Am. J. Hum. Genet. 55:981–990, 1994

Genetic Epidemiology of Vitiligo: Multilocus Recessivity Cross-validated

Swapan K. Nath,[1] Partha P. Majumder,[1] and James J. Nordlund[2]

[1]Anthropometry and Human Genetics Unit, Indian Statistical Institute, Calcutta; and [2]Department of Dermatology, University of Cincinnati Medical Center, Cincinnati

Summary

Vitiligo is a dermatological disorder characterized by hypopigmentary patches that tend to become progressive over time. There are reports of extensive familial aggregation. A genetic model for this disorder was earlier proposed by us. This model postulates that recessive alleles at multiple unlinked autosomal loci interact epistatically in the pathogenesis of vitiligo. The present family study was primarily undertaken to cross-validate the proposed genetic model. Data on 194 families from the United States were collected. Each family was ascertained through an affected proband. Analyses of these data reveal that ~20% of probands have at least one first-degree relative afflicted with vitiligo. All types of first-degree relatives of probands show a significant risk of developing vitiligo. Results of segregation and robustness analyses reveal that the genetic model postulated by us previously is the most parsimonious model for the present family data set.

Introduction

A disorder that shows a high degree of familial aggregation but is not inherited in a simple Mendelian fashion may result from epistatic interactions of alleles at two or more loci. Even when only two recessive loci epistatically interact in the pathogenesis of a disorder, the vast majority of families ascertained through an affected proband will have no other affected member. For example, when the population prevalence of a two-locus recessive disorder is 1/1,000, ~75% of nuclear families and ~65% of extended families of the proband are simplex (Majumder 1993). These figures increase to ~85% and ~75%, respectively, when the prevalence decreases to 1/10,000. Segregation analysis of such a multilocus recessive disorder often results in the simplistic and incorrect inference that the disorder is incompletely penetrant with a large proportion of sporadics (P. P. Majumder, unpublished observation). Several studies (e.g., Vieland et al. 1992) have shown that for the purpose of detecting linkage, misspecification of the two-locus model by a single-locus model does not affect the expected maximum lod score substantially. However, model misspecification leads to loss of power of detecting linkage and of biased estimates of the recombination fraction and other segregation parameters (Vieland et al. 1992; Dizier et al. 1993; Goldin and Weeks 1993; Rice et al. 1993). (Some relevant results pertaining to efficiency are also given by Majumder [1989].) Rice et al. (1993) have also shown that "no single-locus trait can fit the recurrence risks in relatives when the true mode of inheritance is oligogenic" (see also Neuman and Rice 1992). For a two-locus trait, Schork et al. (1993) have also shown that two-trait-locus, two-marker-locus linkage analysis can provide substantially more linkage information than can standard one-trait-locus, one-marker-locus analysis. Identification of the correct genetic model of a disorder by segregation analysis is, therefore, not only useful but is also necessary for both genetic counseling and localization of genes.

For the past few years, we have been involved in developing methods for analysis of family data under multilocus models (Majumder et al. 1988, 1989; Majumder and Nath 1992), in particular under the multilocus recessive model introduced by Li (1987). As a part of these efforts, we have proposed a genetic model for a pigmentary disorder called vitiligo, which shows significant familial aggregation but no simple Mendelian transmission (Das et al. 1985; Majumder et al. 1988). On the basis of analysis of data collected from 274 nuclear families in Calcutta, we postulated that recessive alleles at four unlinked autosomal diallelic loci are involved in the pathogenesis of vitiligo. It may be noted that in some animals, such as the Smyth chicken and the $C57BL/6 - mi^{vit}/mi^{vit}$ mouse, depigmentation seems to be transmitted as a simple autosomal recessive trait (Nordlund 1987). However, since these are highly inbred strains, such findings may not reflect the true genetic basis of the disorder.

The purpose of this paper is to present results of cross-validation of the proposed genetic model for vitiligo. The cross-validation has been done by collecting a new set of data from families belonging to a different ethnic background. We also present relevant epidemiological features of vitiligo in this new data set, some of which have been

briefly reported by Majumder et al. (1993), and extend our previous model and statistical methodology.

While just over a decade ago, the biochemical and molecular picture of melanogenesis was relatively straightforward, recent advances in pigmentation research have revealed many "melanogenic factors" that influence the quality and/or quantity of melanin produced. Several genes encoding such melanogenic factors have been identified (Hearing and Tsukamoto 1991; Hearing 1993). Further, melanocytes, being migratory cells, must migrate appropriately for normal pigmentation. Even after appropriate migration, pigmentary problems can and do arise at the tissue level (Quevado et al. 1987; Wick et al. 1987). In fact, vitiligo is a disorder in which melanocytes seem to be properly produced and distributed in tissues but are rendered nonfunctional later (probably because of melanocyte destruction), leading to hypopigmentary patches that tend to become progressive over time (Mosher et al. 1979). Hypopigmentation is also observed in Waardenburg, Prader-Willi, and Angelman syndromes and in piebaldism. Mutations at specific loci have been found to be associated with the murine analogues of these phenotypes (see the report by Hearing [1993] for a recent review and references). Obviously, therefore, there are multiple points at which the process of normal pigmentation can be disturbed. If these points are under genetic control, then it is possible that multiple genetic loci are responsible for the pathogenesis of a pigmentary disorder. Alleles at all of these loci may act jointly, or the presence of particular alleles at a subset of these loci may suffice in the abnormal phenotypic manifestation. The genetic models investigated subsequently are motivated from these considerations.

The prevalence of vitiligo, on the basis of population surveys, has been estimated to be 0.38% in Denmark (Howitz et al. 1977) and 0.46% in Calcutta (Das et al. 1985). On the basis of clinical records, the prevalence in the United States is estimated to be ~1% (Lerner 1959; El-Mofty 1968). This latter estimate is obviously biased upwardly. To the best of our knowledge, no population-based estimate is available from the United States. Prevalence of vitiligo varies significantly with age; age-specific prevalences between genders are not significantly different (Das et al. 1985). The age at onset of vitiligo is variable.

