

Combined Segregation and Linkage Analysis of Genetic Hemochromatosis Using Affection Status, Serum Iron, and HLA

I. B. Borecki,* G. M. Lathrop,† G. E. Bonney,§ J. Yaouanq,|| and D. C. Rao*†

*Division of Biostatistics, and †Departments of Psychiatry and Genetics, Washington University School of Medicine, St. Louis;
‡Centre d'Etude du Polymorphisme Humain, Paris; §Division of Biostatistics, Howard University Cancer Center, Washington, DC;
and ||INSERM Unité 49, Rennes, France

Summary

Characterizing the distribution of parameters of iron metabolism by hemochromatosis genotype remains an important goal vis-à-vis potential screening strategies to identify individuals at genetic risk, since a specific marker to detect the abnormal gene has not been identified as yet. In the present investigation, we analyze serum iron values in ascertained families using a method which incorporates both segregation of the clinical affection status and the HLA linkage information to identify the underlying genotypes. The analysis is performed using an extension of the model presented by Bonney et al., comprising regressive models for segregation analysis and the multipoint linkage strategy implemented in LINKAGE. The gene was found to be completely recessive with respect to both clinical manifestations and serum iron abnormalities, with significant differences in expression by sex. Clinical manifestations were present for all male homozygotes in this data set, suggesting that the recessive hemochromatosis genotype is fully penetrant at all ages in males. This was not the case for younger females. Significant genotype-specific age and sex effects were found for serum iron values. It is interesting that deletion of the HLA marker information did not affect our ability to resolve the genetic model when we analyzed a bivariate phenotype. This serves as a reminder that a search for relevant biological markers can be equally important in discerning the genetic etiology of a disease trait, as a search for linked genetic markers.

Introduction

Hemochromatosis is an iron metabolism disorder governed by an autosomal recessive gene linked to HLA (Simon et al. 1977). Gene frequency estimates from various studies are about 6%–7%, leading to a recessive homozygote frequency of approximately .0042. Since clinical symptoms appear as a consequence of a progressive accumulation of iron storage aggregates which interfere with normal organ function, the disease penetrance roughly is a function of the net iron overload and of the age and sex of the individual. Phlebotomy therapy instituted in individuals having a positive iron balance has been shown to be effective in preventing future organ damage (Bomford and Williams 1976).

Received March 29, 1990; revision received May 15, 1990.

Address for correspondence and reprints: Ingrid B. Borecki, Division of Biostatistics, Washington University School of Medicine, Box 8067, 660 South Euclid Avenue, St. Louis, MO 63110.

© 1990 by The American Society of Human Genetics. All rights reserved.
0002-9297/90/4703-0018\$02.00

Lacking a specific marker to detect the abnormal gene, and given that the abnormal gene for hemochromatosis has not been identified as yet, screening procedures based on the measurement of some relevant parameters of iron metabolism have been considered for the detection of genetically predisposed individuals, i.e., hemochromatosis homozygotes. While transferrin saturation has been recommended as the single best screening measure (Dadone et al. 1982; Borwein et al. 1983; Bassett et al. 1984), it is still important to understand the genotype-specific distributions of all available biochemical measures. Although previous studies suggest that heterozygous individuals rarely (if ever) go on to exhibit clinical symptoms (Beaumont et al. 1979; Cartwright et al. 1979; Bassett et al. 1981; Borwein et al. 1983), there is some evidence to suggest that the gene may find some biochemical expression in approximately 21%–25% of heterozygotes as compared with homozygous normal individuals (Bassett et al. 1981; Borwein et al. 1983) potentially confounding

the detection of predisposed individuals in the early stages of iron accumulation with heterozygotes.

Our recent segregation analyses of 146 hemochromatosis families from Brittany, France, suggested no evidence of heterozygote expression for any of the biochemical measurements studied. This conclusion was reached for serum iron (SI) (Lalouel et al. 1985), latent capacity and transferrin saturation (Borecki et al. 1989*b*, 1990*a*), and serum ferritin (Borecki et al. 1990*b*). However, admixture analyses of these variables suggested the possibility of sex-dependent mean genotypic effects for transferrin saturation and latent capacity and of genotype-dependent age effects for serum ferritin (Borecki et al. 1989*a*). The transformations used to ameliorate these effects might have reduced the distinction among genotypes. In addition, in light of the known linkage with the HLA complex (Simon et al. 1977), the simultaneous consideration of HLA haplotypes can aid significantly in the assessment of probable genotype.

