GENETIC EPIDEMIOLOGICAL ANALYSIS OF COMPLEX DISORDERS IN MAN WITH SPECIAL REFERENCE TO VITILIGO

SWAPAN KUMAR NATH

A thesis submitted
to the Indian Statistical Institute
in the partial fulfilment of
the requirements for the
award of the degree of
Doctor of Philosophy
1994

ACKNOWLEDGEMENTS

It gives me a great pleasure to express my deep sense of gratitude to Professor Partha P. Majumder, my supervisor, who introduced me to the field of genetic epidemiology. He not only supervised me but he was also a constant source of inspiration and encouragement to me during the whole course of this research work.

I am, especially grateful to Professor James J. Nordlund, Chairman, Department of Dermatology, University of Cincinnati Medical Center, Ohio, U.S.A., for kindly granting permission to use the family data on vitiligo as a part of this thesis work, which were collected by him jointly with Professor Partha P. Majumder.

I am, indebted to Dr. M. K. Das, Dr. K. Das, Dr. P. Dasgupta, Dr. P. Banerjee, Mr. S. Shete, Mr. A. K. Ghosh, Mr. U. K. Mukherjee for their help and encouragement. I record my indebtedness to Mr. A. C. Naolekar providing expert advise on LaTeX— the wordprocessing software used to prepare this thesis in its present form.

The financial and logistical support provided by the Director, Dean of Studies, and Head, Anthropometry and Human Genetics Unit of The Indian Statistical Institute are gratefully acknowledged.

Finally, I must express my deep appreciation to my family — to my parents for their constant encouragement, and to my brothers and sister for their enthusiastic support. I am really hard-pressed to find the right words to express my gratitude to Miss K. Chakraborty for her love, forbearance, and above all, for her emotional support and encouragement.

November, 1994

Swapan K. Nath.

Contents

Chapter 1

Introduction

1.1 General Remarks

The present study can be broadly classified in the area of genetic epidemiology, which is conventionally defined as "a science that deals with the aetiology, distribution, and control of disease in groups of relatives and with inherited causes of disease in populations" (Morton 1982, 1993). One of the major goals of genetic epidemiology is the study of the nature and extent of clustering of a disease in families and in defined ethnic groups. The study of a disorder within the genetic epidemiological framework is performed by testing :

(1) whether the disorder clusters in families?

(2) whether observed familiality is caused by biologically inherited susceptibility factors or environmental exposure or culturally inherited risk factors?

(3) whether the genetic susceptibility factors are inherited in accordance with a specified model of inheritance?

It may be emphasized that familial aggregation, though a necessary condition, is not a sufficient one to establish that a disorder is genetic. This is because even a non-genetic, environmentally determined disorder can aggregate in families due to common familial environmental factors. However, testing for familial aggregation needs to be performed prior to any formal genetic analysis. Several tests of familial aggregation have been proposed in the literature

(Chakraborty *et al.* 1980, 1984; Hill 1980; Thomson 1980; Majumder 1982). Among these, the more widely used ones are designed to test whether :

a) prevalence of the disorder among relatives of an affected individual is higher than the prevalence among relatives of a normal individual (*relative risk*);

b) prevalence of the disorder among relatives of an affected individual is higher than the general population prevalence. *Standardized mortality ratio* (*SMR*) is commonly used. If the *SMR* is significantly greater than unity, indicating familial aggregation, then to examine whether the observed aggregation is due to genetic factors, one proceeds to test whether affected members of a family are significantly 'closer' — in the sense of kinship coefficient (Hill 1980) or some other measure of biological distance (Chakraborty *et al.* 1980; Thomson 1980) than a pair of randomly chosen affected individuals from the population;

c) age and gender specific correlation of the disorder between pairs of related affected individuals is significantly greater than zero (Chakraborty *et al.* 1984).

For many chronic disorders, an individual with a higher genetic risk often has a lower age at onset. Thus, in a high-risk family one may not observe a significant increase in the *number* of affected individuals, but an aggregation of a few early-onset cases. To take this into account, Majumder (1982) and Chakraborty *et al.* (1984) devised a test procedure to detect familial aggregation, designed broadly to detect excess prevalence of a disorder in a family.

If a disorder clusters in families more frequently than is expected by chance, then determining the reason for familial clustering can contribute both to the understanding of disease aetiology and to its control (King *et al.* 1984). There are three mechanisms that may explain familiality :

(1) a gene or genes increasing susceptibility to disease may be biologically inherited;

(2) behaviours or life styles increasing disease susceptibility may be culturally inherited; or,

(3) related individuals may be exposed together to environmental disease agents.

To discover the mode of inheritance of genetic susceptibility factors of a familial disorder, the usual approach is the fitting of Mendelian genetic models on family data. It is crucial to determine whether one or a few genes can explain a substantial fraction of the susceptibility, and if so to determine the mode(s) of inheritance of the susceptibility gene(s). The goal of segregation analysis is to determine how a susceptibility allele is inherited; whether it is X-linked or autosomal, and whether it is dominant or recessive. This general endeavour is known as 'formal genetics', a term introduced by Haldane in 1948. He defined 'formal genetics of man' as "the study of heredity and variation, based on description and counting of individuals .... enumeration of the results obtained from various matings, and deductions drawn from such enumeration ....". Further, he stated that "the first step in formal genetics is to establish that certain characters are inherited in accordance with Mendel's laws, and in particular that segregation occurs in Mendelian ratios .... The final aim, perhaps asymptotic, should be the enumeration and location of all the genes found in normal human beings". The methods of formal genetics do not distinguish between disorders and other traits (Morton 1993).

Generally, for the purpose of estimating genetic and environmental contributions to the aetiology of a disorder, two types of analyses are performed — segregation analysis and variance components or path analysis. Although these methods are complementary to each other, there are some differences between these two methods. Segregation analysis generally pays greater attention to the role of genes, especially of major genes contributing to the aetiology of a disease (Elston and Stewart 1971). Path analysis, on the other hand, pays equal emphasis to both genes and environment, but does not explicitly consider major genes effects (Li 1975).

Segregation analysis can be defined as the statistical methodology used in the analysis of family data to determine the mode of inheritance of a particular phenotype, especially with a view to uncovering Mendelian segregation at a single locus (Elston 1993). A genetic disorder which has no environmental contribution, usually expresses itself at birth and is determined by a completely penetrant allele at a single autosomal biallelic locus, is generally termed as simple Mendelian disorder. The common recognizable modes of inheritance and criteria for a simple Mendelian pattern are as follows :

Autosomal dominant

(1) Transmission continues from generation to generation, without skipping

(2) Except for mutants, every affected child has an affected parent

(3) In marriage of an affected heterozygote to a normal homozygote, the segregation probability is 1/2

(4) The two genders are affected in equal numbers.

Autosomal recessive

(1) If the trait is rare, parents and relatives, except siblings, are usually normal

(2) All children of two affected parents are affected

(3) In a marriage of two normal heterozygotes, the segregation probability is 1/4

(4) The two genders are affected in equal numbers

(5) If the trait is rare, the parental consanguinity is elevated.

Sex-linked dominant

(1) Heterozygous mothers transmit to both genders in equal frequency with a segregation probability of 1/2

(2) Hemizygous, affected males transmit the trait only to their daughters, the segregation probabilities being 1 to daughters and 0 to sons

(3) Except for mutants, every affected child has an affected parent

(4) If the trait is rare, its frequency in females is approximately twice as large as in males.

Sex-linked recessive

(1) If the trait is rare, parents and relatives, except maternal uncles and other male relatives in the female line, are usually normal

(2) Hemizygous, affected males do not transmit the trait to children of either sex, but all their daughters are heterozygous carriers

(3) Heterozygous, carrier women are normal but transmit the trait to their sons with a segregation probability of 1/2, and 1/2 of the daughters are carriers

(4) Except for mutants, every male comes from a carrier female.

1.2 What is Complex Disorder

Genetics of many human disorders are, however, more complex in nature and do not follow the simple patterns listed above. Such disorders exhibit consistent and significant familial aggregation, and have a genetic component in their aetiologies, but do not exhibit simple Mendelian patterns of inheritance. Often no single pattern of inheritance can explain all observed types of aggregation of such a disorder in families. For example, a complex 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 have no other affected member. For example, when the population prevalence of a two-locus recessive disorder is 1/1000, about 78% of nuclear families and

about 65% of three-generational extended families of the proband are simplex (Majumder 1993). These figures increase to about 82% and 78%, respectively, when the prevalence decreases to 1/10000. Segregation analysis of such a multi-locus recessive disorder often results in incorrectly inferring that the disorder is incompletely penetrant with a large proportion of sporadics (Majumder 1993). 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 to biased estimates of the recombination fraction and other segregation parameters (Majumder 1989; Vieland *et al.* 1992; Rice *et al.* 1993; Dizier *et al.* 1993; Goldin and Weeks 1993). Neuman and Rice (1992) have also shown that "no single-locus trait can fit the recurrence risks in relatives when the true model of inheritance is oligogenic." 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 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 counselling and localization of genes.

1.3 Sources of Complexity

The complexity of a genetic disorder can arise in a variety of ways. For example, a disorder may be determined by the joint action of genes and environment, such as Insulin Dependent Diabetes Mellitus (IDDM). This disorder has a variable age at onset. It aggregates in families, but does not segregate in a simple Mendelian fashion from parents to offspring (Tiwari and Terasaki 1985; Thom-

son 1988). There are also possible environmental effects or effects of other loci (*e.g.*, HLA) on the expression of this disorder. Although no claim is made that the following list is mutually exclusive or exhaustive, some of the more common causes of complexity of a disorder are :

(a) **Variable age at onset** — all individuals with the appropriate genotype do not manifest the disorder either at birth or at the same age later in life. The classic example is Huntington's disease, a degenerative disease of the nerve cells in the basal ganglia (nucleus caudatus and putamen) leading to involuntary extra pyramidal movements, personality changes, and a slow deterioration of mental abilities. Wendt and Drohm (1972) showed that the age at onset ranges from 6 years to 75+ years with the modal age at onset between 41-45 years.

(b) **Reduced penetrance** — some individuals with the appropriate genotype manifest the disorder while some others do not. Such reduced penetrance may be due to random, stochastic factors or due to modification of the susceptible genotype(s) by other gene(s) (Haldane 1941).

(c) **Phenotypic heterogeneity** — all individuals of same genotype do not manifest the same phenotype. Phenotypic heterogeneity can, however, be artifactual. For example, one of the difficult aspects of studying the genetics of psychiatric disorders relates to phenotype definition. Because of the large number of both major and minor psychiatric diagnoses, a primary problem in conducting genetic studies of psychiatric disorders is knowing which ones to include as affected, and which ones to exclude (Risch 1990). Inability to define the phenotype homogeneously may give a false indication of phenotypic heterogeneity.

(d) **Allelic/genetic heterogeneity** — different alleles either at the same locus or at the different loci, give rise to the same phenotype. In classical terms, the former type is known as intra-locus heterogeneity and the latter type is known as inter-locus heterogeneity. An example of intra-locus heterogeneity is cystic fibrosis (CF), a common hereditary disorder among Caucasians, which

is due to a recessive mutant gene on chromosome 7. In the United States, its frequency is 1/2000. Of all CF patients, 70% carry the same mutation, a deletion of a specific codon that causes the CF protein to lack an amino acid phenylalanine at amino acid position number 508. It is known as $\Delta F508$ mutation (Riordan 1989). Additionally, more than 300 mutations of the CF gene have been reported. The phenotypic effects of some of these mutations can be distinguished, while of some others seem indistinguishable. For example, the $M348K$ mutant allele, which is characterized by a T to A substitution at nucleotide position 1175 in exon 7 of the CF gene leading to a methionine to lysine amino acid substitution, is reported to have a phenotypic effect that is indistinguishable from that of $\Delta F508$ (Audrezet *et al.* 1993). Neurofibromatosis (NF), a neurological condition, can be cited as an example for which there is inter-locus heterogeneity. NF can be of two types (Riccardi and Eichner 1986). The most common form is the von Recklinghausen type (NF1), which is linked to markers on chromosome 17 (Barker *et al.* 1987). The other rare form is acoustic type (NF2), which is linked to markers on a different chromosome 22 (Seizinger *et al.* 1987).

(e) **Involvement of multiple loci** — disorder is determined by the joint action of genes at more than one locus. The genetic mechanism for a specific form of prelingual deafness may be cited as an example. This disorder manifests itself only in individuals who are recessive homozygotes at two involved loci (Majumder *et al.* 1989).

(f) **Environmental influence** — environment, together with genotype, jointly influence the manifestation of the disorder. Coronary Heart Disease (CHD) can be cited as an example. The contribution of genetic factors to the development of CHD has been estimated (Slack and Evans 1966; Goldbourt and Neufeld 1986). From these studies it appears that there is no single gene locus responsible for CHD. Rather, different environmental and genetic factors jointly act and interact in a highly complex fashion in the pathogenesis of CHD.

1.4 Problems in the Analysis of Family Data of a Complex Disorder

Methods of segregation analysis of family data are well established when the disorder is primarily determined by alleles at a single locus (Emery 1976). Suitable modifications have also been made to take into account incomplete penetrance, variable age at onset, etc. However, when a disorder is not transmitted in a simple Mendelian fashion, that is, when the observed segregation probability/ratio (which is defined as the conditional probability of an affected offspring given a parental mating type) is vastly different from that expected under a one-locus model, the disorder is usually described as 'polygenic' (meaning that the disorder is determined by alleles, each with a small undetectable effect, at a very large number of loci, which act additively to produce the disorder phenotype). Heritability of the disorder is then estimated. However, with the identification of single genes in the so-called complex disorders, the concept of polygenic inheritance is now beginning to be challenged (Passarge 1993). A recent example is Hirschsprung disease, a complex disorder hitherto considered to be of polygenic origin. Patients with this disease suffer from severe constipation and abdominal distension due to congenital megacolon. A gene for this disease, which was traditionally assumed to be polygenic, has now been successfully localized to the pericentromeric region of the short arm of chromosome 10 (Angrist *et al.* 1993; Lyonnet *et al.* 1993). Thus, there is an increasing realisation and documentation of the fact that many, so called complex disorders may really be due to effects genes at one or a small number of loci. Having rejected the single-locus model of inheritance, no attempt is usually made to analyse the data under multilocus models, assuming the involvement of a small number of loci. This is because of the intrinsic problems associated with multilocus models. For example, under a single locus model, if the disorder is recessive, then

in a family, ascertained through an affected child, if both parents are normal, then both of them are obligate heterozygotes. This fact simplifies family data analysis to a great extent, because in all such families, the segregation ratio is 1/4. Such simplicity vanishes even when two unlinked biallelic loci, with alleles (A,a) and (B,b), are considered. In this case, each normal parent can be of one of the genotypes AaBb, Aabb, aaBb. Thus, a Normal × Normal mating type many be any one of the genotypic matings AaBb × AaBb; AaBb × Aabb; AaBb × aaBb; Aabb × Aabb; Aabb × aaBb; aaBb × aaBb with segregation ratios 1/16; 1/8; 1/8; 1/4; 1/4; 1/4, respectively. This heterogeneity of mating types and segregation ratios introduce considerable complexity in deriving the likelihood function of a set of phenotypic observations on members of a family, and in carrying out computations. Further, because of the small family sizes of humans, low segregation ratios imply that in a large proportion of families only one affected member (the proband) is usually observed. This, in turn, creates problems in data analysis because the members in these families appear as sporadic (nongenetic) cases and/or the normal individuals mimic incompletely penetrant cases. To avoid the confusion regarding whether the affected individual in a single-case family is genetic (but, chance isolated) or sporadic, it may be more practical to select multicase families. However, Chakravarti (1993) has shown that even though such a selection procedure enriches for segregation at a few loci, these segregants will be at multiple, independent loci. This implies that there will be an intrinsic heterogeneity among such families which will adversely affect linkage-mapping efforts. Variable age at onset adds further complexity, because in this case an individual may be of the 'affected genotype', but may not have expressed the disorder at the time of study.

1.5 Development of Multilocus Models

A number of human disorders and congenital malformations show strong familial aggregation but do not conform to the expected recurrence risks in sibs, or are not transmitted from parents to offspring in a simple, single locus, Mendelian fashion. Various models have been proposed to describe the way in which gene(s) affect the liability of individuals to a disorder. They range from models of the effects of alleles at a single locus to multifactorial/polygenic models representing the effects of genes at many loci and the effects of environment. The multifactorial model, while descriptive, sheds very little light on possible underlying biological mechanism. Thus, extending simple single locus Mendelian models to more than one locus represents the next logical step for exploring possible genetic mechanisms for diseases which show strong familial aggregation. However, inherent in analyzing models based on two loci is the question of the biological and statistical aspects of interactions between alleles at the different loci. Straight additivity is rarely a good assumption for any biological mechanism, while epistatic interaction represents a plausible mechanism for many disorders.

To the best of our knowledge, the first two-locus model proposed for a human disorder — psoriasis, a dermatological disorder — was by Steinberg *et al.* (1951). They performed a simple-minded analysis of data from various mating types and concluded that recessive alleles at two unlinked, autosomal, biallelic loci interacted in the manifestation of the disorder, and that individuals who were recessive homozygotes at both loci were affected. Although psoriasis has a variable age at onset, this fact was not rigorously incorporated in Steinberg *et al.*'s (1951) analyses. Be that as it may, it is interesting to note that the first suggestion of a two-locus model for a human disorder pertained to a dermatological disorder, as is vitiligo — the disorder considered in the present

study, and also that the first two-locus model proposed recessive homozygosity at both loci, which is the primary model considered in the present study.

Li (1953) considered the two-locus recessive homozygosis model and derived many useful theoretical results that included showing that Snyder's ratios can be generalized in a straightforward manner to multiple loci. These results are useful for analysing nuclear family data for a disorder that is expressed at birth, provided that families are ascertained randomly or through an affected parent. Li (1987) provided further generalizations and results. Some of these results are described in subsequent chapters of this thesis.

Merry *et al.* (1979) performed a theoretical study of a two-locus model for a familial disease. They considered two autosomal, linked, biallelic loci and a selection scheme in which the double heterozygotes were assumed to be most fit. They derived conditions for the existence of a stable equilibrium and showed that it could be used to explain a wide range of disease frequencies and patterns of inheritance. However, even though this study considered some diseases with variable ages at onset, the theoretical investigations were carried out without taking ages at onset into account.

A simple graphical method for testing two-locus models was proposed by Greenberg (1981). The model assumes two alleles at each locus where both loci exhibit dominance (DD-model), both exhibit recessivity (RR-model), or one locus exhibits dominance and other exhibits recessivity (DR-model). This method is based on two sets of contour graphs — one is the graph of population segregation ratio, which can be viewed as Snyder's ratio averaged over all parental mating types, and other is the graph of population prevalence. Both contour graphs are plotted at different values of the allele frequencies at the two loci. The segregation ratio is calculated from the data; the one s.d. band around this estimated value is called the allowed area. The allowed area of the population prevalence graph is also similarly defined. If the allowed areas of the population prevalence and the population segregation ratio contours

corresponding to a particular genetic model do not overlap, then the model is rejected, since this implies that the gene frequencies corresponding to the population prevalence are inconsistent with those the population for segregation ratio. If the allowed areas do overlap, the model is not rejected. Greenberg (1981) also viewed the simple single locus Mendelian dominant and recessive models as special cases of the more general two-locus models. As the allele frequency at one of the two loci approaches unity, the two-locus model reduces to a single-locus model.

The testing procedure described by the graphical representation is not a test of significance or fit in the statistical sense. It is rather a test of the consistency of the model with biological parameters, namely, the gene frequencies at the two loci. If the gene frequencies predicted by the population segregation ratio and trait prevalence are non-overlapping, the model is obviously not consistent with the observations. Further, this method can only be applied when a disease is expressed at birth, but is inapplicable for diseases which show variable ages at onset in their expression. Greenberg and Rotter (1981) applied this method to data on coeliac disease and obtained a consistent fit to an epistatic, two-locus, recessive model.

A maximum likelihood test of the two-locus model for coeliac disease was subsequently developed by Greenberg and Lange (1982). They derived the likelihood of a sibship given a model (RR or DR) as :

$$L(t,\pi) = \frac{\sum_i M_i(\alpha_i t)^a(1-\alpha_i t)^b \pi^c (1-\pi)^{(a-c)}}{1-\sum_i M_i(1-\pi\alpha_i t)^{(a+b)}}$$

where i indicates the i-th mating type ($i = 1, 2, \ldots, 81$), M_i = probability of mating type i, t = parameter such that $t = 1$ under the null hypothesis, α_i = segregation ratio for mating type i, π = ascertainment probability, a = number of affected offspring, b = number of unaffected offspring, c = number of probands. The numerator of the likelihood equation represents the probability of an ascertained sibship of size $(a + b)$, and the denominator represents the

probability of ascertaining such a sibship.

The proposed maximum likelihood test was exhaustively examined using simulation techniques by Greenberg (1984). The power of this method to distinguish between single-locus and two-locus models, with and without environmentally caused reduced penetrance, was tested. The effects of ascertainment probability on the analysis and the proband-conditioned ascertainment correction proposed by Cannings and Thompson (1977) were also investigated. It was shown that the method has sufficient power to distinguish between the fully penetrant double recessive (RR) model and the fully penetrant single locus dominant and recessive models. The method can also distinguish fairly well between the dominant-recessive (DR) and the RR models, even when population prevalence is not taken into account. It was also found that this method has much less power to distinguish between the fully penetrant RR model and single-locus models with reduced penetrance. When environmental penetrance is taken into account, then the power of the method to distinguish between one-locus and two-locus models improved substantially. However, the estimates of ascertainment probability, π, were robust, regardless of the model under which the data were generated. The Cannings-Thompson (1977) approach to ascertainment correction worked well only when π was < 0.1. It must, however, be noted that the method of Greenberg and Lange (1982) and the subsequent inferences of Greenberg (1984) pertain to a disorder that is expressed at birth, and are not directly applicable to a disorder with a variable age at onset.

