Segregation Analysis of Diabetic Nephropathy in Pima Indians

Giuseppina Imperatore, William C. Knowler, David J. Pettitt, Sayuko Kobes, Peter H. Bennett, and Robert L. Hanson

Familial aggregation of diabetic nephropathy suggests the existence of genes determining susceptibility to nephropathy in addition to those leading to diabetes. In the present study, complex segregation analysis was performed in diabetic members of Pima Indian families to determine whether familial aggregation of nephropathy in this population could reflect the action of a single major gene. Nephropathy, defined by a urinary protein-to-creatinine ratio (PCR) ≥500 mg/g, was analyzed as a discrete trait in a class C regressive logistic model. Individuals with PCR <500 mg/g were considered unaffected. Segregation analysis was performed both for nephropathy at the last examination (prevalent cases) and for duration of diabetes at the onset of nephropathy (incident cases). The REGD program was used for the analysis of the prevalent cases and the REGTL program for the incident cases, both from the Statistical Analysis for Genetic Epidemiology package (Case Western University, Cleveland, OH). The analysis of prevalent cases included 2,107 Pima Indians from 715 nuclear families. A subset of 504 of these families containing 1,403 individuals was used in the analysis of incident cases. Analysis of prevalent cases supported the existence of a gene with a major role, in that hypotheses of no major effect and of no transmission of a major effect were rejected (P = 0.00001; P = 0.003), whereas Mendelian transmission was not rejected (P = 0.85). A dominant model provided the best fit, but a recessive model could not be rejected. The analysis of incident cases, however, did not support a major gene effect on duration of diabetes at the onset of nephropathy, and analyses of lifetime occurrence of nephropathy were inconclusive. The analysis of prevalent cases supports the hypothesis of a major genetic effect on susceptibility to diabetic nephropathy in Pima Indians, but the analysis of incident cases does not support a genetic effect on duration of diabetes at the onset of nephropathy. The discrepancy may reflect the difficulty in precisely dating onset of nephropathy. The parameters of the model derived from segregation analysis of prevalent cases may be useful in linkage studies to detect nephropathy susceptibility loci. *Diabetes* 49:1049–1056, 2000

From the Diabetes and Arthritis Epidemiology Section (G.I., W.C.K., S.K., P.H.B., R.L.H.), Phoenix Epidemiology and Clinical Research Branch, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Phoenix, Arizona; and the Sansum Medical Research Institute (D.J.P.), Santa Barbara, California.

Address correspondence and reprint requests to Robert L. Hanson, MD, National Institute of Diabetes and Digestive and Kidney Diseases, 1550 E. Indian School Rd., Phoenix, AZ 85014. E-mail: rhanson@phx.niddk.nih.gov.

Received for publication 1 June 1999 and accepted in revised form 1 February 2000.

ACR, albumin-to-creatinine ratio; AIC, Akaike's information criterion; PCR, protein-to-creatinine ratio; SAGE, Statistical Analysis for Genetic Epidemiology.

Diabetic nephropathy is the most common cause of end-stage renal disease in the Western world and is associated with considerable morbidity and mortality (1–3). The pathogenesis of diabetic nephropathy is likely to be multifactorial: Nephropathy is strongly dependent on the duration of diabetes (4,5); other risk factors include poor glycemic control, hypertension, and hypercholesterolemia (6–13). In addition, familial aggregation of diabetic nephropathy in Pima Indians and other populations suggests that the genes determining susceptibility to nephropathy are additional to those for diabetes (14–18). However, the mode of inheritance of diabetic nephropathy and the specific genetic factors involved in its etiology are largely unknown. Nonparametric linkage analyses in Pima Indians have identified a potential nephropathy-susceptibility locus on chromosome 7 (19), but parametric linkage analyses that account for the mode of inheritance of nephropathy and its duration-specific onset might have greater power and may identify additional loci. Such analyses require estimates of model parameters (disease allele frequency and duration-specific penetrance) that can be derived from segregation analysis (20). Consequently, segregation analysis of diabetic nephropathy in Pima Indians, a population with a high prevalence of type 2 diabetes and diabetic nephropathy (4,21,22), was performed.

RESEARCH DESIGN AND METHODS

Subjects and phenotypes. The residents of the Gila River Indian Community have participated in a longitudinal study of type 2 diabetes and its complications since 1965 (23). Approximately every 2 years, each resident ≥5 years old is asked to have a standardized medical examination that includes a glucose tolerance test in which glucose concentration is determined in plasma drawn 2 h after the ingestion of 75 g of glucose. Diabetes is diagnosed at the first examination at which the 2-h postload plasma glucose concentration is ≥11.1 mmol/l or in the course of routine medical care (22). Genealogical information for each individual has been collected since 1965, which allows for the construction of pedigrees for family and genetic studies.

Subjects are asked to void before ingesting the glucose, and a urine specimen is collected 2 h later. This urine specimen is tested for protein by dipstick. In all specimens containing at least a trace of protein, total protein is quantified by precipitation (24). Urinary creatinine concentration is measured and the urinary protein-to-creatinine ratio (PCR) (mg protein/g creatinine) is calculated (4). In the present analysis, nephropathy was defined as the presence of a urinary PCR ≥500 mg/g; individuals with a PCR <500 mg/g were considered unaffected.

Since 1982, urinary albumin concentration has been assessed quantitatively for all participants by immunonephelometry (25); however, the albumin concentration is below the limit of detectability by this assay in a small percentage of individuals. For analyses of these data, an inverse Gaussian transformation of the rank of urinary albumin-to-creatinine ratio (ACR) was used (people with undetectable urinary albumin were ranked below all others by the reciprocal of urinary creatinine concentration). This results in a continu-

ous variable suitable for estimating heritability, but not for segregation analysis, since the mixture of distributions produced by a major gene will be obscured by the normalizing transformation. Heritability was estimated as twice the sib-sib intraclass correlation (26) in 813 diabetic sib-pairs from all nuclear families (n = 283) who had at least 2 diabetic siblings with this measurement (n = 767).

The investigation was approved by the institutional review board of the National Institute of Diabetes and Digestive and Kidney Diseases and the Tribal Council of the Gila River Indian Community, and subjects gave informed consent.