The genetic epidemiologic analysis of common familial diseases such as hemochromatosis can profit greatly from a multivariate approach incorporating all available information, including relevant covariates that potentially modify gene expression. Bonney (1984, 1986) has provided a convenient framework in which to integrate these data in the regressive models for segregation analysis, and this method has been merged recently (Bonney et al. 1988) with the multilocus linkage analysis method as implemented in the LINKAGE program (Lathrop et al. 1984). Motivated especially by the suggestion of genotype-specific mean and variance effects for the iron metabolism parameters (Borecki et al. 1989*a*), we sought to investigate these data in this highly informative analytic context. The extended model is applied here to SI values, which show the least distributional complications; that is, the distribution is comparable between males and females, there is no attenuation in range as exhibited by measures of transferrin saturation, and SI values do not appear to be strongly correlated with the progress and/or duration of the disease, as are serum ferritin levels.

Material and Methods

The Model

Combined segregation and linkage analysis has been implemented in a PASCAL program, REGRESS, on the basis of the class A regressive model (Bonney 1984, 1986) and the multipoint linkage strategy used in LINKAGE (Lathrop et al. 1984). This method has been described and applied elsewhere (Bonney et al. 1988). Sev-

eral enhancements have been implemented in the current version of the program. The trait under investigation can be represented by as many as three phenotypic measurements: an affection status indicator and two quantitative measures that relate to the definition of the trait. It is assumed that a single major gene affects all the specified phenotypic outcomes and is potentially linked to a marker or series of linked markers. An arbitrary number of covariates can be entertained in the penetrance functions; the regressors may be genotype specific with respect to their effect on a quantitative trait. An ascertainment correction for single selection has been implemented also.

The model extensions in this version follow from the basic formulations provided by Bonney et al. (1988). For the present application which motivated these extensions, the data consist of a quantitative indicator of the disease (Y_{QU}), affection status based on clinical considerations (Y_{QL}), and phenotypes at a linked marker locus (M). The generalized penetrance function is $P(Y_{QU}, Y_{QL}, M | g, Y_A, X)$, where the probability of observing the phenotypes is conditional on the underlying genotypes (g), phenotype observations on relevant ancestors (Y_A), and adjustment for relevant covariates (X). This expression can be factored as $P(Y_{QU}, Y_{QL} | g, Y_A, X) \cdot P(M | Y_{QU}, Y_{QL}, g, Y_A, X)$. For HLA and RFLP-determined markers in general, the second factor is unity because the marker phenotypes are completely determined given the genotypes. Thus, we only need to specify the penetrance function for the genotype underlying the disease trait phenotypes. The first term in the expression above can be factored as $P(Y_{QU} | g, Y_A, X) \cdot P(Y_{QL} | Y_{QU}, g, Y_A, X)$. Assuming that the value of the quantitative trait does not *directly* affect the affection status (after conditioning on the major genotype), the dependence on Y_{QU} in the second term can be dropped. This assumption can be tested easily.

In general, since the underlying trait genotypes (g) are unobserved, their prior probabilities are taken to be in Hardy-Weinberg proportions for individuals with no parents in the pedigree and are based on segregation probabilities from parental mating types for individuals with ancestors in the pedigree (for a complete discussion, see Bonney 1984). In the present case, the probability of a genotype in any individual is modified by the linkage relationship and the segregation of genotypes at the marker locus. The dependency on phenotypes of ancestors (Y_A) is modeled under class A assumptions (Bonney 1984), which essentially posits that sibs are conditionally independent, given their parents' genetic contribution. The quantitative penetrance can

be modeled by a normal density function. For each individual, the quantitative penetrance function is assumed to follow a normal distribution:

$$\phi(z, w) = \frac{1}{\sqrt{2\pi w}} \exp\left[-z^2/(2w)\right],$$

where

$$z = (Y - \mu_g) - b_S(Y_S - \mu_{gS}) - b_F(Y_F - \mu_{gF}) \\ - b_M(Y_M - \mu_{gM}) - \sum_{j=1}^m b_{gj}X_j,$$

and where Y is the phenotype Y_{QU} , μ_g is the genotype-specific mean, b is a regression coefficient, b_g is a genotype-specific regression coefficient, S refers to the spouse of this individual, F refers to the father of this individual, M refers to the mother of this individual, and j is the j th of m covariates. It should be noted that the dependencies are modeled on residual phenotypes, after account is taken of the mean effects of the major locus. The variance can be written as $w = \sigma_g^2 (1 - b_{SPFM} - b_{FPFC} - b_{MPMC})$, with no adjustment for the covariates, since they are treated here as fixed effects. If father, mother, or both parents are missing, the z and w are modified according to formulas of G. E. Bonney, C. Murigande, F. M. Demenais, and A. Sorant (unpublished data).

In order to enhance the speed of likelihood function evaluations, an approximation in the calculation of the mean of the normal density function is employed, in which each genotype mean appearing in the equation is substituted with the genotype mean for the index case:

$$z = (Y - \mu_g) - b_S(Y_S - \mu_g) - b_F(Y_F - \mu_g) \\ - b_M(Y_M - \mu_g) - \sum_{j=1}^m b_{gj}X_j,$$

corresponding to case (2) of Demenais et al. (in press).