Methods for segregation analyses of data on ascertained families in respect of multilocus recessive homozygosis models, have primarily been developed by Majumder *et al.* (1988, 1989) and applied to vitiligo and prelingual deafness. These methods developed are applicable to disorders which are expressed at birth or which have variable ages at onset. Since these methods form the starting point of the present study, they are described in detail in subsequent chapters and suitably extended for use in the present study.

Recently, Neuman and Rice (1992) have derived formulas for the recurrence risk to various classes of relatives in terms of penetrances and gene frequencies for two locus models.

1.6 Purpose of the Present Study

Disorders which are not inherited in a simple Mendelian fashion are generally termed as complex disorders. Since a large number of complex disorders show significant familial aggregation but are not inherited in a clear-cut, single-locus Mendelian fashion, exploration of models and methods for understanding the genetics of such complex disorders is important. The complexity of a trait may arise in a number of ways, some of which are mentioned earlier. It is, however, useful to investigate whether concrete multilocus genetic models provide adequate fits to family data on such complex disorders because, among other things, one can then make theoretical (probabilistic) risk predictions rather than rely on empirical risk estimates. Further, if the estimated number of loci is found to be small, it may not be too difficult to map all the component loci involved in the aetiology of such a complex disorder. But, available statistical methods for the analyses of family data on such complex traits with a view to finding the mode of inheritance and the number of major loci involved (segregation analysis) are not very satisfactory. During the last few years, attempts have been made to develop methods for the study of disorders that are determined by the epistatic action of recessive alleles at multiple loci (Li 1987; Majumder *et al.* 1988). The present study is an attempt to further explore the population characteristics of such a disorder, to improve upon the methods of segregation analysis available for the study of such a disorder especially when its onset age is variable, and to use these methods to cross-validate a

proposed model (Majumder *et al.* 1988) for a dermatological disorder — vitiligo. Since the strategy of "double cross-validation" (Mosteller and Tukey 1977), that is, collection of new data and estimation of model parameters from the fresh data, has been adopted in the present study, detailed descriptions of pigmentation and vitiligo, methodology used for collection of new data and results of epidemiological and genetic analyses are also presented.

1.7 Structure and Main Findings of The Thesis

This thesis comprises six chapters. In the present chapter (Chapter 1) the background and an introduction to this study have been provided. The nature and some characteristic features of complex disorders, possible sources of complexity, problems of genetic epidemiological analyses of data on complex disorders, and development of multilocus models have also been discussed in this chapter. Finally, an outline and purpose of the present study has been given in this chapter.

Chapter 2 presents an overview of pigmentation and vitiligo primarily to provide biological justifications to the genetic models considered. This overview includes descriptions of biochemistry and genetics of pigmentation, clinical manifestations, prevalence and hypotheses regarding aetiology of vitiligo, and findings of some animal models used in investigating the pathogenesis of this disorder. Detailed descriptions about the data set on vitiligo used for cross-validation are given in this chapter. Each family in this study was ascertained through a single individual afflicted with vitiligo (the proband) and was chosen from an unsorted list of all patients registered with the National Vitiligo Foundation, Tyler, Texas, U.S.A. Data on 160 Caucasian families were collected.

While data on all families have been used for epidemiological analyses, data on only 147 families were used for genetic analyses because of missing information on some key variables.

Chapter 3 deals with epidemiological analyses of the data. The prevalence of vitiligo is $\approx 0.5\%$. Using a non-parametric run test, it was found that the age at onset distributions among affected males and females were not significantly different. The mean ($\pm$ s.e.) ages at onset (in years) for affected males (n = 73) and females (n = 143) were 18.64 ± 1.63 and 24.19 ± 1.27, respectively. The intra-class correlation in ages at onset of affected individuals within families was found to be high (0.6). About 19% of probands had at least one affected first-degree relative and about 31% of probands had at least one affected relative. Relative risks (RR) among various categories of relatives of probands were found to be significantly greater than unity, which is an indication of significant familial aggregation. For example, the RR (95% confidence interval of RR) values for proband fathers was 7.20 (2.89-14.84), mothers was 7.16 (2.88-14.76), brothers was 13.36 (5.77-26.82) and sisters was 11.68 (5.34-22.17). No statistically significant effect of proband's age at onset on relative risks was detected. The effects of some alleged causal factors (*e.g.*, parental ages at first childbirth, stress, thyroid disorders, sun tanning, etc.) on the probability of affection to vitiligo have been examined. However, no significant association of vitiligo with any of these alleged factors was found.

Chapter 4 provides (i) descriptions of the multilocus models used, (ii) statistical evaluations of some characteristics of these models, (iii) derivations of likelihood functions of data on different types of nuclear families, and (iv) some comments on computations of likelihood functions. Two multilocus recessive epistatic models have been considered. In both models, a set of unlinked, biallelic loci have been considered. The first model (Model I) postulates that individuals who are homozygotes for recessive alleles at *every one* of these loci are affected; individuals of other multilocus genotypes are normal. The

second model (Model II) postulates that individuals who are recessive homozygotes at *any one* of these loci are affected; individuals of other genotypes are normal. The first model was proposed earlier by Majumder *et al.* (1988) for vitiligo. In this study, model I is considered for cross-validation. The second model is prompted by the present knowledge of pigmentary pathways. Some relevant population, randomly mating and inbred, and sample characteristics (*e.g.*, population prevalence, types and proportions of various kinds of families in population and in sample) when the disorder is expressed at birth have been derived theoretically for model I, which is the central model of this study. It was seen that prevalence sharply decreases as the number of loci increases. The proportion of Normal × Normal families increases with an increment in the number of loci for a fixed prevalence; the proportions of Normal × Normal and Normal × Affected families correspondingly decrease. For a fixed sibship size and prevalence, the probability that a family contains only one affected child monotonically increases with increase in the number of loci and correspondingly, there is a monotonic decrease in the probability of a family containing a larger number of affected children. While for the one-locus recessive model, the probability distribution of affected children within a sibship depends only on sibship size and not on prevalence, for multilocus recessive models this probability distribution also depends on prevalence. Likelihood functions of data on various types of nuclear families have been derived for the following situations : (i) the disorder is expressed at birth, and (ii) the disorder has a variable age at onset. In both cases, appropriate corrections for biases due to non-random ascertainment (*e.g.*, a family is sampled through an affected parent or an affected offspring) have been incorporated into the likelihood functions.

The results of segregation analysis of the nuclear family data are presented in Chapter 5. Likelihood computations have been performed for varying numbers of loci. Model I was found to be vastly more likely than model II. [If only a single locus is involved, then it can be analytically shown that any data will

have the same likelihood value under both models I and II. For the two-locus case, the observed odds in favour of model I was about 10^8.] For model I it was observed that the data were 10^{13} times more likely (statistically significant) if two loci are involved than if one locus is involved; about 10^6 times more likely (statistically significant) if three loci are involved; but only about 18 times more likely (statistically non-significant) if four loci are involved. Therefore, it was concluded that three epistatically interacting autosomal biallelic loci are involved in the pathogenesis of this disorder. The robustness of this inference with respect to variations in prevalence and age-specific affection probabilities has also been investigated. The inference was found to be robust and valid.

Chapter 6 deals with segregation analysis of the pedigree data (*i.e.*, of data on extended families of probands). Segregation analysis was performed by using the Pedigree Analysis Package (PAP), revision 3.0 (Hasstedt 1989). A non-genetic/sporadic model, three one-locus models [autosomal dominant, recessive and general (allowing for incomplete penetrance and sporadic cases)] and the two-locus recessive homozygosis model (Model I) have been considered. Because of the enormous computational complexity and computer time requirement, it was not possible to fit multilocus models involving more than two loci. Because of this limitation, the present pedigree analysis only confirms whether a multilocus model provides a better fit than a single-locus model, but does not provide an estimate of the number of loci involvement. It was assumed that an individual of a specific genotype and age will be affected if the individual's liability to the disorder exceeded a certain threshold. The data on the 147 families were randomly split into two approximately equal subsets with 75 and 72 families, respectively. This was done for cross-checking of estimates and for computational convenience. For each model, the likelihood function was numerically maximized with respect to the parameters of the model. It was found that the two-locus model provided a best fit to the data, although the general one-locus model that allows for incomplete penetrance and sporadic

case comes close. In spite of the fact that the general one-locus model provided a significantly better fit to the data than the one-locus recessive model, the general model turned out to be effectively a recessive model. The one-locus dominant model was clearly rejected. The non-genetic model yielded an unacceptably high estimate of prevalence. These results of segregation analysis of the pedigree data support that major conclusion derived from analysis of data of nuclear families that recessive alleles at more than one locus are involved in vitiligo. This study, therefore, provides significant evidence in favour of the genetical model proposed earlier for vitiligo by Majumder *et al.* (1988).

The present thesis is a contribution to quantitative genetic analysis of family data on the multilocus recessive disorders with variable ages at onset. It provides some new statistical techniques, and cross-validates a model proposed earlier for a pigmentary dermatological disorder known as vitiligo. The inferred genetical model makes eminent biological sense. The implicated loci may be viewed as controls of key points of the pigmentary pathways, and recessive homozygosis at any of these loci may be viewed as complete disruption or blockage at the corresponding point. If one or two of these control points are blocked then bypass routes in the pigmentary pathways are possibly used and there is no precipitation of disorder state. However, if a greater number of control points are blocked, then possibly there is failure of the entire pathway because of non-availability of further bypass routes, which results in a disruption of the end product or process and consequently in the clinical manifestation of vitiligo.

Chapter 2

Pigmentation, Vitiligo and the Present Data Set

2.1 Biochemistry and Genetics of Pigmentation

Almost all eukaryotic forms of life, animals or plants have a pigmentary system that seems essential for survival, either to capture the energy of the sun to protect against the sun's toxicity or to allure the opposite sex. The pigmentary system is essential for the proper development and function of many organ systems like the skin, eyes, hairs and ears. Its functions are regulated by a large number of genes. Progress in understanding the anatomy, biochemistry, physiology and embryology of pigment cells in subhuman primates and other vertebrate species has provided a basis for understanding the genetic regulation of the human pigmentary system and its disorders.

The human pigmentary system develops during the first two months of embryogenesis. Cells within the neuroectodermal analage differentiate into melanoblasts that migrate to the epidermis, the eyes, ears, gastrointestinal tract and leptomerings. By 8 to 10 weeks of gestation, pigment cells reach the epidermis and begin to mature into melanocytes capable of producing melanosomes and synthesizing melanin. In the eyes, the melanoblasts form the choroid which lies below the cornea. Just below the choroid layer, another kind of pigment

system develops which includes the iris and ciliary body and is known as pigment epithelium. In the middle and inner ear, the pigment cells are essential for the development of the cochlea, stria vascularis and organ of corti.

Within the melanocyte, pigment is synthesized and deposited on specialized membrane bound cytoplasmic organalles called melanosomes (Seiji *et al.* 1973; Moyer 1966). Melanosomes are membrane bound particles containing an internal matrix (Breathnach *at al.* 1966) on which the enzyme tyrosinase attaches. The copper-containing enzyme tyrosinase is synthesized in the endoplasmic reticulum, contained in the golgi and transferred to the melanosome matrix (Maul 1969; Maul and Brumbaugh 1971).

Within the melanosome the enzymatic conversion of the amino acid tyrosine to melanin is catalysed by tyrosinase as the initial critical step in the formation of both brown-black pigments, *i.e.*, eumelanin and red-yellow pigments, *i.e.*, pheomelanin. Melanin is formed by the polymerization of indole-5,6-quinone, which originates from tyrosine through the action of tyrosinase, via 3,4-dihydroxyphenylalanine (dopa) through other intermediate stages of decarboxilation and oxidation. In man, tyrosine serves as the precursor substance of melanin.

Disorders of pigmentation in humans may result from abnormalities at each of the various steps in melanogenesis. 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" which 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 (perhaps because of melanocyte destruction), leading to hypopigmentary patches that tend to become progressive over time (Mosher *et al.* 1979).

Hypopigmentation is also consistently observed in Waardenburg syndrome, Prader-Willi syndrome, Angelman syndrome and in piebaldism. Mutations at specific loci have been found to be associated with the murine analogues of these phenotypes (Hearing 1993). In the mouse, more than 150 different pigment mutations have been identified; these map to 60 distinct genetic loci (Silvers 1979). Mammalian pigmentation is regulated at many different developmental and cellular levels and is influenced by many genes, either directly or indirectly. 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 likely that multiple genetic loci 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.

2.2 Vitiligo

2.2.1 Clinical manifestations

Vitiligo is an idiopathic hypomelanosistic dermatological disorder that is characterised by pale, milk white macules that tend to become progressive over time (Mosher *et al.* 1979). It is highly variable in its manifestations. The lesions may range in size from a single, small circumscribed area of depigmentation to 'complete' vitiligo in which virtually the entire body is involved. Commonly there are multiple patches which may or may not be symmetrical or segmental in distribution (Kugelman and Lerner 1961) Depigmentation may be partial

or complete. The exposed areas are usually involved, but the lesions may appear anywhere on the body surface, even on the hair. Vitiligo may begin at any age. Onset of the disease has sometimes been reported to follow physical or emotional trauma. Some patients have co-existent organic or emotional illness. An increased prevalence of vitiligo in patients with pernicious anaemia, and a number of diseases allegedly of autoimmune aetiology, including diabetes mellitus, halo nevi, thyroid disease, Addison's disease, etc. has been reported (Lerner 1959; Cunliffe *et al.* 1968; Dawber 1968; Bor *et al.* 1969; McGregor *et al.* 1972). The most important disease with which vitiligo is associated is melanoma. Among patients with melanomas, 20% have vitiligo (Lerner and Nordlund 1978).

2.2.2 Prevalence

Although universal in occurrence, the prevalence of vitiligo varies considerably among different regions and ethnic groups. Primarily on the basis of clinical records of hospitals and skin clinics, the prevalence is estimated to be $\sim 1\%$ in Egypt, $\sim 0.39\%$ in Switzerland, $\sim 0.14\%$ in Russia, $\sim 0.24\%$ in London, $\sim 1.64\%$ in Japan, and $\sim 1\%$ in United States (Lerner 1959; El-Mofty 1968;). On the basis of population surveys, in the Isle of Bornholm in Denmark the prevalence is estimated to be $\sim 0.38\%$ (Howitz *et al.* 1977), in western India the prevalence varies from near absence to $\sim 3.6\%$ among different communities (Mehta *et al.* 1973) and in Calcutta, eastern India the prevalence is estimated to be $\sim 0.46\%$ (Das *et al.* 1985a). The estimates based on hospital records are obviously biased upwardly. To the best of our knowledge no population based estimate is available from the USA. Prevalence of vitiligo varies significantly with age; age-specific prevalences between genders are not significantly different (Das *et al.* 1985a). The age at onset of vitiligo is variable. The mean age at onset of vitiligo among male probands was 38.7 ± 19.5 years and among female

probands was 36.7 ± 18.9 years in Denmark (Howitz *et al.* 1977). But in Calcutta, the mean age at onset for male and female probands, respectively, were 24.77±1.32 years and 19.28±1.18 years (Das *et al.* 1985a). It is seen from both studies that the manifestation of vitiligo in women seems to be somewhat earlier than in men, but the difference is not statistically significant. The modal age at onset among probands was about 15 years, and by 30 years of age the disorder manifested in about 75% of the cases regardless of family history (Das *et al.* 1985a).

2.2.3 Hypotheses regarding aetiology

To date, there is no convincing theory for the aetiology of vitiligo. From the clinical and experimental work on vitiligo, two reasonable hypotheses have evolved. One is that a neurogenetic factor is involved and that a cytotoxic agent is released close to or within the melanocyte, destroying the cell (Lerner 1959). Although this hypothesis was not a favoured one, in recent times it is gaining attention. The main role of the melanocyte is thought to be transformation of L-tyrosine into melanin in a precise and efficient way without destruction of the melanocyte by the highly toxic reaction sequence, whereby melanin acts as a scavenger of free radicals and toxic intermediates of melanogenesis. If melanin synthesis is inhibited, then the toxic products can destroy the keratinocytes to which pigment granules (melanosomes) are transported after melanogenesis. In mammals, there is a naturally occurring factor, melatonin, which inhibits melanin synthesis without affecting tyrosinase activity. It is, therefore, hypothesized (Slominski *et al.* 1989) that in vitiligo patients there is an increased concentration of melanin caused by hyperproduction of melatonin in the pineal gland or other extrapineal sites (*e.g.*, gastrointestinal tract). No experimental evidence in support of this hypothesis is yet available. A similar hypothesis involving anti-oxidant defence enzymes has been

proposed by Schallreuter *et al.* (1991). The second hypothesis is that vitiligo is an autoimmune disorder (Cunliffe *et al.* 1968) resulting from the formation of an anti-melanocyte autoantibody. The evidence supporting this hypothesis is based mainly on the clinical association of vitiligo with a number of disorders that are thought to be autoimmune (*e.g.*, diabetes). Light and electron microscopy of vitiliginous skin has shown not only a striking lack of melanin but also a considerable reduction in the population density of secretory and functional melanocytes (Birbeck *et al.* 1961; Lerner and Nordlund 1978). More recently, Cui *et al.* (1992) have shown that a higher proportion of patients with vitiligo have circulating antibodies to pigment cells than normal controls. However, these antibodies are non-specific in that they are directed to several cell surface antigens, some of which preferentially expressed on pigment cells. Ramaiah *et al.* (1989) have proposed a third hypothesis based on some cell culture studies performed by them. They have found that levels of mitogenic factors in sera of vitiligo patients are significantly lower than those found in sera of normal individuals. They have hypothesized that depigmentation in vitiligo results from a reduction of levels of growth factor(s), originating from keratinocytes, fibroblasts and other tissues, which are necessary for normal proliferation and maintenance of melanocytes. These findings are yet to be confirmed.

2.2.4 Genetics

Positive family history and familial aggregation of vitiligo have been reported for a long time (Cockayne 1933; Merelender and Rywlin 1940; Behl 1955; Levai 1958; Lerner 1959; Mehta *et al.* 1973; Goudie *et al.* 1980; Hafez *et al.* 1983; Das *et al.* 1985a).

Both members of two monozygotic twin pairs have also been found to be affected (Mohr 1951; Siemens 1953). Vitiligo has been reported in children —

both boys and girls — below 12 years of age (Halder *et al.* 1987; Jaisankar *al.* 1992). A genetic involvement is supported by observations on familial aggregation of the disorder and is in concordance in monozygotic twins. The mode of inheritance of the disorder is still debated. However, because of the lack of systematic family studies, almost all conceivable modes of inheritance have been proposed. Some believe that vitiligo is due to an incompletely penetrant autosomal dominant gene (Cockayne 1933; Lerner 1959), while some others claim that the disorder gene is autosomal recessive or that the disorder is polygenic/multifactorial in nature (Mehta *et al.* 1973; Carnevale *et al.* 1980). Hafez *et al.* (1983) studied families of 150 vitiligo patients. Their data did not support either a single-locus autosomal dominant or an autosomal recessive model, but a multifactorial model provided an adequate fit. They treated vitiligo as a quantitative liability threshold trait (Falconer 1965). The heritability of liability was estimated to be 72.4%. Their data also supported Edward's (1960) finding that the frequency of a multifactorial trait among relatives approximates the square root of its frequency in the general population. Hafez *et al.* (1983), however, did not take into account the variability of age at onset of vitiligo. Therefore, their conclusions are to be accepted as tentative. In a systematic study of 298 families conducted in Calcutta, India, Das *et al.* (1985a) also assumed the polygenic liability threshold model for vitiligo. Their assumption was based on the lack of consanguinity among parents of vitiligo patients and the fact that recurrence risks were much lower than those expected under single locus inheritance. They estimated the heritability of liability to the disease to be 46.06%.

Unfortunately, in vitiligo, as is true for most rare diseases that are not determined by a single gene, the number of affected individuals in each family is very small, even when it is ascertained through an affected proband. One of the major problems in fitting of concrete genetic models is the lack of multiplex families. Even among multiplex families, most families have only one affected

relative of the proband. Li (1987) developed a concrete genetic model for a rare recessive genetic trait, where he proposed that some traits or disorders are determined by a particular combination of genes at a number of loci, thus forming a gene 'configuration' or 'constellation'. Absence of any one gene in this configuration destroys the manifestation of the trait or disorder. Majumder *et al.* (1988) investigated whether the model proposed by Li (1987) provided an adequate fit to the vitiligo family data collected from Calcutta, India. Their analysis, based on likelihood methods, showed that the multilocus recessive model provided an adequate fit to the data. They proposed that recessive alleles at four unlinked autosomal biallelic loci are involved in the pathogenesis of vitiligo.

