Am. J. Hum. Genet. 57:160–170, 1995

# Segregation Analysis of Non-Insulin-Dependent Diabetes Mellitus in Pima Indians: Evidence for a Major-Gene Effect

Robert L. Hanson,[1] Robert C. Elston,[2] David J. Pettitt,[1] Peter H. Bennett,[1] and William C. Knowler[1]

[1]Diabetes and Arthritis Epidemiology Section, National Institute of Diabetes and Digestive and Kidney Diseases, Phoenix; and [2]Department of Biometry and Genetics and Center for Molecular and Human Genetics, Louisiana State University Medical Center, New Orleans

## Summary

Non-insulin-dependent diabetes mellitus (NIDDM) has a high prevalence in Pima Indians. The disorder is familial, but the extent to which genetic factors are involved in its etiology is largely unknown. Segregation analysis was used to determine whether familial aggregation of NIDDM in this population could reflect the action of a single major gene. The analysis included 2,697 subjects from 653 nuclear families in which both parents and at least one offspring had been examined in the course of a longitudinal epidemiological study. The REGTL program of the SAGE package was used to fit models in which age at onset of NIDDM is transmitted from parent to offspring under the unified model for segregation analysis. Likelihood-ratio tests were used to test hypotheses related to genetic transmission. The hypothesis of no major effect was strongly rejected ($P < .01$), as was that of no transmission of the major effect ($P < .01$). Mendelian transmission was not rejected ($P = .91$). Similar results were obtained when covariates for obesity and birth cohort were added to the models and when a power transformation of age at onset was estimated. A strong effect of birth cohort with earlier age at onset in the later born cohorts was observed ($P < .01$). The findings are consistent with the hypothesis that a major gene influences the risk for NIDDM in Pima Indians by affecting age at onset. The expression of this gene may depend on environmental factors that have become more prevalent in recent-birth cohorts.

## Introduction

Non-insulin-dependent diabetes mellitus (NIDDM) is common in Pima Indians (Bennett et al. 1971; Knowler et al. 1978). NIDDM is generally believed to have important genetic determinants (Rich 1990; Knowler 1993), and the disorder in Pimas is strongly familial (Knowler et al. 1981). Molecular genetic studies of NIDDM in this population may, thus, be warranted, but the design and analysis of such studies would be facilitated by knowledge of whether the familial aggregation of diabetes is consistent with segregation of a major gene.

Segregation analysis of NIDDM, however, is complicated by the fact that disease expression is strongly age dependent and, among Pimas, by the fact that the incidence has increased in recent years (Knowler et al. 1990). Obesity is also a powerful risk factor for NIDDM that influences the familial relationships (Knowler et al. 1981), and its prevalence has also been increasing (Knowler et al. 1991; Price et al. 1993). The present analysis, therefore, was undertaken to determine whether segregation of a major gene can account for the inheritance of NIDDM in Pimas, with allowance for the effects of age, birth cohort, and obesity.

## Subjects and Methods

### *Families*

Since 1965, a longitudinal epidemiological study of NIDDM has been conducted in the Gila River Indian Community in central Arizona (Bennett et al. 1971). In this study, all community residents who are ≥5 years old have been invited biennially to research examinations. Informed consent was obtained, and a 75-g oral glucose tolerance test was administered. The diagnosis of diabetes was made according to World Health Organization (WHO) criteria for epidemiological studies: i.e., a glucose concentration of ≥11.1 mmol/liter observed in the 2-h postload venous sample or in the course of routine medical care (WHO Study Group on Diabetes Mellitus 1985). Age at onset of diabetes was determined from review of each subject's medical record, including biennial examinations. In the present analysis, diabetes status was taken as that at the most recent examination for each subject. Height and weight were measured at each visit and were used to calculate body mass index (BMI) (kg/m$^2$). Because individuals tend to lose weight after onset of diabetes (Knowler et al. 1991), BMI used in the analyses was that at the last nondiabetic examina-

Received December 29, 1994; accepted for publication April 11, 1995.

Address for correspondence and reprints: Dr. Robert L. Hanson, NIH/NIDDK/DAES, 1550 E. Indian School Road, Phoenix, AZ 85014. E-mail: hin@cu.nih.gov



0002-9297/95/5701-0019$02.00

**Table 1**

**Characteristics of Subjects**

| | Mothers | Fathers | Offspring |
|---|---|---|---|
| No. | 653 | 653 | 1856 |
| Age at examination (years)[a] | 49 (14–87) | 48 (9–88) | 23 (5–71) |
| Body Mass Index (kg/m$^2$)[a] | 33.3 (15.6–72.6) | 30.3 (16.2–56.4) | 30.1 (11.8–72.6) |
| Birth year[a] | 1938 (1882–1967) | 1935 (1879–1964) | 1961 (1910–1986) |
| Male (%) | .0 | 100.0 | 48.7 |
| Prevalence of diabetes (%) | 55.6 | 45.3 | 21.0 |

[a] The median value is reported, with the range in parentheses.

tion or, if the subject had been examined only after the diagnosis of diabetes, that at the earliest examination.

Information on familial relationships has been obtained from all participants. This has allowed for the construction of pedigrees in which phenotypic data are based on direct examination of all individuals, though the examinations may have occurred years apart. The present analysis considers nuclear families in which both parents and $\geq 1$ offspring had been examined and in which the heritage of all family members is full Pima or Tohono O'odham (a closely related tribe). Most individuals have had determination of erythrocyte antigen types at the ABO and MNSs loci. Of the 747 families identified, 94 were excluded from analysis because at least one family member was incompatible with the others at one of these loci. There were, thus, 653 families included in the present analysis. These families contained 3,162 members, whose characteristics are shown in table 1. Because of computational problems posed by multiple loops in extended pedigrees, the analysis was restricted to nuclear families. Therefore, some individuals ($n$ = 405) appeared in two or more families; the 3,162 persons represented in the families comprised 2,697 unique individuals.

### *Segregation Analysis*

Three criteria have been proposed for inferring a major gene (Elston et al. 1975); the specific hypotheses can be tested under the unified model for segregation analysis (Lalouel et al. 1983):

1. Rejection of the hypothesis of no major effect.
2. Rejection of the hypothesis of no transmission of the major effect.
3. Failure to reject the hypothesis of Mendelian transmission.

