

Influence of Genotype-dependent Effects of Covariates on the Outcome of Segregation Analysis of the Body Mass Index

Ingrid B. Borecki,* George E. Bonney,[‡] Treva Rice,* Claude Bouchard,[§] and D. C. Rao*[†]

*Division of Biostatistics and [†]Departments of Psychiatry and Genetics, Washington University School of Medicine, St. Louis; [‡]Department of Biostatistics, Fox Chase Cancer Center, Philadelphia; and [§]Physical Activity Sciences Laboratory, PEPS, Laval University, Ste-Foy, Quebec

Summary

Several recent studies of the body mass index (BMI) have provided support for a recessive major gene influencing heaviness in humans. Segregation analysis of the BMI was carried out recently in a series of randomly sampled French-Canadian families to determine whether we could replicate the major gene finding by using a residual phenotype adjusted for the effects of age and sex. The best model included a recessive major effect for high BMI values with residual familial resemblance; however, Mendelian transmission could not be confirmed, and the no-transmission hypothesis (where all the τ 's are constrained to be equal) was not rejected. Considering that the BMI is a complex phenotype affected by many factors and that there are known variations in body composition during growth and aging, we undertook a reanalysis of the data, using a model that allowed the estimation of genotype-specific age and gender effects. New tests on the transmission parameters satisfy the criteria for inferring Mendelian segregation. The results suggest that individuals with the "high" recessive genotype show the greatest degree of heaviness at birth, with a subsequent trend toward lower values throughout life, while individuals with the dominant "normal" genotypes show no appreciable trends with age. In addition, the "high" genotype appears to confer a greater degree of heaviness in females as compared with males. These results, along with other observations from the data, suggest that, while a recessive single gene influence may be discernible, the phenotypic expression of the BMI is likely to be complicated by genotype $\times$ environment interactions and, possibly, by the action of other loci. Further, the data also are consistent with the hypothesis that modifying factors may include the adoption of a more prudent life-style by individuals genetically predisposed to heaviness and a secular increase in the incidence, prevalence, and potency of environmentally based triggers leading to a higher penetrance of the "heavy" genotype in the young.

Introduction

Phenotypes devised to quantitate the degree of obesity have been the subject of many recent investigations, motivated by the increased risk of cardiovascular morbidity and mortality (Burton et al. 1985; Kral 1985; van Itallie and Abraham 1985), largely through increased risk of diabetes, hypertriglyceridemia, changes in plasma lipoprotein profile and composition, and hyper-

tension (Royal College of Physicians 1983; Bray 1985). Specifically of interest is the identification and characterization of both genetic and environmental determinants of obesity, so as to foster an understanding of the heterogeneous processes leading to human obesity and to provide an opportunity to devise rational intervention strategies aimed at reducing health risks.

Obesity can be considered as a complex phenotype that is multifactorially determined, involving genes, environmental exposures, and gene $\times$ environment interactions (Bouchard 1989). Perhaps the most widely studied phenotype is the body mass index (BMI, weight [kg]/height [m]²). The BMI is a measure of "heaviness" (Garn et al. 1986; Bouchard and Pérusse 1988) rather than of "fatness," reflecting the combined effects of body build, proportions, and lean body mass as well as

Received December 1, 1992; final revision received May 21, 1993.

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

© 1993 by The American Society of Human Genetics. All rights reserved.
0002-9297/93/5303-0012\$02.00

body fat (Bouchard 1989); however, it exhibits substantial correlations with fat mass, on the order of .81 and .83 in adult males and females, respectively (Borecki et al. 1991), and has been validated as a risk factor for cardiovascular disease (van Itallie and Abraham 1985).

The observation that the prevalence of obesity has been increasing in developed countries during the past half-century (Price 1992) is consistent with studies suggesting the importance of environmental influences. While the frequency and magnitude of precipitating factors may be rising, the phenotypic response at the individual level possibly could be a function of genetic predisposition. Direct evidence to support this hypothesis has been presented recently for various obesity phenotypes by measuring the response, in identical twins, to a 100-d overfeeding regimen (Bouchard et al. 1990). The changes in absolute and relative fat mass in response to the homogeneous, controlled environment were significantly more similar within each MZ twin pair than among nonrelated subjects, suggesting that gene $\times$ environment interaction effects may account for an appreciable amount of the variance in obesity phenotypes. However, the presence of secular environmental effects could affect the genetic analysis of obesity phenotypes and become manifest in an analysis of cross-sectional data, either as intergenerational differences in heritability or as shared environmental effects for individuals close in age and belonging to the same cohort, depending on the age structure of the families. In an analysis of possible major gene determinants, such unaccounted temporal trends could result in apparently non-Mendelian transmission of the major effect.

In addition to the complication of secular trends, there are developmental trends for most obesity phenotypes studied to date. That is, the absolute and relative amounts of fat mass, fat-free mass, and amount of fat in regional depots change through early childhood and the growth spurts that characterize adolescence, before attaining more stable adult values (Malina and Bouchard 1988). For example, the absolute and relative amounts of fat in truncal/abdominal subcutaneous depots increase progressively with time, until about 60–70 years of age, particularly and more frequently in males (Bouchard et al. 1991). In recognition of this variability, it is common practice to adjust obesity phenotypes for the effects of age, usually by multiple regression, prior to analysis. While this statistical procedure can correct adequately for age trends in the overall mean and optionally in the variance as well, there may remain temporal trends in the covariance structure within families

that reflect these developmental phases. A common assumption made in most studies is that the effects of the relevant determinants remain constant over all ages, with variations being attributed to random or unique environmental effects. Failure to take account of temporal trends in either the genetic or environmental components can be one factor contributing to the heterogeneity of reported results and can result in altered segregation patterns for major loci, besides being a well-motivated biological question.