Segregation analyses

Families. The present study included 715 nuclear families in whom at least 2 diabetic relatives (i.e., at least 1 parent and 1 offspring or 2 offspring) had been examined. These families included 1,430 diabetic offspring and their parents. Segregation analyses of nephropathy were restricted to diabetic members of these families (i.e., the affection status of the nondiabetic members was considered unknown). Clinical data were thus available for 435 diabetic mothers and 242 diabetic fathers of the 1,430 diabetic offspring: a total of 2,107 subjects. Because of computational problems caused by multiple loops in extended pedigrees, the analysis was restricted to nuclear families. For this reason, the phenotypic data for 487 individuals were used in >1 family; the 2,107 members, therefore, correspond to 1,620 unique individuals. The sibship size of the nuclear families ranged from 1 to 8, with a median size of 2 siblings per family. Given the low prevalence of nephropathy among nondiabetic subjects in this population (4,21), segregation analysis of nephropathy in this group was not performed.

The longitudinal nature of the Pima study permits estimation of the duration of diabetes at the onset of nephropathy for incident cases. Because individuals are often examined intermittently, the precision of the estimate will vary among individuals, and some people have nephropathy at their first diabetic examination. Because such cases may strongly influence estimation of the genetic parameters, segregation analyses were conducted both for incidence of diabetic nephropathy (first occurrence) and for prevalence of nephropathy at the last examination for which urinary protein data were available.

Prevalence analysis. Segregation analysis of prevalence of nephropathy was performed using the REGD program of the Statistical Analysis for Genetic Epidemiology (SAGE) package (27). Using a logistic model, this program analyzes a dichotomous trait. The probability that an individual is affected is modeled conditionally on the genotype at the putative disease locus, familial effects, and covariates. The parameters of the putative major gene that the program estimates include the following: q_A equals the frequency of the putative disease allele (A); β_{AA}, β_{AB}, and β_{BB} equal coefficients representing the effects of the AA, AB, and BB genotypes, respectively, on disease prevalence; τ_{AA}, τ_{AB}, and τ_{BB} equal the probability that 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).

To account for sources of familial resemblance in addition to the major gene (multifactorial inheritance), a regressive model was used (28). Thus, 6 additional parameters, $\delta_j(Y_j)$, were included in the model; they measure 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). (These effects are in comparison with the affection status of the relative in question being unknown.) If the sibling-sibling resemblance is greater than the parent-offspring resemblance, this can be the result of either the action of a major gene or of shared sibling environment. To avoid false inference of a major gene, therefore, it may be necessary to model a multifactorial sibling effect in addition to the parental effects (29,30). This was accomplished with an additional parameter, $\delta_{Sb}(Y_{Sb})$, which represents the residual effect of having an affected index sibling, where the index sibling for each individual represents the immediately preceding sibling, when siblings are ordered according to birth date. This model corresponds to the class C regressive model of Bonney (28).

Additional parameters representing the effects of covariates can also be included (ζ_k = covariate coefficients, for covariates k = 1 to n). In the present analysis, age, sex, and duration of diabetes were included as covariates in all models. Since the occurrence of nephropathy depends on the degree of glycemia (4,7–11), the maximum 2-h glucose observed during the longitudinal study was included as a covariate in some models to assess to extent to which glycemia influences estimates of the genetic effects. Moreover, examination date was entered in some models to account for the effect of possible temporal trends. The role of hypertension is difficult to assess, since it can be both a risk factor for and a consequence of renal disease, and since pharmacologic treatment for hypertension can delay the occurrence of nephropathy (12). Covariates related to hypertension and its treatment were not included, because these complex biologic relationships cannot be adequately accounted for by simply including such covariates in the model.

Given the parameters representing the genotype (β_i), multifactorial [$\delta_j(Y_j)$], and covariate (ξ_k) effects described above, the probability that an individual is affected is modeled as the following logistic function: $e^f/(1+ e^f)$, where $f = \beta_i + \delta_S (Y_S) + \delta_M (Y_M) + \delta_F (Y_F) + \delta_{Sb} (Y_{Sb}) + \zeta_1 (x_1) + \ldots \zeta_N (x_n)$.

Incidence analysis. Individuals who had prevalent nephropathy at their first diabetic examination or whose only examination was at the onset of diabetes were not considered in this analysis. The analysis, therefore, included 504 nuclear families representing a subset of the families included in the prevalence analysis. These families contain 1,403 diabetic people (969 offspring, 289 mothers, and 145 fathers), of whom 453 individuals developed nephropathy. Segregation analysis was conducted using the REGTL program of SAGE (27). As in the prevalence analysis, nephropathy was analyzed as a binary trait under the class C regressive logistic model. The major difference is that the incidence model is a survival analysis of time until first occurrence of nephropathy; thus, duration of diabetes at the onset of nephropathy is not a covariate, but part of the phenotype. The logistic function used in the REGTL program describes the probability that an individual is affected by duration of diabetes d as: $\gamma_i [e^{\Phi}/(1+ e^{\Phi})]$, where $\Phi = \beta_i + \alpha (d) + \delta_S (Y_S) + \delta_M (Y_M) + \delta_F (Y_F) + \delta_{Sb} (Y_{Sb}) + \zeta_1 (x_1) + \ldots \zeta_N (x_n)$. The mean duration at the onset of nephropathy can be calculated as $\mu = -\beta_i/\alpha$ with variance equal to $\sigma^2 = \pi^2/3\alpha^2$ (27,31). Either the β or the γ parameter can be dependent on the genotype. The first case corresponds to a model in which the genotype influences duration of diabetes at the onset of nephropathy and the susceptibility parameter γ is assumed to be the same for the 3 genotypes; i.e., the 3 putative genotypes have a different distribution of duration of diabetes at the onset of nephropathy, but lifetime cumulative incidence will be asymptotically the same. In the second model, the susceptibility parameter γ is assumed to vary with genotype and the β parameter is not. In this model, the mean duration of diabetes at the onset of nephropathy is the same for the 3 genotypes, whereas the lifetime cumulative incidence varies. The duration of diabetes at the onset of nephropathy was estimated in each subject by calculating the slope of the value of PCR versus the duration of diabetes between the date of the last examination with no nephropathy (PCR <500 mg/g) and the first examination with nephropathy (PCR ≥500) and interpolating the duration in which PCR = 500 mg/g. Individuals who did not develop nephropathy during the period of observation were considered "censored" at the last examination (i.e., duration of diabetes at the onset of neuropathy was treated as unknown but was greater than the duration at the last examination).