Evidence of association with some genetic loci can provide helpful clues regarding the genetics of the disease. Earlier studies have indicated that among vitiligo patients there is a significant excess of M blood group (Wasfi *et al.* 1980); G-6-PD deficient males (Saha *et al.* 1982); some HLA antigens, for example, with HLA-DR4 (Foley *et al.* 1982; Dunston and Halder 1990), HLA-B13 (Metzker *et al.* 1980), HLA-DR1 (Poloy *et al.* 1991) and ACP1 and Rh loci (Das *et al.* 1985b). Some specific HLA supratypes have also been reported to be associated with early and late onset forms of vitiligo (Finco *et al.* 1991). It has also been suggested, based on association study (Venneker *et al.* 1992), that abnormalities of the C4B gene of the fourth (C4) human complement system may be risk factors in vitiligo. Although such association studies cannot provide as strong evidence as linkage studies, nevertheless, the reported association of vitiligo with genetic markers on several chromosomes, if true, may indicate the involvement of several loci.

2.2.5 Animal models for vitiligo

Animal models provide an excellent tool for studying the biological aberrations in human disorders (Lerner *et al.* 1986; Naughton *et al.* 1986; Boissy and Lamoreux 1988). Rapid progress can be made in the understanding of a disorder that occurs in human beings when an experimental animal model is available for study. Animal models can be classified into two types —(1) induced, and (2) genetic. Induced animal models result from specific treatments that cause a pathologic condition and can thus be used to evaluate the biochemical responses to the causative agent. Genetic models, on the other hand, are traditionally the result either of a spontaneous mutation, or of selective breeding that brings together a group of genes that cause an aberrant phenotype in a particular stock or breed of animals. Inbred strains are often extremely advantageous in evaluating complex, genetically controlled physiological systems because often there is only a single genetic difference between the mutant animal and controlled animal. However, inbred strains can provide false evidence of involvement of a single locus in a disorder when the disorder is under multilocus control. This may happen because of genomic homozygosity achieved by inbreeding over several generations.

It is now clear that probably all vertebrates with melanin pigmentation can develop vitiligo. Animal models for vitiligo have been identified or developed and are being analysed to understand the underlying mechanism involved in the expression of vitiligo. Some well described animal lines are — (1) the Smyth chicken (Smyth *et al.* 1981), (2) the vitiligo mouse (Lerner *et al.* 1986), (3) the gray horse (Naughton *et al.* 1986) and (4) the Sinclair pig (Millikan *et al.* 1974; Hook *et al.* 1979).

Smyth line chickens are phenotypically characterized by a posthatch depigmentation of feathers. Vitiligo in Smyth chicken begins with an inherent melanocyte defect which, during the pre-amelanotic stage, is manifested as

an increase in the activity of tyrosinase and a modification of its intracellular distribution (Boissy *et al.* 1986). One of the prevalent hypotheses for the aetiology of vitiligo in humans is the involvement of melanin precursors (Lerner 1971) that have been demonstrated to be toxic to melanocytes (Wick *et al.* 1977; Pawelek *et al.* 1980), and may selectively induce pigment cell death when in excess. Results of preliminary experiments on cultured Smyth melanocytes suggest that such a phenomenon may occur in this animal model. Other direct evidence that treatment with immunosuppressive agents decreases the incidence and severity of depigmentation suggests that the expression of vitiligo in the Smyth chicken is related to an autoimmune response. Austin *et al.* (1992) have detected circulating melanocyte-specific autoantibodies in the sera of depigmented Smith chicks which are not present in sera from Light Brown Leghorn control chicks. Cyclosporin-A administration, which inhibits the cellular component of the immune system, suppresses the expression of amelanosis (Pardue *et al.* 1987).

The $C57BL/6-mi^{vil}/mi^{vil}$ (vitiligo) mouse, originally developed by Lerner *et al.* (1986), is homozygous for an autosomal recessive gene. The *vit* (vitiligo) gene has been mapped to the microphthalmia (*mi*) locus on mouse chromosome 6 (Lamoreux *et al.* 1992). Mice of this genotype are born with white spotting and develop, beginning at about five weeks of age, loss of melanocytes from the epidermis of the ears and tail and a selective cell mediated immune deficiency to epicutaneous-administered allergens. This observation is consistent with that observed in humans with vitiligo, who also exhibit loss of contact hypersensitivity (CHS), that appears to be associated with loss of pigment cells from the epidermis (Amornsiripanitch *et al.* 1988). There does not appear to be a significant immune component to the aetiology of vitiligo in this mouse model. The absence of inflammatory cells during depigmentation and the inability of Cyclosporin-A treatment to influence the expression of amelanosis, supports this assumption. In the $C57BL/6-mi^{vit}/mi^{vit}$ mouse, depigmenta-

tion seems to be transmitted as a simple autosomal recessive trait (Nordlund 1987). However, since there are highly inbred strains, such findings may not reflect the true genetic basis of the disorder. Besides, work on the genetics of pigmentation in the laboratory mouse have revealed the influence of about 70 genes at approximately 40 loci (Fitzpatrick and Quevedo 1971; Quevedo *et al.* 1987).

The gray horse and the Sinclair pig may also support the genetic basis for the causation of vitiligo (Gebhart *et al.* 1977, Naughton *et al.* 1986). So it is clear from the animal models that the origin, spread, and differentiation of melanoblasts, morphology of melanocytes and melanosomes, biosynthesis of tyrosinase and melanogenesis, and, finally, transfer to and arrangement of melanosomes in keratinocytes are each under the direction and influence of different genes.

Before closing this section, we review some recent findings on mutations that lead to pigmentary anomalies since these may eventually turn out to be important in the pathogenesis of vitiligo. Mutations at two loci in the mouse are known to lead to pigmentation defects, in addition to other defects. These loci are *W* and *Sl*. Dominant white spotting associated with the *W* locus results from mutations in the *c-kit* membrane receptor gene. The *Steel* phenotype (hypopigmentation and anaemia, primarily) results from mutations in the gene encoding the legend for *kit*. The *c-kit* proto-oncogene encodes a transmembrane receptor belonging to the class III family of tyrosine kinases, and is thus obviously related to pigmentation. Recently, Spritz *et al.* (1992) have reported heterozygous mutations of the *c-kit* gene in families with piebald trait, which is an autosomal dominant defect characterised by white hair forelock, by stable areas of hypopigmentation on the anterior trunk and extremities and by the absence of other manifestations. Lee *et al.* (1994) have reported that mutations of the P gene, which produces a transmembrane polypeptide that is thought to transport small molecules such as tyrosine, are associated with Type

II (tyrosinase-positive) occlutaneous albinism, ocular albinism and Prader-Willi or Angelman syndromes. While the clinical and phenotypic manifestations of vitiligo are different from piebaldism or from the disorders associated with P gene mutations, since all of these disorders have hypopigmentation in common, it is possible that *c-kit* or the P genes may be involved in vitiligo, although no direct evidence is yet available.

2.3 The Present Data Set

2.3.1 Data structure and sampling procedure

The data used in this study were collected by Professor James J. Nordland, Chairman, Department of Dermatology, University of Cincinnati Medical Center, U.S.A., and Professor Partha P. Majumder, Anthropometry and Human Genetics Unit, Indian Statistical Institute, Calcutta. They have kindly permitted the use of these data, and have provided some of the information given below.

Each family in this study was 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, Texas, U.S.A. Every fifth patient on the list was sent a questionnaire requesting relevant information on himself/herself and his/her family members, including first and second degree relatives. Returned questionnaires were scrutinized for missing and/or inconsistent information. In the few instances when inconsistent observations were noted, telephone interviews or follow-up letters were used to resolve inconsistencies. Prior to receiving the data back from the patients, no information on gender, ethnicity, age, number of affected relatives or other variables were available. The questionnaire also sought information on the names

and qualifications of the persons who had performed clinical evaluation 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. 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 preponderance of familial cases among those registering themselves with foundations such as the National Vitiligo Foundation. However, as we shall report later, about 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 in India (Das *et al.* 1985a), 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 due to sampling probands from the list of the National Vitiligo Foundation may be small and negligible.

2.3.2 Items of information collected

The questionnaire sought information regarding presence of vitiligo in a fixed set of relatives of the proband (children, grandchildren, spouse(s), parents, siblings, grandparents, uncles and aunts); current age; age at onset of vitiligo; 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.

2.3.3 Sample sizes and other descriptive information

In all, 300 questionnaires were sent. A total of 194 completed questionnaires were returned. The response rate, therefore, was about 65%. 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. Of these 160 families, 5 families were nuclear families; therefore, about 97% of sampled families were extended. [Pedigree charts of a selection of 10 sampled families are provided in the Appendix.] Prevalence estimates from the non-Caucasian (e.g., Black, Hispanic, etc.) population groups are not available. Because pooling data of families different ethnic backgrounds with possibly differing prevalences of vitiligo might have yielded distorted results of the analyses, it was necessary to treat families of different ethnic backgrounds separately. The number of families of non-Caucasian ethnicity also too small for performing any reasonable statistical analysis of the data from these families. The Statewise distribution of the place of residence (U.S. postal State codes are used) of the 160 probands is: AK=1, AL=1, 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, WY=1. This distribution shows that there is no notable geographical clustering of probands.

For epidemiological analyses, we have used the data of 160 families. The total number of males and females in these families were 1136 and 1120, respectively. In addition to the 53 male and 107 female probands, 43 first-degree relatives of probands were affected, of whom 16 were male and 27 were female. In the extended families of probands, there were additionally 13 affected individu-

als of whom 4 were male and 9 were female. For genetic analyses we have used the data of 147 families; 13 families were excluded from these analyses because of unresolved multiple missing information (*e.g.*, age at onset of affected relatives of the proband). Of these 147 families, 67 (45.58%) were 3-generational, 57 (38.77%) were 4-generational and 23 (15.65%) were 5-generational families. Among nuclear families of the 147 probands, 86 (58.50%) (were Normal × Affected (Nor × Aff) families, ascertained through an affected parent and 61 (41.50%) were Normal × Normal (Nor × Nor) families, ascertained through an affected offspring.

Chapter 3

Epidemiological Results

3.1 Age at Onset

Each family was ascertained through exactly one proband. Of the 160 probands, 53 (33.12%) were male and 107 (66.88%). 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. The percent frequency distributions of proband's age at onset by gender are given in table 3.1.

To test whether the frequency distributions of age at onset are the same for males and females, we performed a non-parametric run test (Chakravarti *et al.* 1967). The test statistic, is :

$$T = [U - E(U)]/\sqrt{V(U)},$$

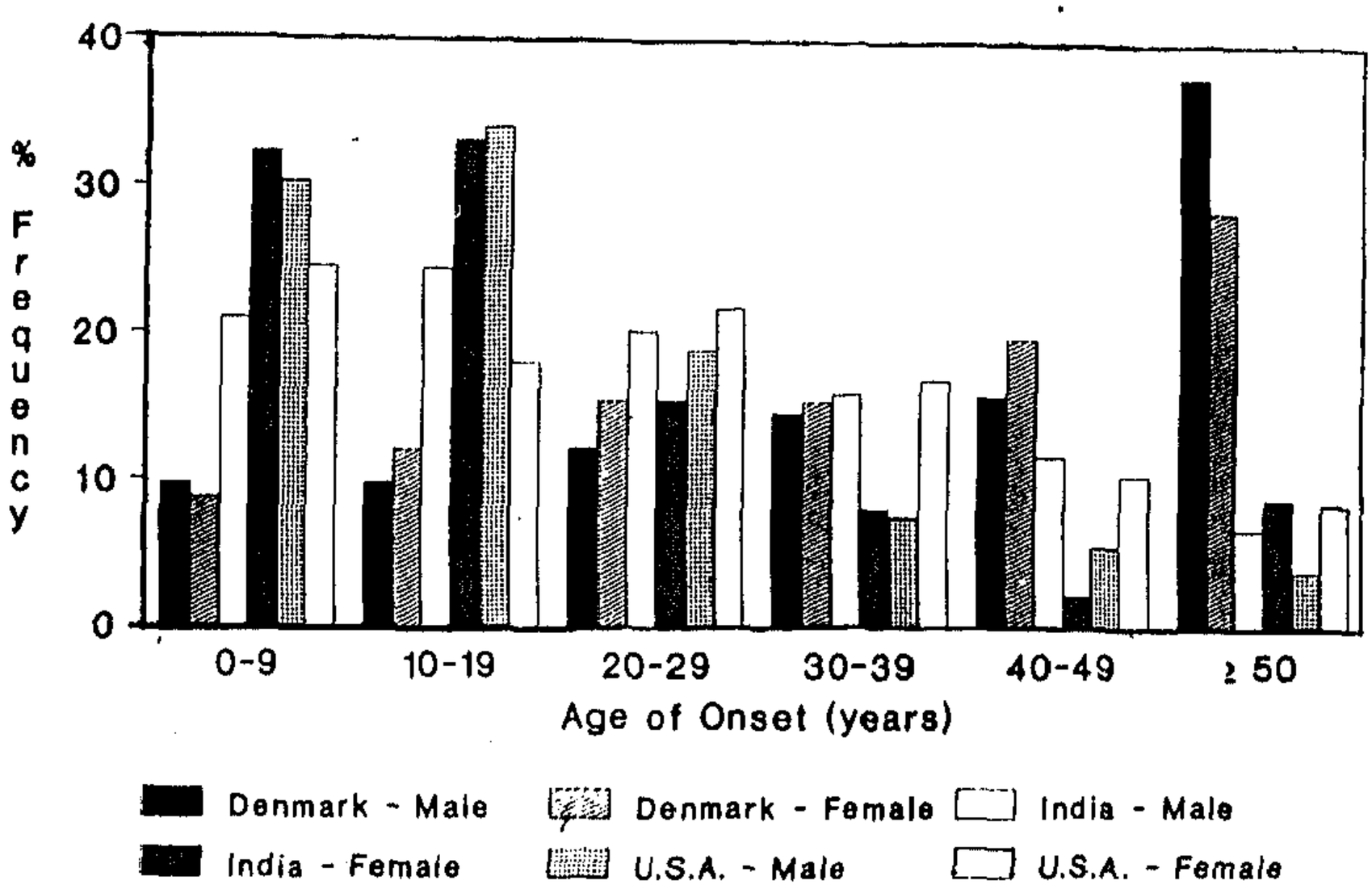
where U = observed number of runs, $E(U)$ = expected number of runs, and $V(U)$ = variance of the number of runs. For the present data, the values of $U, E(U), V(U)$ and T were, respectively, 63, 71.2, 30.6 and -1.48. Since the observed value of T exceeded -1.65 (the cut-off point corresponding to the 5% tail probability of a standard normal distribution), the null hypothesis that there is no difference in the frequency distributions of age at onset of males and females was accepted.

The age at onset distributions observed in the present study and in two previous studies (Howitz *et al.* 1977 and Das *et al.* 1985a) are presented, for

Table 3.1: Percent frequency distributions of proband's age at onset by gender

Age group	Male	Female	Total
≤ 4	3.77	6.54	5.62
5 - 9	28.30	17.76	21.26
10 - 14	24.53	9.35	14.38
15 - 19	7.55	8.41	8.12
20 - 24	7.55	14.96	12.50
25 - 29	11.32	7.48	8.76
30 - 34	3.77	9.34	7.50
35 - 39	3.77	7.48	6.25
40 - 44	1.89	4.67	3.75
45 - 49	3.77	4.67	4.37
50 - 54	1.89	6.54	5.00
55 - 59	0.00	1.87	1.25
60 - 64	1.89	0.00	0.62
≥ 65	0.00	0.93	0.62
Total	100.0	100.0	100.0

Figure 3.1: Frequency distributions of age at onset of vitiligo among genders in three countries



comparison, in figure 3.1. The most significant feature evident from this figure is that for both males and females, the age at onset of vitiligo in Denmark is later than that found in either India or the U.S.A. For the present data, the mean and the s.d. of age at onset among probands were 21.76 years and 14.67 years, respectively. The observed cumulative age at onset distribution is plotted in figure 3.2. Among male probands, the range of age at onset was 4 to 60 years; among females the range was 1 to 69 years. The mean ages at onset among male and female probands are given in table 3.2. It is seen that male probands, on an average, manifested vitiligo about seven years earlier than female probands. The mean ages at onset among all affected males and

Figure 3.2: Observed cumulative probability distribution of age at onset of vitiligo

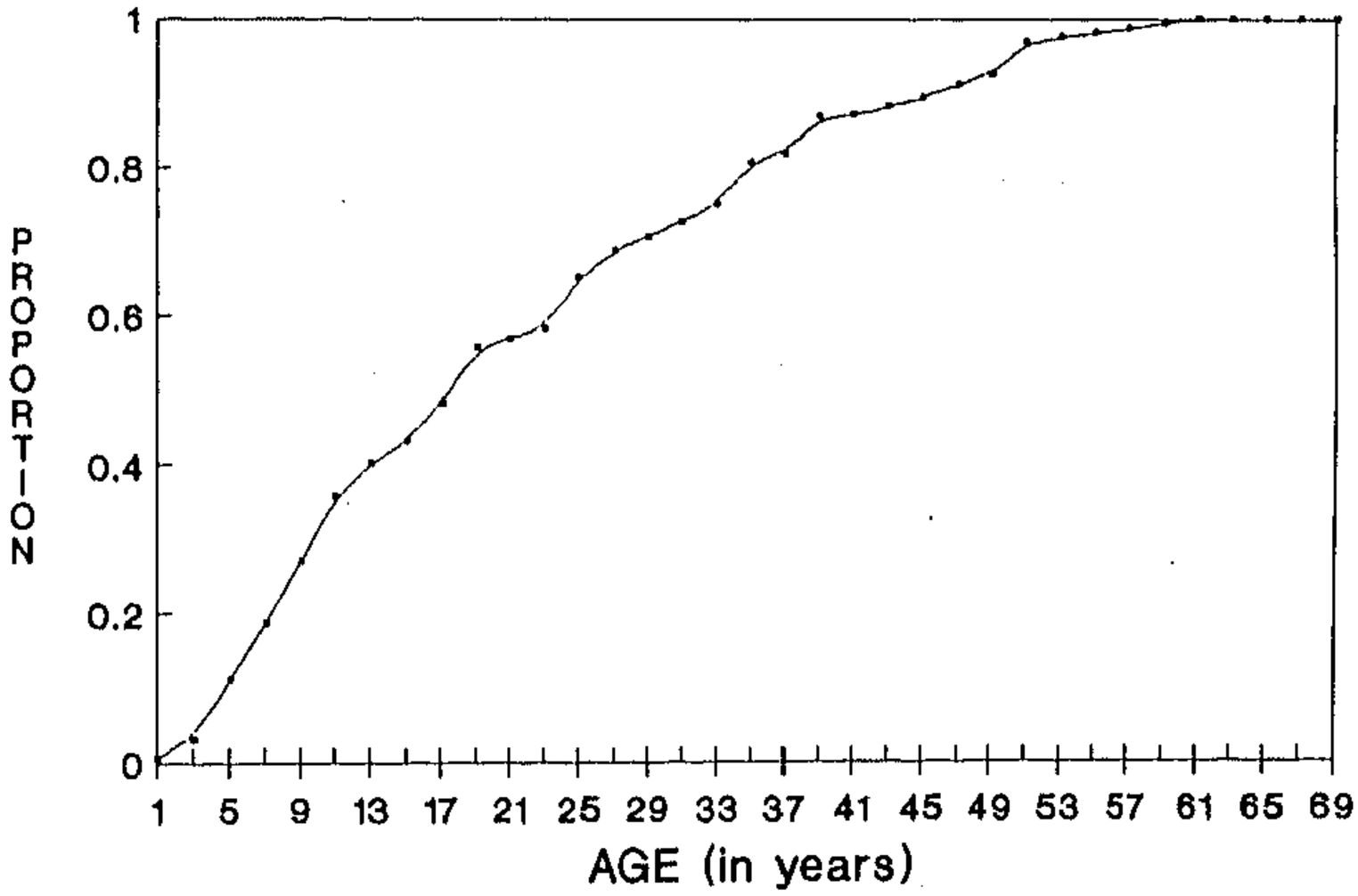


Table 3.2: Sample sizes and mean ages at onset (in years) $\pm$ s.e. among different categories of affected individuals

Gender of affected individual	Probands only	All affected individuals (including probands)	
		only first-degree relatives	all relatives
Male	53 17.62 ± 1.64	69 18.64 ± 1.63	73 18.64 ± 1.63
Female	107 23.81 ± 1.51	134 22.07 ± 1.40	143 24.19 ± 1.27
Total	160 21.76 ± 1.16	203 22.37 ± 1.03	216 22.31 ± 1.02

females are also given in table 3.2. To test the null hypothesis of the equality of mean ages at onset for males and females, we employed the test statistic :

$$T = \frac{(\bar{x}_1 - \bar{x}_2)\sqrt{(n_1 n_2)/(n_1 + n_2)}}{\sqrt{(n_1 s_1^2 + n_2 s_2^2)/(n_1 + n_2 - 2)}},$$

where n_1, $\bar{x}_1$ and s_1 denote, respectively, the sample size, mean age at onset and standard deviation of age at onset for males, and n_2, $\bar{x}_2$ and s_2 denote the corresponding values for females. The test statistic follows a t-distribution with $(n_1 + n_2 - 2)$ degrees of freedom under the null hypothesis. For the present data, the observed value of T for probands was 2.52; not significant at the 5% level with 158 degrees of freedom.

In the set of 160 families, there were 56 additional relatives (20 males and 36 females) of probands who also had vitiligo. Of these 56 affected relatives, 48 were first-degree relatives (20 males and 28 females). Therefore, including the probands, in the 160 families, there were 73 male and 143 female vitiligo

patients. The mean ages at onset among all affected individuals are also given in table 3.2. (Data on age at onset were unavailable for 4 male and 1 female affected first-degree relatives of probands.) These mean ages at onset are nearly identical to those observed among probands only.