Subjects and Methods

Families in this study were ascertained through a single individual affected with vitiligo (the proband). The probands were chosen from an unsorted list of all patients registered with the National Vitiligo Foundation, Tyler, TX. Every fifth patient on the list was sent a questionnaire requesting relevant information on himself or herself and his or her family members. In all, 300 questionnaires were sent. A total of 194 completed questionnaires were returned. The response rate, therefore, was ~65%. Prior to the data received from the patients, no information on gender, ethnicity, age, number of affected relatives, or other variables was available to us. The questionnaire sought information regarding presence of vitiligo in a fixed set of relatives of the proband; current ages; ages at onset; consanguinity; participation in sunbathing, tanning, etc.; handling of chemicals, pesticides, etc.; thyroid problems; premature graying of hair; hearing loss; ocular abnormalities; diabetes; and other clinically important variables. Returned questionnaires were scrutinized for missing and/or inconsistent information. When incomplete or inconsistent observations were noted, telephone interviews or follow-up letters were used to resolve inconsistencies. The questionnaire also sought information on the names and qualifications of the persons who had performed clinical evaluations of the probands. It was found that clinical evaluations of all the probands and also of the vast majority of their affected relatives were performed by professional dermatologists. In fact, many of the probands reported that their initial clinical evaluations had been performed by one of us (J.J.N.). Therefore, the possibility of false-positive reporting is minimal in this data set. While the possibility of false-negative reporting remains, it must be mentioned that the probands, being members of the National Vitiligo Foundation, are very well informed about the clinical features of vitiligo, which reduces the possibility of false-negative reportings. It has sometimes been reported that there is a preponderence of familial cases among those registering themselves with foundations such as the National Vitiligo Foundation. However, as we shall report later, ~80% of the probands have no first-degree relative afflicted with vitiligo. This proportion also closely agrees with the finding of a previous study conducted by us in India, in which probands were not selected from a registry list. While these facts by themselves do not prove that the present data set is free of any bias, they indicate that any biases that may exist because of sampling probands from the list of the National Vitiligo Foundation may be small and negligible.

Of the 194 families from whom data were collected, 160 families were Caucasian; the remaining 34 families were Black, Hispanic, Asian, or of other ethnic backgrounds. No two probands belonged to the same family. For the analyses performed, it was necessary to use age- and gender-specific estimates of population prevalence of vitiligo. We therefore restricted our attention in this report to the set of 160 Caucasian families. Prevalence estimates from the non-Caucasian (e.g., Black, Hispanic, etc.) population groups are not available. Pooling of data from populations of different ethnic backgrounds with possibly differing prevalences of vitiligo might yield distorted results of the analyses. Thus, it was necessary to treat families of different ethnic backgrounds separately. The number of families of non-Caucasian ethnicity was also too small for performing any reasonable statistical analysis of the data from these families. The statewise distribution of the places of residence of the 160 probands is as follows: AK, 1; AL, 1;

Table 1

Age at Onset (years [mean ± standard error]) of Vitiligo of Affected Individual among Probands and Their First-Degree Relatives

Gender of Affected Individual	Probands Only	All Affected Individuals (Including Probands and Their First-Degree Relatives)
Male	17.62 ± 1.86 (n=53)	18.61 ± 1.64 (n=69)
Female	23.81 ± 1.51 (n=107)	22.07 ± 1.40 (n=134)
Overall	21.76 ± 1.16 (n=160)	22.37 ± 1.03 (n=203)

AR, 2; AZ, 3; CA, 6; CO, 1; CT, 5; DC, 1; FL, 7; GA, 5; IL, 17; IN, 2; KS, 1; KY, 4; LA, 2; MA, 7; ME, 1; MI, 6; MN, 2; MO, 2; NC, 1; NH, 1; NJ, 6; NY, 8; OR, 1; OH, 11; PA, 8; TN, 21; TX, 20; VA, 1; WA, 2; WI, 2; WV, 1; and WY, 1.

Table 2

Distribution of Age at Onset of Vitiligo among Affected Probands

Age at Onset (years)	Cumulative Proportion (p_i)	$z_i = 1-p_i$
≤4	.056	.944
5–9	.268	.732
10–14	.412	.588
15–19	.494	.506
20–24	.619	.381
25–29	.707	.293
30–34	.782	.218
35–39	.844	.156
40–44	.882	.118
45–49	.926	.074
50–54	.976	.024
55–59	.988	.012
60–64	.994	.006
65–69	1.000	.000

Results

Epidemiology

Age at onset.—Each of the families was ascertained through exactly one proband. Of the 160 probands, 53 were male and 107 were female. This gender difference is merely a reflection of the membership of the National Vitiligo Foundation as determined by the first names of members. Among male probands, the range of age at onset was 4–60 years; among females the range was 1–69 years. The mean ages at onset among male and females are given in table 1. It is seen that male probands, on average, manifested vitiligo ~7 years earlier than did female probands. To test whether the frequency distributions of age at onset are the same for males and females, we performed a nonparametric run test (Chakravarti et al. 1967). The test statistic, T, is $T = [U - E(U)]/\sqrt{V(U)}$, where U = observed number of runs, $E(U)$ = expected number of runs, and $V(U)$ = variance of the number of runs. For the present data, the values of U, $E(U)$, $V(U)$ and T are, respectively, 63, 71.2, 30.6, and −1.48. Since the observed value of $T > -1.65$ (the cutoff point corresponding to the 5% tail probability of a standard normal distribution), the null hypothesis that there is no difference in the frequency distributions of age at onset in males and females was accepted. The pooled cumulative distribution of age at onset is presented in table 2.

In the set of 160 families, there were 48 additional first-degree relatives of probands who also had vitiligo. Of these 48 individuals, 20 were male, and 28 were female. Therefore, including the probands, in the 160 families, there were 73 male and 135 female vitiligo patients. The mean ages at onset among all affected individuals are also given in table 1. (Data on age at onset were unavailable for four male and one female affected first-degree relatives of probands.) These mean ages at onset are nearly identical to the ages at onset observed among probands only.