While the majority of genetic studies have investigated the BMI under a multifactorial model, there have been a few investigations testing a major gene hypothesis; that is, after allowance is made for the possibility of many genes and environmental factors, is there evidence that a single gene has a major effect on the distribution of BMI values? Earlier studies using Structured Exploratory Data Analysis (SEDA) concluded that there was no evidence to support a major gene hypothesis (Karlin et al. 1981; Zonta et al. 1987), although there was suggestion of possible dominance effects. Several recent studies using large samples and parametric mixed-model segregation analysis have provided support for a recessive major gene whose transmission patterns conform to the Mendelian hypothesis (Price et al. 1990; Province et al. 1990; Moll et al. 1991).

Analysis of the BMI has been carried out recently by using data on 1,630 individuals in 375 randomly sampled French-Canadian families living in the environs of Quebec City, to determine whether we could replicate the major gene finding (Rice et al. 1993a). Prior to analysis, raw BMI values were adjusted for the effects that age had on the mean and variance within four groups (fathers, mothers, sons, and daughters), by multiple regression (Borecki et al. 1991). There was minimal heteroscedasticity. Commingling analysis carried out on the adjusted BMI values revealed that the parents and offspring were identically distributed under any distributional model, although under parsimony there was admixture of normal components in the parental distribution, while only residual skewness under a single distribution was found in the offspring (Borecki et al. 1991). Thus, while there was no significant indication of distributional heterogeneity by generation, the differing parsimonious models were consistent with a developmental hypothesis.

Segregation analysis was carried out by using the unified mixed model as implemented in the computer program POINTER. Hypothesis tests indicated that high BMI values were determined by a recessive genotype and that residual familial resemblance could be ac-

counted for by a multifactorial component with no intergenerational differences. However, tests on the transmission parameters, carried out to assess the tenability of the major gene model, showed that it was possible to reject the Mendelian hypothesis, while the no-transmission hypothesis (where all the τ 's are constrained to be equal) was not rejected. Thus, not only was the Mendelian major gene hypothesis not confirmed, but it appeared that the major effect was not transmitted within families, indicating a probable environmental origin. In addition, the major effect did not appear to be in equilibrium between the parental and the offspring generations, since the equal τ 's did not approximate the quantity $(1-q)$, where q is the frequency of the allele leading to high BMI values.

In addition to both the known variations in body composition during growth and aging and the increasing prevalence of obesity, the suggestion of possible intergenerational differences in the major effect in these data comes from the observation that the parsimonious distributional models in parents and children differed and that, while there was solid statistical evidence of a major effect from segregation analysis, the transmission apparently did not depend on parental outisotype and suggested a lack of intergenerational equilibrium. Thus, we hypothesize that evidence for a Mendelian major gene is affected by possible developmental and/or secular trends. Here, a reanalysis of the data was undertaken by using a model that allowed the estimation of genotype-specific covariate effects. This provides an opportunity to entertain the hypothesis that genotypes may exhibit different age- and/or sex-dependent penetrances.

Material and Methods

Sample and Phenotype Adjustments

The Quebec Family Study (QFS) consists of 1,628 individuals from randomly sampled French-Canadian families living in the environs of Quebec City and recruited through the media during the years 1978–81 to study the genetic effects on several physiological and biochemical traits. The age range in parents was 30–60 years, and in offspring it was 8–26 years. Certain exclusions were made for the purposes of this analysis, on the basis of several relevant factors. First, adopted offspring of no genetic relation to either of the parents, as well as cousins and friends living with the family, were excluded. In instances where no offspring with valid data remained or in families with no natural issue, the parents also were excluded from the analysis. One twin

Table 1

Descriptive Statistics of Raw BMI and Age in the QFS

GROUP (N)	BMI		AGE	
	Mean	SD	Mean	SD
Fathers (357)	25.4	3.3	44.2	5.1
Mothers (368)	23.4	3.6	42.1	5.0
Sons (484)	19.2	2.8	14.5	3.3
Daughters (419)	19.1	2.8	14.8	3.4

was excluded at random from each MZ twin pair, although verified DZ twin pairs were retained. In all, there remained 301 nuclear families in which 282 fathers, 294 mothers, and 647 offspring had valid data.

The means and SDs of the raw BMI within four groups defined by fathers, mothers, sons, and daughters are shown in table 1, along with the age distribution. Raw BMI values were adjusted prior to analysis, for the effects that age and its higher-order terms had on the mean and variance within the four groups, by multiple regression (Borecki et al. 1991); there was minimal heteroscedasticity. The residual phenotypes also were standardized within the four groups. This adjustment procedure was undertaken to produce residual phenotypes suitable for analysis under the mixed model (Rice et al. 1993a). We use the same standardized residual phenotype in the present analysis, in order to enable a comparison of the results; in addition, utilization of data adjusted in this manner allows a separate identification of covariate effects that are solely genotype dependent, since any parallel effects of age and sex that are common to all genotypes already are absent.

Segregation Analysis Methodology

Segregation analysis with genotype-dependent covariate effects was carried out by using REGRESS (Bonney et al. 1988), which is based on the regressive models of Bonney (1984). The extension including the simultaneous estimation of covariate effects (with or without genotypic dependencies) is easily carried out in this context and has been done for another application (Borecki et al. 1990). The quantitative phenotype is denoted Y (e.g., BMI). The likelihood function takes the form $P(Y|g, Y_A, X)$, where the probability of observing the phenotypes depends on the underlying major genotypes (g , where g is AA, Aa, or aa), phenotype observations on relevant ancestors (Y_A), and adjustment for an arbitrary number of covariates (X). The likelihood of

observing the phenotypes for all family members, conditional on major genotypes, is summed over all genotypic combinations, weighted by the probability of the major genotypes. 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, by using transmission probabilities as follows: τ_1 , τ_2 , and τ_3 are the probabilities that an A allele is transmitted from a parent whose genotype is AA, Aa, or aa, respectively. Under Mendelian transmission, $\tau_1 = 1$, $\tau_2 = 1/2$, and $\tau_3 = 0$. It is assumed that the distribution of phenotypes, conditional on genotypes, follows a multivariate normal distribution with mean μ_g and variance-covariance matrix $V = (\sigma^2, \rho_{ij})$.