3.2 Family History

The percentage of families with affected first-degree relatives of probands is an indicator of the extent of familial aggregation. In the 160 families, positive family history was observed in 49 families. That is, at least one relative of the proband in each of these 49 families had vitiligo. Of these 49 families, in 31 families at least one first-degree relative (offspring, parent, sibling) of the proband had vitiligo. Thus, the percentage of probands with at least one affected relative is about 30%, and with at least one affected first-degree relative is about 20%.

To determine whether there are any differences in characteristics of families when ascertained through a male proband or a female proband, we classified the families by gender of the proband. The data are summarized in table 3.3, from which it is seen that there is striking similarity in the proportions of families with positive family history whether the family was ascertained through a male or a female proband.

To test the null hypothesis of equality of proportions of multiplex families ascertained through a male and a female proband, we employed the test statistic :

$$Z = \frac{(\widehat{p}_1 - \widehat{p}_2)}{\sqrt{\widehat{p}(1-\widehat{p})(1/n_1 + 1/n_2)}},$$

where, $\widehat{p} = (y_1 + y_2)/(n_1 + n_2)$, y_1 and y_2 denote the numbers of multiplex families ascertained, respectively, through male and female probands, and n_1

Table 3.3: Proportions of multiplex family types classified by gender of proband

Proband's Gender	Number of Probands	Number (%) of probands with	
		at least one affected relative	at least one affected first degree relative
Male	53	16 (30.19)	11 (20.75)
Female	107	33 (30.84)	20 (18.69)
Total	160	49 (30.62)	31 (19.37)

and n_2 denote the total numbers of male and female probands, respectively. We reject the null hypothesis if the observed value of the test statistic is greater than $Z(\alpha)$, the $\alpha\%$ cut-off point corresponding to a $N(0,1)$ distribution. The observed Z values are small : $Z = 0.330$ for multiplex families with at least one affected relative and $Z = 0.139$ for multiplex families with at least one affected first-degree relative. Neither of these two values exceeded $Z(0.05) = 1.645$. We, therefore, accept the null hypothesis at the 5% level.

Further, to determine the nature of the distribution of affected relatives of the probands, we classified all individuals by age and relationship to proband. These data are presented in table 3.4. It is seen that the proportions of affected male and female relatives in various comparable relationship categories (*e.g.*, father vs. mother, brother vs. sister, etc.) are also remarkably similar. The notable exception is between uncles (0.6%) and aunts (2.02%). The observed difference (1.67% vs. 0%) between grandsons and granddaughters is perhaps attributable to small sample sizes. Among first-degree relatives of probands, for parents and siblings the percentage of affecteds is about 5%. This percentage among sons is about 8% and among daughters is about 9%. Thus, children seem to be afflicted between 1.6-1.8 times more commonly than other first-degree relatives.

Table 3.4: Numbers of normal and affected relatives of probands in different relationship categories and age groups

Relationship to proband	Affection status	Age group[a] (in years)									Total	
		1	2	3	4	5	6	7	8	9	No.	%
Father	Nor	0	0	0	15	19	24	45	29	16	148	95.48
	Aff	0	1	2	0	1	0	2	1	0	7	4.52
Mother	Nor	0	0	1	17	19	25	40	27	20	149	95.51
	Aff	1	0	0	3	1	1	0	1	0	7	4.49
Brother	Nor	8	11	24	49	27	16	14	6	2	157	95.15
	Aff	1	2	1	1	0	2	1	0	0	8	4.85
Sister	Nor	4	8	24	50	20	22	19	4	1	152	94.41
	Aff	1	2	2	1	1	1	1	0	0	9	5.59
Grandfather	Nor	0	0	3	5	6	36	54	55	52	211	96.35
	Aff	0	1	0	0	0	3	1	1	2	8	3.65
Grandmother	Nor	0	0	1	5	10	26	50	58	70	220	96.07
	Aff	2	1	0	3	0	0	0	2	1	9	3.93
Uncle	Nor	5	3	12	38	42	48	78	76	27	329	99.40
	Aff	0	1	0	0	0	0	0	0	1	2	0.60
Aunt	Nor	5	3	14	38	24	59	84	63	50	340	97.98
	Aff	0	1	2	1	0	0	1	0	2	7	2.02
Son	Nor	12	17	37	33	7	0	0	0	0	106	92.17
	Aff	3	3	1	1	1	0	0	0	0	9	7.83
Daughter	Nor	20	24	32	24	8	0	0	0	0	108	90.76
	Aff	2	4	3	1	1	0	0	0	0	11	9.24
Grandson	Nor	41	10	8	0	0	0	0	0	0	59	92.33
	Aff	1	0	0	0	0	0	0	0	0	1	1.67
Granddaughter	Nor	40	10	4	0	0	0	0	0	0	54	100.00
	Aff	0	0	0	0	0	0	0	0	0	0	0.00

[a]Age groups (in years) are : 1 = 0-9; 2 = 10-19; 3 = 20-29; 4 = 30-39; 5 = 40-49; 6 = 50-59; 7 = 60-69; 8 = 70-79; 9 = $\geq$ 80

3.3 Relative Risk

To determine whether being biologically related to an affected individual results in an increased risk of being affected with vitiligo compared to an individual of the same age and gender drawn randomly from the population, we computed the population relative risk (Penrose 1953; Weiss *et al.* 1982). The population relative risk (RR), also known as the *standardized mortality ratio* [SMR] (Breslow and Day 1980), is defined as : $RR = n_a/n_e$, where n_a = observed number of affected individuals; n_e = expected number of affected individuals $= \sum_{i=1}^{G} N_i p_i$; G = number of age groups; N_i = total number of individuals in $i-$th age group ($i = 1, 2, \ldots, G$) and p_i = affection probability for an individual belonging to the $i-$th age group ($i = 1, 2, \ldots, G$) [It may be recalled that no significant gender difference in age-specific prevalences was found.] Since the age- and gender-specific estimates of population prevalence of vitiligo among U.S. Caucasians are, to the best of our knowledge, unavailable, we have used two available sets of estimates derived from the (Caucasian) Danish (Howitz *et al.* 1977) and Indian (Das *et al.* 1985a) epidemiological studies. The use of two sets of prevalence estimates additionally permitted the investigation of the variation in relative risks caused by using different prevalence estimates of the two diverse ethnic groups. The confidence intervals of RR estimates were obtained using the method given by Ulm (1990). Ulm's method is based on the relation between the Poisson distribution and the χ^2 distribution :

$$\begin{aligned} \sum_{i=0}^{m-1} e^{-\lambda}\lambda^i/i! &= 1 - \text{Prob}\{\chi^2_{2m} \leq 2\lambda\} \\ &= \text{Prob}\{\chi^2_{2m} > 2\lambda\} \end{aligned}$$

The lower (λ_L) and upper (λ_U) boundaries of the (1-α)% confidence interval of the Poisson-distributed variable Y = observed number of affected individuals,

are defined as :

$$\begin{aligned} \text{Prob}(\lambda_L) &= \sum_{i=Y}^{\infty} e^{-\lambda_L} \lambda_L^i / i! = \alpha/2 \\ \text{Prob}(\lambda_U) &= \sum_{i=0}^{Y} e^{-\lambda_U} \lambda_U^i / i! = \alpha/2. \end{aligned}$$

Based on relation between Poisson and χ^2 distributions given above, the values λ_L and λ_U can easily be obtained from tables of χ^2 distribution :

$$\begin{aligned} \text{Prob}(\lambda_L) &= \text{Prob}\{\chi^2_{2Y} \leq 2\lambda_L\} = \alpha/2 \\ 1 - \text{Prob}(\lambda_U) &= \text{Prob}\{\chi^2_{2(Y+1)} \leq 2\lambda_U\} \\ &= 1 - \alpha/2 \end{aligned}$$

where, $2\lambda_L$ is simply the $\alpha/2$ fractile of a χ^2 variable with 2Y degrees of freedom, while $2\lambda_U$ is the (1 - $\alpha/2$) fractile of a χ^2 variable with $2(Y + 1)$ degrees of freedom. These values, λ_L and λ_U, divided by the expected number of events are the limits of the 95% confidence interval of the corresponding *RR*. If the value of 1.0 is included in the confidence interval, then the corresponding *RR* is not significant.

The observed numbers of affected individuals, the estimates of *RR*, and their confidence intervals are given in table 3.5 separately for various observed types of relatives of probands. Offspring of probands were found to have the highest risk of developing vitiligo, followed by siblings, parents and grandparents. There was a remarkable similarity in estimates of relative risk between genders within any relationship category. This implies that the risk of developing vitiligo for any relative of the proband is independent of the gender of the relative, but is strongly dependent on his/her kinship coefficient with the proband. It was also interesting to note that the estimates of relative risk are similar whether the Danish prevalence data or the Indian prevalence data are used. From the confidence intervals of relative risks, it was seen that the value of unity was outside the confidence intervals for all relationship categories except for uncles and grandsons. (For grandsons, the lower bound of the 95% confidence interval derived using the Indian prevalence estimate was

1.05, which is slightly greater than 1.) Since $RR = 1$ implies no elevation of risk for developing vitiligo, and since the value of unity was 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 sharing of genetic factors and/or exposure to common familial environmental factors.

Table 3.5: Observed and expected numbers of affected relatives of probands, and relative risks

Relationship to proband	Total no.	No. of affecteds			Relative Risk[a]		95% confidence interval of Relative Risk[a]	
		Observed	Expected[a]					
			(1)	(2)	(1)	(2)	(1)	(2)
Father	155	7	0.97	0.67	7.20	10.41	2.89-14.84	4.19-21.45
Mother	156	7	0.98	0.88	7.16	7.93	2.88-14.76	3.19-16.33
Brother	165	8	0.60	0.72	13.36	11.04	5.77-26.32	4.76-21.74
Sister	161	9	0.77	0.83	11.68	10.77	5.34-22.17	4.93-20.45
Grandfather	219	8	1.38	0.92	5.81	8.65	2.51-11.45	3.74-17.05
Grandmother	229	9	1.46	1.34	6.16	6.72	2.82-11.69	3.07-12.75
Uncle	331	2	1.93	1.43	1.03	1.40	0.12-3.74	0.17-5.06
Aunt	347	7	2.11	1.73	3.32	4.05	1.33-6.83	1.63-8.35
Son	115	9	0.25	0.47	35.27	19.21	16.13-66.95	8.78-36.47
Daughter	119	11	0.29	0.55	37.61	20.07	18.77-67.29	10.02-35.90
Grandson	60	1	0.06	0.24	16.18	4.20	0.41-90.16	1.05-23.38
Granddaughter	54	0	0.07	0.27	-	-	-	-

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

3.4 Impact of Proband's Age at Onset on Relative Risk

Age at onset can, conceptually, provide indication of genetic involvement. An affected individual with an early age at onset is often viewed as a genetic case, while an affected individual with a late age at onset is viewed as a sporadic/nongenetic case. This viewpoint seems to hold in the case of cancers of some particular sites, such as breast (Skolnick *et al.* 1990). If this be true, then one would expect that risks to relatives of probands with early ages at onset will be significant, while the risks to relatives of probands with late ages at onset will be smaller and non-significant.

To investigate the impact of proband's age at onset on risk of developing vitiligo to relatives of a proband, we have classified the families into two classes based on proband's age at onset : ≤ 17 years and > 17 years. The cut off point of 17 years has been used because the proband's age at onset distribution showed an antimode at approximately in this age. The observed numbers of affected individuals in various relationship categories, estimated relative risks (RR), and their confidence intervals are given in table 3.6. From the confidence intervals, it is seen that RR values are all significant whether they pertain to relatives of probands with early ages at onset or late ages at onset, because the value of unity is outside the corresponding confidence intervals.

Further, for all types of relatives, the difference of RR values between the two classes of probands is not statistically significant at the 5% level, as is evident from the overlaps of the confidence intervals of RR values. The present data, therefore, do not provide any evidence for a statistically significant effect of proband's age at onset on the risk of developing vitiligo to first degree relatives of the proband.

Table 3.6: Observed and expected numbers of affected relatives of probands and risks to first-degree relatives of probands with early and late ages at onset

Relationship to proband	Age at onset of proband (in years)	Total no.	No. of affecteds			Relative risk[a]		95% confidence interval of relative risk[a]	
			Observed	Expected[a]					
				(1)	(2)	(1)	(2)	(1)	(2)
Father	≤ 17	75	4	.412	.3398	9.71	11.77	2.64-24.86	3.21-30.14
	> 17	80	3	.586	.3323	5.12	9.03	1.05-14.96	1.86-26.38
Mother	≤ 17	76	4	.4424	.4126	9.04	9.69	2.46-23.15	2.64-24.82
	> 17	80	3	.538	.4701	5.61	6.38	1.15-16.30	1.50-18.65
Brother	≤ 17	76	3	.2267	.3319	13.23	9.04	2.73-38.67	1.86-26.41
	> 17	89	5	.367	.393	13.62	12.72	4.43-31.79	4.13-29.69
Sister	≤ 17	76	4	.3086	.3729	12.96	10.72	3.53-33.17	2.92-27.46
	> 17	85	5	.4468	.440	11.19	11.36	3.63-26.11	3.69-26.52
Son	≤ 17	35	4	.0683	.1494	58.56	26.78	15.96-149.95	7.30-68.56
	> 17	80	5	.1869	.3191	26.75	15.67	8.69-62.43	5.09-36.57
Daughter	≤ 17	31	3	.0661	.1120	45.38	21.11	9.36-132.64	5.52-78.33
	> 17	88	8	.2264	.4063	35.33	19.69	15.26-69.62	8.50-38.80

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

3.5 Age at Onset Correlations Among Affected Relatives

Clues regarding possible aetiological heterogeneity can be obtained from the distribution of ages at onset among affected members in families. Presence of many young affected individuals in a family may be indicative of genes segregating, compared to 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 ages at onset; while in those families in which there is no gene segregating, ages at onset of affected individuals may be highly variable. One would, therefore, expect that there will be a high intra-class correlation in the ages at onset of affected family members if there are segregating genes in most families.

To investigate this, we calculated intra-class correlation coefficient of ages at onset of vitiligo in multiplex families taking (i) only affected first-degree relatives of probands, and (ii) all affected relatives of probands.
The intra-class correlation coefficients were estimated using the analysis of variance technique (Chakravarti *et al.* 1967), results of which are presented table 3.7. From this table, it is seen that the intra-class correlation coefficient for ages at onset among first-degree relatives is fairly high (0.61). The value of the correlation coefficient changed only minimally when all affected relatives are considered (0.57). These results indicate that there is possibly little heterogeneity in the nature of vitiligo families.

Table 3.7: Results of analysis of variance and estimates of intra-class correlation coefficient for ages at onset among affected family members

Affected family members considered	Source of variation						Intra-class correlation coefficient
	Between families			Within families			
	Sum of squares	d.f	Mean square	Sum of squares	d.f.	Mean square	
First-degree relatives	15685.23	33	475.31	6295.53	55	114.46	0.61
All relatives	16895.18	40	422.38	7361.20	63	116.84	0.57

3.6 Effect of Parental Age at First Childbirth

It is relevant to test whether the age of father/mother at first childbirth has any effect on the frequency of vitiligo among offspring. For many chromosomal syndromes (*e.g.*, Down syndrome), the probability of having an affected offspring is known to increase with increase in the ages of the parents. To investigate whether such a relationship exists for vitiligo, the data on age at first childbirth were classified separately for each parent of each proband. To eliminate gross generational effects on ages at marriage and/or first childbirth, parents born during the period 1900-1930 and those born after 1930 have been considered separately. For purposes of comparison, one control per proband was selected from this data set. For choosing a control, we first enumerated a set of 'eligible' controls for each proband. An eligible control was defined as an unaffected individual whose age was within one year of the age of the proband under consideration, and whose parents were born in the same period (1900-1930 or 1930+) as the parents of the proband. From the list of eligible controls, one control was randomly chosen for each proband. Probands with

no eligible controls or for whose parents the age at birth at first childbirth was unavailable had to be excluded from this analysis. In computing the proportion of affected offspring, the proband was excluded. (Inclusion of probands would have resulted in ascertainment bias since the families were ascertained through the probands.) The relevant data are presented in table 3.8.

Table 3.8: Effect of parental age at first childbirth on proportion of offspring with vitiligo

Father's age at first childbirth	Father born during 1900-1930								Father born after 1930							
	probands				controls				probands				controls			
	n_f	n_c	n_a	p_a	n_f	n_c	n_a	p_a	n_f	n_c	n_a	p_a	n_f	n_c	n_a	p_a
$<$ 25 Years	13	51	14	.274	18	84	12	.143	27	68	29	.426	29	73	11	.151
$\geq$ 25 Years	33	111	38	.342	28	113	19	.168	22	55	25	.454	20	46	6	.130
Total	46	162	52	.321	46	197	31	.157	49	123	54	.439	49	119	17	.143
Mother's age at first childbirth	Mother born during 1900-1930								Mother born after 1930							
	probands				controls				probands				controls			
	n_m	n_c	n_a	p_a	n_m	n_c	n_a	p_a	n_m	n_c	n_a	p_a	n_m	n_c	n_a	p_a
$<$ 25 Years	27	96	31	.323	25	117	20	.171	40	111	45	.405	40	111	14	.126
$\geq$ 25 Years	14	49	15	.306	16	58	9	.155	14	29	15	.517	14	30	5	.167
Total	41	145	46	.317	41	175	29	.166	54	140	60	.428	54	141	19	.135

n_f = number of fathers; n_m =number of mothers; n_c = total number of offspring; n_a = number of affected offspring; $p_a = n_a/n_c$ = proportion of affected offspring.

To avoid vagaries of small sample sizes, we grouped parents into two classes : those with ages at first childbirth $\leq$ 25 years, and those with ages at first childbirth $>$ 25 years. We tested the null hypothesis that the difference in proportions of affected offspring is not different for parents belonging to these two classes. This hypothesis was tested separately for fathers and mothers, separately for the two year of birth intervals 1900-1930 and 1930+ , and also separately for probands and controls. A $N(0,1)$ test of equality of proportions (described in the section 3.2 of this chapter) was performed in each case. None of the observed values of the test statistic was significant at the 5% level.

We also used a cut-off point of 30 years for classifying parents by age at first childbirth. Even in this case, the null hypotheses were all accepted at the 5% level. We, therefore, found no evidence of the effect of increased parental age at first childbirth on the probability of having an affected offspring.

We further tested the null hypothesis that the differences in proportions of affected offspring for parents with age at first childbirth ≤ 25 years and > 25 years are equal for proband parents and control parents. This null hypothesis was again tested separately, using a $N(0,1)$ test statistic, for the two periods of year of birth and also separately for fathers and mothers. None of these tests was significant at the 5% level. Therefore, even though the proportions of affected offspring were greater for proband parents than for control parents, the differences in proportions of affected offspring observed between these two groups of parents based on the age at first childbirth were not significant.

3.7 Associated Factors

In addition to parental ages at first childbirth, we have studied some other factors that are allegedly positively associated with the probability of affection to vitiligo. In table 3.9, we present the proportions of individuals reporting these factors in various subclasses of families. These were derived from the responses of probands. In spite of the limitation that there were multiple responses from a large number of probands, this table does not indicate any strong association of any of the alleged factors with family history of vitiligo. While the list of correlates considered is not exhaustive, indications are that the causes of elevation of risk of vitiligo to biological relatives of an affected individual is primarily genetic. This, of course, does not rule out the possibility that an environmental insult/exposure is required to trigger phenotypic expression in

genetically predisposed individuals. No statistical analyses of these data were performed because of small cell frequencies.

Table 3.9: Responses of 76 probands to various problems/factors implicated in *vitiligo*

Reported Problem or participation	Category	Type of family		
		No positive family history (n = 93)	Positive family history but no affected 1^0 relative (n = 14)	Positive family with at least one affected 1^0 relative (n = 23)
Sunbathing, Sunburning, Suntanning, etc.	Frequent	16(17.2)*	1 (7.1)	-
	Occasional	47 (50.5)	8 (57.1)	14 (60.9)
Stress		35 (37.6)	8 (57.1)	8 (34.8)
Thyroid disorder	Hypo	10 (10.8)	-	6 (26.1)
	Hyper	4 (4.3)	-	-
	Other	2 (2.2)	-	1 (4.4)
Early hearing loss		7 (7.5)	1 (7.1)	1 (4.4)
Ocular problem		-	1 (7.1)	2 (8.8)
Alopecia areata		1 (1.1)	-	1 (4.4)
Childhood anemia		3 (3.2)	-	-
Exposure to chemicals		4 (4.3)	1 (7.1)	-
Other (Asthma, Melanoma, Arthritis, Skinning, Burn, etc.)		19 (10.4)	4 (28.6)	1 (4.4)

* figure in parentheses indicates percentages

Chapter 4

Multilocus Models with Special Reference to the Multilocus Recessive Model

4.1 Models Considered

Many of the challenging problems in the study of hereditary disorders involve use of mathematical/statistical modeling to describe the transmission of a disorder within families. Many complex disorders have not been amenable to genetic analysis under the assumption of single locus or multifactorial models. The observed familial risks are often inexplicable under any single locus or multifactorial models and also segregation analysis has often not been decisive (Neuman and Rice 1992). Consequently, interest has turned to the consideration of the properties of oligogenic models, *i.e.*, genetic models involving a small number of genes. Development of multilocus models have been discussed in Chapter 1. The interaction of nonallelic genes is referred to as epistasis. When the phenotypic expression of a trait is determined by independent genetic mechanisms, the term heterogeneity is used.