The penetrance function for the qualitative trait is modeled by a multiple logistic function. It is convenient to transform the affection status variable (Y_{QL}) such that the resulting variable (Z) is 1 if affected, 0 if missing, and -1 if normal. In this way, missing values do not contribute to the estimates of the regression coefficients (γ). Now we define an individual's logit score as

$$\lambda = \alpha + \gamma_S Z_S + \gamma_F Z_F + \gamma_M Z_M + \sum_{j=1}^m \beta_j X_j,$$

where α is the genotype-specific intercept, γ 's are the logistic regression coefficients, and β 's are the regres-

sion coefficients for the m covariates. Thus, the probability that an individual is affected (AF) given the genotype, the phenotypes of his spouse, father and mother, and the values of his covariates is

$$\Pr(Y=AF|g, Y_A, X_1, \dots, m) = e^\lambda / (1 + e^\lambda).$$

The ascertainment correction implemented in the current version of the program essentially entails conditioning the likelihood of observing a family on the likelihood of observing the proband affected under the parameters of the model

$$L(Y_{QU}, Y_{QL}, X; \omega) / L(Y_{QLPR} = AF; \omega),$$

where ω denotes the parameters of the model. This has been shown to be a valid correction when pedigrees of arbitrary structure are ascertained under single selection (Cannings and Thompson 1977).

The parameters of the model as concerns the quantitative trait include major gene effects (genotype means and gene frequency), residual variance and residual correlations (spouse-spouse, father-child, and mother-child), and the regression coefficients for any covariates (sex and age) which may be genotype-specific. The qualitative trait is assumed to be under the influence of the same biallelic major gene, whose effect is expressed as the probability of being affected given each possible genotype (i.e., via penetrances). The degree of dependency on a respective spouse, father, and mother is estimated as part of the multiple logistic regression function. The effects of covariates also can be made part of the logistic function. Finally, the strength of the linkage relationship is expressed by the recombination fraction (θ) between the trait and marker loci. Parameter estimation is carried out using maximum likelihood methods, with the routine implemented in GEMINI (Lalouel 1979).

The Family Data

The data being investigated comprise 146 single-proband families ascertained from consecutive patients referred to the Cliniques Médicales A and B, Centre Hospitalier Régional, Rennes (LeMignon 1982). First-degree relatives and spouses of probands were screened; whenever possible, the investigation was extended to more remote relatives. For 43 of the 146 families there was information (clinical and laboratory assessment and HLA typing) on non-first-degree relatives, usually nephews and nieces; in 15 families, data on half-siblings, cousins, grandchildren, and grandnephews or grand-

Table 1**Distribution of SI Values ($\mu\text{M}/\text{liter}$) in Controls and Probands**

Group (N)	Mean	SD	Skewness	Excess
Controls:				
Males (163)	21.341	6.231	.288	.331
Females (235)	20.438	6.898	.174	.422
Probands:				
Males (230)	43.809	9.643	.190	-.257
Females (43)	42.302	6.243	.151	-.757

nieces were also available. Note that the likelihood for a family of arbitrary structure can be written as a series of recursive parent-child relationships, as described in detail by Elston and Stewart (1971).

Clinical and laboratory screening, including several biochemical measures of iron metabolism, was performed on all individuals entering the study. In the present analysis, we will be concerned only with the measure of SI. HLA typing of the A and B loci was performed on all individuals by the standard microlymphocytotoxicity method.

Raw SI values were used in this analysis, and age and sex were entertained as covariates potentially affecting SI values and the disease penetrance function. In addition to data on the family members of probands recruited into the study, data were also collected on a series of unrelated probands and on unaffected controls (Lalouel et al. 1985). The distribution of raw SI values by sex are shown in table 1 for all controls (controls plus marry-ins to the families) and for all probands (probands for the family data plus the unrelated probands). None of the distributions were significantly skewed, and only for female probands was there significant excess (γ_2). Detailed information regarding the distribution of SI values can be found in the work of Borecki et al. (1989a). Linkage equilibrium between the hemochromatosis and HLA loci was assumed.

Results

Initial modeling was carried out under certain assumptions in order to reduce the number of estimated parameters. Close linkage between the hemochromatosis locus and HLA was assumed, with $\theta = .001$. On the basis both of the estimates presented by Lalouel et al. (1985) and of some preliminary explorations of the data, the penetrance of the homozygous recessive hemochromatosis genotype was fixed at .959, 1.000, and 1.000, and at .929, .999, and 1.000, in males and

females, respectively, for the age groups 20–39, 40–59, and 60+ years. The penetrance of the heterozygous genotype was assumed to be zero for all age groups in both sexes. The “general” model consisted of a recessive effect, residual spouse-spouse and parent-child correlations, genotype-specific effects of sex and age on mean SI values, and residual spouse and parent dependencies in the disease penetrance function. The results of hypothesis testing against this general model are summarized in table 2. Residual familial dependency in the penetrance function was not significantly different from zero and, accordingly, was dropped from the model. However, residual familial correlations for SI values were significant. Sex and age both had a significant effect on SI values, and the relationship appears to be different for hemochromatosis homozygous individuals and individuals of other genotypes. There was no evidence of heterozygote expression either in disease penetrance or in elevated SI values. The maximum likelihood estimate of θ between the hemochromatosis and HLA loci was .015.