Family history.—Since the present data pertain to first-degree relatives of each proband, the percentage of families with positive family history is an indication of the extent of familial aggregation. This information is provided in table 3, from which it is seen that ~20% of probands have at least one affected first-degree relative.

Relative risk (RR).—To determine whether being biologically related to an affected individual results in an increased risk of being affected with vitiligo, compared with an individual of the same age and gender drawn randomly from the population, we computed the population RR (Penrose 1953; Weiss et al. 1982). The population RR, also known as the "standardized mortality ratio" (Breslow and Day 1980), is defined as n_a/n_e, where n_a = observed number of affected individuals; n_e = expected number of affected individuals = $\sum_{i=1}^{G} N_i \cdot p_i$, where G = number of age groups; N_i = total number of individuals in ith age group ($i = 1, 2, \ldots, G$); p_i = prevalence of vitiligo in ith age group ($i = 1, 2, \ldots, G$). For the present study, we have computed RR values for first-degree relatives of probands. Since age- and gender-specific estimates of population

Table 3

Probands with at Least One Affected First-Degree Relative

Gender of Proband	No. (%) of Probands with at Least One First-Degree Relative
Male (n=53)	11 (20.75)
Female (n=107)	20 (18.69)
Total (n=160)	31 (19.37)

Am. J. Hum. Genet. 55:981–990, 1994

Table 4

Risk of Vitiligo among Relatives of Probands

Relationship to Proband	Total No.	Observed No. of Affecteds	Relative Risk		95% Confidence Interval of Relative Risk	
			(1)[a]	(2)[b]	(1)[a]	(2)[b]
Father	155	7	7.20	10.41	2.89–14.84	4.19–21.45
Mother	149	7	7.16	7.93	2.88–14.76	3.19–16.33
Brother	157	8	13.36	11.04	5.77–26.32	4.76–21.74
Sister	152	9	11.68	10.77	5.34–22.17	4.93–20.45
Son	106	9	35.27	19.21	16.13–66.95	8.78–36.47
Daughter	108	11	37.61	20.07	18.77–67.29	10.02–35.90

[a] (1) is based on age- and gender-specific prevalence estimates given by Howitz et al. (1977).
[b] (2) is based on age- and gender-specific prevalence estimates given by Das et al. (1985).

prevalence of vitiligo among U.S. Caucasians are, to the best of our knowledge, unavailable, we have additionally used two available sets of estimates derived from the (Caucasian) Danish (Howitz et al. 1977) and Indian (Das et al. 1985) epidemiological studies, to investigate the variation in RRs that is caused by using different prevalence rates. The confidence intervals of RR estimates were obtained using the method given by Ulm (1990). The observed numbers of affected individuals, the estimates of RR, and their confidence intervals are given in table 4.

Offspring of probands are found to have the highest risk of developing vitiligo, followed by siblings and parents. There is a remarkable similarity in estimates of RR between genders within any relationship category. It is also interesting to note that the estimates of RR are similar whether the Danish prevalence data or the Indian prevalence data are used. From the confidence intervals of RRs, it is seen that the value of unity is outside the confidence intervals for all first-degree relatives of probands. Since RR = 1 implies no elevation of risk for developing vitiligo, and since the value of unity is excluded from the confidence intervals, it can be concluded that there is a significant elevation of risk of developing vitiligo if one is biologically related to a vitiligo patient. This is indicative of strong familial aggregation, which may be due to the sharing of genetic factors and/or exposure to common familial environmental factors.

Correlations among ages at onset in affected relatives.—Clues regarding the possible etiologic heterogeneity can be obtained from the distribution of ages at onset among members in families. Presence of many young affected individuals in a family may be indicative of genes segregating, compared with another family in which there are a few young affected individuals and many older affected individuals. Put differently, it is possible that most affected individuals in a family in which there is a segregating gene may have low age at onset; while in those families in which there is no gene segregating, ages at onset in affected individuals may be highly variable. One would, therefore, expect that there will be a strong intraclass correlation in the ages at onset in affected individuals within those families in which there are segregating genes. If this is indeed the case, then it may be possible to investigate etiologic heterogeneity and to classify families into "genetic" and "nongenetic" categories based on ages at onset in affected individuals in the families.

To investigate this, we calculated intraclass correlation coefficients of ages at onset of vitiligo for affected first-degree relatives of probands. By the analysis-of-variance technique, the intraclass correlation coefficient was estimated to be .61. This estimated value is only moderate, indicating that the classification of families into subgroups based on ages at onset may not be possible with a high probability of success.

Genetic Models

Previous family studies have shown that vitiligo does not segregate in a simple Mendelian fashion, even when the variation in age at onset of the disorder is taken into account (Hafez et al. 1983; Das et al. 1985; Majumder et al. 1988). It is known that the number of possible models for a multilocus system is large (Hartl and Maruyama 1968), which precludes the exhaustive investigation of all possible models. On the basis of our previous study (Majumder et al. 1988) and the findings of recent molecular genetic studies on pigmentation, we have considered two multilocus recessive models in the present study. These models are described below.

Model I.—An individual is affected if the individual is a recessive homozygote at *all* the loci involved in the pathogenesis of the disorder. The loci are assumed to be autosomal, unlinked, and diallelic. For example, if the disorder is caused by the action of L unlinked loci and at each locus there are two alleles—A,a; B,b; C,c; etc. (a, b, c, . . . denoting the recessive alleles)—affected individuals are of genotype aabbcc . . . ; individuals of all other genotypes are phenotypically normal. Thus, of the 3^L genotypes, only one genotype leads to the affected phenotype; individuals

of the remaining $3^L - 1$ genotypes are phenotypically normal. If q denotes the frequency of the recessive allele at each locus, and if each locus is assumed to be in Hardy-Weinberg equilibrium, then the frequency of the trait in the general population is $\delta = q^{2L}$. Thus, given the prevalence δ, the allele frequency q can be obtained as $q = \delta^{1/2L}$, for any specific value of the number of loci, L. This model was introduced by Li (1987), who also derived many population characteristics of this model. Methods of segregation analysis of some types of family data under this model have been developed by Majumder et al. (1988, 1989). Some additional properties of this model have been derived by Majumder and Nath (1992). This model yielded an adequate fit to the family data collected earlier by us from Calcutta (Majumder et al. 1988). Obviously, this is the model that is being considered from cross-validation in the present study.