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. For example, for a disorder controlled by two

loci, 50 phenograms (genotype-phenotype relationship descriptions) are possible. However, not all of these phenograms are biologically meaningful. Therefore, usually only a subset of these phenograms are considered. Neuman and Rice (1992) have discussed nine of these two-locus models. Of them, six are models of epistasis and three are models of heterogeneity.

To understand the behaviour of oligogenic models, we shall first consider the simplest case — two autosomal, biallelic, unlinked loci. Locus 1 has alleles A and a with frequencies p_1 and q_1 $(= 1 - p_1)$, respectively; the two alleles at locus 2, B and b, have frequencies p_2 and q_2 $(= 1 - p_2)$, respectively. It is assumed that the underlying population is in Hardy-Weinberg equilibrium with respect to each of the two loci and that there is no linkage disequilibrium between the loci. We also assume that penetrances are equal (say f for epistatic models) for all at-risk genotypes and that only those individuals with an at-risk genotype may become affected (*i.e.*, no phenocopies). In heterogeneity models, the penetrance (say g) of a genotype with respect to both loci is computed from marginal penetrances : $g = f_1 + f_2 - f_1 f_2$, where f_1 and f_2 denote the marginal penetrances at loci 1 and 2, respectively. The genotypes and their corresponding population frequencies for the general two locus model are given in table 4.1. The notation used to denote penetrances of the corresponding two-locus genotypes are given in table 4.2. Now for example, consider a disorder that expresses itself in an individual if either (s)he is a recessive homozygote at each of the two loci (epistatic model) or is a recessive homozygote at any one of the two loci (heterogeneity model). The penetrances of the genotypes for these models are presented in table 4.3. We have focussed on the epistatic and the heterogeneity models because these two models have been considered in the present study in view of the findings of a previous family study on vitiligo (Majumder *et al.* 1988), and the findings of recent molecular genetic studies on pigmentation (reviewed in Chapter 3). These models, particularly the epistatic model, are described in greater detail and generality below. The

Table 4.1: Two-locus genotypes and their frequencies in a population in Hardy-Weinberg equilibrium

Loci		Locus 2		
		BB	Bb	bb
Locus 1	AA	AABB	AABb	AAbb
		$(p_1^2 p_2^2)$	$(2p_1^2 p_2 q_2)$	$(p_1^2 q_2^2)$
	Aa	AaBB	AaBb	Aabb
		$(2p_1 q_1 p_2^2)$	$(4p_1 q_1 p_2 q_2)$	$(2p_1 q_1 q_2^2)$
	aa	aaBB	aaBb	aabb
		$(q_1^2 p_2^2)$	$(2q_1^2 p_2 q_2)$	$(q_1^2 q_2^2)$

Table 4.2: Penetrances of two-locus genotypes

Loci		Locus 2		
		BB	Bb	bb
Locus 1	AA	h_1	h_2	h_3
	Aa	h_4	h_5	h_6
	aa	h_7	h_8	h_9

Table 4.3: Penetrances for the genotypes for two different models

Model	h_1	h_2	h_3	h_4	h_5	h_6	h_7	h_8	h_9
Epistatic	0	0	0	0	0	0	0	0	f
Heterogeneity	0	0	f_2	0	0	f_2	f_1	f_1	g

epistatic model is emphasized because this is the model that is being considered for cross-validation.

4.1.1 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 biallelic. 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 1 genotype gives rise to the affected phenotype; individuals of the remaining $3^L - 1$ genotypes are phenotypically normal. This model was introduced by Li (1953, 1987), who also derived many population characteristics of this model. Some additional properties of this model have been derived by Majumder and Nath (1992). This model yielded an adequate fit to family data on vitiligo collected earlier from Calcutta, India (Majumder *et al.* 1988). Some properties of this model are discussed in detail in the next section.

4.1.2 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, AaBb are phenotypically normal.

4.2 Some Population and Sample Characteristics When a Multilocus Recessive Disorder is Expressed at Birth

Since Model I is the major focus of this study, we shall mainly concentrate on this model in this section.

4.2.1 Population prevalence

Consider a disorder determined by the epistatic action of recessive alleles at multiple unlinked loci (Model I). Suppose, q_i denotes the frequency of the recessive allele at the biallelic locus i $(i = 1, 2, \ldots, L)$ in a population. If the population practises random mating, then the prevalence (δ) of the disorder in the population is :

$$\delta = \prod_{i=1}^{L} q_i^2$$

If $q_i = q$ (for all $i = 1, 2, 3, \ldots, L$), then,

$$\delta = q^{2L}, \qquad [0 < \delta,\ q < 1].$$

For a disorder which is due to recessive homozygosity at any one of the L loci involved (Model II)

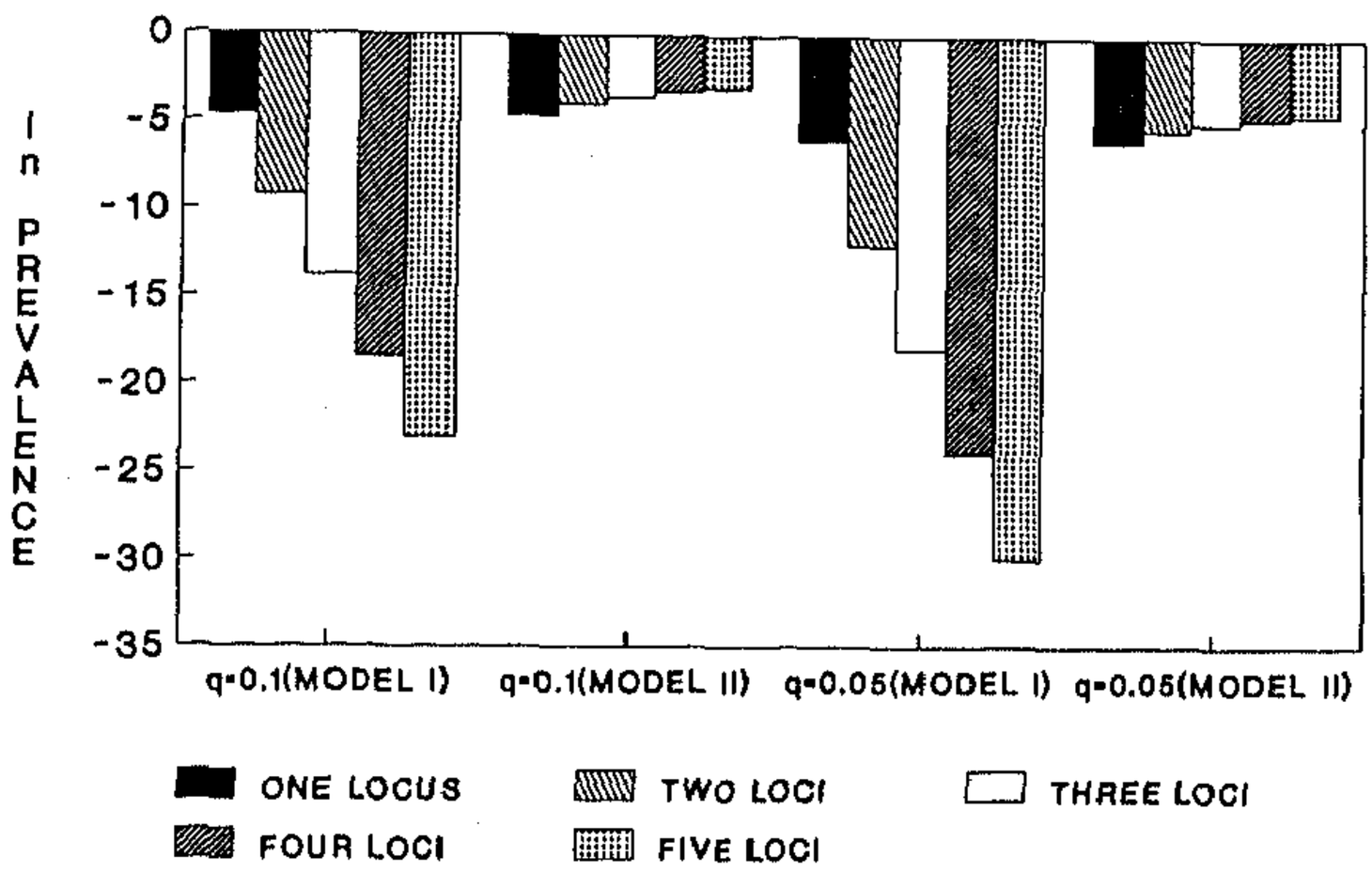
$$\delta = [1 - \prod_{i=1}^{L}(1 - q_i^2)].$$

If $q_i = q$ (for all $i = 1, 2, \ldots, L$), then,

$$\delta = [1 - (1 - q^2)^L], \qquad [0 < \delta,\ q < 1].$$

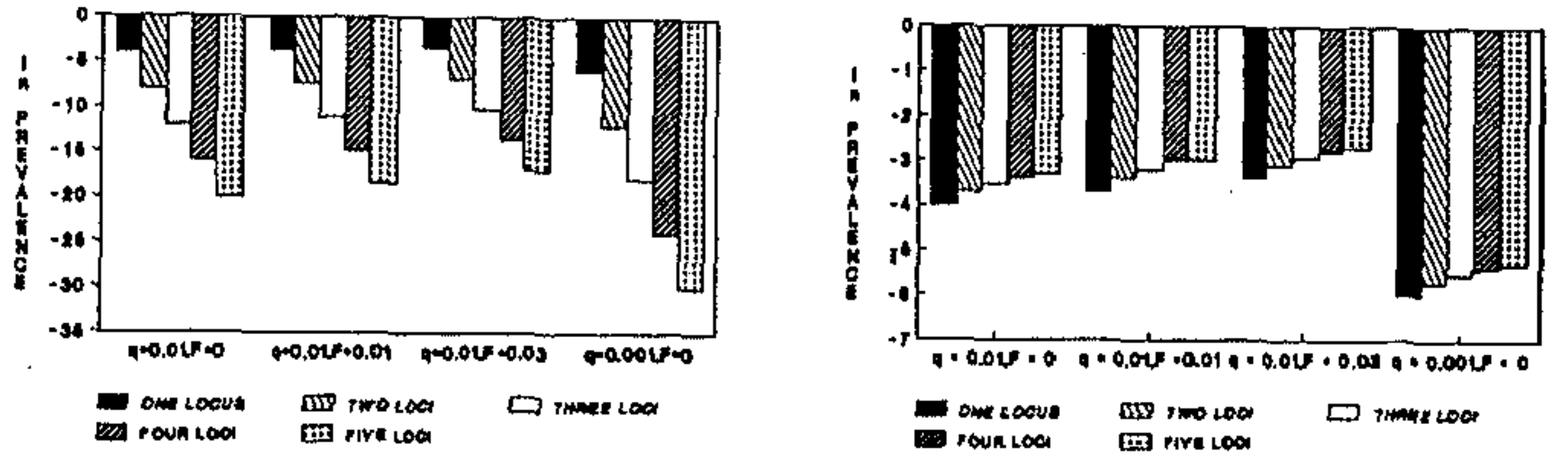
For Model I, the prevalence, for a fixed value of the allele frequency q, decreases sharply with the increase in the number of loci L. For model II, however,

Figure 4.1: *ln*-Prevalence in a random-mating population under models I and II as functions of L and q



the prevalence increases with the increase of the number of loci (figure 4.1). The prevalence, for a fixed value of L, increases with increase in the allele frequency q for both models. It is seen analytically, as also, from this figure, that for both models the prevalence is exactly same when $L = 1$. For a recessive disorder, it is of interest to examine its characteristics in an inbred population. Suppose, a population practises inbreeding and the inbreeding coefficient is $F(F > 0)$. Then, at the i-th locus $(i = 1, 2, 3, \ldots, L)$, the frequency of the recessive homozygote will be $q_iF + q_i^2(1-F)$ for model I and $1 - [1 - \{q_iF + q_i^2(1-F)\}]$ for model II. If $q_i = q$ $(i = 1, 2, 3, \ldots, L)$, then, prevalence of a disorder for models

Figure 4.2: *ln*-Prevalence in an inbred population under models I and II as functions of q, L and F



I and II will be :

$$\delta = [qF + q^2(1-F)]^L$$

and,

$$\delta = 1 - [1 - \{qF + q^2(1-F)\}]^L.$$

The prevalence under models I and II as functions of q, L and F are depicted in figure 4.2. As is expected, for fixed values of q and L, prevalence increases with increase in the population inbreeding coefficient F. Further, the rate of decrease in population prevalence under model I, or the rate of increase in population prevalence under model II, with increase in L for a fixed value of q is dependent on the value of the population inbreeding coefficient, F.

Table 4.4: Mating types, their frequencies and segregation probabilities in the population

Mating type	No. of genotypic mating	Frequency*	P(offspring is affected given parental mating)
Nor × Nor	$3^L(3^L-1)/2$	$(1-Q^2)^2$	$S_2 = S_1^2$
Nor × Aff	$3^L - 1$	$2Q^2(1-Q^2)$	$S_1 = Q/(1+Q)$
Aff × Aff	1	1	$S_0 = 1$
Total	$3^L(3^L+1)/2$	1	-

* $Q^2 = \prod_{i=1}^{L} q_i^2$

4.2.2 Types and proportions of families in population and in sample

Since there are only two phenotypes (Normal and Affected), there will be three different types of phenotypic matings in the population : Normal × Normal (Nor × Nor); Normal × Affected (Nor × Aff); Affected × Affected (Aff × Aff). However, because individuals of the normal phenotype comprise $3^L - 1$ genotypes, Nor × Nor matings comprise a large number of different genotypic matings. Li (1987) worked out the phenotypic and genotypic mating frequencies and their segregation ratios in the general population. These are provided in table 4.4.

While the above results are a description of characteristics of the disorder in families drawn at random from the population, the mode of inheritance of such a disorder cannot be effectively studied by sampling families at random from the population. For such studies, families need to be sampled non-randomly, because the vast majority of randomly drawn families will not even be capable of producing an affected child. The standard methods of sampling families non-

randomly are : (i) to ascertain families through an affected parent, and (ii) to ascertain families through an affected offspring. When a family is ascertained through an affected parent, no correction of any kind is necessary to take into account the effect of non-random sampling. The observed proportion of affected offspring in the sampled Nor × Aff families can be compared with the expected proportion S_1 given in table 4.4 to check agreement.

When families are ascertained through an affected offspring, the observed proportion of affected offspring is grossly upwardly biased because not only are these families capable of producing an affected offspring, but only those families that have actually produced an affected child come under study. Later we shall discuss some methods of correcting for this ascertainment bias in the analysis of family data for determining the mode of inheritance of a complex disorder. For now, we discuss some more characteristics of the model under consideration.

In a sample of families ascertained through an affected offspring, the proportions of the 3 mating types (Nor × Nor; Nor × Aff; Aff × Aff), which are functions of the number of loci involved in the determination of the phenotype, may yield clues to the number of loci involved.

Since ascertainment is through an affected offspring, the father has to be heterozygous at i loci ($i = 1, 2, 3, \ldots, L$) and recessive homozygous at $(L - i)$ loci and the mother heterozygous at j loci ($j = 1, 2, 3, \ldots, L$) and recessive homozygous at $(L - j)$ loci. Therefore,

(1) if $i = 0$ and $j = 0$, both parents are affected;

(2) if $i \neq 0$ and $j \neq 0$, both parents are normal;

(3) if $i \neq 0$ and $j = 0$, father is normal and mother is affected;

(4) if $i = 0$ and $j \neq 0$, father is affected and mother is normal.

Now, the total mating frequency given that the family is ascertained through

an affected child is :

$$T = \sum_{i=0}^{L}\sum_{j=0}^{L}\binom{L}{i}\binom{L}{j}H^{i+j}R^{2L-i-j}$$

where, $H = 2q(1-q)$ denotes heterozygosity and $R = q^2$ denotes recessive homozygosity. [Although H and R are assumed to be same at all loci for algebraic simplicity, generalization for unequal values is straightforward.] Now suppose,

$$\begin{aligned} p_{11} &= \text{Prob (both parents are affected);} \\ p_{00} &= \text{Prob (both parents are normal);} \\ p_{10} &= \text{Prob (one parent is affected and other is normal).} \end{aligned}$$

Then, it is easy to show that

$$\begin{aligned} p_{11} &= R^{2L}; \\ p_{00} &= \sum_{i=1}^{L}\sum_{j=1}^{L}\binom{L}{i}\binom{L}{j}H^{i+j}R^{2L-i-j}; \\ p_{10} &= R^{L}[\sum_{j=1}^{L}\binom{L}{j}H^{j}R^{L-j} + \sum_{i=1}^{L}\binom{L}{i}H^{i}R^{L-i}]; \\ &= 2R^{L}[\sum_{i=1}^{L}\binom{L}{i}H^{i}R^{L-i}]. \end{aligned}$$

Hence, in a set of ascertained families, the probabilities of various mating types are :

$$\begin{aligned} \text{Prob (Nor} \times \text{Nor)} &= p_{00}/T; \\ \text{Prob (Nor} \times \text{Aff)} &= p_{10}/T; \\ \text{Prob (Aff} \times \text{Aff)} &= p_{11}/T. \end{aligned}$$

The percentage distributions of the different types of matings in a sample of families ascertained through an affected offspring are given in figure 4.3 for

different values of prevalence and number of loci. As is expected, the proportion of Nor × Nor families increases with an increase in the number of loci for a fixed prevalence and the proportions of Nor × Aff and Aff × Aff matings correspondingly decrease. The trend is similar if the prevalence decreases for a fixed number of loci. Unfortunately, however, for a fixed prevalence, the proportions of the three mating types do not change drastically with the increase of number of loci. This means that it is virtually impossible to estimate the number of involved loci by using the data on proportions of different mating types.

4.2.3 Likelihood of a Normal × Normal family ascertained through an affected offspring

To determine the likelihood of a Normal × Normal family ascertained through an affected offspring, we first note that each of the normal parents must either be heterozygous at each of the L loci or recessive homozygous at all L loci except at least one. This is because, to produce an affected (aabbcc...) offspring, each parent must be capable of transmitting an abc... gamete, and the reason why neither parent can be recessive homozygous at all the L loci is that each parent is known to be phenotypically normal. Hence, for any such family in which the father is heterozygous at i loci and the mother is heterozygous at j loci, the mating frequency, M_{ij} is :

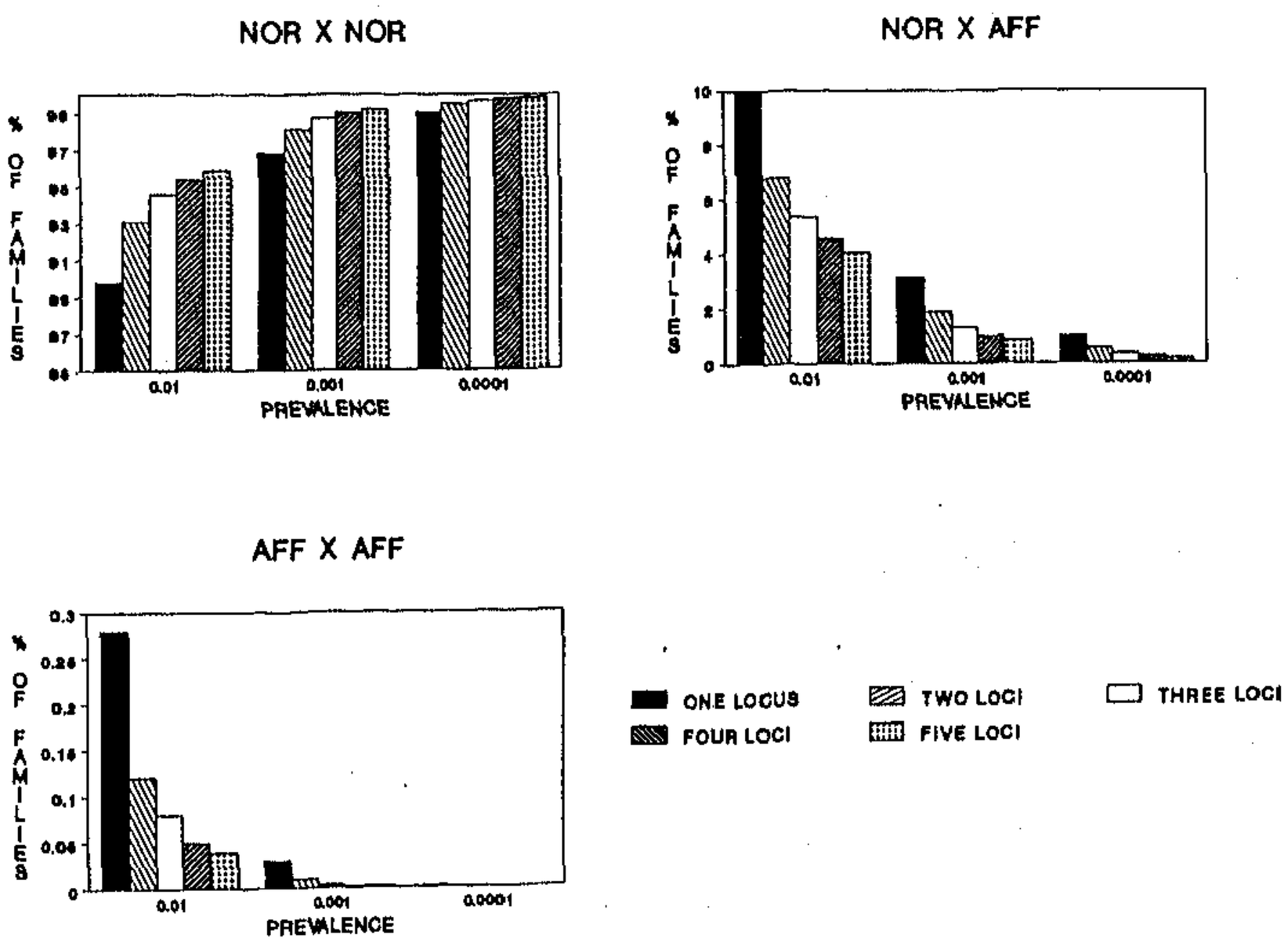
$$M_{ij} = \frac{\binom{L}{i}\binom{L}{j} H^{i+j} R^{2L-i-j}}{\sum_{i=1}^{L}\sum_{j=1}^{L}\binom{L}{i}\binom{L}{j} H^{i+j} R^{2L-i-j}}$$

The probability θ_{ij} that this family produces an produces an affected offspring is :

$$\theta_{ij} = \frac{1}{2^{i+j}}$$

Now, the probability, α_r, that a family with r affected offspring will have

Figure 4.3: Percentage distributions of mating types in families ascertained through an affected offspring for different levels of prevalence and number of loci



at least one proband is (Elandt-Johnson 1971) :

$$\alpha_r = 1 - (1 - \pi)^r$$

where π denotes the conditional probability that an offspring is a proband given that (s)he is affected, which we assume is independent of the parental mating type.