Models and hypothesis tests. In segregation analysis, a variety of genetic and nongenetic models are fit to the data to determine which model is most consistent with the observed data. The simplest model is the multifactorial model in which the familial resemblance is solely a function of the multifactorial effects with no major gene (or type) effect. This can be compared with a Mendelian model, which, in addition to multifactorial inheritance, includes a major gene effect with transmission probabilities fixed at their Mendelian expectations. An environmental model is also fit that includes a major type effect that is not transmitted from parent to offspring (i.e., all transmission probabilities are set equal to the allele frequency, q_A). All models can be compared with a general model in which the transmission probabilities are estimated from the data.

Models were compared using the likelihood ratio test, assuming that the negative value of twice the difference in natural logarithms for hierarchical models has a χ^2 distribution with degrees of freedom equal to the difference in the number of parameters independently estimated in the 2 models. The null hypothesis of no major effect was tested by comparing the multifactorial model (H_0: q_A = 1) with the Mendelian model (H_A: τ_{AA} = 1.0, τ_{AB} = 0.5, τ_{BB} = 0.0; $q_A \neq 1$; $\beta_{AA} \neq \beta_{AB} \neq \beta_{BB}$; 3 df). The null hypothesis of no transmission of the major effect was tested by comparing the environmental model (H_0: $q_A = \tau_{AA} = \tau_{AB} = \tau_{BB}$) with the general model (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 Mendelian model (H_0: τ_{AA} = 1.0, τ_{AB} = 0.5, τ_{BB} = 0.0) with the general model (H_A: $\tau_{AA} \neq 1.0$ or $\tau_{AB} \neq 0.5$ or $\tau_{BB} \neq 0.0$, 3 df). When covariates and major gene effects are considered simultaneously, the number of potential hypothesis tests is large and the natural hierarchy of the models may not be clear; in this case, models can also be compared with use of Akaike's information criterion (AIC) (32). A lower value of AIC represents a better fitting model.

RESULTS

Among the subset of 767 diabetic siblings with measurements of ACR, the heritability of this index of nephropathy estimated by sib-sib correlation was 21% (P = 0.005), adjusted for age, sex, and duration of diabetes, consistent with the hypothesis of a genetic effect. This hypothesis was tested further by means of segregation analyses of prevalence of nephropathy (total PCR ≥500 mg/g) in the full set of 2,107 diabetic subjects

TABLE 1
Clinical characteristics of diabetic individuals in segregation analysis of nephropathy

	Prevalence analysis	Incidence analysis
Families (*n*)	715	504
Individuals (*n*)	2,107	1,403
Men (%)	37.8	35.6
Individuals with nephropathy (%)	31.6	32.3
Parents with known affection status (%)	47.3	43.1
Duration of diabetes at most recent examination (years)	11.2 ± 8.7	13.0 ± 8.1
Duration of diabetes at onset of nephropathy (years)	—	12.4 ± 6.3
Age at most recent examination (years)	52.2 ± 14.9	53.6 ± 14.6
Age at onset of diabetes (years)	41.0 ± 13.6	40.6 ± 13.5
Year of most recent examination	1986 ± 9.3	1987 ± 8.7
Maximum 2-h postload plasma glucose (mmol/l)	22.5 ± 7.1	23.2 ± 6.9
Systolic blood pressure (mmHg)	135.8 ± 25.3	135.0 ± 24.8
Diastolic blood pressure (mmHg)	79.0 ± 12.6	78.4 ± 12.8
Individuals treated with antihypertensive medicines (%)	29.9	31.8

Data for continuous variables are means ± SD. Individuals and families in the incidence analysis are subsets of those in the prevalence analysis. Nephropathy is reported at the most recent examination for the prevalence analysis and at any examination for the incidence analysis. Systolic blood pressure, diastolic blood pressure, and treatment with antihypertensive medicines are shown as assessed at the most recent examination.

and by analyses of incident nephropathy in the subset of 1,403 diabetic subjects informative for these analyses.

Clinical characteristics of people included in the segregation analysis are shown in Table 1. Duration of diabetes at the most recent examination was almost 2 years longer for individuals in the incidence analysis compared with those in the prevalence analysis, a finding that reflects the exclusion of those who were only examined at the onset of diabetes. Other characteristics were generally similar for the subset included in the incidence analysis compared with those in the prevalence analysis.

Segregation analysis

Prevalence analysis. Table 2 shows parameter estimates for the 4 models in which age at examination, sex, and duration of diabetes were included as covariates. In the general model, the maximum likelihood estimates of the transmission probabilities were 1.00, 0.43, and 0.07, which are not significantly different from the Mendelian expectations of 1.0, 0.5, and 0.0 (Table 2). In the Mendelian model, the parameter estimates were very close to those of the general model, whereas the fit of both the multifactorial and environmental models was worse than the fit of the general model. The hypothesis of no major effect was strongly rejected ($\chi^2 = 32.5$, $P = 0.00001$, 3 df). The hypothesis of no transmission of the major effect was also rejected ($\chi^2 = 15.2$, $P = 0.003$, 3 df). The null hypothesis of Mendelian transmission was not rejected ($\chi^2 = 0.9$, $P = 0.85$, 3 df). The effect of duration of diabetes was highly statistically significant ($P < 0.001$, 1 df), whereas the effects of age ($P = 0.28$, 1 df) and sex ($P = 0.62$, 1 df) were not. In the absence of the major gene, the parameters representing multifactorial inheritance contributed significantly to the model ($\chi^2 = 34.8$, $P < 0.0001$, 7 df), but once the effect of the major gene was taken into account, they were not statistically significant ($\chi^2 = 9.8$, $P = 0.20$, 7 df).