The dependency of the phenotypes, adjusted for major genotypes (residuals from major genotypes [RMGs]), on those of ancestors is modeled under class A assumptions (Bonney 1984), which essentially posits that sibs' RMGs are conditionally independent, given their parents' RMGs. Thus, the sib-sib residual correlation is due solely to the fact of common parentage and is determined by the parent-offspring correlation and spouse correlation, if any: $\rho_{CC} = 2\rho_{PC}^2/(1+\rho_{FM})$. The distribution of an individual's phenotype, conditional on major genotype and on the genotypes and phenotypes of the spouse, father, and mother, is assumed to follow a normal density whose mean, u , is given by $u = \mu_g + b_S(Y_S - \mu_{gS} - \sum b_{gSj} X_{Sj}) + b_F(Y_F - \mu_{gF} - \sum b_{gFj} X_{Fj}) + b_M(Y_M - \mu_{gM} - \sum b_{gMj} X_{Mj}) + \sum b_{gij} X_{ij}$, where 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, M refers to the mother, C refers to a child, and the genotype-specific covariate adjustment is summed over all covariates. The variance can be written as $w = \sigma^2(1 - b_S\rho_{FM} - b_F\rho_{FC} - b_M\rho_{MC})$.

The parameters of the model are as follows: the mean of the overall distribution or, under a major gene model, the mean of low homozygotes (m); the displacement between the means of the two homozygote distributions (t); the relative position of the heterozygote distribution or dominance (d); the frequency of the allele associated with lower BMI values (p); the within-genotype variance (σ^2); ρ_{PC} , the residual parent-child correlation; ρ_{CC} , the dependent sib-sib correlation constrained to be equal to $2\rho_{PC}^2/(1+\rho_{FM})$; and regression coefficients for age, up to third-order terms, which can

be subscripted by the homozygous-low (LL), heterozygous (LH), or homozygous-high (HH) genotypes. Parameter estimation and tests of hypotheses are carried out by using the maximum-likelihood method.

Results

Initially, certain hypotheses were tested by using REGRESS without genotype-dependent covariate effects, to determine the comparability between the present analysis and the previous mixed-model analysis. It should be noted that we found no evidence against the equality of the father-child and mother-child correlations, and the spouse correlation was not significantly different from zero. The results suggested a major gene effect, and the mode of inheritance was found to be recessive, a result similar to that of the POINTER analysis. The most parsimonious model without genotype-specific effects of covariates is presented as model I in table 2 and involves the following parameters: m , σ , p , t , and ρ_{PC} , with $d = 0$.

Next, age terms were added to the model, where m , σ , p , t , and ρ_{PC} were estimated in addition to the regression coefficients for the age terms. The results are shown in table 2. As expected, age and higher-order terms were not significant (model II vs. model I and model III vs. model I) when the preadjusted data were used, suggesting that our prior data adjustments were adequate. However, allowing the regression on age to be genotype specific resulted in a significant improvement in the fit of the model (model IV vs. model I). Higher-order genotype-specific age terms were not necessary (model V vs. model IV and model VI vs. model IV). Under the model with first-order genotype-specific effects, neither the Mendelian ($\chi^2 = 2.67$, 3 df, $P = .445$) nor the equal τ model could be rejected ($\chi^2 = 0.92$, 2 df, $P = .631$). Spouse correlation remains nonsignificant (model VII vs. model IV).

Finally, even though our initial interest was in age trends per se, we decided to investigate also possible sex differences in the genotypic effect. To this end, the next series of models had m , σ , p , t , ρ_{PC} , age_{LL} , and age_{HH} estimated, with $d = 0$ (model IV in table 2), in addition to the parameters for sex and age $\times$ sex interaction, with the possibility of genotype-specific effects. The results are shown in table 3. Again, the adjusted BMI does not depend on aggregate sex effects (model VIII vs. model IV), since the data were standardized within sex, prior to analysis. However, it appears that there are significant genotype-dependent sex effects (model IX vs. model IV). While the mean for the LL or

Table 2**Tests on Age and Higher-Order Terms, with m , σ , p , t , and ρ_{PC} Estimated and $d = 0$**

Model	Term(s) ^a	-2 Log Likelihood	Likelihood-Ratio Test ^b
I	No age effects	1,099.52	...
II	age	1,098.01	1.51 (1 df)
III	age, age ² , and age ³	1,097.94	1.58 (3 df)
IV	age _{L-} and age _{HH}	1,090.44	9.08 (2 df)
	General τ 's	1,087.77	2.67 (3 df)
	Equal τ 's	1,088.69	.92 (2 df)
V	age _{L-} , age _{HH} , age _{L-} ² , and age _{HH} ²	1,089.11	1.33 (2 df)
VI	age _{L-} , age _{HH} , age _{L-} ² , age _{HH} ² , age _{L-} ³ , and age _{HH} ³	1,087.10	3.34 (4 df)
VII	age _{L-} , age _{HH} , and ρ_{FM}	1,088.27	2.17 (1 df)

^a L- represents either the LL genotype or the LH genotype.^b χ^2 shown is that associated with testing the significance of added model parameters (see text).

LH genotypes is little affected by sex differences, the effect of the HH genotype is greater in females than in males. There is no evidence to support a role for age $\times$ sex interaction, neither overall nor genotype specific. Tests on the transmission probabilities suggest that the Mendelian hypothesis was of borderline significance ($\chi^2 = 7.90$, 3 df, $P = .048$), while the no-transmission hypothesis was rejected ($\chi^2 = 8.67$, 3 df, $P = .013$). Parameter estimates for the most parsimonious recessive model (model IX), under Mendelian, general, and equal transmission probabilities, are shown in table 4.