The probability, $\tau_{sr}^{(ij)}(r = 1, 2, \ldots, s)$, that a family of ijth type (that is, in which the father is heterozygous at i loci and the mother is heterozygous at j loci) of size s will have r affected offspring is :

$$\tau_{sr}^{(ij)} = \binom{s}{r} \theta_{ij}^r (1 - \theta_{ij})^{s-r}.$$

Therefore, the probability, $\varphi_{sr}^{(ij)}$, that a family of ijth type of size s will have r affected offspring and will be ascertained is :

$$\varphi_{sr}^{(ij)} = \tau_{sr}^{(ij)} \cdot \alpha_r; \qquad r = 1, 2, 3, \ldots, s.$$

Hence, the probability, $\Phi_s^{(ij)}$, of a family of ijth type of size s having at least one affected child and being ascertained is :

$$\Phi_s^{(ij)} = \sum_{r=1}^{s} \varphi_{sr}^{(ij)} = 1 - (1 - \pi\theta_{ij})^s.$$

It, therefore, follows that the probability that a family of size s will have at least one affected offspring and will be ascertained is :

$$\sum_{i=1}^{L} \sum_{j=1}^{L} M_{ij} \Phi_s^{(ij)} = \sum_{i=1}^{L} \sum_{j=1}^{L} M_{ij} [1 - (1 - \pi\theta_{ij})^s].$$

The likelihood, $\mathcal{L}$, of an ascertained family of size s having r affected offspring is :

$$\mathcal{L} = \frac{\sum_{i=1}^{L} \sum_{j=1}^{L} M_{ij} \varphi_{sr}^{(ij)}}{\sum_{i=1}^{L} \sum_{j=1}^{L} M_{ij} \Phi_s^{(ij)}}.$$

Table 4.5: Parental genotypic mating classes, segregation probabilities and mating frequencies for model I with $L = 2$

Class	Genotypic mating	No. of heterozygous loci		Probability of affected offspring	Mating probability for class	
	Father × Mother	Father	Mother		Uncondi-	Condi-
		(i)	(j)	(θ_{ij})	tional	tional
1	AaBb × AaBb	2	2	1/16	$16p^4q^4$	p^2
2	AaBb × aaBb	2	1	1/8	$32p^3q^5$	$2pq$
	aaBb × AaBb	1	2	1/8		
	AaBb × Aabb	2	1	1/8		
	Aabb × AaBb	1	2	1/8		
3	aaBb × aaBb	1	1	1/4	$16p^2q^6$	q^2
	aaBb × aaBb	1	1	1/4		
	Aabb × aaBb	1	1	1/4		
	Aabb × Aabb	1	1	1/4		

Under single ascertainment, $\alpha_r \approx r\pi$. Hence, $\mathcal{L}$ reduces to :

$$\mathcal{L} = \binom{s}{r}\binom{r}{s}\frac{\sum_{i=1}^{L}\sum_{j=1}^{L}\binom{L}{i}\binom{L}{j}2^{(i+j)(1-s)}(2^{i+j}-1)^{s-r}p^{i+j}q^{4L-i-j}}{\sum_{i=1}^{L}\sum_{j=1}^{L}\binom{L}{i}\binom{L}{j}p^{i+j}q^{4L-i-j}}$$

where $p = 1 - q$.

Although the above equation looks formidable, it can be considerably simplified because several mating types have the same values of i and j, and consequently the same value of θ_{ij}. This is exemplified in table 4.5. While the above likelihood equation has been derived for unrelated parents, extension to

the situation when the parents are related is straightforward. The likelihood function remains valid; the only modification that is necessary is in the mating probabilities. These changed probabilities can be derived using the I-T-O method (Li and Sacks 1954). For the two-locus model, the unconditional mating probabilities as given in table 4.5 for unrelated parents change to :
when parents are an uncle-niece pair :

$$\begin{aligned} \text{class} \quad 1: \quad & p^2q^2(1/2+2pq)^2; \\ \text{class} \quad 2: \quad & p^2q^3(1+4pq)(1+2q); \\ \text{class} \quad 3: \quad & 2pq^4\{[(1+q)(1+4q)+p(1+2q)^2]/4\}; \end{aligned}$$

when parents are a pair of first cousins :

$$\begin{aligned} \text{class} \quad 1: \quad & 1/16+3p^3q^3(1/2+3pq); \\ \text{class} \quad 2: \quad & p^2q^3[1/4+3q/2+3pq(1+6q)]; \\ \text{class} \quad 3: \quad & pq^4[(1+q)+12pq(1+3q)]/4. \end{aligned}$$

4.2.4 Distribution of the number of affected children in Normal × Normal families under single ascertainment

The probability distribution of the number of affected children in families where both parents are normal, can easily be worked out for the case of single ascertainment by using the likelihood function derived in the previous section. These distributions are presented in figure 4.4 for various values of sibship size (s), prevalence (δ) and number of loci (L). It is seen that for fixed values of s and δ, the probability that a family contains only one affected child monotonically increases with increase in the number of loci. Correspondingly, there is a monotonic decrease in the probability of a family containing a large number of affected children. While for the one-locus recessive model, the probability distribution of the number of affected children depends only on sibship size and

not on prevalence, for the multilocus model (model I) this probability distribution also depends on prevalence. As is expected, the probability that a family contains a single affected child increases with decrease in the prevalence.

4.2.5 Sample size considerations

From the probability distribution given in section 4.2.3, the mean and variance of number of affected children in Nor × Nor families can easily be worked out. Equality of mean values for any two values of L may be tested by using the test statistic :

$$T = \frac{\sqrt{n}(\overline{X_1} - \overline{X_2})}{\sqrt{s_1^2 + s_2^2}}$$

where $\overline{X_1}$ and $\overline{X_2}$ denote the mean numbers of affected offspring for the two values of L under comparison, s_1^2 and s_2^2 denote the corresponding variances, and n denotes the number of families. This statistic can be used to answer a question of relevance : what is the number of families required to distinguish between two cases with different values of L based on the mean number of affected children? We have investigated this question; the results are given in table 4.6. It is clear from this table that while the sample sizes (*i.e.*, number of families to be sampled) for distinguishing between one-locus and two-locus models are within feasible limits at different levels of prevalence, this is not so for discriminating among multilocus models based on mean number of affected offspring.

Figure 4.4: Probability distributions of the number of affected children in families ascertained through an affected offspring for different sibship sizes and levels of prevalence

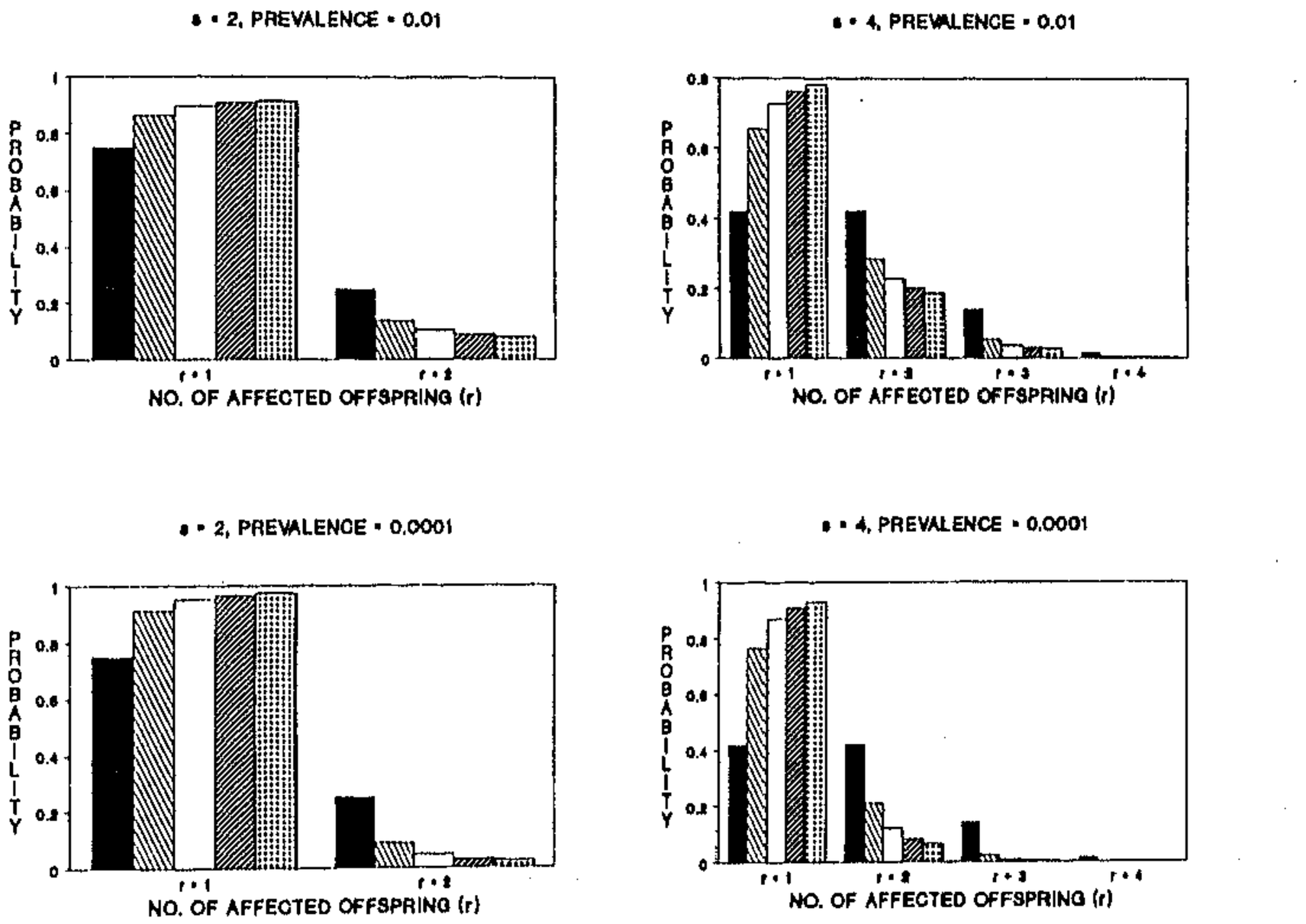


Table 4.6: Number of Nor × Nor families each ascertained through an affected offspring, required to distinguish between various cases based on mean number of affected offspring per family

Cases to be distinguished	Prevalence								
	$\delta = 0.01$			$\delta = 0.001$			$\delta = 0.0001$		
	Sibship size			Sibship size			Sibship size		
	2	4	6	2	4	6	2	4	6
One-locus vs. two-loci	92	32	19	52	18	11	39	13	8
Two-loci vs. three-loci	824	299	194	434	155	99	293	103	65
Three-loci vs. four-loci	3518	1301	859	1922	701	457	1341	477	505
Four-loci vs. five-loci	10288	3844	2555	5871	2162	1420	4301	1544	992

4.3 Incorporation of Variable Age at Onset : Derivation of Likelihood Functions

4.3.1 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 8 genotypes (AABB, AABb, AAbb, AaBB, AaBb, Aabb, aaBB, aaBb) if two loci are considered, while under model II, such an individual can be of any one of 4 genotypes (AABB, AABb, AaBB, 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 5 genotypes (aaBB, aaBb, aabb, AAbb, 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 4.7 , the list of various possible genotypic matings, mating probabilities, phenotypic mating types and segregation probabilities, separately for models I and II. While listing 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 of 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 $\times$ aabb may phenotypically either be normal $\times$ affected or normal $\times$ normal.

For the formulation of likelihood of phenotypic observations on offspring given the parental phenotypic 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 parent(s) who is(are) phenotypically normal, then the current age(s) of the parent(s) also need to be taken into consideration while enumerating the possible genotypic matings. We shall denote as g_f and g_m, the current ages of father and mother respectively.

2. We shall denote as : z_x = Prob {an individual of age x is phenotypically normal given that (s)he 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 denote as : μ_k = Prob {genotypic mating type is k given the phenotypic mating type and age(s) of the phenotypically normal parent(s)}; $k = 1, 2, \ldots, K$ = number of genotypic matings for a specified phenotypic mat-

ing. These are calculated straightforwardly from the mating probabilities given in table 4.7. However, these probabilities need to multiplied by appropriate z_i values in specific cases. For example, under model I, given a normal $\times$ affected mating, K should equal 8 (corresponding to serial numbers 9, 17, 24, 30, 35, 39, 42 and 44 of table 4.7) if the disorder expresses itself at birth. However, when the disorder has a late age at onset, a normal $\times$ affected mating may also be of type aabb $\times$ aabb (serial number 45 of table 4.7). Thus, $K = 9$. The mating probability of the aabb $\times$ aabb mating given that the phenotypic mating type is normal $\times$ affected, and that the normal individual belongs to i-th 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 denote as : θ_k = 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 or Aabb or aaBB or aaBb or aabb given the parental genotypic mating of type k}. These are also given in table 4.7.

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 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 denote as : n_i = total number of offspring in age-group i; m_i = number of affected offspring in age-group i; $(n_i - m_i)$ = number of normal offspring in age-group i.

Table 4.7: Genotypic and phenotypic mating types, their probabilities and segregation probabilities when two loci are involved

Sl. No.	Genotypic Mating Type	Mating Prob.	Segregation Prob.		Phenotypic Mating Type	
			Model I	Model II	Model I	Model II
1	AABB × AABB	p^8	0	0	Nor × Nor	Nor × Nor
2	AABB × AABb	$4p^7$q	0	0	Nor × Nor	Nor × Nor
3	AABB × AAbb	$2p^6q^2$	0	0	Nor × Nor	Nor × Aff
4	AABB × AaBB	$4p^7$q	0	0	Nor × Nor	Nor × Nor
5	AABB × AaBb	$8p^6q^2$	0	0	Nor × Nor	Nor × Nor
6	AABB × Aabb	$4p^5q^3$	0	0	Nor × Nor	Nor × Aff
7	AABB × aaBB	$2p^6q^2$	0	0	Nor × Nor	Nor × Aff
8	AABB × aaBb	$4p^5q^3$	0	0	Nor × Nor	Nor × Aff
9	AABB × aabb	$2p^4q^4$	0	0	Nor × Aff	Nor × Aff
10	AABb × AABb	$4p^6q^2$	0	1/4	Nor × Nor	Nor × Nor
11	AABb × AAbb	$4p^5q^3$	0	1/2	Nor × Nor	Nor × Aff
12	AABb × AaBB	$8p^6q^2$	0	0	Nor × Nor	Nor × Nor
13	AABb × AaBb	$16p^5q^3$	0	1/4	Nor × Nor	Nor × Nor
14	AABb × Aabb	$8p^4q^4$	0	1/2	Nor × Nor	Nor × Aff
15	AABb × aaBB	$4p^5q^3$	0	0	Nor × Nor	Nor × Aff
16	AABb × aaBb	$8p^4q^4$	0	1/4	Nor × Nor	Nor × Aff
17	AABb × aabb	$4p^3q^5$	0	1/2	Nor × Aff	Nor × Aff
18	AAbb × AAbb	p^4q^4	0	1	Nor × Nor	Aff × Aff
19	AAbb × AaBB	$4p^5q^3$	0	1/2	Nor × Nor	Aff × Nor
20	AAbb × AaBb	$8p^4q^4$	0	1/2	Nor × Nor	Aff × Nor
21	AAbb × Aabb	$4p^3q^5$	0	1	Nor × Nor	Aff × Aff
22	AAbb × aaBB	$2p^4q^4$	0	0	Nor × Nor	Aff × Aff
23	AAbb × aaBb	$4p^3q^5$	0	1/2	Nor × Nor	Aff × Aff

Table 4.7 (continued)

Sl. No.	Genotypic Mating Type	Mating Prob.	Segregation Prob.		Phenotypic Mating Type	
			Model I	Model II	Model I	Model II
24	AAbb × aabb	$2p^2q^6$	0	1	Nor × Aff	Aff × Aff
25	AaBB × AaBB	$4p^6q^2$	0	1/4	Nor × Nor	Nor × Nor
26	AaBB × AaBb	$16p^5q^3$	0	1/4	Nor × Nor	Nor × Nor
27	AaBB × Aabb	$8p^4q^4$	0	1/4	Nor × Nor	Nor × Aff
28	AaBB × aaBB	$4p^5q^3$	0	1/2	Nor × Nor	Nor × Aff
29	AaBB × aaBb	$8p^4q^4$	0	1/2	Nor × Nor	Nor × Aff
30	AaBB × aabb	$4p^3q^5$	0	1/2	Nor × Aff	Nor × Aff
31	AaBb × AaBb	$16p^4q^4$	1/16	7/16	Nor × Nor	Nor × Nor
32	AaBb × Aabb	$16p^3q^5$	1/8	1/2	Nor × Nor	Nor × Aff
33	AaBb × aaBB	$8p^4q^4$	0	1/2	Nor × Nor	Nor × Aff
34	AaBb × aaBb	$16p^3q^5$	1/8	5/8	Nor × Nor	Nor × Aff
35	AaBb × aabb	$8p^2q^6$	1/4	3/4	Nor × Aff	Nor × Aff
36	Aabb × Aabb	$4p^2q^6$	1/4	1	Nor × Nor	Aff × Aff
37	Aabb × aaBB	$4p^3q^5$	0	1/2	Nor × Nor	Aff × Aff
38	Aabb × aaBb	$8p^2q^6$	1/4	3/4	Nor × Nor	Aff × Aff
39	Aabb × aabb	$4pq^7$	1/2	1	Nor × Aff	Aff × Aff
40	aaBB × aaBB	p^4q^4	0	1	Nor × Nor	Aff × Aff
41	aaBB × aaBb	$4p^3q^5$	0	1	Nor × Nor	Aff × Aff
42	aaBB × aabb	$2p^2q^6$	0	1	Nor × Aff	Aff × Aff
43	aaBb × aaBb	$4p^2q^6$	1/4	1	Nor × Nor	Aff × Aff
44	aaBb × aabb	$4pq^7$	1/2	1	Nor × Aff	Aff × Aff
45	aabb × aabb	q^8	1	1	Aff × Aff	Aff × Aff

4.3.2 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 $n_i - m_i$; $i = 1, 2, \ldots, G$. 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 taking into account the age of the normal parent. 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 :

$$\mathcal{L} = \sum_{k=1}^{K}\mu_k\prod_{i=1}^{G}\binom{n_i}{m_i}[\theta_k(1-z_i)]^{m_i}[1-\theta_k(1-z_i)]^{n_i-m_i}.$$

4.3.3 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 $(i, j = 1, 2, \ldots, G)$. 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, $\mathcal{L}$, can be written as :

$$\mathcal{L} = [\alpha_m \cdot \lambda(\mathbf{n}, \mathbf{m})]/\beta(\mathbf{n}, \mathbf{m}),$$

where the form of the function $\lambda(\mathbf{n}, \mathbf{m})$; $\mathbf{n}$ and $\mathbf{m}$ being vectors, 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 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 $\mathcal{L}$, $\alpha_m \cdot \lambda(\mathbf{n}, \mathbf{m})$, denotes the likelihood that in a family with n_i offspring there will be m_i affected in age group $i(i = 1, 2, \ldots, G)$, and that such a family will be ascertained. The denominator of $\mathcal{L}$, $\beta(\mathbf{n}, \mathbf{m})$, 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 :

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

where,

$$N = \sum_{i=1}^{G} n_i.$$

When, $\pi \approx 0$, the likelihood function simplifies to :

$$\mathcal{L} = r \cdot \lambda(\mathbf{n}, \mathbf{m})/\beta(\mathbf{n}, \mathbf{m}),$$

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.

4.3.4 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 onset age compared to one which is expressed at birth. Thus, for a disorder with a variable onset age, 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 4.7. 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.

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).

Chapter 5

Segregation Analysis of Nuclear Family Data

Under each of the two multilocus models considered in this study, the conditional likelihood of data on offspring given the parental mating type, current ages of normal parents and method of ascertainment, were derived analytically, in Chapter 4, as functions of number of loci and recessive allele frequencies. In evaluating the likelihoods, the numbers of affected and normal offspring need to be classified in various age groups under consideration for the various types of families. 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 (n = 86); and, (ii) families in which both parents are normal, each family ascertained through an affected offspring (n = 61). The analysis of data on nuclear families involved a total of 674 individuals of whom 165 were affected.