This analysis provided an initial assessment of which model parameters were significant in the context of fixed penetrances with respect to the homozygous recessive genotype for six age- and sex-specific groups (liability classes). In the subsequent model-fitting procedure, the approach with liability classes was abandoned, and the logistic penetrance function parameters were estimated directly. The model is shown in table 3, along with standard errors of the parameters. The parameters were named as follows: θ is the recombination fraction between the hemochromatosis and HLA loci; q is the frequency of the hemochromatosis allele (H) versus the normal allele (O); α , β_{sex} , and β_{age} are parameters of the logistic penetrance function; $\mu_{g(i)}$ is the mean SI for the i th genotype; σ is the within-genotype SD; r_{FM} and r_{PC} are the father-mother and parent-child SI residual

Table 2**Results of Hypothesis Testing**

Hypothesis	χ^2	df
No residual familial dependency for disease penetrance	1.93	2
No residual familial correlations for SI	6.32	2
No effect of covariates on SI	41.01	4
No genotype dependency for effects of covariates	15.78	2
No heterozygous disease penetrance	3.57	6
No elevation of SI in heterozygotes	1.79	1

Table 3**Parameter Estimates $\pm$ Standard Errors under Various Models**

PARAMETER	MODEL					
	General	No Quantitative SI	No Linked Marker	Affection Status Only	No Ascertainment Correction	No Likelihood Approximation
General:						
θ	.017 $\pm$.007	.000 ^a	.5 ^b	...	.000 ^a	.015
q	.065 $\pm$.010	.081 $\pm$.018	.065 $\pm$.011	.093 $\pm$.022	.506 $\pm$.017	.098
Logistic penetrance function:						
α	27.111 $\pm$ 2.793	2.989 $\pm$ 2.527	22.028 $\pm$ 15.685	.903 $\pm$ 1.527	.620 $\pm$.999	24.623
β_{sex}	-16.870 $\pm$ 4.925	-3.623 $\pm$ 1.565	-13.196 $\pm$ 7.775	-2.843 $\pm$ 1.039	-2.526 $\pm$.547	-15.197
β_{age}	.237 $\pm$.095	.156 $\pm$.026	.134 $\pm$.039	.140 $\pm$.023	.161 $\pm$.011	.219
Quantitative penetrance function:						
$\mu_{\text{OO/HO}}$	24.806 $\pm$.872	...	24.854 $\pm$.861	...	25.148 $\pm$.878	28.802
μ_{HH}	58.898 $\pm$ 1.015	...	59.095 $\pm$ 1.007	...	53.143 $\pm$ 2.251	57.061
σ	7.663 $\pm$.129	...	7.487 $\pm$.128	...	7.657 $\pm$.135	7.906
r_{FM}	.035 $\pm$.019	...	.035 $\pm$.019	...	.027 $\pm$.023	.040
r_{PC}	.066 $\pm$.021	...	.074 $\pm$.020	...	.029 $\pm$.026	.124
$\beta_{\text{sex/O}}$	-2.385 $\pm$.423	...	-2.707 $\pm$.417	...	-2.637 $\pm$.437	-2.492
$\beta_{\text{sex/HH}}$	-5.771 $\pm$.916	...	-5.860 $\pm$.900	...	-7.282 $\pm$.995	-6.126
$\beta_{\text{age/O}}$	.047 $\pm$.013	...	.053 $\pm$.013	...	.051 $\pm$.013	-.016
$\beta_{\text{age/HH}}$	-.136 $\pm$.036	...	-.136 $\pm$.036	...	.020 $\pm$.063	-.087

^a Parameter estimated to a boundary value.^b Parameter fixed at value shown.^c Sex coded "1" for males, "2" for females.

correlations, respectively; and $\beta_{\text{sex}/g(i)}$ and $\beta_{\text{age}/g(i)}$ are genotype-specific sex effects and genotype-specific age effects, respectively, on SI values. The results of the general model are shown graphically in figure 1A. For these data, there was no putative male hemochromatosis homozygote who was not clinically affected; therefore, the gene appears to be fully penetrant over all ages in males. In females, the genotype appears to reach full penetrance by approximately 50 years of age. Mean SI values, by sex and genotype, as a function of age are shown in figure 2. The linear effects of age were significantly different between hemochromatosis homozygotes and nonhomozygotes; higher-order terms and interaction effects were not entertained.