Model II.—An individual is affected if the individual is a recessive homozygote at *any one* of the L loci involved. In this case, of the 3^L genotypes, $3^L - 2^L$ genotypes lead to the affected phenotype; the remaining 2^L genotypes lead to the normal phenotype. Thus, for $L = 2$, individuals of genotypes AAbb, Aabb, aaBB, aaBb, and aabb are phenotypically affected, and those of genotypes AABB, AABb, AaBB, and AaBb are phenotypically normal. The frequency of the trait in the general population is $\delta = 1 - (1 - q^2)^L$. Therefore, $q = \sqrt{1 - (1 - \delta)^{1/L}}$.

Under each of these models, the conditional likelihood of data on offspring, given the parental mating type, ages of parents, and method of ascertainment, can be derived analytically for fixed values of L and q. It must be noted that data on offspring comprise numbers of affected and normal individuals in the various age groups under consideration. In the present study, only two types of families were observed: (i) families in which one parent is normal and the other affected, each family ascertained through the affected parent and (ii) families in which both parents are normal, each family ascertained through an affected offspring. The derivations of the likelihood functions for these two types of families are given in the appendix, along with a description of a simplified computation of likelihoods.

In the present study, we have calculated values of the likelihood functions for the two types of families under each of the two models for different values of L, under the assumption that $\delta = .005$. For this value of δ, under model I, the recessive allele frequency, q, turns out to be .0707 (for $L = 1$), .2659 (for $L = 2$), .4135 (for $L = 3$), and .5157 (for $L = 4$). Under model II, $q = .0707$ (for $L = 1$) and .0500 (for $L = 2$). We note that when $L = 1$, q is the same for both models I and II.

Results of Segregation Analysis

Under model I, we have calculated the value of the log-likelihood function of the data separately for normal × affected families ascertained through an affected parent and for normal × normal families ascertained through an affected offspring. The value of z_i, which denotes the probability that an individual of a susceptible genotype belonging to the ith age group is unaffected, which is required in the likelihood computations, was obtained from the data on age at onset in probands. (We note that if an individual is of a susceptible genotype, the individual will, if he or she lives long enough, eventually become affected.) These values are given in table 2. Likelihood computations were performed for various values of the number of loci, L. Although each normal × normal family had exactly one proband, indicating that the ascertainment probability $\pi \approx 0$ (single selection), we have performed computations for different values of π. The results are presented in table 5. It is seen from this table that the likelihood of the data increases with number of loci. However, the rate of increase decreases with increasing values of L. For normal × normal families, the likelihood values for various values of π are not very different for a fixed value of L.

In calculating the joint likelihood of all families, we have used the values corresponding to $\pi \sim 0$. The joint-likelihood values are presented in the last column of table 5. From these values it is seen that the data are $\sim 10^{13}$ times (log-likelihood ratio = 12.86) more likely if two loci are involved than if only one locus is involved, $\sim 10^6$ times more likely (log-likelihood ratio = 5.79) if three loci are involved, but only ~ 18 times more likely (log-likelihood ratio = 1.25) if four loci are involved. The difference in likelihoods between the three-locus and the four-locus models seems to be nonsignificant at the 5% level, indicating that the three-locus model is the most parsimonious.

Similar results for model II are presented in table 6. For brevity, we have presented results only for $L = 1$ and $L = 2$. As is expected, the joint likelihood under this model when $L = 1$ is exactly the same as the corresponding value under model I. When $L = 2$, the likelihood of the data under model II is $\sim 10^9$ times less likely (log-likelihood ratio = −8.42) than it is under model I. Similar results are obtained when the value of L is increased. Thus, it is clear that between models I and II, model I is more parsimonious.

We therefore conclude that three epistatically interacting autosomal diallelic loci are involved in the pathogenesis of vitiligo. Individuals afflicted with the disorder are those who are recessive homozygotes at each of these three loci. These results obtained from U.S. Caucasians are completely consistent with the results obtained earlier, by Majumder et al. (1988), from a vastly different geographic and ethnic setting, i.e., the Hindus of India.

Robustness of Inferences

There are two potential sources of variation that may affect the results and inferences derived in the previous section. These are (i) variation in estimates of z_i's presented in table 2 and (ii) variation in the estimate of δ. We have investigated the effects of both these sources of variation

Table 5

Log-Likelihood Values under Model I, for Different Numbers of Loci, for Data of 86 Normal × Affected Families Ascertained through an Affected Parent, and 61 Normal × Normal Families Ascertained through an Affected Child

No. of Loci	Normal × Affected Families	Normal × Normal Families			All Families
		$\pi = .01$	$\pi = .05$	$\pi \approx 0$	
1	−38.1338	−73.8925	−72.8980	−74.1403	−112.2741
2	−28.2017	−70.9667	−69.9722	−71.2144	−99.4161
3	−23.0454	−70.3291	−69.3346	−70.5768	−93.6222
4	−22.0301	−70.0961	−69.1016	−70.3439	−92.3740

on the values of the likelihood functions and on the inferences derived earlier.

Effect of variation in age-specific affection probabilities.—To study this effect, we first computed the standard deviation, s_i, of the estimate of z_i for each age group i (1, 2, . . . , G). Then, for each i (1, 2, . . . , G − 1), we derived a new estimate of z_i by drawing a random number from the uniform distribution $U[z_i - s_i, z_i + s_i]$. The new estimate of z_G was taken to be .0, since the cumulative proportion to the last age group G must equal 1. Further, for obvious reasons, for the first age group (i.e., $i = 1$), the new estimate of z_1 was actually derived by drawing a random number from $U[0, z_1 + s_1]$.