In the present analysis, models were fitted by using the REGTL program of the Statistical Analysis for Genetic Epidemiology package (SAGE 1992). This program incorporates multifactorial inheritance in a class A logistic regression model (Bonney 1986). Diabetes is defined as a dichotomous variable ($Y$), where $Y = 1$ if affected and $Y = 0$ if unaffected, and the probability that an individual is affected is modeled, conditional on age, genotype at the putative disease locus, and a set of other covariates (Elston and George 1989). The program estimates the following parameters: $q_A$ = the frequency of the putative disease allele ($A$); $\tau_{AA}$, $\tau_{AB}$, $\tau_{BB}$ = the probability an individual of type AA, AB, or BB transmits the $A$ allele to an offspring. (For the Mendelian case these correspond to 1.0, 0.5, and 0.0, respectively); $\beta_i$ = baseline parameters, where $i$ represents an individual's type (AA, AB, or BB); $\alpha_i$ = age coefficients; $\gamma_i$ = susceptibilities; $\zeta_k$ = covariate coefficients, for covariates k = 1 to $N$; $\delta_j(Y_j)$ = coefficients measuring the residual multifactorial effects of having an affected ($Y_j = 1$) or unaffected ($Y_j = 0$) spouse ($j$ = S), mother ($j$ = M), or father ($j$ = F). The logistic function describes the probability an individual is affected by age "a" as: $\gamma_i[1/1 + e^{-\Phi})]$, where:

$$\Phi = \beta_i + \alpha_i(a) + \delta_S(Y_S) + \delta_M(Y_M) + \delta_F(Y_F) + \zeta_1(x_1) + \ldots + \zeta_N(x_N) .$$

The $\alpha$, $\beta$, or $\gamma$ parameters can be made dependent on genotype at the putative disease locus; the first two possibilities correspond to transmission of age at onset and the last one to transmission of susceptibility (Elston et al. 1978). In the present analysis, the highest likelihoods were obtained with models incorporating a type-specific effect on the age coefficient ($\alpha$). Models with effects on both $\alpha$ and the baseline parameter ($\beta$) did not fit significantly better than those with an effect on $\alpha$ alone ($\chi^2 = 2.1$; df = 2; $P = .35$) but did fit better than those with $\beta$ alone ($\chi^2 = 41.7$; df = 2; $P < .01$). For the most part, therefore, results from models with a type specific effect on the age coefficient ($\alpha$) are presented, but similar results were obtained with an effect on the baseline parameter ($\beta$). Models with an effect on $\beta$ are presented in order to assess hypotheses concerning dominance, as the program does not allow the necessary

restrictions for dominant and recessive models when the type-specific effect is on α. Likelihoods were substantially lower with models in which the type-specific effect was on susceptibility (γ).

Models in which all six multifactorial effects (δ) were included were associated with large variances for these estimates. Therefore, some parameters may be imprecisely estimated in these models, but hypotheses can still be tested by comparing models. To obtain more precise parameter estimates, models were also analyzed under the restriction of equal magnitude of effect for affected and unaffected classes of relatives ($\delta_S(0) = -\delta_S(1)$; $\delta_M(0) = -\delta_M(1)$; $\delta_F(0) = -\delta_F(1)$), when appropriate.

Models that were fitted included: multifactorial inheritance only; Mendelian inheritance only; an "environmentally" commingled model with a major type effect and equal transmission probabilities in addition to multifactorial inheritance; a mixed Mendelian model with a major Mendelian locus and a multifactorial effect; a non-Mendelian model in which $\tau_{AB}$ was estimated; and a non-Mendelian model in which all transmission probabilities were estimated. Birth cohort was represented by categorical variables (born before 1930, born 1930–45, born 1946–55, and born after 1955). These categories were chosen to correspond approximately to periods of economic and social change. The categories represent the pre-Depression era, a particularly difficult time for Pimas, due to loss of their traditional water supplies, the Depression–World War II era, the immediate postwar period, and the era after the initiation of the Indian Health Service. Similarly, BMI categories were also represented by categorical variables.

For segregation analysis of a quantitative trait, a single distribution with skewness can result in false inference of a major gene (MacLean et al. 1975). The analogous situation for the present analysis is skewness in age at onset of diabetes. To assess this possibility, the data were also analyzed with a transformation of age according to the method of Box and Cox (1964). Transformed age is equal to $(a^{\lambda} - 1)/(\lambda a_{G1}^{(\lambda - 1)})$, where $a$ is age in years, $a_{G1}$ is the geometric mean of all ages of onset (36.6 years), and λ is a parameter to be estimated.

### *Hypothesis Tests*

Hypotheses were assessed by the likelihood ratio test, under the assumption that the negative of twice the difference in natural logarithms for hierarchical models has a $\chi^2$ distribution (Elston 1981). The null hypothesis of no major type effect was assessed by comparison of the model with multifactorial inheritance alone ($H_0$: $q_A = 1$) with that containing both a Mendelian major gene and multifactorial inheritance ($H_A$: $0 < q_A < 1$; $\tau_{AA} = 1.0$, $\tau_{AB} = 0.5$, $\tau_{BB} = 0.0$). The model with both effects estimates three additional parameters over the reduced model ($q_A$ and two α parameters), so the $\chi^2$ test has 3 df. The null hypothesis of no transmission was tested by comparing the commingled model ($H_0$: $q_A = \tau_{AA} = \tau_{AB} = \tau_{BB}$) with the model in which all transmission probabilities were estimated ($H_A$: $q_A \neq \tau_{AA} \neq \tau_{AB} \neq \tau_{BB}$, 3 df). The null hypothesis of Mendelian transmission was tested by comparing the mixed Mendelian model ($H_0$: $\tau_{AA} = 1.0$, $\tau_{AB} = 0.5$, $\tau_{BB} = 0.0$) with the model where $\tau_{AB}$ was estimated ($H_A$: $\tau_{AB} \neq 0.5$, 1 df) or with a model where all transmission probabilities were estimated ($H_A$: $\tau_{AA} \neq 1.0$, $\tau_{AB} \neq 0.5$, $\tau_{BB} \neq 0.0$, 3 df). The null hypothesis of no multifactorial effect was tested by comparing the model with Mendelian inheritance only with the mixed Mendelian model—a test with 3 or 6 df, depending on the restrictions placed on δ.

Because of the duplication of individuals that was necessitated by analyzing the pedigrees as nuclear families, the families are not completely independent. *P*-values obtained from these analyses may, therefore, be too small. To correct for this, certain analyses were also performed conditioning on the parental phenotypes (Morton and MacLean 1974; Elston and Sobel 1979; Lalouel and Morton 1981). Such an approach may result in more appropriate *P*-values, but estimates of some parameters, such as frequency of the disease allele, are less precise.

In order to assess the adequacy of the fit of the models and to determine how representative estimates derived from these families are of the occurrence of NIDDM in the Pima population, average cumulative incidence predicted by the model was compared with that observed in the population. Cumulative incidence in the population was determined in 5,311 subjects by the product-limit method (Kaplan and Meier 1958). This approach assumes that the age at onset determined at the biennial examination is estimated without bias in all individuals, including those diabetic at their first examination.

## Results

### *No Covariates*

Parameter estimates for models with no extraneous covariates are shown in table 2. The hypothesis of no major effect was rejected ($\chi^2 = 205.1$; df = 3; $P < .01$), as was the hypothesis of no transmission of the major effect ($\chi^2 = 113.3$; df = 3; $P < .01$). The null hypothesis of Mendelian transmission was not rejected, whether it was assessed by estimating $\tau_{AB}$ ($\chi^2 = 0.6$; df = 1; $P = .46$) or by estimating all transmission probabilities ($\chi^2 = 0.6$; df = 3; $P = .91$). The estimate of $\tau_{AB}$ was 0.59 (95% confidence interval (CI) = 0.37–0.81). The hypothesis of no multifactorial effect was rejected ($\chi^2 = 199.1$; df = 6; $P < .01$).