These analyses were carried out by using existing computer programs, which are based on class A assumptions. To look into the validity of the class A assumptions, we obtained maximum-likelihood estimates of the familial correlations in the QFS data. There is a significant sib-sib correlation ($\rho_{CC} = .24 \pm .05$), which is approximately equal to the parent-child correlation

($\rho_{PC} = .23 \pm .03$); the spouse correlation is not significant ($\rho_{FM} = .07 \pm .06$). If these were residual correlations, the class A model would imply a sibling correlation of .11; therefore, it appears likely that there is excess sibling resemblance in these data, above that which is modeled under class A assumptions. Whether our inference, especially that of Mendelian inheritance of the major effect, is affected by this potential underestimation of the sibling correlation can be checked only by using a more general, class D model. Although implementation of the class D models has been computationally daunting, recent advances in identifying suitable approximations (Demenais et al. 1990) and alternative formulations (Bonney 1992) of the likelihood function have made it feasible. We have used one such computer program, REGX, which is under the final stages of development. By fitting a compound regressive model (Bonney 1992), which is equivalent to a class

Table 3**Tests on Sex and Age $\times$ Sex Interaction Terms, with m , σ , p , t , ρ_{PC} , age_{L-}, and age_{HH} estimated and $d = 0$**

Model	Term(s)	-2 Log Likelihood	Likelihood-Ratio Test
VIII	sex	1,089.52	.92 (1 df)
IX	sex _{L-} and sex _{HH}	1,076.77	13.67 (2 df)
	General τ 's	1,068.87	7.90 (3 df)
	Equal τ 's	1,077.47	8.60 (2 df)
X	sex _{L-} , sex _{HH} , and age $\times$ sex	1,075.75	1.02 (1 df)
XII	sex _{L-} , sex _{HH} , age $\times$ sex _{L-} , and age $\times$ sex _{HH}	1,075.24	1.53 (2 df)

NOTE.—See footnotes to table 2.

Table 4

Parameter Estimates of the Most Parsimonious Recessive Model, under Mendelian, General, and Equal Transmission Probabilities

PARAMETER	τ 's		
	Mendelian ^a	General	Equal
p	.764	.737	.740
σ	.698	.677	.676
ρ_{PC}	.169	.169	.181
ρ_{CC}^b	.057	.057	.066
μ_L	-.274	-.279	-.301
μ_{HH}	1.917	2.176	2.171
$\beta_{age(L-)}$	.002	.002	.002
$\beta_{age(HH)}$	-.023	-.028	-.028
$\beta_{sex(L-)}^c$	.045	.049	.056
$\beta_{sex(HH)}^c$	1.042	.952	.970
τ_1	[1]	1.000	.774
τ_2	[1/2]	.550	.774
τ_3	[0]	.440	.774
-2 log likelihood	1,076.77	1,068.87	1,077.47

^a Parameters in brackets are fixed at values shown.

^b Sibling correlation is a dependent parameter calculated as $2\rho_{PC}$.

^c Sex is coded as 1 for males and as 2 for females.

D formulation, we were able to perform some validity checks on the possible effects of class A assumptions.

Analysis using REGX focused on the case where no covariates were entertained. A class A model can be formed by making the sibling correlation dependent on the parent-child correlation; thus, comparison with a class D model yields a 1-df test wherein the sibling correlation is estimated. Results under class A assumptions using REGX were comparable to those obtained by using REGRESS: a recessive major gene was verified ($\chi^2 = 30.5$, 2 df, $P < .001$), whose transmission was compatible with Mendelian expectations ($\chi^2 = 1.64$, 3 df, $P = .650$), and the no-transmission model was rejected ($\chi^2 = 6.36$, 3 df, $P = .041$). Estimating the residual sibling correlation by assuming a Mendelian recessive gene showed significant excess sibling resemblance over that predicted under a class A model, as expected ($\rho_{CC} = .187 \pm .051$, $\chi^2 = 10.9$, 1 df, $P < .0001$). Following up with tests on the transmission parameters under the class D formulation showed a remarkably good fit to the Mendelian hypothesis ($\chi^2 = 0.08$, 3 df, $P = .994$) and definite rejection of the no-transmission model ($\chi^2 = 13.74$, 2 df, $P < .0001$). Apparently, although there is significant excess sibling resemblance, the magnitude of the effect was not sufficient to prevent the inference of Mendelian transmission under class A assumptions, in

this case. The results obtained by accounting for the residual sibling correlation give stronger support for the conclusions of our original analysis and provide a better fit to the Mendelian gene hypothesis. It should be noted that parameter estimates are virtually identical under either class A assumptions or class D assumptions; aside from the obvious difference in the sibling correlation, only the parent-child correlations are slightly affected (table 5).

Recalling that the test of the Mendelian hypothesis, under class A assumptions and incorporating genotype-dependent covariates, was of borderline significance ($P = .048$; table 4), we also endeavored to see whether allowing for a sibling correlation could clarify the conclusions regarding transmission patterns. Fitting a class D model again showed a significant sibling residual correlation ($\rho_{CC} = .264 \pm .100$), and, more important, the Mendelian hypothesis was not rejected ($\chi^2 = 6.32$, 3 df, $P = .10$), while the no-transmission hypothesis was rejected ($\chi^2 = 23.38$, 2 df, $P < .0001$). Thus, it appears that the borderline fit of the Mendelian hypothesis under the class A model could be attributed, at least in part, to unaccounted sib-sib resemblance.