In Table 3, parameter estimates for best-fitting general, dominant, and recessive Mendelian models are reported. The dominant model was similar to the more general Mendelian model, both in terms of parameter estimates and fit of the data. However, the recessive model was not rejected ($P = 0.19$). The prevalence of nephropathy by duration of diabetes and putative genotype on the basis of the parameters from the Mendelian model is shown in Fig. 1. The model predicts an important genetic influence on the prevalence of nephropathy in all categories of diabetes duration; for example, the predicted prevalence of nephropathy at 20 years of duration of diabetes is 85, 80, and 5% for the 3 putative genotypes

Both an earlier date of examination and a higher maximum 2-h glucose were significantly associated with nephropathy prevalence ($P < 0.001$ for each). In the models including date of examination as a covariate, however, the evidence for a major gene effect persisted. In models using the maximum 2-h postload plasma glucose as a covariate in addition to age, sex, and duration of diabetes, the estimates of the genetic effect were also generally similar to those of models that did not include glycemia as a covariate (data not shown).

Incidence analysis. The results of segregation analysis using the incidence data are shown in Tables 4 and 5. Under the duration-of-onset model (Table 4), the hypothesis of no major effect was rejected ($\chi^2 = 39.5$, $P < 0.001$, 3 df), suggesting the existence of >1 duration-of-onset distribution in the population. However, the null hypothesis of no transmission of the major effect was not rejected ($\chi^2 = 4.3$, $P = 0.23$, 3 df), whereas the null hypothesis of Mendelian transmission was ($\chi^2 = 8.0$, $P = 0.05$, 3 df). Under the susceptibility model (Table 5), estimates of the Mendelian model were similar to those of the general model. However, the null hypothesis of no major effect was not rejected ($\chi^2 = 2.7$, $P = 0.44$, 3 df), and neither the environmental nor the Mendelian model was rejected compared with the general model ($\chi^2 = 4.1$, $P = 0.25$, 3 df; $\chi^2 = 1.4$, $P = 0.71$, 3 df, respectively). The ability of the susceptibility model to discriminate among the various hypotheses in these data thus appears to be limited. Hypothesis tests for models including the date of diabetes diagnosis as a covariate did not differ from analyses that did not contain this covariate for both duration-of-onset and susceptibility models.

TABLE 2
Maximum likelihood estimates of parameters of segregation models for prevalence of diabetic nephropathy

Parameters	Multifactorial	Mendelian	Environmental	General
q_A	(1.0)	0.46	0.52	0.45
τ_{AA}	—	(1.0)	—	1.00
τ_{AB}	—	(0.5)	—	0.43
τ_{BB}	—	(0.0)	—	0.073
β_{AA}	–0.744	0.768	–0.217	1.713
β_{AB}	—	–0.428	–0.217	–0.685
β_{BB}	—	–26.318	–24.679	–28.619
δ_S (0)	0.078	–0.047	–0.018	–0.102
δ_S (1)	–0.328	–0.779	–0.624	–0.970
δ_M (0)	–0.343	–0.016	–0.379	0.127
δ_M (1)	0.064	–0.475	0.023	–0.683
δ_F (0)	–0.479	–0.338	–0.598	–0.244
δ_F (1)	0.163	–0.283	0.190	–0.488
δ_{Sb} (1)	0.356	0.081	0.425	0.024
ζ Age	–0.007	0.118	0.059	0.138
ζ Duration of diabetes	1.041	1.630	1.414	1.939
ζ Sex	–0.102	–0.003	–0.042	0.039
–2 ln likelihood	2,157.8	2,125.3	2,139.6	2,124.4
AIC	2,179.8	2,153.3	2,167.6	2,158.4

Parameters in parentheses are fixed at the listed values. Covariates age and duration of diabetes were standardized to a mean of 0 and an SD of 1 before the analyses. Sex was coded as 1 = male and 2 = female. q_A, Frequency of the putative disease allele (A); τ_{AA}, probability an individual of type AA transmits the A allele; τ_{AB}, probability an individual of type AB transmits the A allele; τ_{BB}, probability an individual of type BB transmits the A allele; β_{AA}, parameter representing AA genotype effect; β_{AB}, parameter representing AB genotype effect; β_{BB}, parameter representing BB genotype effect; δ_S (0), coefficient for residual multifactorial effect of having an unaffected spouse; δ_S (1), coefficient for residual multifactorial effect of having an affected spouse; δ_M (0), coefficient for residual multifactorial effect of having an unaffected mother; δ_M (1), coefficient for residual multifactorial effect of having an affected mother; δ_F (0), coefficient for residual multifactorial effect of having an unaffected father; δ_F (1), coefficient for residual multifactorial effect of having an affected father; δ_{Sb} (1), coefficient for residual multifactorial effect of having an affected previous sibling; ζ, covariate coefficient. AIC, lowest value corresponds to the best fitting model.

DISCUSSION

The hypothesis that diabetic renal disease has important genetic determinants is supported by the observation of strong familial aggregation of this disorder (14–18). Although it has been suggested that the sibling recurrence risk for diabetic nephropathy is compatible with a major genetic effect (17), the present study tested this hypothesis by a formal segregation analysis. Using prevalent cases of diabetic nephropathy, the analysis supported the hypothesis that the familial aggregation of diabetic nephropathy in Pima Indians is due, in large part, to the action of a major genetic locus. The longitudinal nature of the Pima study also allowed for an analysis of incident cases of diabetic nephropathy, but this analysis did not support a major gene influencing duration of diabetes at the onset of nephropathy (Mendelian transmission was rejected) and gave inconclusive results for a gene influencing lifetime susceptibility to nephropathy.

The discrepancy in the conclusions of prevalence and incidence analyses could arise from differences in the family data available for each analysis, differences in the assumptions of the models used, and differences in the way nephropathy is defined. In the longitudinal analysis, due to the exclusion of cases of nephropathy that were prevalent at the first examination and in individuals without follow-up, the sample was a subset (1,403 individuals) of that used in the prevalence analysis (2,107 individuals). To test the hypothesis that exclusion of these individuals from the incidence analysis may have resulted in the selection of a subset of families with a different segregation pattern, we also conducted the prevalence analysis using only individuals who were included in the incidence analysis. The results of segregation analysis in this subset of families (n = 504) were still consistent with a major gene effect. Therefore, the selection of families for the incidence analysis does not account for the discrepant conclusions between the incidence and prevalence analyses.