Predicted cumulative incidence of diabetes by age and genotype at the putative disease locus for the best fitting

**Table 2**

**Maximum Likelihood Estimates of Parameters of Segregation Models for Inheritance of Age at Onset of Diabetes: No Covariates**

| | Multifactorial Only[a] | Mendelian Only | Multifactorial[a] + Commingled | Multifactorial[a] + Mendelian | Multifactorial[a] + Free $\tau_{AB}$ | Multifactorial[a] + Free $\tau$'s |
|---|---|---|---|---|---|---|
| $q_A$ | (1.0) | .38 | .42 | .39 | .37 | .37 |
| $\tau_{AA}$ | ... | (1.0) | ... | (1.0) | (1.0) | 1.00 |
| $\tau_{AB}$ | ... | (.5) | ... | (.5) | .59 | .59 |
| $\tau_{BB}$ | ... | (.0) | ... | (.0) | (.0) | .00 |
| $\alpha_{AA}$ | .14 | .26 | .29 | .33 | .33 | .33 |
| $\alpha_{AB}$ | ... | .18 | .20 | .23 | .22 | .22 |
| $\alpha_{BB}$ | ... | .13 | .15 | .16 | .16 | .16 |
| $\delta_S$ (0) | −.31 | (.0) | −.52 | −.60 | −.58 | −.58 |
| $\delta_S$ (1) | −.47 | (.0) | −.98 | −1.13 | −1.10 | −1.10 |
| $\delta_M$ (0) | −.94 | (.0) | −.47 | −.07 | .04 | .41 |
| $\delta_M$ (1) | .17 | (.0) | .93 | .85 | .93 | 1.30 |
| $\delta_F$ (0) | .53 | (.0) | −.01 | .64 | .33 | −.04 |
| $\delta_F$ (1) | .99 | (.0) | .72 | .89 | .58 | .21 |
| $\beta$ | −6.40 | −7.52 | −8.82 | −9.54 | −9.36 | −9.36 |
| $\gamma$ | .85 | .91 | .90 | .89 | .88 | .88 |
| −2 lnL | 9,844.42 | 9,838.37 | 9,752.06 | 9,639.29 | 9,638.74 | 9,638.74 |

NOTE.—Models were fitted using the SAGE program REGTL (SAGE 1992). Values in parentheses were fixed at the listed values.
[a] Multifactorial effect.

Mendelian model is shown in figure 1. The model predicts an important genetic influence on NIDDM that occurs at younger ages and a high phenocopy rate at older ages. It can be seen by comparing parameter estimates that an almost identical figure would result from using the estimates of the most general model (last column of table 3). When the above analyses were performed conditioning on the parental phenotypes to correct for duplication of individuals, the hypotheses of no major effect ($\chi^2$ = 62.3; df = 3; $P < .01$) and no transmission ($\chi^2$ = 125.0; df = 3; $P < .01$) were still rejected and Mendelian transmission was not ($\chi^2$ = 0.11; df = 3; $P$ = .99).

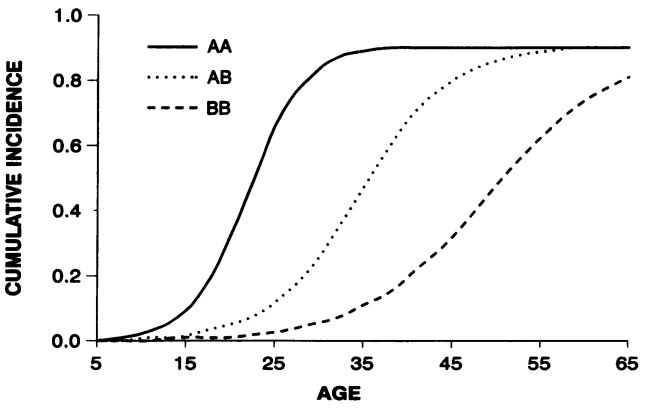


**Figure 1** Cumulative incidence of diabetes in Pima Indians predicted by the best-fitting mixed Mendelian segregation model, by age (years) and genotype at the putative disease locus. AA = homozygous for disease allele A; AB = heterozygote; and BB = persons without the A allele.

### *Birth Cohort Effect*

The addition of a birth cohort effect to the above models resulted in significant improvement in the fit of the models ($\chi^2$ = 165.6; df = 3; $P < .01$, for the mixed Mendelian model). In this case, the hypothesis of a restricted multifactorial effect [$\delta_S(0) = -\delta_S(1)$, $\delta_M(0) = -\delta_M(1)$, $\delta_F(0) = -\delta_F(1)$] was not rejected ($\chi^2$ = 6.8; df = 3; $P$ = .08). Because of the more precise parameter estimates, models with this restriction are shown (table 3), but results were similar when all six $\delta$ parameters were estimated. The hypotheses of no major effect ($\chi^2$ = 141.0; df = 3; $P < .01$) and no transmission ($\chi^2$ = 71.0; df = 3; $P < .01$) were strongly rejected. Mendelian transmission was not rejected, whether it was tested by estimating $\tau_{AB}$ ($\chi^2$ = 3.2; df = 1; $P$ = .07) or by estimating all transmission probabilities ($\chi^2$ = 3.2; df = 3; $P$ = .36). The estimate of $\tau_{AB}$ was 0.64 (95% CI, 0.48–0.80). The hypothesis of no multifactorial effect was rejected ($\chi^2$ = 19.3; df = 3; $P < .01$). When models conditioned on parental phenotypes were fitted, the hypotheses of no major effect ($\chi^2$ = 118.7; df = 3; $P < .01$) and no transmission ($\chi^2$ = 89.3; df = 3; $P < .01$) were still rejected, and the hypothesis of Mendelian transmission was not rejected ($\chi^2$ = 5.4; df = 3; $P$ = .14).

Parameter estimates for the best-fitting dominant, recessive, and general Mendelian models are shown in table 4. The effect of the putative disease allele is assumed to be on the baseline parameter ($\beta$) in these mod-