Having obtained a new set of estimates of z_i (i = 1, 2, . . . , G), we computed the likelihood functions under the assumption that $\pi \approx 0$, separately for normal × affected families and for normal × normal families, under model I for L = 1, 2, 3, and 4. This procedure was repeated five times. The mean ± SD of the values of the joint log-likelihood function (i.e., based on data of normal × affected and normal × normal families pooled) for L = 1, 2, 3, and 4 are, respectively, −112.3581 ± 0.6976, −97.7132 ± 0.7737, −93.9385 ± 0.7782, and −92.6724 ± 0.7647. When these values are compared, with the values presented in the last column of table 5, it is seen that these mean values are nearly identical. It was also found that model I is more parsimonious than model II, and the magnitudes of the likelihood ratios are similar to those reported in the previous section. (For brevity, numerical results are not presented.)

Effect of variation in prevalence.—To investigate the effect of variation in prevalence, we computed the joint-likelihood functions, under the assumption that $\pi \sim 0$, for $\delta = .004$ and .01. For $\delta = .004$, the values of the likelihood function under model I for L = 1, 2, 3, and 4 are −112.6346, −97.6292, −93.7929, and −92.5413, respectively. The corresponding values for $\delta = .01$ are −111.1756, −96.8335, −93.1913, and −91.9634. From these values and from the values presented in the last column of table 5, we find that there is no significant effect of variation in δ on the values of the likelihood function for a fixed value of L. Similar results are also obtained for model II. (For brevity, numerical results are not presented.) Hence, the previous inferences on genetic models remain unaltered, even after allowance for variation in δ within an appropriate range.

Discussion

The present study, which to the best of our knowledge is the first extensive family study on vitiligo in the United States, was undertaken with the primary aim of cross-validating a genetic model proposed by us earlier. The method adopted in the present study, in which new data have been collected and model parameters estimated from the fresh data, is the most reliable method of cross-validation. In analogy with a double-blind study, this method has been termed "double cross-validation" (Mosteller and Tukey 1977). In performing the genetic analyses, we have also extended the relevant statistical methodology relating to

Table 6

Log-Likelihood Values under Model II, for Different Numbers of Loci, of Data of 86 Normal × Affected Families Ascertained through an Affected Parent, and 61 Normal × Normal Families Ascertained through an Affected Child

No. of Loci	Normal × Affected Families	Normal × Normal Families			All Families
		$\pi = .01$	$\pi = .05$	$\pi \approx 0$	
1	−38.1338	−73.8925	−72.8980	−74.1403	−112.2741
2	−33.3822	−74.2087	−73.2132	−74.4554	−107.8376

analysis of family data on a complex disorder with variable onset age, which was developed by us earlier (Majumder et al. 1988, 1989).

We have analyzed the data from an epidemiological perspective. Some of our epidemiological findings agree with those obtained in two previous population studies (Howitz et al. 1977; Das et al. 1985), while some are at variance. Our finding that males have a lower mean onset age than do females differs from that obtained in the previous studies. This perhaps indicates that the notion that vitiligo is related to puberty and that females show an earlier onset because they attain puberty earlier than males might be incorrect. About 20% of probands have at least one affected first-degree relative. Risks of developing vitiligo for all types of first-degree relatives of probands were found to be significant. This indicates that there is significant familial aggregation, which is in agreement with the findings of previous studies. No significant familial clustering of affected individuals with early ages at onset was detected, indicating that age at onset cannot be used as a criterion for classifying families into subgroups. We have also reported earlier (Majumder et al. 1993) that the various alleged correlates of vitiligo—such as parental age at first childbirth, stress, thyroid disorders, suntanning, use of pesticides, etc.—are not significantly associated with vitiligo in the present data set.

We have performed a segregation analysis of the data on nuclear families. Likelihood functions have been analytically derived, taking into account the complexities arising out of delayed age at onset and biases of ascertainment. The present derivation is an extension of the methods presented in our previous studies (Majumder et al. 1988, 1989) and takes into account the possibility of parents and offspring being of the susceptible genotype(s) but of not having manifested the disorder at the time of study. The results of the present segregation analysis are in complete agreement with those found in our earlier study (Majumder et al. 1988). In the present study, we have not only cross-validated our previous model, which postulates that recessive alleles at multiple unlinked loci act epistatically in the manifestation of vitiligo, but we have, in fact, additionally shown that this model is more parsimonious than a competing model in which a recessive allele at any one of a number of unlinked loci is stated to produce the disorder. The competing model considered was chosen by taking into consideration the biochemical and molecular picture of melanogenesis. In the present study the best estimate of the number of loci involved in the pathogenesis of vitiligo is three, while this estimate was four in our previous study. This, we believe, is not a serious difference and may have occurred because in our previous study we had not taken into account the possibility—albeit small, because most parents were >40 years of age—that an unaffected parent may be of the susceptible genotype. While we have not explicitly considered the possibility of the existence of nongenetic (sporadic) cases in our models, we do not believe the standard assumption that most probands in single-case families (80%) are nongenetic cases. For the type of multilocus models considered, it is not possible to estimate the proportion of nongenetic cases through statistical analyses; refinement of methods of clinical evaluation is necessary to identify nongenetic cases, if there are any.

We have also performed a statistical study of robustness to investigate the effect of variation of certain parameters on our inferences. We have found that our results are quite robust with respect to variations in prevalence and age-specific probabilities of affection.

Acknowledgments

This study was supported in part by the National Vitiligo Foundation, Tyler, TX. We are grateful to Prof. Robert C. Elston for useful comments on an earlier version of this paper.

Appendix

Derivation of Likelihood Functions

Preliminaries and Notations

Under the multilocus models considered, an individual of a given phenotype may potentially be of any one of several genotypes. For example, under model I, a phenotypically normal individual can be of any one of eight genotypes (AABB, AABb, AAbb, AaBB, AaBb, Aabb, aaBB, and aaBb) if two loci are considered, while under model II, such an individual can be of any one of four genotypes (AABB, AABb, AaBB, and AaBb). Although under model I an affected individual is necessarily of genotype aabb if two loci are involved, under model II such an individual can be of any one of five genotypes (aaBB, aaBb, aabb, AAbb, and Aabb). Late age at onset adds to the list of potential genotypes of normal individuals. For example, under model I, a normal individual may also be of genotype aabb but may not have manifested the disorder at the age at examination.