Segregation analysis can determine which of a limited set of models is most consistent with the observed data, but this does not constitute definitive evidence for or against a major genetic effect. It is thus possible for unmeasured familial environmental risk factors to have caused the false inference of a major genetic effect on prevalence of diabetic nephropathy. Alternatively, rejection of the Mendelian model in the incidence analysis may simply reflect a failure of the model used to capture the underlying genetic architecture. The prevalence and incidence models make different distributional assumptions and thus assess somewhat different hypotheses. This could result in a different degree of robustness to potential errors in estimating duration of diabetes at onset of nephropathy; these errors arise because individuals are often examined at irregular intervals. In the incidence analysis of duration of onset of nephropathy, the probability that an affected individual carries the putative disease allele is determined by the ratio of the density function of the genotypic duration-at-onset distributions and, consequently, a long duration of diabetes at the onset of nephropathy can strongly mitigate against a genetic cause of the disease. On the other hand, in the prevalence model, this probability is

TABLE 3
Maximum likelihood estimates of parameters of dominant, additive, recessive, and general Mendelian models for prevalence of diabetic nephropathy

Parameters	Dominant	Additive	Recessive	General Mendelian
q_A	0.46	0.84	0.84	0.46
τ_{AA}	(1.0)	(1.0)	(1.0)	(1.0)
τ_{AB}	(0.5)	(0.5)	(0.5)	(0.5)
τ_{BB}	(0.0)	(0.0)	(0.0)	(0.0)
β_{AA}	–0.055	–0.076	–0.076	0.768
β_{AB}	$(=\beta_{AA})$	$[=(\beta_{AA}+\beta_{BB})/2]$	$(=\beta_{BB})$	–0.428
β_{BB}	–26.268	–55.07	–27.697	–26.318
δ_S (0)	–0.017	–0.012	–0.012	–0.047
δ_S (1)	–0.734	–0.766	–0.766	–0.779
δ_M (0)	–0.146	–0.080	–0.080	–0.016
δ_M (1)	–0.286	–0.390	–0.390	–0.475
δ_F (0)	–0.448	–0.391	–0.391	–0.338
δ_F (1)	–0.125	–0.148	–0.148	–0.283
δ_{Sb} (1)	0.165	0.146	0.146	0.081
ζ Age	0.109	0.101	0.101	0.118
ζ Duration of diabetes	1.541	1.522	1.523	1.630
ζ Sex	–0.025	–0.023	–0.023	–0.003
–2 ln likelihood	2,125.9	2,127.6	2,127.6	2,125.3
AIC	2,151.9	2,153.6	2,153.6	2,153.3

Parameters in parentheses are fixed at the listed values. Covariates age and duration of diabetes were standardized to a mean of 0 and an SD of 1 before the analyses. Sex was coded as 1 = male and 2 = female. Symbols are defined as in legend of Table 2.

determined by the ratio of the duration-specific genotypic prevalence of the disease, and an affected person with a long duration of diabetes may provide little information regarding the genetic cause of the disease but is not evidence against a genetic cause (33). Inaccurate estimation of the duration of diabetes at the onset of nephropathy, as will occur with intermittent examinations, will tend to produce a greater distortion in the transmission pattern for the incidence than for the prevalence model.

The diagnosis of nephropathy in the present study was made on the basis of a single urine specimen, and the data were not available to exclude nondiabetic causes of proteinuria. Intraindividual fluctuation in urinary protein excretion could thus result in misclassification of some individuals. Because the incidence analysis depends on determining the first of multiple examinations at which nephropathy occurred, it may be more susceptible to such misclassification. There were 44 individuals who met the criteria for nephropathy at 1 examination, and thus were considered incident cases, who had a lower level of proteinuria at the last examination and were not considered prevalent cases. In 33 (75%) of these individuals, no proteinuria was detectable by dipstick at the last examination, suggesting that the initial proteinuria was not due to diabetic nephropathy, but to a transient cause of proteinuria (e.g., infection). When these 44 individuals were considered to have nephropathy at the last examination, a major genetic effect on the prevalence of nephropathy was not supported (Mendelian transmission was rejected). Moreover, when these 44 individuals were not considered to be cases of nephropathy in the incidence analysis, Mendelian transmission was no longer rejected. These analyses suggest that the differences between the incidence and prevalence analyses are due, in part, to individuals whose affection status is potentially misclassified. Although one might expect the incidence analysis to be less robust to such misclassification errors, it is difficult to assess whether this has obscured the major genetic effect in the incidence analysis or whether the prevalence analysis falsely implicated such an effect.

The present segregation analyses were conducted on nuclear families. Many of these families could be connected into a few large pedigrees, but analysis of these pedigrees is made computationally difficult by the presence of multiple loops. It is not clear, however, that complex segregation

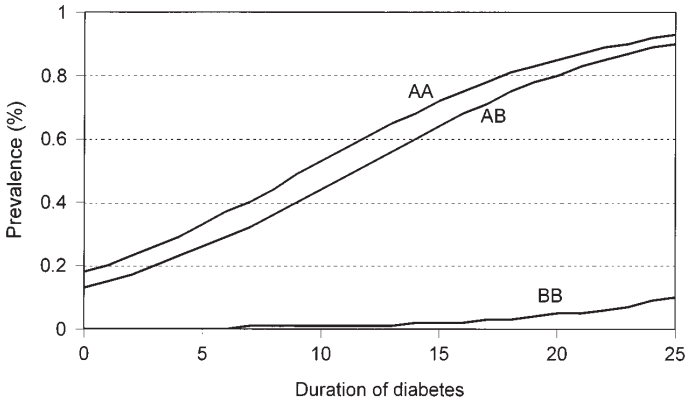

FIG. 1. Prevalence of diabetic nephropathy predicted by the best-fitting Mendelian model by duration of diabetes and genotype at the putative disease locus. AA, homozygous for the disease allele A; AB, heterozygote; BB, subjects without the A allele. The model estimated a very low value of β_{BB}, suggesting a low prevalence of nephropathy for the BB genotype, but with a wide standard error. For plotting purposes, therefore, the value of β_{BB} was fixed at –4.59 (i.e., a prevalence of 0.01 at the mean value of duration of diabetes [11.2 years]).