In order to assess the impact of certain aspects of the data, the basic model, inferred as described above, was refit under various conditions. The resulting parameter estimates are shown in columns 2–5 of table 3. The first variation was to remove the information on SI values, using only clinical affection status and HLA (fig. 1B); then removing the information on HLA haplotypes, using only clinical affection status and SI

values (fig. 1C); and, finally, considering only clinical affection status (fig. 1D). The highest penetrance values, for both males and females, are obtained when all of the information is used together (depicted in fig. 1A), and the lowest penetrances are observed when only clinical affection status is considered in the segregation analysis. There is a trend toward higher penetrance of the homozygous hemochromatosis genotype as the amount of information utilized in the analysis increases. Presumably, the increased information (from SI measurement, HLA typing, or both) serves to alter the genotypic classification of some individuals, from unaffected hemochromatosis homozygotes to either heterozygotes or normal homozygotes. Particularly in the case where the HLA information is deleted, there is a substantial loss of precision in the estimates of the parameters describing the disease penetrance function, as evidenced by the standard errors (table 3); however, the shape of the penetrance curve (fig. 1C) remains much the same as that which was inferred with the additional information from the marker cosegregation (fig. 1A). This would suggest that the simultaneous consideration of

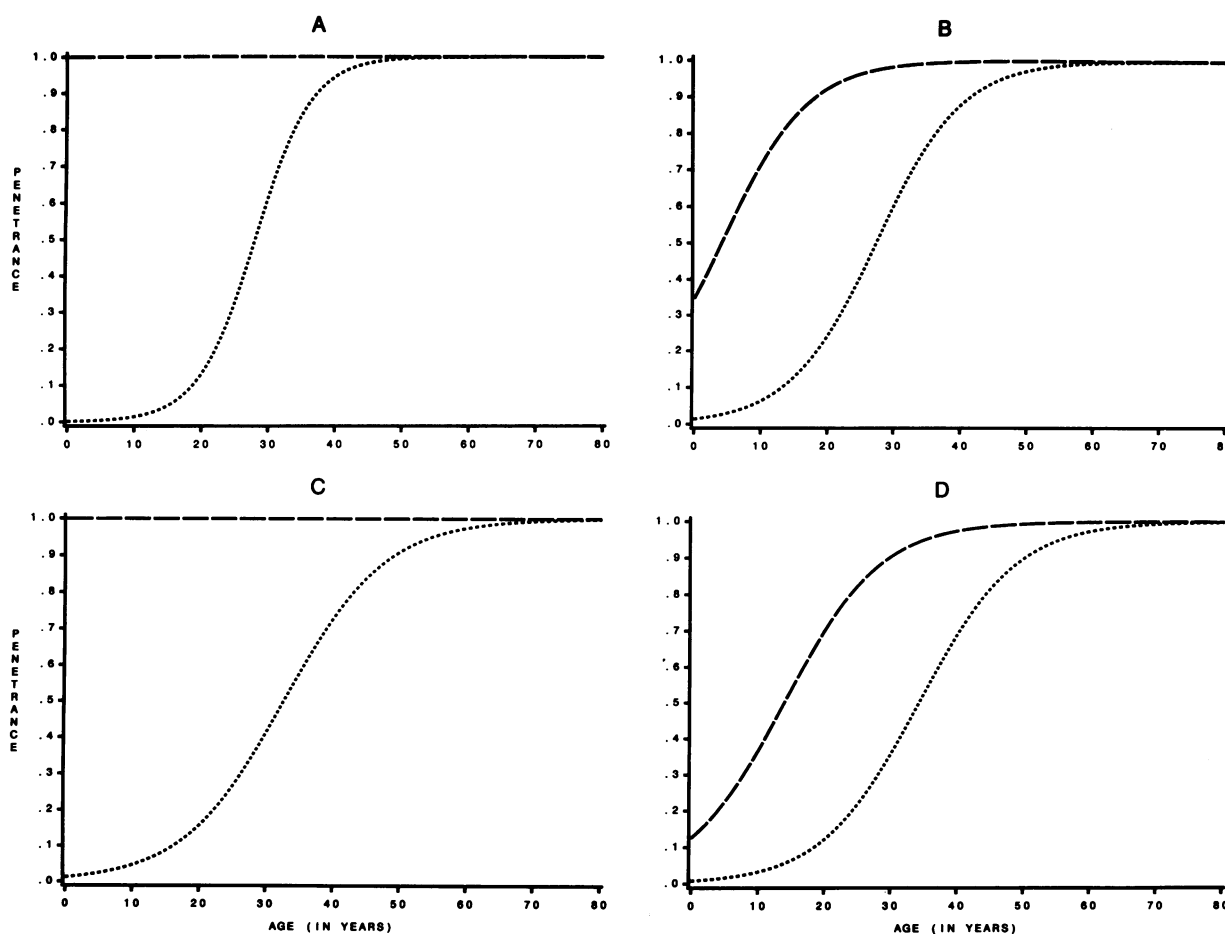


Figure 1 Penetrance of the recessive homozygous hemochromatosis genotype as a function of age (x -axis) and by sex. There was no evidence to suggest that clinically detectable disease could result in individuals with any other genotype. A, With use of clinical affection, SI, and HLA. B, With use of clinical affection and HLA only. C, With use of clinical affection and SI only. D, With use of only clinical affection. — — = Males; ---- = females.

a quantitative marker can provide considerable information for this type of analysis, even in the absence of a linked genetic marker.