Discussion

Genetic analysis of phenotypes such as obesity requires development and application of models that begin to capture the complexity of the phenotypic expression documented by observational studies. This strategy can clarify the evidence for single-gene determination and can provide insight as to the interaction of genotype with environment, and it also can provide parameter estimates that permit a better assessment of the probable genotype, an essential component for a suc-

Table 5

Comparison of Parameter Estimates of the Mendelian Recessive-Gene Model That Were Obtained from Analysis Under Class A and Class D Model Assumptions

Parameter	Class A	Class D
p	.753 $\pm$.020	.757 $\pm$.019
σ	.687 $\pm$.027	.687 $\pm$.026
ρ_{PC}	.142 $\pm$.025	.136 $\pm$.027
ρ_{MC}	.135 $\pm$.022	.128 $\pm$.022
ρ_{CC}	.038 ^a	.187 $\pm$.051
μ_L	-.151	-.156
μ_{HH}	2.758	2.760

^a Dependent parameter, not iterated.

cessful linkage study. Entertaining possible genotype-dependent variations by age and gender represents a first step in that direction and has been undertaken elsewhere, both in an analysis of serum iron levels in hereditary hemochromatosis (Borecki et al. 1990) and in the evaluation of a major gene determinant for systolic blood pressure (Pérusse et al. 1991). In both cases, phenotypic expression was found to be a function not only of the underlying major genotype but also of genotype-specific age and gender effects. The hypothesis that expression of the BMI is influenced by age and gender is well motivated (e.g., see Malina and Bouchard 1988; Bouchard et al. 1991). Since a previous analysis of age- and sex-adjusted BMI in the QFS data resulted in rejection of Mendelian transmission (Rice et al. 1993a), we hypothesized that there may remain variations by age and gender that differ among genotypes.

Our reanalysis indicated evidence for a recessive major gene effect where the frequency of the high genotype, HH, was 5.6%, similar to results of previous reports (Price et al. 1990; Province et al. 1990; Moll et al. 1991). There was a residual parent-child correlation of .17, with no evidence of spouse correlation. The genotypic means are shown, as a function of age, by sex, in figure 1 and table 6. We should recall that age effects in the mean that are common to all genotypes have been removed prior to analysis, so trends such as that most people tend to get heavier as they get older no longer exist in the data. For the most part, it appears that individuals of the LL genotype or the LH genotype do not exhibit much difference from the general population trend, as the phenotypic mean does not change appreciably over the course of time, showing only a slight increase. The more pronounced trend can be observed in relation to the HH genotype, where the difference with respect to the other genotypes is the largest at birth and decreases over time. The specific reason for this is not immediately apparent. These data would seem to suggest that the unique environmental variations experienced by these HH individuals act in such a way as to systematically decrease the degree of obesity over time. This could be the case if genetically obese individuals were, with some degree of success, adopting a more prudent life-style by dieting or taking some other measures to reduce their heaviness. Alternatively, a secular increase in the incidence, prevalence, or potency of relevant environmental factors or triggers, interacting specifically with the permissive genotype to produce greater degrees of heaviness in younger individuals, also is consistent with these findings. In any case, the results suggest that the strategy of selecting families

through obese children, for the identification of specific genes relevant to obesity or heaviness, is a sound one.

The HH genotype also appears to confer a greater degree of heaviness in females as compared with males. In these data and without regard to latent genotype, the BMI is more highly correlated with fat mass than with fat-free mass (Borecki et al. 1991), and the correlation with fat mass is stronger in females than in males, especially within the offspring generation ($r = .806$ and $.644$ among daughters and sons, respectively). In the parental generation, fathers show a slightly higher correlation between BMI and fat-free mass ($r = .654$) than do mothers ($r = .542$). However, it is not immediately apparent whether the sex difference in "heaviness" among individuals with the HH genotype is attributable to the amount of fat mass, fat-free mass, or a function of both.

An investigation of a related trait, weight adjusted for height, has been reported recently by Tired et al. (1992), who used family data on healthy individuals attending the Preventive Medicine Center in Nancy. Although the phenotype under analysis is not identical, it reflects the level of ponderosity, as does the BMI. The familial correlations were slightly different from those of the present study, notably exhibiting a higher degree of sibling resemblance: $\rho_{FM} = .094 \pm .040$, $\rho_{PC} = .198 \pm .023$, and $\rho_{CC} = .327 \pm .034$. Analysis using a class D regressive model revealed that an additive major effect with age- and generation-dependent effects best fitted the data; however, both the Mendelian hypothesis and the no-transmission hypothesis were rejected. The effect of the latent ousiotype differed between parents and offspring; the effects of the LH and HH ousiotypes increased with age in offspring, and, while the effect seemed to stabilize with age in the LH ousiotype, the effect of the HH ousiotype appeared to decrease over the course of adulthood. It is interesting to note the similarity of this finding to the one reported here—i.e., the decreasing effect of HH over adulthood—despite the fact that, in our analysis of the BMI, it was possible to infer a Mendelian effect with no evidence of heterozygote expression. However, the finding by Tired et al. that the age dependence differed by generation prompted us to test an additional hypothesis in the QFS data by including generation and age $\times$ generation interaction as potential covariates. There was no evidence to suggest that these effects are significant in our data ($\chi^2 = 1.36$, 4 df, $P = .851$).

While the primary focus of this investigation is on the genetic etiology of obesity and, specifically, on the