TABLE 4
Maximum likelihood estimates of parameters of segregation models for incidence of diabetic nephropathy: effect on duration of diabetes at the onset of nephropathy

Parameters	Multifactorial	Mendelian	Environmental	General
q_A	(1.0)	0.68	0.66	0.30
Transmission probability				
τ_{AA}	—	(1.0)	—	0.95
τ_{AB}	—	(0.5)	—	0.25
τ_{BB}	—	(0.0)	—	0.29
Mean duration of onset				
μ_{AA}	16.25	10.85	10.27	7.63
μ_{AB}	—	20.75	20.16	12.15
μ_{BB}	—	27.70	26.83	22.72
σ^2	54.09	24.77	21.83	21.63
γ	0.82	0.86	0.86	0.86
δ_S (0)	–0.301	–0.658	–0.727	–0.715
δ_S (1)	0.132	0.013	–0.142	0.242
δ_M (0)	–0.064	–0.014	–0.234	–0.216
δ_M (1)	0.204	0.160	0.478	0.381
δ_F (0)	–0.096	0.0748	–0.164	0.039
δ_F (1)	0.706	0.156	0.517	0.510
δ_{Sb} (1)	0.427	0.241	0.584	0.408
ζ Age	0.079	0.058	0.058	0.091
ζ Sex	0.008	–0.330	–0.402	–0.495
–2 ln likelihood	3,860.4	3,820.9	3,817.2	3,812.9
AIC	3,884.4	3,850.9	3,847.2	3,848.9

Parameters in parentheses are fixed at the listed values. The covariate age was standardized to a mean of 0 and an SD of 1 before the analyses. Sex was coded as 1 = male and 2 = female. q_A, Frequency of the putative disease allele (A); τ_{AA}, probability an individual of type AA transmits the A allele; τ_{AB}, probability an individual of type AB transmits the A allele; τ_{BB}, probability an individual of type BB transmits the A allele; μ_{AA}, mean duration of diabetes at the onset of nephropathy for individuals of type AA; μ_{AB}, mean duration of diabetes at the onset of nephropathy for individuals of type AB; μ_{BB}, mean duration of diabetes at the onset of nephropathy for individuals of type BB; σ^2, variance of duration of onset; γ, susceptibility; δ_S (0), coefficient for residual multifactorial effect of having an unaffected spouse; δ_S (1), coefficient for residual multifactorial effect of having an affected spouse; δ_M (0), coefficient for residual multifactorial effect of having an unaffected mother; δ_M (1), coefficient for residual multifactorial effect of having an affected mother; δ_F (0), coefficient for residual multifactorial effect of having an unaffected father; δ_F (1), coefficient for residual multifactorial effect of having an affected father; δ_{Sb} (1), coefficient for residual multifactorial effect of having an affected previous sibling; ζ, covariate coefficient. AIC, lowest value corresponds to the best fitting model.

analysis in extended pedigrees will result in more valid parameter estimates than analyses conducted in nuclear families. In fact, analysis of nuclear families may be preferable, because it is likely to be more robust to violations of model assumptions (34). This approach, however, necessitates the duplication of some individuals and this could inflate the statistical significance of the results. To correct for the fact that a person may appear as a parent in one family and as an offspring in another family, analyses were conducted using an ascertainment correction that removes the contribution of the parental phenotypes from the likelihood, and the results were similar to those presented above.

Diabetes in Pima Indians is type 2 diabetes (35) and 92% of the end-stage renal disease in this population occurs in individuals with diabetes (36). The present data, therefore, do not allow for estimation of the potential genetic determinants of nephropathy in type 1 diabetes or in the absence of diabetes. Since few nondiabetic Pima Indians have nephropathy, the segregation analyses were restricted to individuals with type 2 diabetes. In general, this approach should not result in biased parameter estimates, unless the putative nephropathy-susceptibility gene also influences susceptibility to diabetes. In this case, analyses restricted to diabetic individuals may overestimate the frequency of the high-risk allele relative to its frequency in the whole population. With the available data, it is difficult to evaluate whether the major nephropathy-susceptibility gene identified in the present study also influences susceptibility to diabetes, but linkage analyses to date have not identified common loci for diabetes and nephropathy in the Pima Indians (19,37,38). Segregation analysis of urinary ACR in Caucasians with type 2 diabetes has also suggested a major genetic effect (39). This study was not performed on incident cases of nephropathy, but the findings lend support to the hypothesis that in subjects with type 2 diabetes, the familial aggregation of nephropathy is partly due to a major gene. It is unknown, however, whether nephropathy-susceptibility genes in Pima Indians are the same as those conferring susceptibility to nephropathy in other populations.

Inclusion of unaffected individuals in a genetic analysis of diabetic nephropathy necessitates accounting for the effect of duration of diabetes, which is a strong risk factor for complications of diabetes. Duration of diabetes was taken into account in the prevalence analysis by including it as a covariate and in the incidence analysis by including it as part of the phenotype. Other factors that may also influence the onset and progression of nephropathy

TABLE 5
Maximum likelihood estimates of parameters of segregation models for incidence of diabetic nephropathy: effect on susceptibility to nephropathy

Parameters	Multifactorial	Mendelian	Environmental	General
q_A	(1.0)	0.64	0.53	0.55
Transmission probability				
τ_{AA}	—	(1.0)	—	0.71
τ_{AB}	—	(0.5)	—	0.71
τ_{BB}	—	(0.0)	—	0.0
Susceptibility				
γ_{AA}	0.82	0.98	1.00	0.99
γ_{AB}	—	0.87	1.00	1.00
γ_{BB}	—	0.00000001	0.21	0.00000001
μ	16.25	16.10	16.25	16.07
σ^2	54.09	52.65	54.09	52.42
δ_S (0)	–0.301	–0.260	–0.301	–0.233
δ_S (1)	0.132	0.124	0.132	0.130
δ_M (0)	–0.064	0.033	–0.064	0.062
δ_M (1)	0.204	0.144	0.204	0.118
δ_F (0)	–0.096	–0.059	–0.096	–0.056
δ_F (1)	0.706	0.611	0.706	0.579
δ_{Sb} (1)	0.428	0.354	0.428	0.362
ζ Age	0.079	0.094	0.079	0.099
ζ Sex	0.008	0.012	0.008	0.010
–2 ln likelihood	3,860.4	3,857.7	3,860.4	3,856.3
AIC	3,884.4	3,887.7	3,890.4	3,892.3