**Table 3**

**Maximum Likelihood Estimates of Parameters of Segregation Models for Inheritance of Age at Onset of Diabetes: Birth Cohort Effect**

| | Multifactorial Only[a] | Mendelian Only | Multifactorial[a] + Commingled | Multifactorial[a] + Mendelian | Multifactorial[a] + Free $\tau_{AB}$ | Multifactorial[a] + Free $\tau$'s |
|---|---|---|---|---|---|---|
| $q_A$ | (1.0) | .36 | .43 | .37 | .35 | .35 |
| $\tau_{AA}$ | ... | (1.0) | ... | (1.0) | (1.0) | 1.00 |
| $\tau_{AB}$ | ... | (.5) | ... | (.5) | .64 | .64 |
| $\tau_{BB}$ | ... | (.0) | ... | (.0) | (.0) | .00 |
| $\alpha_{AA}$ | .17 | .33 | .28 | .32 | .32 | .32 |
| $\alpha_{AB}$ | ... | .24 | .21 | .24 | .24 | .24 |
| $\alpha_{BB}$ | ... | .18 | .14 | .19 | .18 | .18 |
| $\delta_S$ (1)[b] | .13 | (.0) | .19 | .12 | .11 | .11 |
| $\delta_M$ (1)[b] | .53 | (.0) | .69 | .48 | .36 | .36 |
| $\delta_F$ (1)[b] | .22 | (.0) | .33 | .15 | .09 | .09 |
| $\beta$ | −6.70 | −8.84 | −8.53 | −8.99 | −8.92 | −8.92 |
| $\zeta$ < 1930[c] | −2.21 | −3.25 | −2.92 | −3.05 | −2.88 | −2.88 |
| $\zeta$ = 1930–45[c] | −.58 | −.73 | −.75 | −.69 | −.62 | −.62 |
| $\zeta \geq$ 1956[c] | .33 | .70 | .55 | .65 | .58 | .58 |
| $\gamma$ | .81 | .89 | 1.00 | .86 | .86 | .86 |
| −2 lnL | 9,624.88 | 9,503.22 | 9,551.77 | 9,483.91 | 9,480.73 | 9,480.73 |

NOTE.—Models were fitted using the SAGE program REGTL (SAGE 1992). Values in parentheses were fixed at the listed values.
[a] Multifactorial effect.
[b] $\delta_j$ (0) is constrained to equal $-\delta_j$ (1).
[c] By definition, the parameter is 0 for those born in the period 1946–55.

els. Both dominant ($\chi^2$ = 20.4; df = 1; $P < .01$) and recessive ($\chi^2$ = 19.6; df = 1; $P < .01$) models were rejected in comparison with the general Mendelian model (in which the heterozygote is at an intermediate risk for NIDDM).

Predicted cumulative incidence of NIDDM by age and birth cohort is shown for the three putative genotypes in figure 2, on the basis of the parameters from the mixed Mendelian model in table 3. Overall predicted cumulative incidence for the population from this same model, compared with that observed in the population, is shown in figure 3. The predicted cumulative incidence corresponds well with that observed, but it deviates somewhat at the oldest ages in all birth cohorts, where data tend to be sparse. The cumulative incidence of NIDDM predicted by the model derived from these families is, thus, a reasonable estimate of the occurrence of diabetes in the population.

Cumulative incidence ratios for the putative AA and AB genotypes, relative to the BB homozygote, are shown for various ages in table 5. The model predicts that the risk of NIDDM for gene carriers relative to that of noncarriers is greater at younger ages and, for a given age, in earlier-born cohorts.

### *Transformation of Age at Onset*

In order to determine whether skewness in age at onset could mimic a major gene effect in the above analysis, the analysis was repeated using a power transformation of age. Results of this analysis are shown in table 6. The hypotheses of no major effect ($\chi^2$ = 79.6; df = 3; $P < .01$) and no transmission ($\chi^2$ = 68.6; df = 3; $P < .01$) were rejected. Mendelian transmission was not rejected ($\chi^2$ = 3.4; df = 3; $P$ = .33). In comparison of mixed Mendelian models, the hypothesis $\lambda$ = 1 was not rejected ($\chi^2$ = 1.5; df = 1; $P$ = .22); i.e., transformation of age did not significantly improve the fit of the model.

### *Obesity Effect*

Parameter estimates from models with BMI as a covariate are shown in table 7. The addition of this covariate resulted in significantly higher likelihoods compared with models containing no covariates (for comparison of mixed Mendelian models, $\chi^2$ = 11.2; df = 4; and $P$ = .02). The hypothesis of no major effect was rejected ($\chi^2$ = 194.6; df = 3; $P < .01$), as was that of no transmission ($\chi^2$ = 112.2; df = 3; $P < .01$). Mendelian transmission was not rejected ($\chi^2$ = 0.2; df = 3; $P$ = .97). The estimate of $\tau_{AB}$ was 0.56 (95% CI, 0.30–0.82). The hypothesis of no multifactorial effect was also rejected ($\chi^2$ = 175.1; df = 6; $P < .01$). Parameter estimates and hypothesis tests for models that included covariates for both BMI and birth cohort did not differ substantially from models that contained a cohort effect alone (data not shown).

## Discussion

The present analysis shows that familial transmission of age at onset of NIDDM in Pima Indians corresponds

**Table 4**

**Maximum Likelihood Estimates of Parameters of Dominant, Recessive, and General Mendelian Models for Inheritance of Age at Onset of Diabetes: Birth Cohort Effect**

| | Dominant | Recessive | General |
|---|---|---|---|
| $q_A$ | .32 | .74 | .53 |
| $\tau_{AA}$ | (1.0) | (1.0) | (1.0) |
| $\tau_{AB}$ | (.5) | (.5) | (.5) |
| $\tau_{BB}$ | (.0) | (.0) | (.0) |
| $\beta_{AA}$ | −7.45 | −7.48 | −7.48 |
| $\beta_{AB}$ | (=$\beta_{AA}$) | (=$\beta_{BB}$) | −9.93 |
| $\beta_{BB}$ | −10.60 | −10.64 | −13.45 |
| $\delta_S$ (1)[†] | .15 | .16 | .12 |
| $\delta_M$ (1)[†] | .56 | .57 | .50 |
| $\delta_F$ (1)[†] | .27 | .27 | .13 |
| $\alpha$ | .23 | .23 | .25 |
| $\zeta$ < 1930[‡] | −2.94 | −2.91 | −3.03 |
| $\zeta$ = 1930–45[‡] | −.79 | −.79 | −.65 |
| $\zeta \geq$ 1956[‡] | .39 | .37 | .59 |
| $\gamma$ | .83 | .82 | .84 |
| −2 lnL | 9,521.06 | 9,520.21 | 9,500.62 |

NOTE.—Models were fitted using the SAGE program REGTL (SAGE 1992). Values in parentheses were fixed at the listed values.

[a] $\delta_j$ (0) is constrained to equal $-\delta_j$ (1).

[b] By definition, the parameter is 0 for those born in the period 1946–55.

well with Mendelian expectations. These findings indicate that a single major gene that influences age at onset could be an important determinant of NIDDM in this population. The model predicts a high genetic risk for diabetes with younger age at onset and a high prevalence of phenocopies at older ages. The high prevalence of NIDDM at older ages, therefore, may reflect other genetic and/or environmental factors.