5.1 Results

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, assuming population prevalence of the disorder (δ) is 0.5%. For this value of δ, under model I, the recessive allele frequency, q, turns out to be: q = .0707 (for L = 1), $q = .2659$

(for $L = 2$), $q = .4135$ (for $L = 3$) and $q = .5157$ (for $L = 4$). Under model II, $q = .0707$ (for $L = 1$) and $q = .0500$ (for $L = 2$). [We note that when $L = 1$, the allele frequency, q, is the same for both models I and II.]

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 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 i-th age group is unaffected, which is required in the likelihood computations, was obtained from the data on age at onset of probands. [We note that if an individual is of a susceptible genotype, the individual will, if (s)he lives long enough, eventually become affected.] These values are given in table 5.1.

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 (π) ≈ 0 [single selection], we have performed computations for different values of π. The results are presented in table 5.2. 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 $\pi \approx 0$. The joint likelihood values are presented in the last column of table 5.2. From these values it is seen that the data are about 10^{13} times (log-likelihood ratio = 12.86) more likely if two loci are involved than if only one locus is involved; about 10^6 times more likely (log-likelihood ratio = 5.79) if three loci are involved; but only about 18 times more likely (log-likelihood ratio = 1.25) if four loci are involved. The likelihood differences between the one-locus vs. two-loci and between two-loci vs. three-loci cases are large. However, the difference in likelihoods between the three-locus and

Table 5.1: 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
$\geq$ 65	1.000	.000

Table 5.2: Log-likelihood values under model I, 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 offspring

Number of loci	Normal × Affected families	Normal × Normal families			All families
		$\pi = 0.01$	$\pi = 0.05$	$\pi \approx 0$	
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

the four-locus models seems to be non-significant at the 5% level, indicating that three-locus model is the most parsimonious. [Because of the underlying assumption of equality of allele frequencies at the loci, the number of parameters remains the same for differing values of L. Formal statistical tests of significance, therefore, could not be performed.]

Similar results for model II are presented in table 5.3. 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 about 10^9 times less likely (log-likelihood ratio = -8.42) than under model I. Similar results are obtained when the value of L is increased. Thus, it is clear that between model I and II, model I is more parsimonious.

We, therefore, conclude that three epistatically interacting autosomal dialalic 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.

Table 5.3: 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 offspring

Number of loci	Normal × Affected families	Normal × Normal families			All families
		$\pi = 0.01$	$\pi = 0.05$	$\pi \approx 0$	
1	-38.1338	-73.8925	-72.8980	-74.1403	-112.2741
2	-33.3822	-74.2087	-73.2132	-74.4554	-107.8376

5.2 Robustness of Inferences

There are two potential sources of variation which may affect the results and inferences derived in the previous section. These are: (i) variation in estimates of z_i's presented in table 5.1; and, (ii) variation in the estimate of δ. We have investigated the effects of both these sources of variation on the values of the likelihood functions and on the inferences derived earlier.

5.2.1 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, z_i^* 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^*, z_G was taken to be 0.0, since the cumulative proportion to the last age group G must equal 1. Further, for obvious reasons, for the first age group (that is, $i = 1$), the new estimate z_1^* was actually derived by drawing a random number from U[0,z_1+s_1]. The various estimates of z_i^* values are presented in table 5.4.

Table 5.4: Randomly generated z_i values in five runs

Age group	$z_i \pm s.e.$	z_i^t for run number				
		1	2	3	4	5
≤ 4	$.944 \pm .018$	.951	.938	.955	.959	.962
5-9	$.732 \pm .035$	.697	.759	.709	.721	.747
10-14	$.588 \pm .039$	.584	.588	.559	.567	.612
15-19	$.506 \pm .040$	.540	.503	.480	.474	.534
20-24	$.381 \pm .038$	.393	.347	.381	.398	.357
25-29	$.293 \pm .036$	.294	.261	.296	.304	.299
30-34	$.281 \pm .033$	.248	.231	.217	.235	.209
35-39	$.156 \pm .029$	.146	.146	.163	.140	.182
40-44	$.118 \pm .026$	.113	.129	.117	.126	.138
45-49	$.074 \pm .021$	.081	.075	.092	.090	.065
50-54	$.024 \pm .012$	.015	.034	.029	.034	.027
55-59	$.012 \pm .009$	.005	.006	.013	.013	.010
60-64	$.006 \pm .006$	.002	.001	.003	.011	.008
≥ 65	$.000 \pm .000$	.000	.000	.000	.000	.000

Table 5.5: Joint $\log_{10}$ likelihood values for each run and their mean $\pm$ s.d.

No. of loci	Run Number					Mean $\pm$ s.d.
	1	2	3	4	5	
1	-111.21	-112.38	-112.42	-112.73	-113.06	-112.36 $\pm$ .312
2	-96.57	-98.26	-97.35	-97.87	-98.51	- 97.71 $\pm$ .346
3	-92.84	-94.47	-93.50	-94.09	-94.79	-93.94 $\pm$ .348
4	-91.61	-93.11	-92.24	-92.84	-93.56	-92.67 $\pm$ .342

Having obtained a new set of estimates of z_i $(i = 1, 2, ..., G)$, we computed the likelihood functions assuming $\pi \approx 0$, separately for normal $\times$ affected families and for normal $\times$ normal families, under model I for $L = 1, 2, 3$ and 4. This procedure was repeated 5 times. The mean $\pm$ s.d. of the values of the joint (that is, data of normal $\times$ affected and normal $\times$ normal families pooled) log-likelihood function for $L = 1, 2, 3$ and 4 were evaluated. The results are presented in table 5.5.

By comparing with the values presented in the last column of table 5.2 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. We have not presented these results for brevity.]

5.2.2 Effect of variation in prevalence

To investigate the effect of variation in prevalence, we computed the joint likelihood functions, assuming $\pi \approx 0$, for $\delta = .004$ and .01. The $\log_{10}$ likelihood values are presented in table 5.6 for different values of δ.

Since in the primary analyses $\delta = 0.005$ was assumed, in the last column of

Table 5.6: $\log_{10}$ likelihood values for different δ

No. of loci	Joint $\log_{10}$ likelihood		
	$\delta = .004$	$\delta = .01$	$\delta = .005$
1	-112.63	-111.17	-112.27
2	-97.63	-96.83	-99.42
3	-93.80	-93.19	-93.62
4	-92.54	-91.96	-92.37

this table we have also presented the $\log_{10}$ likelihood values for $\delta = 0.005$, for comparison. There was no significant effect of variation in δ on the values of the likelihood function for a fixed value of L. [Similar results were also obtained for model II; results are not presented.] Hence, the previous inferences on genetic models remain unaltered, even after allowing for variation in δ within an appropriate range.

5.3 Conclusions

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 as double cross-validation (Mosteller and Tukey 1977). In performing the genetic analyses, we have also extended the relevant statistical methodology relating to analysis of family data on a complex disorder with variable onset

age.

We have performed a segregation analysis of the data on nuclear families. Likelihood functions were analytically derived taking into account the complexities arising from delayed age at onset and biases of ascertainment. The results of the present segregation analysis are in complete agreement with those found in the earlier study (Majumder *et al.* 1988). In the present study, we have not only cross-validated the previous model which postulates that recessive alleles at multiple unlinked loci act epistatically in the manifestation of vitiligo, but 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 processes of melanogenesis. In the present study, the best estimate of the number of loci involved in the pathogenesis of vitiligo was found to be three, while this estimate was four in the previous study. This, we believe, is not a serious difference, and may have occurred because the previous study had not taken into account the possibility — albeit small because most parents were over 40 years of age — that an unaffected parent may be of the susceptible genotype. It is, in fact, remarkable that the difference is so small given that the data were collected from two disparate ethnic backgrounds. While we have not explicitly considered the possibility of the existence of non-genetic (sporadic) cases in our models, the standard assumption that most probands in single-case families are non-genetic may not always be valid because in the present data set $\approx 80\%$ of nuclear families are single-case families. For the type of multilocus models considered, it is not possible to estimate the proportion of non-genetic cases through statistical analyses; refinement of methods of clinical evaluation are necessary to identify non-genetic 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.

Before concluding this chapter it is pertinent to point out that some of the assumptions — equality of gene frequencies across loci, complete penetrance, no sporadic cases — have, in fact, been relaxed in the analyses of data on pedigrees, results of which are presented in the next chapter.

Chapter 6

Segregation Analysis of Pedigree Data

As mentioned in Chapter 2, from each proband data were sought on a fixed set of his/her relatives that included parents, spouse(s), children, grandparents, grandchildren, uncles, and aunts. From the data on families of the 147 probands, it was found that 67 (45.6%) were 3-generational, 57 (38.8%) were 4-generational, 23 (15.6%) were 5-generational families. In the previous chapter, results of segregation analysis performed on nuclear families of the probands have been reported. In this chapter, the results of analyses of data on extended families (pedigrees) of probands are presented. The analysis of data on nuclear families involved a total of 674 individuals of whom 165 were affected (including probands). This analysis of data on extended families involves 2256 individuals of whom 216 are affected (including probands).

6.1 Methodology

Segregation analysis of the data on members of 147 pedigrees each ascertained through an affected proband was performed using the Pedigree Analysis Package (PAP), revision 3.0 (Hasstedt 1989), which uses the nonlinear optimization routine GEMINI (Lalouel 1979). PAP uses the 'peeling algorithm' (Cannings *et al.* 1976), a generalization of Elston-Stewart algorithm (1971), to compute

likelihoods of pedigrees. PAP incorporates an ascertainment bias correction by dividing the likelihood of the pedigree by the likelihood of the proband(s). A logistic function with mean 20 and standard deviation 15 was used as an approximation to the age specific prevalence of the disorder. [As stated in Chapter 3, this function with these parameter values provided a reasonable fit to the observed age-specific distribution.] A threshold model, postulating that an individual of a specific genotype and age is affected when her/his liability exceeds a certain threshold, was assumed. Likelihoods corrected for ascertainment bias, were computed for nongenetic, various one-locus and two-locus recessive genetic models.

6.1.1 Elston-Stewart algorithm

The general mathematical formulation of pedigree analysis as developed by Elston and Stewart (1971) and extended by Cannings *et al.* (1978) has been used for pedigree analysis. The joint likelihood of all pedigrees was evaluated under a series of hypotheses and maximum likelihood estimates of parameters were computed through a direct numerical search of the likelihood surface. Under the assumption of random mating and random sampling of the pedigrees, the likelihood of observing the data on a particular family is dependent upon

(1) the distribution of the different genotypes in the population;

(2) the distribution of phenotypes for each genotype; and

(3) the distribution of offspring's genotype given the parental genotypes.

The likelihood computation for a pedigree for the one-locus, two alleles (D and d) case involves several parameters. Given the pedigree structure and the phenotypic observations (*Normal* or $Affected$) on individuals of a pedigree, the likelihood of these observations is computed for given values of the following parameters :

(1) Frequency, p, of the allele D ;

(2) Transmission probabilities :

$$\tau_{i,D} = \text{Prob } \{\text{an individual transmits } D \text{ allele} \mid \text{(s)he is of genotype } i \} \text{ ; and}$$

(3) Penetrance probabilities :

$$\begin{aligned} Fi,j &= \text{Prob } \{\text{an individual is of phenotype } j \mid \text{(s)he is of genotype } i \} \text{ ;} \\ i &= DD \text{ or } Dd \text{ or } dd \text{ ;} \\ j &= \textit{Normal} \text{ or } \textit{Affected} \text{ .} \end{aligned}$$

In practice, it is assumed that the population is in Hardy-Weinberg equilibrium, so that given the gene frequency p, the genotype probabilities :

$$\begin{aligned} g_i &= \text{Prob } \{\text{a random individual is of genotype } i\} \text{ ;} \\ i &= DD \text{ or } Dd \text{ or } dd \text{ ;} \end{aligned}$$

are easily calculated by the binomial rule. Moreover, in accordance with Mendelian transmission rules, the transmission probabilities are fixed at : $[\tau_{DD,D};\ \tau_{Dd,D};\ \tau_{dd,D}] = [1,\ 1/2,\ 0]$.

If $\tau_{i,D}$'s are fixed, then the segregation probabilities :

$$\begin{aligned} t_{ij,k} &= \text{Prob } \{\text{an individual is of genotype } k \mid \text{the parental genotypes are } i \text{ and } j\}; \\ i,j,k &= DD,\ Dd \text{ or } dd, \text{ are easily computed.} \end{aligned}$$

Now, given a pedigree consisting of N individuals with phenotypes $(x_1, x_2, \ldots, x_N)$, respectively, the likelihood of the pedigree for a specified set of parameter values is computed as follows :

(1) The probability of an individual of phenotype x marrying into the pedigree (including founders) is :

$$\sum_i F_{i,x} \cdot g_i.$$

(2) The likelihood of two parents (of phenotypes x_1 and x_2) and an offspring (*of phenotype* x_3) is :

$$\begin{aligned} & \sum_i F_{i,x_1} \cdot g_i \sum_j F_{j,x_2} \cdot g_j \sum_k F_{k,x_3} \cdot t_{ij,k} \\ = & \sum_i \sum_j \sum_k F_{i,x_1} \cdot F_{j,x_2} \cdot F_{k,x_3} \cdot g_i \cdot g_j \cdot t_{ij,k}. \end{aligned}$$

(3) The likelihood, $\mathcal{L}$, of the entire pedigree can, therefore, be written as :

$$\mathcal{L} = \sum_{i_1} \sum_{i_2} \cdots \sum_{i_N} \prod_j F_{i_j,x_j} \prod_k g_{i_k} \prod_l t_{i_m i_n, i_l}$$

where the summations are taken over all genotypes ($i_1, i_2, \ldots, i_N = DD, Dd$ or dd), the first product over all individuals, the second product over all individuals marrying into the pedigree (*including founders*) and the third product over all offspring with parental genotypes i_m and i_n.

When the age at onset is variable, ages of individuals (current ages for unaffected individuals and ages at onset for affected individuals) need to be taken into account. Ages are incorporated by appropriately modifying the $F_{i,j}$ values.

When a multilocus model is considered, the number of genotypes increases, genotypic probabilities (g_i) are appropriately modified, and transmission probabilities ($\tau_{i,D}$) refers to transmission of gamete types D given a multilocus genotype (i). The form of the likelihood function ($\mathcal{L}$) remains the same.

6.1.2 Testing of hypotheses

Tests of hypotheses are performed by employing the likelihood ratio statistic. A specific hypothesis to be tested corresponds to a restriction on one or more parameters. The likelihood is then maximized both in the unrestricted case with r parameters (the likelihood denoted by $\widehat{\mathcal{L}}_r$) and the hypothesis with k independent restrictions on the r parameters ($\widehat{\mathcal{L}}_{r-k}$). The smallar the likelihood ratio $\Lambda_{r,k} = \widehat{\mathcal{L}}_{r-k}/\widehat{\mathcal{L}}_r$, the less likely the null hypothesis is true; the test thus consists of rejecting the null hypothesis at a specified significance level, if this ratio is smaller than a particular value. The quantity $-2\,ln\Lambda_{r,k}$ is asymptotically distributed as a χ^2 with k degrees of freedom under the null hypothesis. The null hypothesis is rejected if the observed value of this statistic is greater than the appropriate tabulated value of χ^2.

6.2 Models

Before considering specific genetic models it was of interest to examine how well the data can be explained without invoking any genetic basis for the disorder, that is assuming that the disorder is solely due to environmental causes. To examine this, we have considered a nongenetic/sporadic model, where the frequency of the allele causing the disorder (q) was set at 1.0, so that all individuals were of the same genotype. The prevalence of the disorder, which is the conditional probability of affection given the genotype, was set at a fixed value (=.005 and .01) or was estimated. These three cases are denoted as N_1, N_2 and N_3. We have considered a three one-locus models (R, D, and G) and two two-locus models (R_1 and R_2). For one-locus models, we have assumed that the locus is biallelic and denoted the allele causing the disorder as a and the alternate allele as A. Lifetime prevalences of the genotypes were denoted as

l_1, l_2, etc. We have assumed Hardy-Weinberg equilibrium and Mendelian transmission. The model R denotes the recessive model with complete penetrance. In this model we fixed the lifetime prevalences of the genotypes : $l_1(aa) = 1$ and $l_2(Aa) = l_3(AA) = 0$. The only parameter of this model is q.

The model D denotes the dominant model with complete penetrance. In this model we fixed $l_1(aa) = l_2(Aa) = 1$ and $l_3(AA) = 0$. This model also has only one parameter, q.

The model G denotes a general genetic model which allows for both incomplete penetrance in the susceptible genotype (aa) and the presence of phenocopies among individuals of the non-susceptible genotypes (AA, Aa). This model, therefore, has four parameters q, l_1, l_2 and l_3.

For the two-locus model, we considered two unlinked biallelic loci. We denoted the alleles at the two loci, as a and b and their counterparts as A and B, respectively. The frequencies of the alleles a and b were denoted as q_1 and q_2, respectively. Under the postulated two-locus recessive model, out of the ten possible two-locus genotypes, individuals of only one genotype $(aabb)$ are affected; individuals of the other nine genotypes are normal. Under this model, there are ten genotypic lifetime prevalences denoted as $l_1, l_2, \cdots, l_{10}$, of which l_1 corresponds to the genotype $aabb$, and $l_2, l_3, \cdots, l_{10}$ to the remaining nine genotypes. For the present two-locus recessive model under consideration, $l_1 = 1$ and $l_2 = \cdots = l_{10} = 0$. These values were held fixed in all likelihood computations. Likelihood computations were performed in two ways : Model R_1 : assuming $q_1 = q_2 = q$, in which q was held fixed at 0.2659 (which corresponds to a population prevalence of 0.5%), and Model R_2 : treating q as an independent parameter to be estimated from the data. We note that from the available data it is not possible to independently estimate q_1 and q_2. This is because allele frequencies are estimated from information on founders (including individuals marrying into the pedigree); the proportion of affecteds among founders is equal to, except for sampling fluctuations, the population prevalence, which is

a function solely the product of allele frequencies q_1 and q_2.

For the purpose of cross-checking of estimates and inferences, as also for computational convenience, the set of 147 pedigrees was randomly split into two approximately equal subsets (subset I and subset II) comprising 75 and 72 pedigrees, including 1185 and 1071 individuals, respectively. All computations were performed separately for the two subsets.

6.3 Results

The general form of the observed cumulative age at onset distribution appears to be that of the logistic distribution function. The logistic distribution function (d.f.) has the form :

$$F(x) = \{1 + \exp[-(x-a)/k]\}^{-1}$$

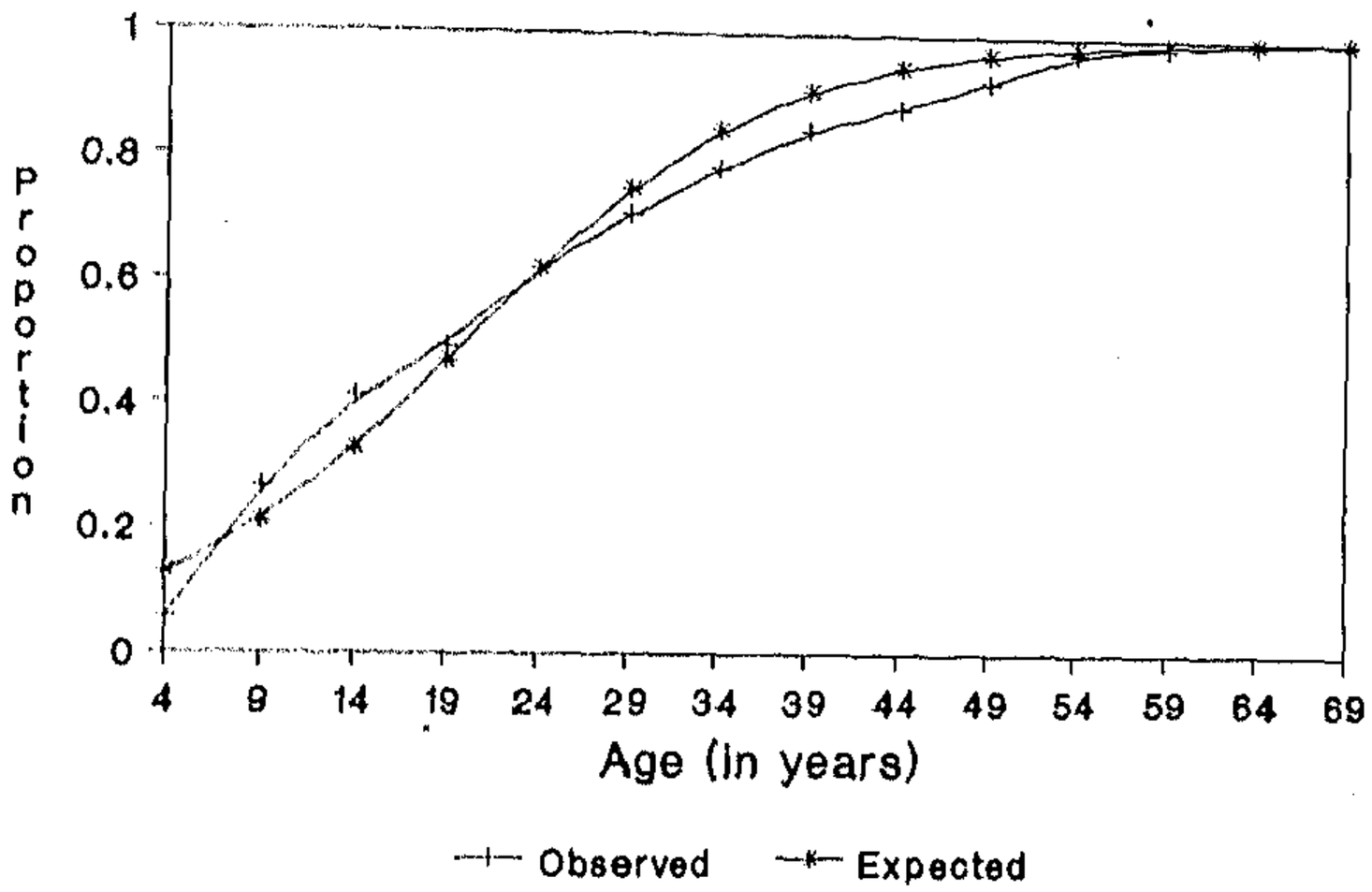
where, $k = \sqrt{3}\,b/\pi$ or $b = k\pi/\sqrt{3}$ and a = mean, b = s.d. The corresponding probability density function (p.d.f.) is :

$$f(x) = \frac{exp\,[(x-a)/k]}{(k\{1 + exp\,[(x-a)/k]\}^2)}$$

In figure 6.1 we have plotted the observed cumulative age at onset distribution and the logistic distribution function with parameters $a = 20$ and $b = 15$. [These parameter values are approximately equal to the observed mean and s.d. of the ages at onset among probands.] We have examined whether the logistic distribution with the above parameter values with the observed cumulative age at onset distribution. This was done by using the Kolmogorov-Smirnov nonparametric test procedure (Chakravarti *et al.* 1967). The test statistic is

$$d = \sqrt{n}\,\sup_x \mid F(x) - \widehat{F}(x) \mid,$$

Figure 6.1: Observed cumulative frequency distribution of age at onset of vitiligo and the logistic distribution function with mean = 20 and s.d = 15



where $F(x)$ is the logistic distribution function as defined above, and $\widehat{F}(x)$ is the observed cumulative age at onset distribution. The observed value of d was 1.32, which is non-significant at the 5% level, implying that the logistic approximation to the observed age at onset distribution is acceptable.