The effect of the ascertainment correction also was of interest. The model parameters were reestimated without conditioning the likelihood on probands' affected status. As shown in table 3, the expected distortion in the estimate of the gene frequency is obtained. This has a significant effect on the resulting disease penetrance function, shown in figure 3. Whereas the curve for females is nominally affected, it seems that the elevated estimated gene frequency predicts considerably more preclinical hemochromatosis homozygotes, suggesting higher degrees of incomplete penetrance in males. Residual familial dependencies in the logistic function remain nonsignificant ($\chi^2_2 = 1.02$).

Finally, recall that an approximation designed to reduce computing time was implemented in the calculation of the residuals from major genotype in the quantitative penetrance function, where the parents were adjusted for the genotype of the i th child. In order to assess the effect of this procedure, the model parameters were reestimated by adjusting each phenotype for the actual parental and spouse genotypes, weighted by their appropriate probabilities. This resulted in an approximately 36-fold increase in the CPU time required by a SUN 4/260 computer to reach the solution. The information matrix associated with this solution could not be factored to obtain standard errors for the parameter estimates; however, the observation can be made that the point estimates are comparable to those obtained by using the approximation. One notable differ-

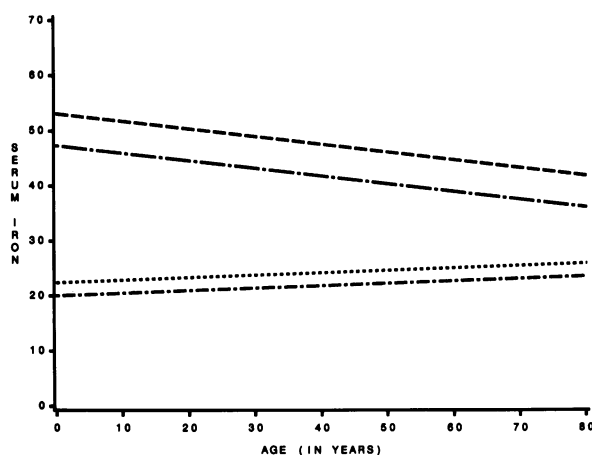


Figure 2 Estimated mean SI values ($\mu\text{M}/\text{liter}$), by genotype and sex, as a function of age. --- = HH males; -.-.- = HH females; = OO/HO males; - - - - = OO/HO females.

ence is the underestimation of the parent-child residual correlation (r_{PC}) under the solution utilizing the approximation; a similar observation was made in simulation experiments carried out by Demenais et al. (in press).

Discussion

The fact that autosomal recessive single-gene inheritance for predisposition to hemochromatosis is well established both from a consensus of studies conducted in various populations and for the Brittany data in particular allows us to validate this newly developed and extended methodology. The assumption of conditional independence of siblings, given their common parents, appears to be justified in this situation, as there is no evidence of excess sibling correlation in these data (Lalouel et al. 1985). In fact, results from likelihood ratio tests support the recessive hypothesis with respect to disease penetrance, as well as biochemical expression in SI values among hemochromatosis homozygotes.

The estimated penetrance of the recessive hemochromatosis genotype under the general model shows a clear difference between males and females. The genotype appears to be fully penetrant among males; this is comparable to the derived penetrances obtained by using the bivariate version of POINTER incorporating age-specific population prevalence data (Lalouel et al. 1985). This conclusion was somewhat surprising; therefore, the data were examined to determine which cases were providing the support for full penetrance in males, even at young ages. The youngest male homozygotes were

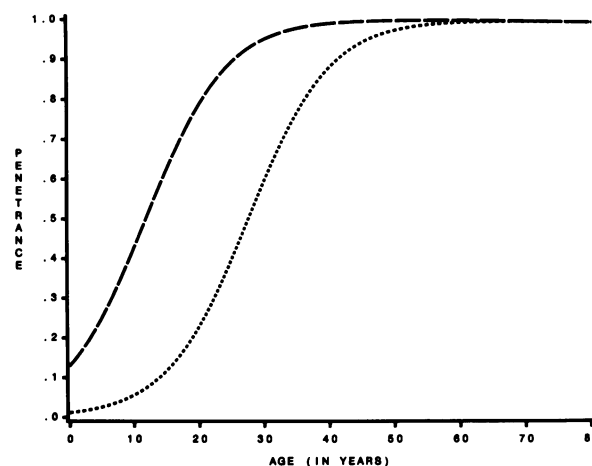


Figure 3 Penetrance of the recessive homozygous hemochromatosis genotype when estimated neglecting the ascertainment of the families. — = Males; = females.