The longitudinal nature of the present study allows for more accurate assessment of age at onset of diabetes than is usually available from a cross-sectional study. However, excess mortality among diabetic individuals could have influenced the availability of subjects for analysis, particularly in the earliest-born cohort. Additionally, in earlier-born cohorts, many subjects already had diabetes at the first examination, so their age at onset is not as well documented as in later-born cohorts. Since the present analysis assumes that age at onset is determined without bias in all birth cohorts, the magnitude of the cohort effect may be overestimated. Inclusion of a cohort effect in the segregation models, however, did not substantially alter the parameter estimates or the inferences regarding segregation of a major gene. Moreover, the increasing *incidence* of NIDDM at all ages (Knowler et al. 1990) suggests that factors that lead to earlier expression of diabetes have become more prevalent in later-born cohorts. Obesity is a strong risk factor for diabetes (Knowler et al. 1981), and its prevalence has also increased considerably in recent years (Knowler et al. 1991; Price et al. 1993). However, the increasing incidence of diabetes in Pimas is not entirely explained by higher BMI (Knowler et al. 1990). Although BMI may not fully account for changes in adiposity, secular trends in other factors may have contributed to the rising incidence of NIDDM. Increasing physical inactivity, which is associated with NIDDM (Kriska et al. 1993), could play a role, as could dietary changes.

Previous segregation analyses of NIDDM in Pimas have sought the parameters of Mendelian models that have not accounted for a cohort effect or for multifacto-

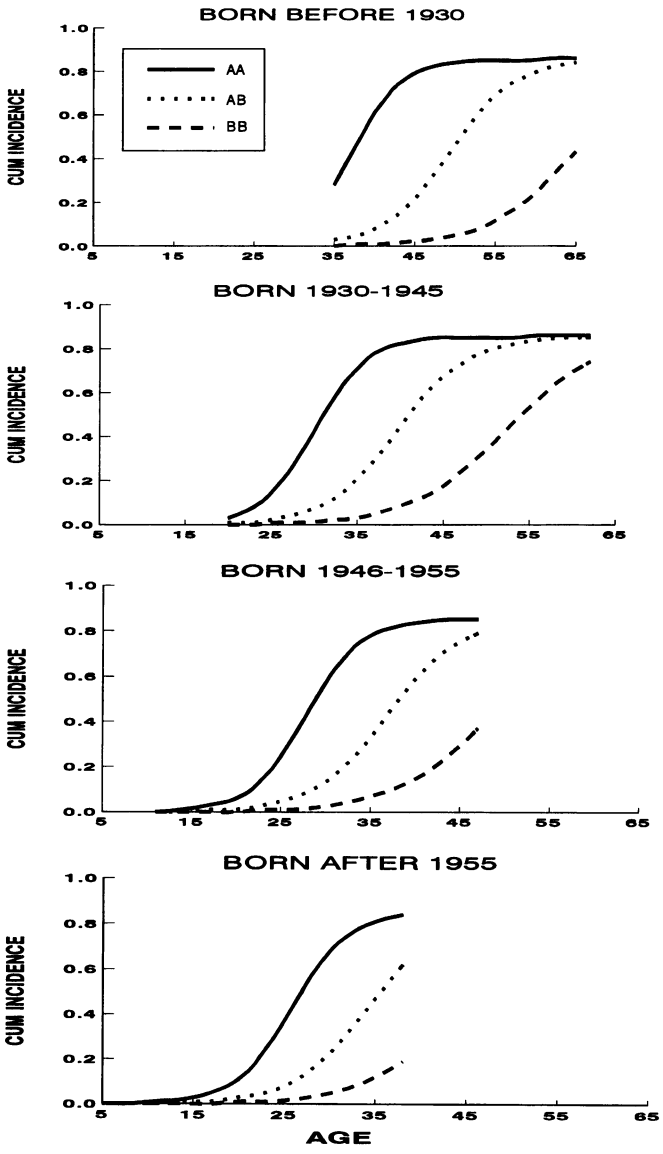


**Figure 2** Cumulative incidence predicted by the best-fitting mixed Mendelian segregation model (table 3) by age (years), birth cohort, and genotype at the putative disease locus. AA = homozygous for disease allele A; AB = heterozygote; BB = persons without the disease allele A.

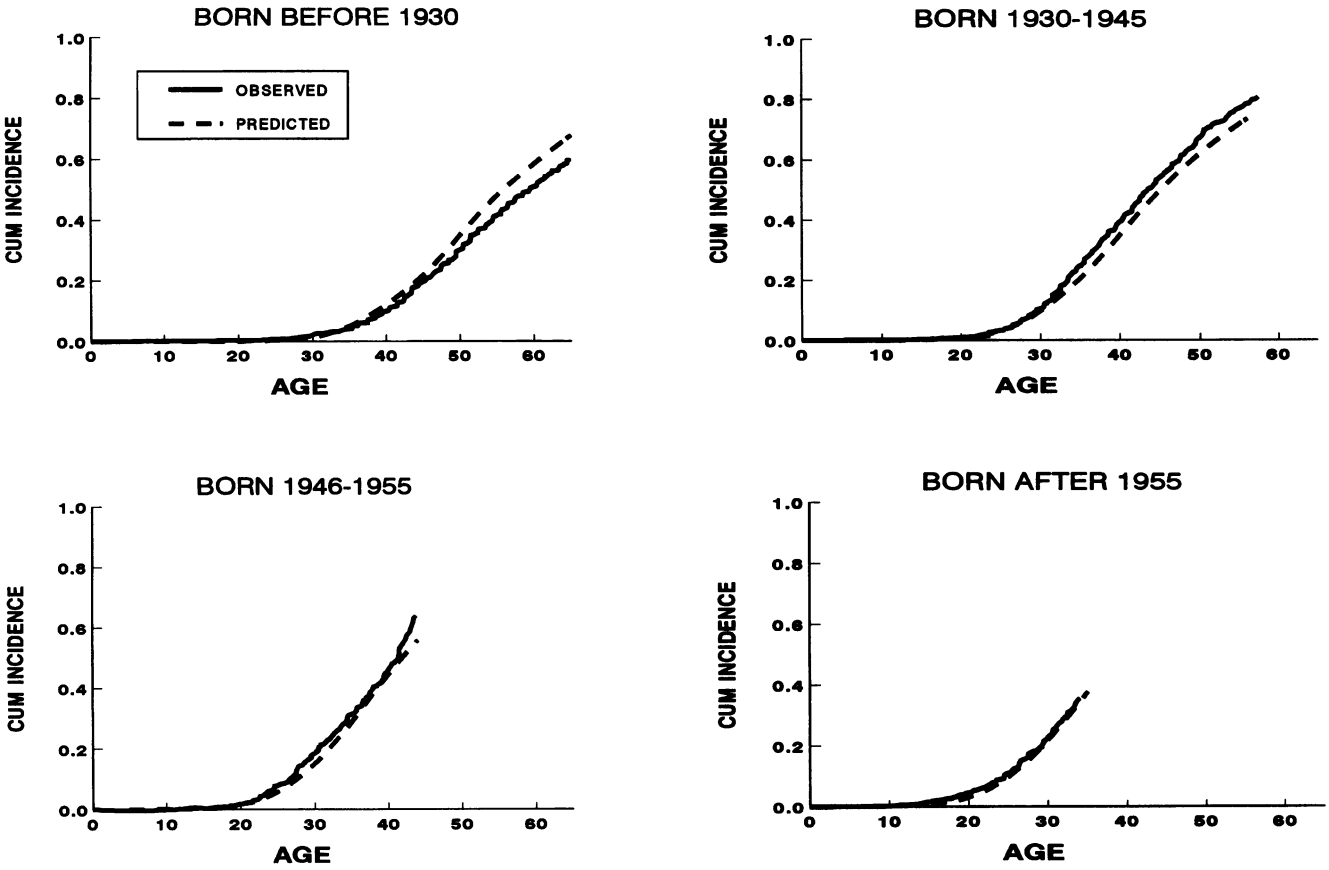


**Figure 3** Cumulative incidence by birth cohort predicted by the best-fitting mixed Mendelian segregation model (dashed lines) and observed in the population (solid lines). Cumulative incidence derived from the model is the sum of those for the three genotypes weighted by the predicted frequency of these genotypes. The population–cumulative incidence was derived by the product-limit method; curves are truncated at the 95th percentile of age (years) in each cohort, on account of sparse data.