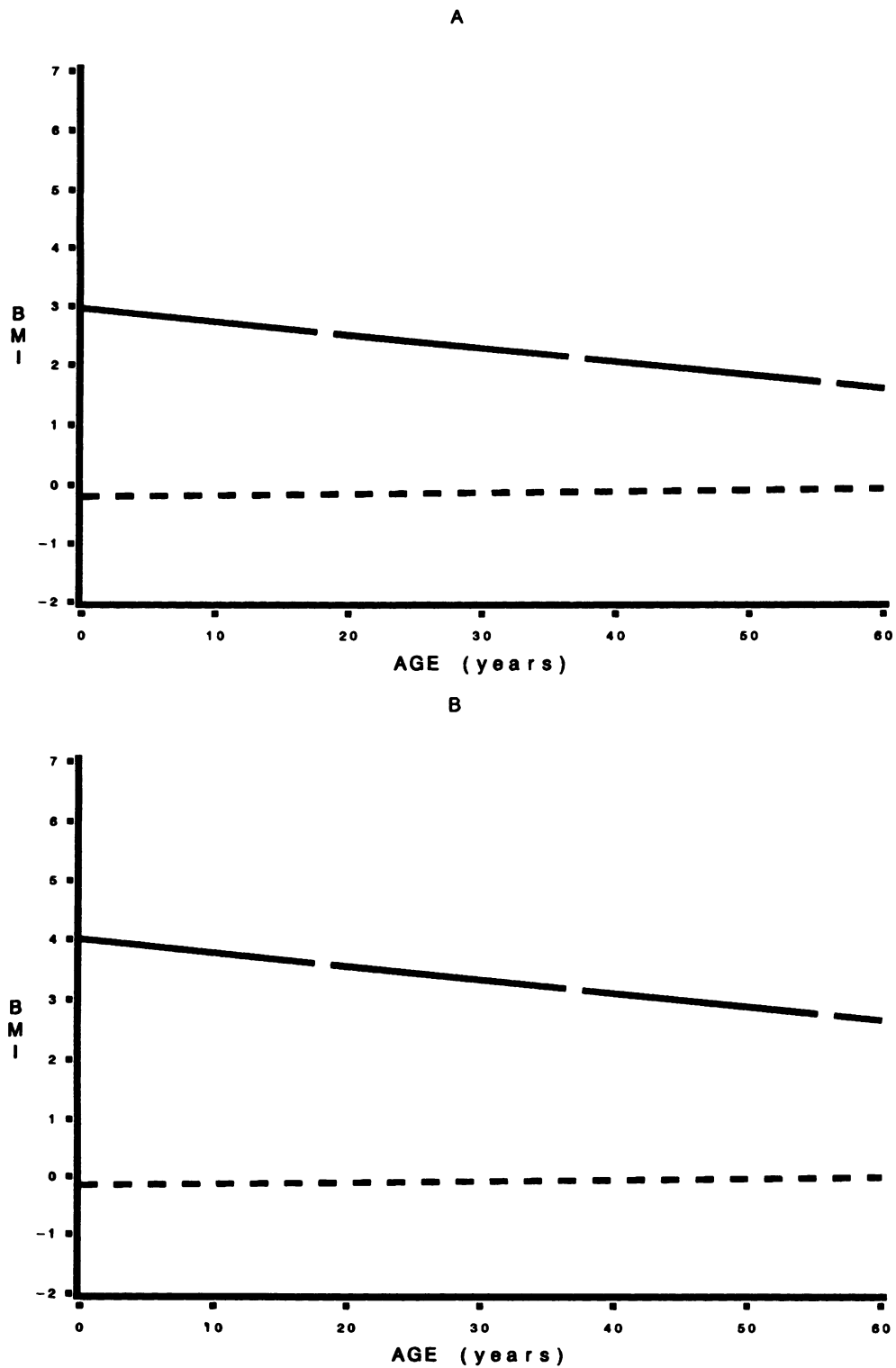


Figure I Mean adjusted BMI plotted as a function of age for individuals with LL or LH “normal” genotypes (short dashed lines) or with HH “heavy” genotype (long dashed lines) in males (A) and females (B), under a recessive model with residual familial correlations estimated under class A assumptions.

Table 6

Comparison of Genotypic Means of Standardized BMI Expressed as a Regression Function on Age, by Sex, under Class A and Class D Model Assumptions

Genotype	Frequency	Males	Females
Class A:			
LL or LH	94.4%	$-.229 + (.002 \times \text{age})$	$-.184 + (.002 \times \text{age})$
HH	5.6%	$2.959 - (.023 \times \text{age})$	$4.001 - (.023 \times \text{age})$
Class D:			
LL or LH	94.9%	$-.278 + (.004 \times \text{age})$	$-.237 + (.004 \times \text{age})$
HH	5.1%	$2.726 - (.012 \times \text{age})$	$3.869 - (.012 \times \text{age})$

potential and adequacy of the BMI as an index of obesity, several methodological issues also are germane. The differences among previous segregation analyses, in terms of sample structure, selection criteria, and alternative indexes of obesity, have been reviewed by Rice et al. (1993a). The present discussion will focus on the use of class A versus class D models and on the tests of the transmission parameters.

Demenais et al. (1990) have investigated the robustness of the class A model against false inference of a major gene in the presence of a sibling correlation exceeding that which is specified by the model. Analysis of data generated from models incorporating polygenic-type effects amounting to sibling correlations that exceeded those predicted in a class A model by .045–.32 yielded 87%–100% false inference rates. However, tests on the transmission probabilities, requiring both nonrejection of Mendelian transmission and rejection of no transmission, reduced these rates substantially, to 0%–13%. Thus, some protection can be afforded by relying on tests on the transmission parameters. These tests were carried out in this analysis first by using the class A formulation without covariate effects, and the results satisfied the expectations from a major gene etiology: Mendelian transmission was verified ($\chi^2 = 6.34$, 3 df, $P = .096$), and the no-transmission hypothesis (where the τ 's are constrained to be equal) was rejected ($\chi^2 = 10.55$, 2 df, $P = .005$). This result in itself should provide some confidence in concluding that there is a major gene effect. Additional reanalysis, using a compound regressive model (Bonney 1992), which is a class D formulation, led to the inference of the same qualitative model (i.e., a recessive single gene with residual familial correlations and no spouse correlation), in which a significant sibling correlation was estimated at $.19 \pm .05$, as compared with the dependent value of .04 under class A assumptions. Allowing for the sibling correlation resulted in both an improved fit

to the Mendelian hypothesis and stronger rejection of the no-transmission hypothesis, suggesting that the effect of failing to account for excess sibling resemblance in this case was accommodated in the model by the transmission parameters' deviations from their Mendelian values, since all the other parameter estimates remained remarkably stable under either model formulation. It appears that a class A model may be sufficient for the analysis of some data when the Mendelian hypothesis is true, even in the presence of excess sibling resemblance. However, the obstacle is that it is difficult to establish the veracity of the Mendelian hypothesis with confidence when a class A model is used alone.