When two loci are involved, we present in table A1 the list of various possible genotypic matings, mating probabilities, phenotypic mating types, and segregation probabilities, separately for models I and II. In the listing of the phenotypic mating types in this table, the possibility that an individual may be of the susceptible genotype (e.g., aabb under model I) but may not have manifested the disorder because late age at onset has not been taken into account. This possibility introduces a complexity. For example, under model I, when variable age at onset is considered, the genotypic mating AABB × aabb may phenotypically either be normal × affected or normal × normal. Extension to a larger number of loci is made in analogous manner.

For the formulation of likelihood of phenotypic observations on offspring in light of the parental phenotypic

Table A1

Genotypic and Phenotypic Mating Types, Mating Probabilities, and Segregation Probabilities When Two Loci Are Involved

Serial No.	Genotypic Mating Type	Mating Probability	Segregation Probability		Phenotypic Mating Type	
			Model I	Model II	Model I	Model II
1	AABB × AABB	p^8	0	0	Normal × normal	Normal × normal
2	AABB × AABb	$4p^7q$	0	0	Normal × normal	Normal × normal
3	AABB × AAbb	$2p^6q^2$	0	0	Normal × normal	Normal × affected
4	AABB × AaBB	$4p^7q$	0	0	Normal × normal	Normal × normal
5	AABB × AaBb	$8p^6q^2$	0	0	Normal × normal	Normal × normal
6	AABB × Aabb	$4p^5q^3$	0	0	Normal × normal	Normal × affected
7	AABB × aaBB	$2p^6q^2$	0	0	Normal × normal	Normal × affected
8	AABB × aaBb	$4p^5q^3$	0	0	Normal × normal	Normal × affected
9	AABB × aabb	$2p^4q^4$	0	0	Normal × affected	Normal × affected
10	AABb × AABb	$4p^6q^2$	0	1/4	Normal × normal	Normal × normal
11	AABb × AAbb	$4p^5q^3$	0	1/2	Normal × normal	Normal × affected
12	AABb × AaBB	$8p^6q^2$	0	0	Normal × normal	Normal × normal
13	AABb × AaBb	$16p^5q^3$	0	1/4	Normal × normal	Normal × normal
14	AABb × Aabb	$8p^4q^4$	0	1/2	Normal × normal	Normal × affected
15	AABb × aaBB	$4p^5q^3$	0	0	Normal × normal	Normal × affected
16	AABb × aaBb	$8p^4q^4$	0	1/4	Normal × normal	Normal × affected
17	AABb × aabb	$4p^3q^5$	0	1/2	Normal × affected	Normal × affected
18	AAbb × AAbb	p^4q^4	0	1	Normal × normal	Affected × affected
19	AAbb × AaBB	$4p^5q^3$	0	1/2	Normal × normal	Affected × normal
20	AAbb × AaBb	$8p^4q^4$	0	1/2	Normal × normal	Affected × normal
21	AAbb × Aabb	$4p^3q^5$	0	1/2	Normal × normal	Affected × affected
22	AAbb × aaBB	$2p^4q^4$	0	1	Normal × normal	Affected × affected
23	AAbb × aaBb	$4p^3q^5$	0	1	Normal × normal	Affected × affected
24	AAbb × aabb	$2p^2q^6$	0	0	Normal × affected	Affected × affected
25	AaBB × AaBB	$4p^6q^2$	0	1/4	Normal × normal	Normal × normal
26	AaBB × AaBb	$16p^5q^3$	0	1/4	Normal × normal	Normal × normal
27	AaBB × Aabb	$8p^4q^4$	0	1/4	Normal × normal	Normal × affected
28	AaBB × aaBB	$4p^5q^3$	0	1/2	Normal × normal	Normal × affected
29	AaBB × aaBb	$8p^4q^4$	0	1/2	Normal × normal	Normal × affected
30	AaBB × aabb	$4p^3q^5$	0	1/2	Normal × affected	Normal × affected
31	AaBb × AaBb	$16p^4q^4$	1/16	7/16	Normal × normal	Normal × normal
32	AaBb × Aabb	$16p^3q^5$	1/8	1/2	Normal × normal	Normal × affected
33	AaBb × aaBB	$8p^4q^4$	0	1/2	Normal × normal	Normal × affected
34	AaBb × aaBb	$16p^3q^5$	1/8	5/8	Normal × normal	Normal × affected
35	AaBb × aabb	$8p^2q^6$	1/4	3/4	Normal × affected	Normal × affected
36	Aabb × Aabb	$4p^2q^6$	1/4	1	Normal × normal	Affected × affected
37	Aabb × aaBB	$4p^3q^5$	0	1/2	Normal × normal	Affected × affected
38	Aabb × aaBb	$8p^2q^6$	1/4	3/4	Normal × normal	Affected × affected
39	Aabb × aabb	$4pq^7$	1/2	1	Normal × affected	Affected × affected
40	aaBB × aaBB	p^4q^4	0	1	Normal × normal	Affected × affected
41	aaBB × aaBb	$4p^3q^5$	0	1	Normal × normal	Affected × affected
42	aaBB × aabb	$2p^2q^6$	0	1	Normal × affected	Affected × affected
43	aaBb × aaBb	$4p^2q^6$	1/4	1	Normal × normal	Affected × affected
44	aaBb × aabb	$4pq^7$	1/2	1	Normal × affected	Affected × affected
45	aabb × aabb	q^8	1	1	Affected × affected	Affected × affected

mating type, the following further preliminaries and notations are in order:

1. For a particular phenotypic mating type, several genotypic mating types are possible. If the mating involves a parent(s) who is (are) phenotypically normal, then the current age(s) of the parent(s) also needs to be taken into consideration in the enumeration of the possible genotypic matings. We shall denote as g_f and g_m the current ages of the father and mother, respectively.

2. We shall define z_x as Prob{an individual of age x is phenotypically normal, given that he or she is of the susceptible genotype(s)}. These probabilities are estimated from age-at-onset data of affected individuals. In practice, it may be necessary to form age groups to avoid vagaries of small sample sizes. In the present study, this has been done. When, from practical considerations, age groups are formed, z_i will denote the above conditional probability for an individual belonging to age group i; $i = 1, 2, \ldots, G$.