rial inheritance (Yamashita et al. 1984; Lewis 1991). The analyses of Yamashita et al. (1984) on 2-h glucose concentration suggested a model in which the mean value for the putative heterozygote was intermediate to the other two values, a finding consistent with the present analysis. On the other hand, Lewis (1991) found one maximum likelihood at a heterozygote-affected model and another at a heterozygote-unaffected model. In the present analysis, a heterozygote affected model was the best fitting Mendelian model when transmission of susceptibility ($\gamma$), rather than age at onset ($\alpha$, $\beta$), was assumed, but the evidence for a major effect was weaker ($\chi^2 = 8.2$; df = 3; $P = .04$). The support for a major gene is, thus, dependent on the parameterization of the model. Segregation analysis of BMI in Pimas, accounting for birth cohort, also suggests a major gene effect (Price et al. 1994). It seems possible that the present analysis reflects the same biological phenomenon: i.e., a diabetes-obesity gene. However, the addition of BMI to the segregation models in the present analysis did not affect the evidence for a major diabetes susceptibility gene. Furthermore, the finding that relatives of less obese diabetic Pima Indians have a higher prevalence of NIDDM than relatives of more obese diabetic subjects suggests that the familial determinants of NIDDM and obesity are separate (Hanson et al., in press).

Segregation analyses of NIDDM or related phenotypes in other populations have provided equivocal support for a major gene. Among Seminoles, another Native American population, analyses of glucose tolerance were consistent with a major gene in the Oklahoma branch of the tribe but not in the Florida branch (Elston et al. 1974). Analysis of glycemia among Nauruans also suggested a major gene with dominant inheritance (Serjeantson and Zimmet 1991), but this analysis did not formally test the hypothesis against non-Mendelian models. On the other hand, in analysis of fasting plasma glucose in European Americans, an environmental model provided the best fit (Rice et al. 1992). In other segregation analyses of NIDDM, models incorporating a major gene effect did not fit significantly better than those with multifactorial inheritance, and more complex

**Table 5**

**Predicted Cumulative Incidence Ratios, by Genotype, Age, and Birth Cohort**

| Birth Cohort and Genotype | Age (years) | | |
|---|---|---|---|
| | 25 | 45 | 65 |
| Born before 1930: | | | |
| AA | ... | 37.9 | 1.98 |
| AB | ... | 10.6 | 1.93 |
| BB | ... | 1.00[a] | 1.00[a] |
| Born 1930–45: | | | |
| AA | 26.1 | 4.74 | ... |
| AB | 4.23 | 3.76 | ... |
| BB | 1.00[a] | 1.00[a] | ... |
| Born 1946–55: | | | |
| AA | 22.5 | 2.88 | ... |
| AB | 4.14 | 2.54 | ... |
| BB | 1.00[a] | 1.00[a] | ... |
| Born after 1955: | | | |
| AA | 18.0 | ... | ... |
| AB | 4.00 | ... | ... |
| BB | 1.00[a] | ... | ... |

NOTE.—Ratios computed from the mixed Mendelian model in table 3, as the predicted cumulative incidence for the relevant genotype, divided by that of the BB genotype.

[a] Equal to 1, by definition.

models were required to explain the data (Cook et al. 1994; McCarthy et al. 1994).

In segregation analysis of a quantitative trait, skewness in a single distribution can result in the false inference of a major gene (MacLean et al. 1975). The analogous situation in the present analysis is skewness in age at onset of NIDDM. Prior use of a power transformation can reduce the probability of falsely detecting a major effect, although this approach can reduce the power to detect a major gene if it is truly present (Demenais et al. 1986). In the present analysis, the power transformation was estimated simultaneously, and evidence for a major gene affecting age at onset of diabetes remained.

Analysis of the entire Pima population as a small number of very large pedigrees is theoretically possible, but this is made computationally difficult by the presence of multiple loops. For this reason, nuclear families were analyzed in the present study. This approach necessitates duplicating some individuals, and this results in lack of independence among pedigrees. Therefore, the *P*-values obtained in the present analysis may be too small, but, given the results, this is not likely to affect the conclusions. Moreover, similar findings were obtained when the analysis was performed conditioning on the parental phenotypes in order to correct for the duplication of individuals.

A potential limitation of the present analysis is that the models do not allow the covariate effects to be genotype specific. In the presence of a gene-environment interaction, analyses that do not allow for genotype-specific covariates can lead to loss of power and to biased parameter estimates (Tiret et al. 1993). On the other hand, incorporating genotype-specific covariates may complicate estimation of a power transformation (Khoury et al. 1993). The fact that the estimates of $\tau_{AB}$ were somewhat different from 0.50 could reflect gene-environment interaction. However, it is noteworthy that parameters determining the type-specific age at onset distribution ($\alpha_i$, $\beta$, and $\gamma$) were estimated to have very similar values whether or not the transmission probabilities were Mendelian. On this basis, one might expect any bias due to genotype-environment interaction to be modest.

The class A model is a flexible parameterization of multifactorial inheritance because it allows for different maternal and paternal effects. The present analysis shows a stronger maternal than paternal effect ($\delta_M > \delta_F$, in table 3); this may, in part, reflect an effect of the diabetic intrauterine environment (Pettitt et al. 1988). It could also reflect mitochondrial inheritance, but mtDNA mutations commonly associated with diabetes have not been found in Pimas (Sepehrnia et al., in press). The class A model is limited, however, because it constrains the resemblance among siblings to be solely a function of the parental phenotypes. If the resemblance among siblings is higher than the parent-offspring resemblance, failure to account for this effect (e.g., in a class D model) can lead to false inference of a major gene (MacLean et al. 1975; Elston 1981). To guard against this possibility, it has been suggested that one should be suspicious of results that support both dominant and recessive models over intermediate models or in which estimates of transmission probabilities are markedly non-Mendelian despite failure to reject Mendelian transmission (Go et al. 1978). In the present analysis, neither of these situations occurred. To further assess the adequacy of class A models, odds ratios (OR) for affection of different classes of relatives were compared. If the first-born sibling in each sibship was considered the index person, the age-adjusted OR for diabetes for subjects having a diabetic index sibling was 2.2 (95% CI, 1.6–3.1). This is only slightly higher than the risk for having a diabetic father (OR = 1.9; 95% CI, 1.3–2.7) and is lower than that for having a diabetic mother (OR = 3.4; 95% CI, 2.1–5.2). Although the program used does not allow a formal test of a class A against a class D model, this result and the comparison of predicted with observed cumulative incidence suggest that the class A model fits the data reasonably well.

