis CA, Hoofnagle JH, Bacon BR. Measurements of iron status in patients with chronic hepatitis. Gastroenterology 1992;102:2108–13.
22. Wong K, Adams PC. The diversity of liver diseases among outpatient referrals for an elevated serum ferritin. Can J Gastroenterol 2006;20:467–70.
23. McGrath JS, Deugnier Y, Moirand R, Jouanolle AM, Chakrabarti S, Adams PC. A nomogram to predict C282Y hemochromatosis. J Lab Clin Med 2002;140:6–8.
24. Namboodiri KK, Elston RC, Glueck CJ, Fallat R, Buncher CR, Tsang R. Bivariate analyses of cholesterol and triglyceride levels in families in which probands have type IIb lipoprotein phenotype. Am J Hum Genet 1975;27:454–71.
25. Elston RC, Namboodiri KK, Glueck CJ, Fallat R, Tsang R, Leuba V. Study of the genetic transmission of hypercholesterolemia and hypertriglyceridemia in a 195 member kindred. Ann Hum Genet 1975;39:67–87.
26. Cartwright GE, Wintrobe MM. Chemical, clinical, and immunological studies on the products of human plasma fractionation. Xxxix. The anemia of infection. Studies on the iron-binding capacity of serum. J Clin Invest 1949;28:86–98.
27. Gordeuk VR, Brittenham GM, McLaren GD, Spagnuolo PJ. Hyperferremia in immunosuppressed patients with acute nonlymphocytic leukemia and the risk of infection. J Lab Clin Med 1986;108:466–72.