3. We shall define μ_k as Prob {genotypic mating type is k, given the phenotypic mating type and age(s) of the phenotypically normal parent(s)}; k = 1, 2, ... K = number of genotypic matings for a specified phenotypic mating. These are calculated straightforwardly from the mating probabilities given in table A1. However, these probabilities need to be multiplied by appropriate z_i values in specific cases. For example, under model I, given a normal × affected mating, K should equal eight (corresponding to serial numbers 9, 17, 24, 30, 35, 39, 42, and 44 of table A1) if the disorder expresses itself at birth. However, when the disorder has a late age at onset, a normal × affected mating may also be of type aabb × aabb (serial number 45 of table A1). Thus, K = 9. The mating probability of the aabb × aabb mating, given that the phenotypic mating type is normal × affected and that the normal individual belongs to the ith age group, is $q^8 z_i$. The conditional probabilities, μ_k's, are obtained by dividing the unconditional probabilities by the sum of the unconditional probabilities of all genotypic matings corresponding to the given phenotypic mating.

4. We shall define θ_k as Prob {offspring is of a susceptible genotype, given that the parental genotypic mating is of type k}. For example, under model I, θ_k = Prob {offspring is of genotype aabbcc . . . , given that the parental genotypic mating is of type k}. But under model II, θ_k = Prob {offspring is of AAbb, Aabb, aaBB, aaBb, or aabb, given that the parental genotypic mating type is k}. These are also given in table A1.

5. Consider an offspring of age x in a family in which parental genotypic mating is of type k. The probability of this offspring being phenotypically affected is $\theta_k(1 - z_x)$, and that of being phenotypically normal is $1 - \theta_k + \theta_k z_x = 1 - \theta_k(1 - z_k)$.

6. For a particular nuclear family, we shall define n_i as the total number of offspring in age group i, m_i as the number of affected offspring in age group i, and $n_i - m_i$ as the number of normal offspring in age group i.

Likelihood Function for a Normal × Affected Family, Ascertained through an Affected Parent

The data comprise numbers of affected offspring belonging to each of the G age groups, that is, m_i and $n_i - m_i$; i = 1, 2, . . . , G. Given that parents are normal × affected, one can enumerate all possible genotypic matings that can give rise to a normal × affected phenotypic mating, under either model I or model II. Suppose K such genotypic matings are possible. For each genotypic mating, k, the conditional mating probability μ_k can be worked out as indicated in the previous section after the age of the normal parent is taken into account. For a given genotypic mating, k, the likelihood of phenotypic observations of offspring belonging to age group i is $\binom{n_i}{m_i}[\theta_k(1 - z_i)]^{m_i}[1 - \theta_k(1 - z_i)]^{n_i - m_i}$. Thus, the conditional likelihood function of phenotypic observations on all offspring, given the parental mating type, is $L = \sum_{k=1}^{K} \mu_k \prod_{i=1}^{G} \binom{n_i}{m_i}[\theta_k(1 - z_i)]^{m_i}[1 - \theta_k(1 - z_i)]^{n_i - m_i}$.

Likelihood Function for a Normal × Normal Family, Ascertained through an Affected Offspring

In comparison with the previous case, a normal × affected family ascertained through an affected offspring raises two problems. First, the ages of both normal parents have to be considered in determining μ_k's. For example, under model I, for L = 2, a normal × normal mating may actually be of type aabb × aabb. That is, both parents can be of the susceptible genotype (aabb), without manifesting the disorder at the time of data collection. The unconditional probability of this genotypic mating will be $q^8 z_i z_j$, when the parents belong to age groups i and j (i, j = 1, 2, . . . , G). Second, while no correction for bias of ascertainment was required in the previous case (normal × affected family ascertained through an affected parent), when a family is ascertained through an affected offspring, the likelihood has to be corrected for ascertainment bias. Correction for ascertainment bias can be made following Elandt-Johnson (1971) and Majumder et al. (1988). The likelihood function, L, can be written as $L = [\alpha_m \cdot \lambda(\mathbf{n}, \mathbf{m})]/\beta(\mathbf{n}, \mathbf{m})$,where the form of the function $\lambda(\mathbf{n}, \mathbf{m})$ is the same as the likelihood function of the previous case. (Of course, enumeration of genotypic matings and calculation of conditional mating probabilities will correspond to a normal × normal phenotypic mating rather than to a normal × affected mating.) If $m = \sum_{i=1}^{G} m_i$ denotes the total number of affected offspring in the family, then α_m = Prob {a family with r affected offspring will have at least one proband} $= 1 - (1 - \pi)^m$, where π denotes the probability of ascertainment. Thus, the numerator of L, $\alpha_m \cdot \lambda(\mathbf{n}, \mathbf{m})$, denotes the likelihood that in a family with n_i offspring there will be m_i affected in age group i (i = 1, 2, . . . , G) and that such a family will be ascertained. The denominator of L, $\beta(\mathbf{n}, \mathbf{m})$, denotes the probability that a family with n_i offspring in age group i has at least one affected offspring and is ascertained. This term is obtained as:

$$\beta(\mathbf{n}, \mathbf{m}) = \sum_{r=1}^{N} \left[\alpha_r \cdot \sum_{\substack{l=(l_1, l_2, \ldots, l_G) \\ l_i \le n_i \\ \sum l_i = r}} \lambda(\mathbf{n}, \mathbf{l})\right],$$

where $N = \sum_{i=1}^{G} n_i$. When $\pi \approx 0$, the likelihood function simplifies to $L = r \cdot \lambda(\mathbf{n}, \mathbf{m})/\beta(\mathbf{n}, \mathbf{m})$, where $\beta(\mathbf{n}, \mathbf{m}) = \sum_{r=1}^{N} r \cdot \sum \lambda(\mathbf{n}, \mathbf{l})$, and the range and constraints of the second summation are those of $\beta(\mathbf{n}, \mathbf{m})$, given earlier.