Studies of DNA should more fully characterize the genetics of NIDDM in Pimas. While segregation analysis does not prove or disprove the existence of a major

## Table 6

**Maximum Likelihood Estimates of Parameters of Segregation Models for Inheritance of Age at Onset of Diabetes: Birth Cohort Effect and Transformation of Age at Onset**

| | Multifactorial Only[a] | Mendelian Only | Multifactorial[a] + Commingled | Multifactorial[a] + Mendelian | Multifactorial[a] + Free $\tau_{AB}$ | Multifactorial[a] + Free $\tau$'s |
|---|---|---|---|---|---|---|
| $q_A$ | (1.0) | .35 | .76 | .35 | .32 | .32 |
| $\tau_{AA}$ | ... | (1.0) | ... | (1.0) | (1.0) | 1.00 |
| $\tau_{AB}$ | ... | (.5) | ... | (.5) | .67 | .67 |
| $\tau_{BB}$ | ... | (.0) | ... | (.0) | (.0) | .00 |
| $\alpha_{AA}$ | .16 | .30 | .20 | .30 | .29 | .29 |
| $\alpha_{AB}$ | ... | .23 | .15 | .24 | .23 | .23 |
| $\alpha_{BB}$ | ... | .18 | .24 | .19 | .19 | .19 |
| $\delta_S$ (1)[b] | .12 | (.0) | .17 | .13 | .12 | .12 |
| $\delta_M$ (1)[b] | .51 | (.0) | .63 | .48 | .34 | .35 |
| $\delta_F$ (1)[b] | .22 | (.0) | .29 | .16 | .09 | .09 |
| $\beta$ | −14.47 | −9.97 | −10.90 | −10.13 | −10.33 | −10.24 |
| $\zeta <$ 1930[c] | −1.89 | −3.08 | −2.50 | −2.88 | −2.66 | −2.67 |
| $\zeta =$ 1930–45[c] | −.55 | −.72 | −.65 | −.68 | −.60 | −.60 |
| $\zeta \geq$ 1956[c] | .48 | .71 | .54 | .67 | .57 | .57 |
| $\lambda$ | .25 | .81 | .60 | .80 | .77 | .78 |
| $\gamma$ | .88 | .90 | 1.00 | .86 | .87 | .87 |
| −2 lnL | 9,562.00 | 9,501.75 | 9,547.57 | 9,482.42 | 9,478.99 | 9,478.99 |

NOTE.—Transformed age = $(age^{\lambda} - 1)/\lambda(36.6^{\lambda-1})$. Parameters in parentheses are fixed at the listed values.

[a] Multifactorial effect.

[b] $\delta_j$ (0) is constrained to equal $-\delta_j$ (1).

[c] By definition, the parameter is 0 for those born in the period 1945–56.

## Table 7

**Maximum Likelihood Estimates of Parameters of Segregation Models for Inheritance of Age at Onset of Diabetes: Effect of Body Mass Index**

| | Multifactorial Only[a] | Mendelian Only | Multifactorial[a] + Commingled | Multifactorial[a] + Mendelian | Multifactorial[a] + Free $\tau_{AB}$ | Multifactorial[a] + Free $\tau$'s |
|---|---|---|---|---|---|---|
| $q_A$ | (1.0) | .23 | .48 | .39 | .38 | .38 |
| $\tau_{AA}$ | ... | (1.0) | ... | (1.0) | (1.0) | 1.00 |
| $\tau_{AB}$ | ... | (.5) | ... | (.5) | .56 | .56 |
| $\tau_{BB}$ | ... | (.0) | ... | (.0) | (.0) | .00 |
| $\alpha_{AA}$ | .14 | .32 | .26 | .32 | .32 | .32 |
| $\alpha_{AB}$ | ... | .20 | .18 | .22 | .22 | .22 |
| $\alpha_{BB}$ | ... | .14 | .12 | .16 | .16 | .16 |
| $\delta_S$ (0) | −.26 | (.0) | −.49 | −.52 | −.49 | −.49 |
| $\delta_S$ (1) | −.40 | (.0) | −.64 | −1.02 | −1.00 | −1.00 |
| $\delta_M$ (0) | −.50 | (.0) | −.40 | .09 | .49 | .48 |
| $\delta_M$ (1) | .57 | (.0) | .92 | .93 | 1.31 | 1.31 |
| $\delta_F$ (0) | .14 | (.0) | .01 | .58 | .03 | .04 |
| $\delta_F$ (1) | .57 | (.0) | .73 | .75 | .21 | .22 |
| $\beta$ | −6.93 | −8.45 | −9.20 | −9.98 | −9.82 | −9.82 |
| $\zeta =$ 25–30[a] | .37 | .47 | .58 | .24 | .20 | .20 |
| $\zeta =$ 30–35[c] | .59 | 1.00 | .88 | .59 | .56 | .56 |
| $\zeta =$ 35–40[a] | .51 | 1.06 | .63 | .66 | .62 | .62 |
| $\zeta \geq$ 40[a] | .91 | 1.53 | 1.05 | .83 | .78 | .78 |
| $\gamma$ | .86 | .90 | 1.00 | .90 | .90 | .90 |
| −2 lnL | 9,822.71 | 9,803.19 | 9,740.75 | 9,628.08 | 9,627.86 | 9,627.86 |

NOTE.—Models were fitted using the SAGE program REGTL (SAGE 1992). Parameters in parentheses were fixed at the listed values.

[a] Multifactorial effect.

[b] Those with BMI < 25 kg/m$^2$ are the reference category; by definition, the parameter is 0.

genetic locus for a trait, it does provide an estimate of the prior probability of such a locus and can facilitate the design and analysis of genetic linkage studies. It might be thought that the high frequencies of the putative disease allele observed in the present analysis, which are not surprising in light of the high prevalence of NIDDM, would compromise the power of a linkage study in this population. However, simulation studies suggest that power to detect a diabetes susceptibility gene in Pimas is high (Hanson et al. 1994).

The findings of the present analysis are consistent with the hypothesis that a major genetic locus influences the risk of NIDDM in Pimas by affecting age at onset. The expression of this gene may depend on environmental factors to which more recent birth cohorts have had greater exposure.

## Acknowledgments

The authors would like to thank the members of the Gila River Indian Community who participated in the study and all of the Diabetes and Arthritis Epidemiology Section staff for their assistance. We would also like to thank Dr. D. Timothy Bishop for advice. Some of the results of this paper were obtained by using the program package SAGE, which is supported by a U.S. Public Health Service Resource Grant (1 P41 RR03655) from the National Center for Research Resources. This work was presented in part at the 44th Annual Meeting of the American Society of Human Genetics, Montreal, Canada, October 22, 1994.