Parameters in parentheses are fixed at the listed values. The covariate age was standardized to a mean of 0 and an SD of 1 before the analyses. Sex was coded as 1 = male and 2 = female. q_A, Frequency of the putative disease allele (A); τ_{AA}, probability an individual of type AA transmits the A allele; τ_{AB}, probability an individual of type AB transmits the A allele; τ_{BB}, probability an individual of type BB transmits the A allele; γ_{AA}, susceptibility to nephropathy for individuals of type AA; γ_{AB}, susceptibility to nephropathy for individuals of type AB; γ_{BB}, susceptibility to nephropathy for individuals of type BB; μ, mean duration of onset; σ^2, variance of duration of onset; δ_S (0), coefficient for residual multifactorial effect of having an unaffected spouse; δ_S (1), coefficient for residual multifactorial effect of having an affected spouse; δ_M (0), coefficient for residual multifactorial effect of having an unaffected mother; δ_M (1), coefficient for residual multifactorial effect of having an affected mother; δ_F (0), coefficient for residual multifactorial effect of having an unaffected father; δ_F (1), coefficient for residual multifactorial effect of having an affected father; δ_{Sb} (1), coefficient for residual multifactorial effect of having an affected previous sibling; ζ, covariate coefficient. AIC, lowest value corresponds to the best fitting model.

include the following: pharmacological treatment for hypertension, especially the use of ACE inhibitors, which may delay the progression of renal disease in subjects with either type 1 (40) or type 2 (12,41) diabetes; lipid levels (8); and glycemic control, which influences the progression of renal disease (4,10,11). Given the extent to which hypertension and glycemic control influence the risk for diabetic nephropathy might be under genetic control, the genetic effects detected in the present study, by segregation analysis, might operate by these mechanisms. However, analyses including the maximum glucose value observed in the longitudinal study as a covariate did not substantially modify the estimates of the genetic effect.

The ultimate test of the utility of segregation analysis is whether the resulting models can be used in linkage analysis to identify disease-susceptibility loci. Given its compatibility with the major gene model, the model derived from the prevalence analysis will likely be the most useful in this regard. Although the major gene model provides the best fit to the data, the models used do not distinguish between monogenic or oligogenic inheritance. Thus, more than 1 gene, with different penetrances and modes of inheritance, may be involved. Even under oligogenic inheritance, the parameters derived in a single locus segregation analysis can provide power for detecting the susceptibility loci (42). The parameter estimates derived from this segregation analysis have been used in linkage analysis using duration-specific penetrances in a subset of diabetic Pima Indians who participated in a genome-wide scan, which has identified suggestive evidence of a nephropathy-susceptibility locus on chromosome 18q (37).

ACKNOWLEDGMENTS

Some of the results were obtained with use of the program package SAGE, which was supported by a U.S. Public Health Service Resource grant (1 P41 RR03655) from the National Center for Research Resources.

We thank the members of the Gila River Indian Community for participating in these studies. Particular thanks are due to Vickie Andre, Sonja Antone, and Linda Phillips for technical assistance.

This work was presented in part at the 47th Annual Meeting of the American Society of Human Genetics, Baltimore, Maryland, 1 November 1997.