Computations of Likelihood Functions: Some Comments

The number of possible genotypic matings, K, for a given phenotypic mating type increases drastically with increase in the number of loci, L. For a fixed value of L, K is also much larger if a disorder has a variable age at onset

compared with one that is expressed at birth. Thus, for a disorder with a variable age at onset, the number of terms to be summed in the likelihood function is usually large. However, several genotypic matings have the same segregation probability, as is evident from table A1. Considerable computational simplification is obtained by pooling all genotypic matings with the same segregation probability. When this is done, the number of terms to be summed in the likelihood function reduces to the number of distinct values of the segregation probability. (For further details, see Majumder et al. [1989].)

When data on a number of nuclear families of a specific mating type are available, the joint likelihood is the product of likelihoods of individual families. Here again, considerable computational simplification is obtained by pooling data of all families in which the normal parent(s) belongs to the same age group(s).

References

Breslow NE, Day NE (1980) The analysis of case control studies. Vol 1 in: Statistical methods in cancer research. IARC, Lyon

Chakravarti IM, Laha RG, Roy J (1967) Techniques of computation, descriptive methods and statistical inference. Vol 1 in: Handbook of methods of applied statistics. John Wiley & Sons, New York

Das SK, Majumder PP, Chakraborty R, Majumdar TK, Haldar B (1985) Studies on vitiligo. I. Epidemiological profile in Calcutta, India. Genet Epidemiol 2:71–78

Dizier M-H, Bonaiti-Pellie C, Clerget-Darpouz F (1993) Segregation analysis of diseases depending on the interactive effect of two genes. Am J Hum Genet Suppl 53:793

Elandt-Johnson RC (1971) Probability models and statistical methods in genetics. John Wiley & Sons, New York

El-Mofty AM (1968) Vitiligo and psoralens. Pergamon, New York

Goldin LR, Weeks DE (1993) Two-locus models of disease: comparison of likelihood and non-parametric linkage methods. Am J Hum Genet Suppl 53:1006

Hafez M, Sharaf L, El-Nabi SMA (1983) The genetics of vitiligo. Acta Derm Venereol (Stockh) 63:249–251

Hartl DL, Maruyama T (1968) Phenogram enumeration: the number of regular genotype-phenotype correspondences in genetic systems. J Theor Biol 20:129–163

Hearing VJ (1993) Unraveling the melanocyte. Am J Hum Genet 52:1–7

Hearing VJ, Tsukamoto K (1991) Enzymatic control of pigmentation in mammals. FASEB J 5:2902–2909

Howitz J, Brodthagen H, Schwartz M, Thomsen K (1977) Prevalence of vitiligo: epidemiological survey of the isle of Bornholm, Denmark. Arch Dermatol 113:47–52

Lerner AB (1959) Vitiligo. J Invest Dermatol 32:285–310

Li CC (1987) A genetical model for emergenesis: in memory of Laurence H. Snyder, 1901–86. Am J Hum Genet 41:517–523

Majumder PP (1989) Strategies and sample-size considerations for mapping a two-locus autosomal recessive disorder. Am J Hum Genet 45:412–423

——— (1993) Segregation analysis in the understanding of complex human disorders: some recent developments. In: Majumder PP (ed) Human population genetics: a centennial tribute to J. B. S. Haldane. Plenum, New York, pp 139–152

Majumder PP, Das SK, Li CC (1988) A genetical model for vitiligo. Am J Hum Genet 43:119–125

Majumder PP, Nath SK (1992) Statistical analysis of family data on complex disorders in man. J Genet 71:89–103

Majumder PP, Nordlund JJ, Nath SK (1993) Pattern of familial aggregation of vitiligo. Arch Dermatol 129:994–998

Majumder PP, Ramesh A, Chinnappan D (1989) On the genetics of prelingual deafness. Am J Hum Genet 44:86–99

Mosher DB, Fitzpatrick TB, Ortonne JP (1979) Abnormalities of pigmentation. In: Fitzpatrick TB, Eisen AZ, Wolff K, Freedberg IM, Austen KF (eds) Dermatology in general medicine. McGraw-Hill, New York, pp 582–590

Mosteller F, Tukey JW (1977) Data analysis and regression. Addison-Wesley, Reading, PA

Neuman RJ, Rice JP (1992) Two-locus models of disease. Genet Epidemiol 9:347–365

Nordlund JJ (1987) Hypopigmentation, vitiligo and melanoma: new data, more enigmas. Arch Dermatol 123:1005–1011

Penrose LS (1953) The genetical background of common diseases. Acta Genet 4:257–265

Quevado WC Jr, Fitzpatrick TB, Szabo G, Jimbow K (1987) Biology of melanocytes. In: Fitzpatrick TB, Eisen AZ, Wolff K, Freedberg IM, Austen KF (eds) Dermatology in general medicine, 3d ed. McGraw-Hill, New York, pp 224–251

Rice JP, Neuman RJ, Burroughs TE, Hampe CL, Daw EW, Suarez BK (1993) Linkage analysis for oligogenic traits. Am J Hum Genet Suppl 53:66

Schork NJ, Boehnke M, Terwilliger JD, Ott J (1993) Two-trait-locus linkage analysis: a powerful strategy for mapping complex genetic traits. Am J Hum Genet 53:1127–1136

Ulm K (1990) A simple method to calculate the confidence interval of standardized mortality ratio (SMR). Am J Epidemiol 131:373–376

Vieland VJ, Hodge SE, Greenberg DA (1992) Adequacy of single-locus approximations for linkage analyses of oligogenic traits. Genet Epidemiol 9:45–59

Weiss KM, Chakraborty R, Majumder PP, Smouse P (1982) Problems in the assessment of relative risk of chronic disease among biological relatives of affected individuals. J Chronic Dis 35:539–553

Wick MM, Hearing VJ, Rorsman H (1987) Biochemistry of melanization. In: Fitzpatrick TB, Eisen AZ, Wolff K, Freedberg IM, Austen KF (eds) Dermatology in general medicine, 3d ed. McGraw-Hill, New York, pp 251–258