14, 17, 19 (two individuals), and 21 years of age. These young men all had some mild symptoms, usually cutaneous manifestations or diabetes, which could be attributed to the disease. While this presentation might not ordinarily be recognized as hemochromatosis, these individuals were assessed to be clinically affected for the purposes of the present study, and the diagnosis was supported by follow-up biochemical tests. It should be noted that the extrapolation of the penetrance curve to age zero is simply a result of the estimated logistic function; there are no actual affected male hemochromatosis homozygotes younger than 14 years old in this data set. Thus, the implication that the hemochromatosis gene is fully penetrant in young males (i.e., those younger than 14 years of age) is not actually supported by the data. In the case of females, the genotype appears to reach full penetrance by approximately 50 years of age.

The distribution of SI values was characterized. Although mean SI values by genotype are expressed here as a function of age, the values obtained from this analysis are quite comparable to those presented in the previous studies by Cartwright et al. (1979) and Bassett et al. (1981). The means are comparable also to those presented by Dadone et al. (1982) in their analysis of a series of Utah pedigrees, with the exception of hemochromatosis homozygous females, whose mean SI value was considerably lower ($32.2 \mu\text{M}/\text{liter}$, with no significant age effects) than the estimate from the present study ($41.9 \mu\text{M}/\text{liter}$, at age 40 years). Genotype-specific covariate effects were demonstrated. Although only linear effects were entertained in this case, higher-

order terms and age $\times$ sex interaction can be entertained easily. The ability to statistically detect and estimate the effects of covariates in the context of the analysis model will be of great importance in subsequent investigations of other parameters of iron metabolism, where we anticipate more substantial and complex covariate effects. Moreover, this methodology may raise some issues about standard methods of adjusting phenotypes for covariate effects by regression prior to analysis, especially when the only available data comprise the ascertained sample of families.

It is well known in the context of a linkage analysis that ascertained families selected for their distribution of affected individuals will not bias either tests of hypotheses or estimates of θ , as long as they have been selected without regard to the marker phenotypes (Ott 1985 p. 121). However, in the present model, segregation parameters are being estimated simultaneously, and neglecting the ascertainment does produce bias in addition to the expected overestimation of the gene frequency. Thus, an ascertainment correction appropriate for the data appears to be necessary at least insofar as parameter estimation is concerned.

Linkage equilibrium constitutes one of the assumptions of this analysis. However, significant allelic associations between hemochromatosis and the HLA antigens A3, B7, and B14 have been reported (Simon et al. 1980), and LeMignon et al. (1983) also have reported a weaker, but significant, association with A11. The effect of ignoring allelic associations in a linkage analysis was studied by Clerget-Darpoux (1982), the problem being formulated to address the analysis of recessive diseases presumed to be linked to the HLA complex. Her results suggest that the lod score may be considerably underestimated (not pertinent to the present analysis) and that θ also may be slightly underestimated; for true θ 's less than 5%, the bias appears to be negligible.

Virtually all of the model fitting and hypothesis testing was carried out utilizing an approximation in the likelihood for calculating the residuals from major genotype, which led to conclusions not different from those of either other studies or previous analyses of these data by another method (Lalouel et al. 1985). Results from simulation studies reported by Demenais et al. (in press) suggest that, in the context of class A model assumptions, use of this approximation may lead to underestimation of the residual parent-child correlation in the case where there is an additional sibship correlation. Under these circumstances, Demenais et al. (in press) also report a spuriously high rejection rate for the "no

major gene" hypothesis when the null hypothesis is true. However, in the present case, Mendelian inheritance of a hemochromatosis gene was justifiably assumed, and use of the approximation expedited the likelihood function evaluations, making the segregation analysis of a multivariate phenotype with a linked marker possible within the practical limitations of available computing resources.

There are far-reaching applications for this sort of combined segregation/linkage analysis approach. This synthesis (Bonney et al. 1988) provides a comprehensive framework in which to conduct tests for candidate genes, as well as to consider several phenotypic outcomes simultaneously in order to detect and characterize the expression of an unmeasured gene. This latter aspect is particularly important; hemochromatosis has served as an excellent example of the power afforded for detecting Mendelian segregation by the simultaneous consideration of a quantitative indicator related to the disease liability and affection status. Although the discovery of a linked genetic marker certainly provides considerable information about the genetic etiology of a trait, careful studies to identify associated biological markers of the disease can be equally significant and powerful.