It is becoming increasingly clear that testing transmission parameters for consistency with Mendelian expectations is a valid and useful procedure (Demenais et al. 1993). While this is an intuitively sound recommendation and has been shown to provide a safeguard against false inference in certain situations (e.g., when skewed distributions are being analyzed [Demenais et al. 1986] or when there is unaccounted excess sibling resemblance [Demenais et al. 1990], when the null hypothesis of no major gene is true), the power and specificity of these tests remain relatively unexplored. Using the present data, the previous segregation analysis using the mixed model in POINTER (Rice et al. 1993a) suggested that the major effect was not Mendelian and also was not transmitted within families. Submitting the same data to analysis using a class A regressive model in the present investigation suggested that the required criteria for inferring a major gene effect were met. However, allowing for the significant genotype-specific age effects resulted in an apparent loss of power to discriminate between the Mendelian hypothesis and the no-transmission hypothesis; neither hypothesis could be rejected. Finally, when incorporating genotype-dependent age and sex effects jointly, the tests on the transmission probabilities suggested a borderline fit for the

Mendelian hypothesis ($P = .048$), with rejection of the no-transmission hypothesis. The borderline fit of the Mendelian hypothesis in this case appears to reflect the underestimation of the sibling resemblance when the class A formulation is used, as suggested by the class D analysis. Thus, it appears that the particular model formulation in itself can affect the inference. Therefore, it would seem that inference based on a strict interpretation of the tests on the transmission parameters has the potential to be conservative insofar as detection is concerned, as any failure of the model assumptions can produce deviations from the Mendelian hypothesis even in the presence of a genuine major gene effect. In this context, it is important to emphasize that rejection of Mendelian segregation does not necessarily mean that a trait is not influenced by a Mendelian gene.

The evidence from the present study consistently suggests a major gene influence on some aspect of the "heaviness" measured by the BMI, although the distinct possibility remains that the etiology involves more than the single-gene determinant posited here. At the very least, there are significant genotype-dependent age and sex effects, and there certainly could be other factors, environmental or possibly genetic, affecting the phenotypic expression. However, it is possible that the phenotypic expression may be still more complicated than what is presently modeled. The latter point also was suggested by Tiret et al. (1992) in the interpretation of their data, where they cited the possibility that sporadic cases of obesity could occur in nongene carriers because of environmental factors or other genes. Certainly, this represents a plausible complication in our data, as well.

It becomes important to consider the adequacy of the phenotype for genetic analysis, especially in the context of seeking linkage relationships between candidate genes and a complex trait such as obesity. In order to achieve the greatest power for detecting and identifying relevant genetic determinants, a phenotype that has good specificity and minimal sources of residual (or error) variance is desirable. While use of the BMI has some advantages, e.g., its ease of measurement in large epidemiological studies, it is a relatively heterogeneous phenotype, reflecting the contributions of both fat and lean mass in unknown proportions. Like body fat content, the BMI is affected by the overall energy balance, varying as a function of habitual energy intake relative to the level of physical activity and other components of energy expenditure. However, the relative effects of energy balance on the lean and fat components are not known. Each of several phenotypes—fat mass (Bou-

chard et al. 1988; Rice et al. 1993*b*), fat-free mass (Bouchard et al. 1988), energy intake (Pérusse et al. 1988), habitual level of physical activity (Pérusse et al. 1989), and resting metabolic rate and thermic response to food (Bouchard et al. 1989)—has been shown to exhibit familial resemblance as well as to be affected by unique environmental determinants, suggesting that, from a biological standpoint, the BMI is a heterogeneous phenotype with a truly multifactorial determination. In light of the complexity of the components and determinants affecting what is finally measured as the BMI, it is not surprising that a clear Mendelian signal is not obtained in every study and varies according to the model.

One approach to clarifying models of the underlying genetic determinants is to incorporate the effects of known covariates, as has been done here, allowing for the possibility that genotypes may be expressed differently in response to environmental factors. Several other approaches to disentangling heterogeneous processes are also promising, including the utilization of phenotypic measures that are more homogeneous in themselves, as we are doing (Borecki et al. 1991, submitted; Rice et al. 1992, 1993*b*). The development and use of intermediate phenotypes can advance efforts at defining genetic and environmental determinants of human obesity, in a manner similar to the paradigm in cardiovascular research, where focusing on the subphenotypes of total cholesterol has fostered a greater understanding of the genetic determinants of cardiovascular disease, although many issues remain to be resolved. Each of several phenotypes—fat mass, fat-free mass, resting metabolic rate, energy intake, thermic response to food, and nutrient partitioning—represent well-justified and interesting intermediate phenotypes. Another approach involves following families over time, where the longitudinal observations can directly provide evidence for temporal effects. However, perhaps the most powerful approach to the investigation of genotype-specific responses to environmental factors is through a controlled intervention trial where one could, for example, observe the phenotypic response to an altered energy balance. This has been done in a study of body composition in male twins after a period of overfeeding (Bouchard et al. 1990) and also in an investigation of apo AI levels in pedigreed baboons under two different dietary regimens (Blangero et al. 1990). Finally, identification and measurement of genes relevant to energy metabolism can be used to elucidate the effect of specific genes on phenotypic variation in body composition. A combination of these approaches likely

will be useful in resolving the etiologic heterogeneity of human obesity.

Acknowledgments

This work was partly supported by NIH grants GM28719, HD18281, and HL47317; NIMH grant MH31302; and Medical Research Council of Canada grants MA-10499 and PG11811. The authors are grateful to Mr. Owen Jiang for carrying out the analyses using REGX.