REFERENCES

1. U.S. Renal Data System: USRDS 1994 annual data report: incidence and causes of treated renal disease. *Am J Kidney Dis* 24 (Suppl. 2):S48–S56, 1994
2. Cordonnier D, Bayle F, Benhamou PY, Milongo R, Zaouli P, Maynard C: Future trends of management of renal failure in diabetics. *Kidney Int* 43 (Suppl. 41):8–13, 1993
3. Raine AEG: Epidemiology, development and treatment of end-stage renal failure in type 2 (non-insulin-dependent) diabetic patients in Europe. *Diabetologia* 36:1099–1104, 1993
4. Kunzelman CL, Knowler WC, Pettitt DJ, Bennett PH: Incidence of proteinuria in type 2 diabetes mellitus in the Pima Indians. *Kidney Int* 35:681–687, 1989
5. Krolewski AS, Warram JH, Christlieb AR, Busik EJ, Kahn CR: The changing natural history of nephropathy in type I diabetes. *Am J Med* 78:785–794, 1985
6. Schmitz A, Vaeth M, Mogensen CE: Systolic blood pressure relates to the rate of progression of albuminuria in NIDDM. *Diabetologia* 37:1251–1258, 1994
7. Ravid M, Brosh D, Ravid-Safran D, Levy Z, Rachmani R: Main risk factors for nephropathy in type 2 diabetes mellitus are plasma cholesterol levels, mean blood pressure, and hyperglycemia. *Arch Intern Med* 158:998–1004, 1998
8. Nelson RG, Knowler WC, Pettitt DJ, Hanson RL, Bennett PH: Incidence and determinants of elevated urinary albumin excretion in Pima Indians with NIDDM. *Diabetes Care* 18:182–187, 1995
9. Gall MA, Hougaard P, Borch-Johnsen K, Parving HH: Risk factors for development of incipient and overt diabetic nephropathy in patients with non-insulin dependent diabetes mellitus: prospective, observational study. *BMJ* 314:783–788, 1997
10. U.K. Prospective Diabetes Study (UKPDS) Group: Intensive blood-glucose with sulphonylureas or insulin compared with conventional treatment and risk of complication in patients with type 2 diabetes (UKPDS 33). *Lancet* 352:837–853, 1998
11. Diabetes Control and Complications Trial Research Group: The effect of intensive treatment of diabetes on the development and progression of long-term complications in insulin-dependent diabetes mellitus. *N Engl J Med* 329:977–986, 1993
12. Nielsen FS, Rossing P, Gall MA, Skott P, Smidt UM, Parving HH: Long-term effect of lisinopril and atenolol on kidney function in hypertensive NIDDM subjects with diabetic nephropathy. *Diabetes* 46:1182–1188, 1997
13. U.K. Prospective Diabetes Study (UKPDS) Group: Tight blood pressure control and risk of macrovascular and microvascular complications in type 2 diabetes: UKPDS 38. U.K. Prospective Diabetes Study Group. *BMJ* 317:703–713, 1998
14. Seaquist ER, Goetz FC, Rich SS, Barbosa J: Familial clustering of diabetic kidney disease: evidence for genetic susceptibility to diabetic nephropathy. *N Engl J Med* 320:1161–1165, 1989
15. Pettitt DJ, Saad MF, Bennett PH, Nelson RG, Knowler WC: Familial predisposition to renal diseases in two generations of Pima Indians with type 2 (non-insulin-dependent) diabetes mellitus. *Diabetologia* 33:438–443, 1990
16. Borch-Johnsen K, Norgaard K, Hommel E, Mathiesen ER, Jensen JS, Deckert T, Parving HH: Is diabetic nephropathy an inherited complication? *Kidney Int* 41:719–722, 1992
17. Quinn M, Angelico MC, Warram JH, Krolewski AS: Familial factors determine the development of diabetic nephropathy in patients with IDDM. *Diabetologia* 39:940–945, 1996
18. Faronato PP, Maioli M, Tonolo G, Brocco E, Noventa F, Piarulli F, Abaterusso C, Modena F, de Bigontina G, Velussi M, Inchiostro S, Santeusanio F, Bueti A, Nosadini R: Clustering of albumin excretion rate abnormalities in Caucasian patients with NIDDM: the Italian NIDDM Nephropathy Study Group. *Diabetologia* 40:816–823, 1997
19. Imperatore G, Hanson RL, Pettitt DJ, Kobes S, Bennett PH, Knowler WC, the Pima Diabetes Genes Group: Sib-pair linkage analysis for susceptibility genes for microvascular complications among Pima Indians with type 2 diabetes. *Diabetes* 47:821–830, 1998
20. Jarvik PG: Complex segregation analysis: uses and limitations. *Am J Hum Genet* 63:942–946, 1998
21. Nelson RG, Kunzelman CL, Pettitt DJ, Saad MF, Bennett PH, Knowler WC: Albuminuria in type 2 (non-insulin-dependent) diabetes mellitus and impaired glucose tolerance in Pima Indians. *Diabetologia* 32:870–876, 1989
22. Knowler WC, Pettitt DJ, Hamman RF, Miller M: Diabetes incidence and prevalence in Pima Indians: a 19-fold greater incidence than in Rochester, Minnesota. *Am J Epidemiol* 108:497–505, 1978
23. Bennett PH, Burch TA, Miller M: Diabetes mellitus in American (Pima) Indians. *Lancet* ii:825–828, 1971
24. Shevsky MC, Stafford DD: A clinical method for the estimation of protein in urine and other body fluids. *Arch Intern Med* 32:222–225, 1923
25. Vasquez B, Flock EV, Savage PJ, Nagulesparan M, Bennion LJ, Baird HR, Bennett PH: Sustained reduction of proteinuria in type 2 (non-insulin-dependent) diabetes following diet-induced reduction of hyperglycemia. *Diabetologia* 26:127–133, 1984
26. Snedecor GW, Cochran WG: One-way classification: analysis of variance. In *Statistical Methods*. 6th ed. Snedecor GW, Cochran WG, Eds. Ames, IA, Iowa State University Press, 1967, p. 294–296
27. Case Western Reserve University: SAGE: *Statistical Analysis for Genetic Epidemiology, release 2.2*. Cleveland, OH, Department of Epidemiology and Biostatistics, Case Western Reserve University, 1994
28. Bonney GE: Regressive logistic models for familial disease and other binary traits. *Biometrics* 42:611–625, 1986
29. MacLean CJ, Morton NE, Lew R: Analysis of family resemblance. IV. Operational characteristics of segregation analysis. *Am J Hum Genet* 27:365–384, 1975
30. Elston RG: Segregation analysis. *Adv Hum Genet* 11:63–120, 1981
31. Elston RC, George VT: Age of onset, age at examination and other covariates in the analysis of family data. *Genet Epidemiol* 6:217–220, 1989
32. Akaike H: Information theory and an extension of the maximum likelihood principle. In *2nd International Symposium of Information Theory*. Budapest, Akademia Kaido, 1973, p. 267–281
33. Terwilliger JD, Ott J: Advanced application of affection status I: incomplete penetrance revisited. In *Handbook of Human Genetic Linkage*. Terwilliger JD, Ott J, Eds. Baltimore, MD, Johns Hopkins University Press, 1994, p. 65–72
34. Lalouel JM, Morton NE: Complex segregation analysis with pointers. *Hum Hered* 31:312–321, 1981
35. Knowler WC, Bennett PH, Bottazzo GF, Doniach D: Islet cell antibodies and diabetes mellitus in Pima Indians. *Diabetologia* 17:161–164, 1979
36. Nelson RG, Newman JM, Knowler WC, Sievers ML, Kunzelman CL, Pettitt DJ, Moffett CD, Teutsch SM, Bennett PH: Incidence of end-stage renal disease in type 2 (non-insulin-dependent) diabetes mellitus in Pima Indians. *Diabetologia* 31:730–736, 1988
37. Imperatore G, Hanson RL, Pettitt DJ, Kobes S, Bennett PH, Knowler WC, the Pima Diabetes Genes Group: Segregation and linkage analyses of diabetic nephropathy in Pima Indians (Abstract). *Am J Hum Genet* 61 (Suppl. 1):A280, 1997
38. Hanson RL, Ehm MG, Pettitt DJ, Prochazka M, Thompson DB, Timberlake D, Foroud T, Kobes S, Baier L, Burns DK, Almasy L, Blangero J, Garvey WT, Bennett PH, Knowler WC: An autosomal genomic scan for loci linked to type II diabetes mellitus and body-mass index in Pima Indians. *Am J Hum Genet* 63:1130–1138, 1998
39. Fogarty DG, Rich SS, Wantman M, Warram JH, Krolewski AS: Albumin excretion in families with NIDDM is strongly influenced by a major gene: results of a segregation analysis. *Diabetes* 47 (Suppl. 1):A12, 1998
40. Lewis EJ, Hunsisker LG, Bain RP, Rohde RD: The effect of angiotensin-converting-enzyme inhibition on diabetic nephropathy: the Collaborative Study Group. *N Engl J Med* 329:1456–1462, 1993
41. Ravid M, Brosh D, Levi Z, Bar-Dayan Y, Ravid D, Rachmani R: Use of enalapril to attenuate decline in renal function in normotensive, normoalbuminuric patients with type 2 diabetes mellitus: a randomized, controlled trial. *Ann Intern Med* 128:982–988, 1998
42. Vieland VJ, Hodge SE, Greenberg DA: Adequacy of single-locus approximations for linkage analyses of oligogenic traits. *Genet Epidemiol* 9:45–59, 1992