## References

Bennett PH, Burch TA, Miller M (1971) Diabetes mellitus in American (Pima) Indians. Lancet 2:125–128

Bonney GE (1986) Regressive logistic models for familial disease and other binary traits. Biometrics 42:611–625

Box GEP, Cox DR (1964) An analysis of transformations. J R Stat Soc B 26:211–252

Cook JTE, Shields DC, Page RCL, Levy JC, Hattersley AT, Shaw JAG, Neil HAW, et al (1994) Segregation analysis of NIDDM in Caucasian families. Diabetologia 37:1231–1240

Demenais F, Lathrop M, Lalouel JM (1986) Robustness and power of the unified model in the analysis of quantitative measurements. Am J Hum Genet 38:228–234

Elston RC (1981) Segregation analysis. Adv Hum Genet 11:63–120

Elston RC, George VT (1989) Age of onset, age at examination and other covariates in the analysis of family data. Genet Epidemiol 6:217–220

Elston RC, Namboodiri KK, Nino HV, Pollitzer WS (1974) Studies on blood and urine glucose in Seminole Indians: indications for segregation of a major gene. Am J Hum Genet 26:13–34

Elston RC, Namboodiri KK, Glueck CJ, Fallat R, Tsang R, Leuba V (1975) Study of the genetic transmission of hypercholesterolemia and hypertriglyceridemia in a 195 member kindred. Ann Hum Genet 39:67–87

Elston RC, Namboodiri KK, Spence MA, Rainer JD (1978) A genetic study of schizophrenia in pedigrees II. One locus hypotheses. Neuropsychobiology 4:193–206

Elston RC, Sobel E (1979) Sampling considerations in the gathering and analysis of pedigree data. Am J Hum Genet 31:62–69

Go RCP, Elston RC, Kaplan EB (1978) Efficiency and robustness of pedigree segregation analysis. Am J Hum Genet 30:38–37

Hanson R, Knowler W, Bennett P (1994) Power to detect diabetes susceptibility genes in Pima Indians. Diabetes Suppl 43:223A

Hanson RL, Pettitt DJ, Bennett PH, Narayan KMV, Fernandes R, de Courten M, Knowler WC (1995) Familial relationships between obesity and non-insulin-dependent diabetes mellitus. Diabetes 44:418–422

Kaplan EL, Meier P (1958) Nonparametric estimation from incomplete observations. J Am Stat Assoc 53:457–481

Khoury MJ, Beaty TH, Cohen BH (1993) Genetic approaches to familial aggregation. II. Segregation analysis. In: Khoury MJ, Beaty TH, Cohen BH (eds) Fundamentals of genetic epidemiology. Oxford University Press, New York, pp 233–311

Knowler WC (1993) Approaches to studying the genetics of type 2 diabetes. In: Belfiore F, Bergman RN, Molinatti GM (eds) Current topics in diabetes research. Frontiers in diabetes. Vol 12. Karger, Basel, pp 1–11

Knowler WC, Bennett PH, Hamman RF, Miller M (1978) Diabetes incidence and prevalence in Pima Indians: a 19-fold greater incidence than in Rochester, Minnesota. Am J Epidemiol 108:497–505

Knowler WC, Pettitt DJ, Saad MF, Bennett PH (1990) Diabetes mellitus in the Pima Indians: incidence, risk factors and pathogenesis. Diabetes Metab Rev 6:1–27

Knowler WC, Pettitt DJ, Saad MF, Charles MA, Nelson RG, Howard BV, Bogardus C, et al (1991) Obesity in the Pima Indians: its magnitude and relationship with diabetes. Am J Clin Nutr 53:1543S–1551S

Knowler WC, Pettitt DJ, Savage PJ, Bennett PH (1981) Diabetes incidence in Pima Indians: contributions of obesity and parental diabetes. Am J Epidemiol 113:144–156

Kriska AM, LaPorte RE, Pettitt DJ, Charles MA, Nelson RG, Kuller LH, Bennett PH, et al (1993) The association of physical activity with obesity, fat distribution and glucose intolerance in Pima Indians. Diabetologia 36:863–869

Lalouel JM, Morton NE (1981) Complex segregation analysis with pointers. Hum Hered 31:312–321

Lalouel JM, Rao DC, Morton NE, Elston RC (1983) A unified model for complex segregation analysis. Am J Hum Genet 35:816–826

Lewis CM (1991) Theoretical and applied problems in pedigree analysis. PhD thesis, University of Sheffield, Sheffield

MacLean CJ, Morton NE, Lew R (1975) Analysis of family resemblance. IV. Operational characteristics of segregation analysis. Am J Hum Genet 27:365–384

McCarthy MI, Hitman GA, Shields DC, Morton NE, Snehalatha C, Mohan V, Ramachandran A, et al (1994) Family

studies of non-insulin-dependent diabetes mellitus in South Indians. Diabetologia 37:1221–1230

Morton NE, MacLean CJ (1974) Analysis of family resemblance III. Complex segregation of quantitative traits. Am J Hum Genet 26:489–503

Pettitt DJ, Aleck KA, Carraher MJ, Bennett PH, Knowler WC (1988) Congenital susceptibility to NIDDM: Role of intrauterine environment. Diabetes 37:622–628

Price RA, Charles MA, Pettitt DJ, Knowler WC (1994) Obesity in Pima Indians: genetic segregation analysis of body mass index complicated by temporal increases in obesity. Hum Biol 66:251–274

Price RA, Charles MA, Pettitt DJ, Knowler WC (1993) Obesity in Pima Indians: large increases among post–World War II birth cohorts. Am J Phys Anthropol 92:473–479

Rice T, Laskarzewksi PM, Rao DC (1992) Commingling and complex segregation analysis of fasting plasma glucose in the Lipid Research Clinics Family Study. Am J Med Genet 44:399–404

Rich SS (1990) Mapping genes in diabetes: genetic epidemiological perspective. Diabetes 39:1315–1319

SAGE (1992) Statistical Analysis for Genetic Epidemiology, release 2.1. Computer program package available from the Department of Biometry and Genetics, Louisiana State University Medical Center, New Orleans

Sepehrnia B, Prezant TR, Rotter JI, Pettitt DJ, Knowler WC, Fischel-Ghodsian N (1995) Screening for mtDNA diabetes mutations in Pima Indians with NIDDM. Am J Med Genet 56:198–202

Serjeantson SW, Zimmet P (1991) Genetics of non-insulin-dependent diabetes mellitus in 1990. Ballières Clin Endocrinol Metab 5:477–493

Tiret L, Abel L, Rakotovao R (1993) Effect of ignoring genotype-environment interaction on segregation analysis of quantitative traits. Genet Epidemiol 10:581–586

WHO (World Health Organization) Study Group on Diabetes Mellitus (1985) Diabetes mellitus. World Health Organ Tech Rep Ser 727

Yamashita T, Mackay W, Rushforth N, Bennett P, Houser H (1984) Pedigree analyses of non-insulin-dependent diabetes mellitus (NIDDM) in the Pima Indians suggest dominant mode of inheritance. Am J Hum Genet Suppl 36:183S