References

- Blangero J, MacCluer JW, Kammerer CM, Mott GE, Dyer TD, McGill HC Jr (1990) Genetic analysis of apolipoprotein A-I in two dietary environments. *Am J Hum Genet* 47:414-428
- Bonney GE (1984) On the statistical determination of major gene mechanisms in continuous human traits: regressive models. *Am J Med Genet* 18:731-749
- (1992) Compound regressive models for family data. *Hum Hered* 42:28-41
- 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, Lathrop GM, Bonney GE, Yaouanq J, Rao DC (1990) Combined segregation and linkage analysis of genetic hemochromatosis using affection status, serum iron, and HLA. *Am J Hum Genet* 47:542-550
- Borecki IB, Rice T, Bouchard C, Rao DC (1991) Commingling analysis of generalized body mass and composition measures: the Québec family study. *Int J Obes* 15:763-773
- Borecki IB, Rice T, Pérusse L, Bouchard C, Rao DC. Segregation analysis of regional fat distribution measures: the Québec family study (submitted)
- Bouchard C (1989) Genetic factors in obesity. *Med Clin N Am* 73:67-81
- Bouchard C, Després J-P, Mauriège P, Marcotte M, Chagnon M, Dionne FT, Bélanger A (1991) The genes in the constellation of determinants of regional fat distribution. *Int J Obes* 15:9-18
- Bouchard C, Pérusse L (1988) Heredity and body fat. *Ann Rev Nutr* 8:259-277
- Bouchard C, Pérusse L, Leblanc C, Tremblay A, Thériault G (1988) Inheritance in the amount and distribution of human body fat. *Int J Obes* 12:205-215
- Bouchard C, Tremblay A, Després J-P, Nadeau A, Lupien PJ, Thériault G, Dussault J, et al (1990) The response to long-term overfeeding in identical twins. *N Engl J Med* 322:1477-1482
- Bouchard C, Tremblay A, Nadeau A, Després J-P, Thériault G, Boulay MR, Lortie G, et al (1989) Genetic effect in resting and exercise metabolic rates. *Metabolism* 38:364-370
- Bray G (1985) Complications of obesity. *Ann Intern Med* 103:1052-1062
- Burton BT, Foster WR, Hirsch J, van Itallie TB (1985) Health implications of obesity: an NIH consensus development conference. *Int J Obes* 9:155-169
- Deménais F, Lathrop M, Lalouel J-M (1986) Robustness and power of the unified model in the analysis of quantitative measurements. *Am J Hum Genet* 38:228-234
- Deménais F, Martinez M, Andrieu N (1993) The transmission probability model is useful to prevent false inference. *Am J Hum Genet* 52:441-442
- Deménais FM, Murigande C, Bonney GE (1990) Search for faster methods of fitting the regressive models to quantitative traits. *Genet Epidemiol* 7:319-334
- Garn SM, Leonard WR, Hawthorne VM (1986) Three limitations of the body mass index. *Am J Clin Nutr* 44:996-997
- Karlin S, Williams PT, Jensen S, Farquhar JW (1981) Genetic analysis of the Stanford LRC family study data. I. Structural exploratory data analysis of height and weight measurement. *Am J Epidemiol* 113:307-324
- Kral JG (1985) Morbid obesity and related health risks. *Ann Intern Med* 103:1043-1047
- Malina RM, Bouchard C (1988) Subcutaneous fat distribution during growth. In: Bouchard C, Johnston FE (eds) *Fat distribution during growth and later health outcomes*. Alan R Liss, New York, pp 63-84
- Moll PP, Burns TL, Lauer RM (1991) The genetic and environmental sources of body mass index variability: the Muscatine Ponderosity Family Study. *Am J Hum Genet* 49:1243-1255
- Pérusse L, Moll PP, Sing CF (1991) Evidence that a single gene with gender- and age-dependent effects influences systolic blood pressure determination in a population-based sample. *Am J Hum Genet* 49:94-105
- Pérusse L, Tremblay A, Leblanc C, Bouchard C (1989) Genetic and environmental influences on level of habitual physical activity and exercise participation. *Am J Epidemiol* 129:1012-1022
- Pérusse L, Tremblay A, Leblanc C, Cloninger CR, Reich T, Rice J, Bouchard C (1988) Familial resemblance in energy intake: contribution of genetic and environmental factors. *Am J Clin Nutr* 47:629-635
- Price RA (1992) Increases in prevalence and extent of obesity in post-world war II birth cohorts: implications for genetic analyses. *Behav Genet* 22:746
- Price RA, Ness R, Laskarzewski P (1990) Common major gene inheritance of extreme overweight. *Hum Biol* 62:747-765
- Province MA, Arnqvist P, Keller J, Higgins M, Rao DC (1990) Strong evidence for a major gene for obesity in the large, unselected, total Community Health Study of Tecumseh. *Am J Hum Genet Suppl* 47:A143
- Rice T, Borecki IB, Bouchard C, Rao DC (1992) Commingling analysis of regional fat distribution measures: the Québec family study. *Int J Obes* 16:831-844
- (1993a) Segregation analysis of body mass index in an unselected French-Canadian sample: the Québec family study. *Obesity Res* 1:288-294

- (1993*b*) Segregation analysis of fat mass and other body composition measures derived from underwater weighing. *Am J Hum Genet* 52:967–973
- Royal College of Physicians (1983) Obesity. *J R Coll Phys* 17:5–65
- Tiret L, André J-L, Ducimetière P, Herbeth B, Rakotovo R, Guegen R, Spyckerelle Y, et al (1992) Segregation analysis of height-adjusted weight with generation- and age-dependent effects: the Nancy family study. *Genet Epidemiol* 9:398–403
- van Itallie TB, Abraham S (1985) Some hazards of obesity and its treatment. In: Hirsch J, van Itallie TB (eds) *Recent advances in obesity research*. Vol 4. J Libbey, London, pp 1–19
- Zonta LA, Jayakar SD, Bosisio M, Galante A, Pennetti V (1987) Genetic analysis of human obesity in an Italian sample. *Hum Hered* 37:129–139