Acknowledgments

This work was partially supported by NIH grants GM 28719 (to D.C.R.), HD 18281 (to D.C.R.), and GM 39573 (to G.E.B.). The authors wish to acknowledge the helpful discussions with Dr. Florence Demenais.

References

- Bassett M, Halliday JW, Ferris RA, Powell LW (1984) Diagnosis of young subjects: predictive accuracy of biochemical screening tests. *Gastroenterology* 87:628-633
- Bassett ML, Halliday JW, Powell LW (1981) HLA typing in idiopathic hemochromatosis: distinction between homozygotes and heterozygotes with biochemical expression. *Hepatology* 1:120-126
- Beaumont C, Simon M, Fauchet R, Hespel JP, Brissot P, Genetet B, Bourel M (1979) Serum ferritin as a possible marker of the hemochromatosis allele. *N Engl J Med* 301:169-174
- Bomford A, Williams R (1976) Long-term results of venesection therapy in idiopathic haemochromatosis. *Q J Med* 45:611-623
- Bonney GE (1984) On the statistical determination of major gene mechanisms in continuous human traits: regressive models. *Am J Med Genet* 18:731-749

- (1986) Regressive models for familial disease and other binary traits. *Biometrics* 42:611–625
- Bonney GE, Lathrop GM, Lalouel J-M (1988) Combined linkage and segregation analysis using regressive models. *Am J Hum Genet* 43:29–37
- Borecki IB, Rao DC, Le Mignon L, Yaouanq J, Simon M, Lalouel J-M (1989a) Genetic hemochromatosis: distribution analysis of six laboratory measures of iron metabolism. *Am J Med Genet* 34:435–441
- Borecki IB, Rao DC, Yaouanq J, Lalouel J-M (1989b) Segregation of genetic hemochromatosis indexed by latent capacity of transferrin. *Am J Hum Genet* 45:465–470
- (1990) Serum ferritin as a marker of affection for genetic hemochromatosis. *Hum Hered* 40:159–166
- Percent transferrin saturation in segregating hemochromatosis. *Am J Med Genet* (in press)
- Borwein ST, Ghent CN, Flanagan PR, Chamberlain MF, Valberg LS (1983) Genetic and phenotypic expression of hemochromatosis in Canadians. *Clin Invest Med* 6:171–179
- Cannings C, Thompson EA (1977) Ascertainment in the sequential sampling of pedigrees. *Clin Genet* 12:208–212
- Cartwright GE, Edwards CQ, Kravitz K, Skolnick M, Amos DB, Johnson A, Buskjaer L (1979) Hereditary hemochromatosis: phenotypic expression of the disease. *N Engl J Med* 301:175–179
- Clerget-Darpoux F (1982) Bias of the estimated recombination fraction and lod score due to an association between a disease gene and a marker gene. *Ann Hum Genet* 46:363–372
- Dadone MM, Kushner JP, Edwards CQ, Bishop DT, Skolnick MH (1982) Hereditary hemochromatosis: analysis of laboratory expression of the disease by genotype in 18 pedigrees. *Am J Clin Pathol* 78:196–207
- Demenaïs FM, Murigande C, Bonney GE. A search for faster methods of fitting the regressive models to quantitative traits. *Genet Epidemiol* (in press)
- Elston RC, Stewart (1971) A general model for the genetic analysis of pedigree data. *Hum Hered* 21:523–542
- Lalouel J-M (1979) GEMINI—a computer program for optimization of general non-linear functions. Tech rep 14, Department of Medical Biophysics and Computing, University of Utah, Salt Lake City
- Lalouel J-M, Le Mignon L, Simon M, Fauchet R, Bourel M, Rao DC, Morton NE (1985) Genetic analysis of idiopathic hemochromatosis using both qualitative (disease status) and quantitative (serum iron) information. *Am J Hum Genet* 37:700–718
- Lathrop GM, Lalouel J-M, Julier C, Ott J (1984) Strategies for multilocus linkage analysis in humans. *Proc Natl Acad Sci USA* 81:3443–3446
- LeMignon L, (1982) Etude génétique de l'hémochromatose idiopathique: le point à propos de 147 enquêtes familiales comportant un groupage HLA. MD thesis, Faculté de Médecine de Rennes, Rennes, France
- LeMignon L, Simon M, Fauchet R, Edan G, LeReun M, Brisot P, Genetet B, et al (1983) An HLA-A11 association with the hemochromatosis allele? *Clin Genet* 24:171–176
- Ott J (1985) Analysis of human genetic linkage. Johns Hopkins University Press, Baltimore
- Simon M, Alexandre JL, Fauchet R, Bourel M (1980) The genetics of hemochromatosis. *Prog Med Genet* 4:135–168
- Simon M, Bourel M, Genetet B, Fauchet R (1977) Idiopathic hemochromatosis: demonstration of recessive transmission and early detection by family HLA typing. *N Engl J Med* 297:1017–1021