The results of model fitting (segregation analysis) are presented in table 6.1, separately for the two subsets. As mentioned earlier, the three non-genetic models, N_1, N_2 and N_3 correspond, respectively, to situations in which the population prevalence of vitiligo was held fixed at 0.005 or 0.01 or was estimated from the data. For all non-genetic models, the frequency of the allele causing the disorder was held fixed at 1. It is seen that the likelihood of the data under the non-genetic model increases with increase in prevalence and the maximum

Table 6.1 : Results of pedigree analysis of vitiligo for non-genetic, one-locus and two-locus models*

Data Subset	Parameter	Non genetic models			Genetic models				
					One-locus models			Two-locus models	
		N_1	N_2	N_3	R	D	G	R_1	R_2
Subset I (n = 75)	q	[1]	[1]	[1]	.0680	.0500	.0230	[.2659]	.2415
	l_1	–	–	–	[1]	[1]	.9975	[1][b]	[1][b]
	l_2	–	–	–	[0]	[1]	.0787	–	–
	l_3	–	–	–	[0]	[0]	.0046	–	–
	Prevalence[a]	[.005]	[.01]	.038	.005	.098	.010	.005	.003
	$\log_{10} \mathcal{L}$	-96.31	-87.22	-77.44	-77.48	-371.91	-69.12	-68.04	-66.87
	$-2ln\ \mathcal{L}$	443.83	401.68	356.64	356.84	1712.7	318.31	313.33	307.97
Subset II (n = 72)	q	[1]	[1]	[1]	.0689	.0400	.0525	[.2659]	.2484
	l_1	–	–	–	[1]	[1]	.6491	[1][b]	[1][b]
	l_2	–	–	–	[0]	[1]	.0597	–	–
	l_3	–	–	–	[0]	[0]	.0108	–	–
	Prevalence[a]	[.005]	[.01]	.038	.005	078	.017	.005	.004
	$\log_{10} \mathcal{L}$	-85.57	-77.46	-68.70	-70.62	-351.37	-64.73	-63.67	-62.88
	$-2ln\ \mathcal{L}$	394.05	356.70	316.37	325.20	1617.7	298.09	293.22	289.56

(*) n denotes number of pedigrees and figures in brackets indicate fixed values of parameters; those not in brackets are maximum likelihood estimates of parameters,

(a) Prevalence was calculated from estimates of parameters,

(b) l_1 refers to lifetime prevalence of genotype aabb; the lifetime prevalences of the remaining nine genotypes, $l_2, l_3, \cdots, l_{10}$, were fixed at 0.

likelihood estimate of prevalence obtained for N_3 is about 4% from both subsets of the data.

Among the one-locus models considered, the dominant model (D) is clearly rejected; compared to other models the -$2ln\Lambda$ (Λ = Likelihood Ratio) varies between 1268.9 and 1404.7 for subset I, and between 1223:7 and 1328.1 for subset II. The likelihoods under the recessive (R) model and the non-genetic model N_3 are of a similar magnitude, but the estimate of prevalence (0.005), derived from the maximum likelihood estimate of the allele frequency, for the recessive model is much closer to observed population prevalences (.005 - .01) than the estimate of 0.04 obtained under the non-genetic model. The general one-locus model (G) provides the best fit to the data and is significantly better than the one-locus recessive model (-$2ln\Lambda$ = 38.5 for subset I and 27.1 for subset II; d.f. = 3; $p < 0.05$ for both subsets). For both subsets, estimated prevalences under the general model are higher than those obtained under the recessive model, but nevertheless, are close to observed population prevalences reported earlier. A careful examination of the estimated lifetime prevalences of the three genotypes obtained under the general one-locus model indicates that these estimates are similar, in particular for subset I, to the lifetime prevalences of the genotypes of the recessive model. In other words, the general model is effectively same as the recessive model.

When compared with the best-fitting one-locus model G, the two-locus model is found to provide a better fit (-$2ln\Lambda \approx 10$). Because of differences in parametric structures of the one-locus and two-locus classes of models, we were unable to perform rigorous statistical tests of significance to compare these two classes of models. Further, because of computational complexities and the enormous computer time requirement it was not possible to perform likelihood computations on pedigree data for genetic models involving more than two loci. Because of this limitation, the present pedigree analysis only confirms whether a multilocus model provides a better fit than a single-locus model, but does

not provide an estimate of the number of loci involved. We were thus unable to verify our previous estimate that three loci are involved in the pathogenesis of vitiligo. Our present analyses, however, show that the pedigree data on vitiligo fits a two-locus recessive model better than a one-locus recessive model. It is also interesting to note that the population prevalences calculated from the estimates of parameters obtained under the two-locus recessive model are same (≈ 0.004) for both subsets of the data. This value of estimated prevalence is also very close to earlier population estimates (Das *et al.* 1985a; Howitz *et al.* 1977).

6.4 Robustness of Inferences

For pedigree analyses, we have used a logistic function with mean = 20 and s.d = 15 to approximate the age specific prevalence distribution. The two parameters of the logistic function are potential sources which may affect the likelihoods. To investigate the effects of these two parameters on inferences, we have performed a robustness study. Because of the enormous computational time requirement, we have performed this analyses only on data of subset II, for non-genetic and one-locus models.

To investigate the effect of the two parameters, we have varied the mean and s.d. values within an appropriate range (mean : 15 to 25 and s.d. : 10 to 20). In table 6.2 we present the likelihood values under different models for subset II. It is seen that for each model, the value of the likelihood function, for a fixed mean value, increases with increase in s.d.; and, for a fixed s.d., decreases with increase in the mean value. However, the relative merits of the various models for any pair of values of mean and s.d. are similar to those discussed in the previous section. The estimates (not presented for brevity) of

Table 6.2: $\log_{10}$ likelihood values of data of subset II for different mean and s.d. values under non-genetic and one-locus models

Mean, S.d.	log_{10} likelihood for models					
	N_1	N_2	N_3	R	D	G
15, 10	-83.50	-75.49	-67.26	-71.80	-316.53	-63.95
15, 15	-83.15	-75.11	-66.74	-71.19	-380.11	-63.18
15, 20	-83.18	-75.10	-66.51	-70.54	-318.30	-62.69
20, 10	-87.05	-78.97	-70.41	-71.33	-467.36	-66.50
20, 15	-85.57	-77.46	-68.70	-70.62	-351.27	-64.74
20, 20	-85.05	-76.90	-67.90	-69.96	-284.07	-63.96
25, 10	-91.70	-83.54	-74.53	-70.72	-427.78	-69.39
25, 15	-88.62	-80.43	-71.19	-69.98	-320.47	-66.68
25, 20	-87.34	-79.10	-69.61	-69.32	-272.53	-64.93
Prevalence range	[.005]	[.01]	.036-.040	.004-.006	.007-.135	.012-.020

parameters of the various models were also similar to those given in table 6.1. Therefore, the inferences presented in the previous section are robust against fluctuations of values of parameters of the logistic function which was used to approximate the distribution of age at onset of vitiligo.

6.5 Conclusions

In conclusion, the present results substantiate our previous finding of epistatic interaction of recessive alleles at multiple unlinked loci in the pathogenesis of vitiligo. This inference makes considerable biological sense. It is known that there are several key points in the human pigmentary pathways (Hearing and King 1993). The implicated loci may serve as controls of these key points and recessive homozygosis at any of these loci may be viewed as complete disruption or blockage at the corresponding point. If one or two of these control points are blocked then bypass routes in the pigmentary pathways are possibly used and there is no precipitation of disorder state. However, if a greater number of control points are blocked, then possibly there is failure of the entire pathway because of non-availability of bypass routes, which results in a disruption of the end product or process and consequently in the clinical manifestation of vitiligo.

REFERENCES

Amornsiripanitch S, Barnes LM, Nordlund JJ, Trinkle LS, Rheins LA (1988) Immune studies in the depigmenting $C57BL/6 - mi^{vit}/mi^{vit}$ mice. J Immunol 140: 3438-3445

Angrist M, Kauffman E, Slaugenhaupt SA, Matise TC, Puffenberger EG, Washington SS, Lipson A, Cass DT, Reyna T, Weeks DE, Sieber W, Chakravarti A (1993) A gene for Hirschsprung disease (megacolon) in the pericentromeric region of human chromosome 10. Nat genet 4: 351-356

Audrezet MP, Novelli G, Mereier B, Sangiuolo F, Maceratesi P, Ferec C, Dallapiccola B (1993) Identification of three novel Cystic Fibrosis mutations in a sample of Italian Cystic Fibrosis patients. Hum Hered 43: 295-300

Austin ML, Boissy RE, Jacobson BS, Smyth JR Jr (1992) The detection of melanocyte autoantibodies in the Smyth chicken model for vitiligo. Clin Immun Immunopathol 64: 112-120

Barker D Wright E, Nguylen K, Cannon L, Fain P, Goldin D, Bishop DT, Carey J, Baty B, Kivlin J, Willard H, Waye JS, Greig G, Leinward L, Nakamura Y, O'Connell P, Leppert M, Lalouel JM, White R, Skolnick M (1987) Gene for von Recklinghausen neurofibrometosis is in the pericentric region of chromosome 17. Science 236: 1100-1102

Behl PN (1955) Leucoderma and its treatment with Ammi Majus. J Ind Med Assoc 24: 615-618

Birbeck MSC, Breathnach AS, Everall JD (1961) An electron microscopic study of basal melanocytes and high level clear cells (Langerhans cells) in vitiligo. J Invest Dermatol 37: 51-63

Boissy RE, Smyth JR Jr, Fite KV (1983) progressive cytological changes during the development of delayed feather amelanosis and associated eye defects in the DAM chicken line: A vitiligo model. Am J Pathol 111: 197-212

Boissy RE, Moellmann G, Trainer AA, Smyth JR Jr, Lerner AB (1986) Delayed-amelanotic (DAM or Smyth) chicken: melanocyte dysfunction *in vivo* and *in vitro*. J Invest Dermatol 86: 149-156

Boissy RE, Lamoreux ML (1988) Vitiligo and antibodies to melanocytes In: Advances in Pigment Cell Research. Alan R Liss, Inc

Bor S, Fiewel M, Chanarin L (1969) Vitiligo and aetiological relationship to organ specific autoimmune disease. Br J Dermatol 81: 83-88

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

Breathnach AS, Bor S, Wyllie LM-A (1966) Electron microscopy of peripheral nerve terminals and marginal melanocytes in vitiligo. J Invest Dermatol 47: 125-140

Bystryn JC and Pfeffer S (1988) Vitiligo and antibodies to melanocytes In: Advances in Pigment Cell Research. Alan R Liss, Inc

Cannings CA, Thompson EA, Skolnick M (1978) Probability functions on complex pedigrees. Adv Appl Prob 10: 26-61

Carnevale A, Zavala C, Castillo VD, Maldonado RR, Tamayo L (1980) Analisis genetico de 127 families con vitilido. Rev Invest clin 32: 37-41

Chakraborty R, Weiss KM, Schull WJ (1980) A test for randomness of occurrence of a disease trait in familial or other similar ordered sequences of epidemiological data. Proc Natl Acad Sci (USA) 77: 2974-2978

Chakraborty R, Weiss KM, Majumder PP, Strong LC, Herson J (1984) A method to detect excess risk of disease in structured data: Cancer in relatives of retinoblastoma patients. Genet Epidemiol 1: 229-244

Chakravarti A (1993) Impact of genetic, somatic, and epigenetic variation on phenotype. In: Sing CF and Hanis CL (eds) Genetics of Cellular, Individual, Family, and Population variability. Oxford University Press, New York, Oxford

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

Cockayne EA (1933) Inherited abnormalities of the skin and appendages. Oxford University Press, London

Cui J, Harning R, Henn M, Bystryn JC (1992) Identification of pigment cell antigens defined by vitiligo antibodies. J Invst Dermatol 98: 162-165

Cui J, Yuko A, Bystryn JC (1992) Cytolic antibodies to melanocytes in vitiligo. J invest Dermatol 100: 812-815

Cunliffe WJ, Hall R, Newell DJ, Stevenson CJ (1968) Vitiligo, thyroid disease and auto-immunity. Br J Dermatol 80: 135-139

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

Das SK, Majumder PP, Majumdar TK, Halder B (1985b) Studies on vitiligo. II. Familial aggregation and genetics. Genet Epidemiol 2: 255-262

Dawber RPK (1970) Clinical association of vitiligo. Postgrad Med J 46: 276-287

Dunston GM, Halder RM (1990) Vitiligo is associated with HLA-DR4 in black patients. Arch Dermatol 126: 56-60

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 53: A793

Edwards JH (1960) The simulation of Mendelism. Acta Genet 10: 63

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

El-Mofty AM and El-Mofty M (1980) Vitiligo: A symptom complex. Int J Dermatol 19:238-244

Elston RC (1993) Some recent developments in the theretical aspects of segregation analysis. In: Majumder PP (ed) Human population genetics: A centennial tribute to J.B.S. Haldane. Plenum Press, New York

Elston RC, Stewart J (1971) A general model for the genetic analysis of pedigree data. Hum Hered 21: 523-542

Emery AEH (1976) Methodology in medical genetics. Churchill-Livingstone, Edinburgh, London and New York.

Falconer DS (1965) The inheritance of liability to certain diseases, estimated from the incidence among relatives. Ann Hum Genet 29: 51-76

Finco O, Cuccia M, Martinetti M, Ruberto G, Orecchia G, Rabbiosi G (1991) Age of onset in vitiligo: relationship with HLA supratypes. Clinical Genet 39: 48-54

Fleischman RA (1993) From white spots to stem cells: the role of the Kit receptor in mammalian development. TIG 9: 285-290

Foley NR, Lowe NJ, Misheloff E, tiwari JL (1983) Association of HLA-DR4 with vitiligo. J Am Acad Dermatol 8: 39-40

George AO (1989) Vitiligo in labdan, Nigeria. Int J Dermatol 28: 385-387

Goldbourt U, Neufeld HN (1986) Genetic aspects of arteriosclerosis. Arteriosclerosis 6: 357-377

Goldin LR, Weeks DE (1993) Two-locus models of disease: Comparison of likelihood and non-parametric linkage methods. Am J Hum Genet 53: A1006

Goudie RB, Goudie DR, Dick HM and Ferguson-Smith MA (1980) Unstable Mutations in vitiligo, organ-specific autoimmune disease and multiple endocrine adenoma/peptic ulcer syndrome. Lancet ii: 285-287

Greenberg DA (1981) A simple method for testing two-locus models of inheritance. Am J Hum Genet 33: 519-530

Greenberg DA, Rotter JI (1981) Two-locus models for gluten sensivity enteropathy: Population genetic consideration. Am J Med Genet 8: 205-214

Greenberg DA, Lange KL (1982) A maximum lilelihood test of the two-locus model for Coeliac disease. Am J Med Genet 12: 75-82

Greenberg DA (1984) Simulation studies of segregation analysis : Application to two-locus models. Am J Hum Genet 36: 167-176

Greenberg DA, Delgado-Escueta AV, Maldanado HM, Widolitz H (1988) Segregation analysis of juvenile Myoclonic epilepsy. Genet Epidemiol 5: 81-94

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

Haldane JBS (1941) The relative importance of principal and modifying genes in determining some human diseases. J Genet 41: 149-157

Haldane JBS (1948) Formal genetics of man. Proc R Soc (Lond) 135 (B): 147-168

Halder RM, Crimes PE, Cowan (1987) Childhood vitiligo J Am Acad Dermatol 16: 948-954

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

Harning R, Cui J, Bystryn JC (1991) Relation between the incidence and level of pigment cell antibodies and disease activity in vitiligo. J Invest Dermatol 97: 1078-1080

Hasstedt SJ (1989) *Pedigree Analysis Package (version 3.0)*. University of Utah, Saltlake City, Utah.

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

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

Hearing VJ, King RA (1993) Determinants of skin color: melanocytes and melanization. In: Levin N (ed) Pigmentation and Pigmentary Disorders C R C Press, Inc Boca Raton, Florida

Hill JR (1980) A survey of cancer sites by kinship in the Utah Mormon population. In: Cairns J, Lyons J, Skolnick M (eds) Banbury report 4 Cancer incidence in defined population. Cold Spring Harbor, New York

Hook RR, Aultman MD, Adelstein EH, Oxenhandler RW, Millikan LE, Middleton CC (1979) Influence of selective breeding on the incidence of melanomas in Sinclair miniature swine. Int J Cancer 24: 668-672

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.

Jaysankar TJ, Baruah MC, Garg BR (1992) Vitiligo in children. Int J Dermatol 31: 621-623

King MC, Lee GM, Spinner NB, Thomson G, Wrensch MR (1984) Genetic epidemiology. Ann Rev Public Health 5: 1-52

Kugelman TP, Lerner AB (1961) Albinism, partial albinism and vitiligo. Yale J Biol Med 33: 407-414

Lalouel JM (1979) GEMINI: A computer program for optimization of general non-linear functions. Tech rep 14, Department of Medical Biophysics and Computing. University of Utah, Salt Lake City.

Lamoreux ML, Boissy RE, Womack JE, Nordlund JJ (1992) The *vit* gene maps to the *mi* (Microphthalmia) locus of the laboratory mouse. J Hered 83: 435-439

Lee ST, Nicholls RD, Bundey S, Laxova R, Musarella M, Spritz RA (1994) Mutations of the *P* gene in oculocutaneous albinism, ocular albinism, and Prader-Willi syndrome plus albinism. N Eng J Med 330: 529-534

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

Lerner AB (1971) On the etiology of vitiligo and gray hair AM J Med 51: 141-147

Lerner AB, Nordlund JJ (1978) Viitligo: What is it? Is it important? JAMA 239: 1183-1187

Lerner AB, Shiohara T, Boissay RE, Jacobson KA, Lamoreux ML, Moellmann GE (1986) A mouse model for vitiligo. J Invest Dermatol 87: 299-304

Levai AB (1958) A study of certain contributory factors in the development of vitiligo in south Indian patients. Arch Dermatol 32: 285-310

Li CC (1953) Some general properties of recessive inheritance. Am J H Genet 5: 269-279

Li CC, Sacks L (1954) The derivation of joint distribution and correlation between relatives by the use of Stochastic metrices. Biomertics 10: 347-360

Li CC (1975) Path analysis — A primer. The Boxwood Press, Pacific Grove, CA

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

Lyonnet S *et al.* (1993) Nature Genet 4: 364-350

Majumder PP (1982) AGGREGATION: A computer package for testing elevation of incidence of a trait in families. Technical report No. 3, University of Texas Health Science Center at Houston, Texas.

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

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

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

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

Majumder PP (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 Press, New York

McGregor BC, Katz HI, Doe RP (1992) Vitiligo and glandular Insufficiencies. JAMA 219: 724-725

Merry A, Roger JH, Curnow RN (1979) A two-locus model for the inheritance of a familial disease. Ann Hum Genet 43: 71-80

Maul GG (1969) Golgi-melanosome relationship in human melanoma *in vitro* J Ultra Res 26: 163

Maul GG and Brumbaugh JA (1971) On the possible function of coated vesicles in melanogenesis of the regenerating fowl feather. J Cell Biol 48: 41

Mehta NR, Shah KC, Theodore C, Vyas VP, Patel AB (1973) Epidemiological study of vitiligo in Surat area, South Gujarat. Ind J Med Res 61: 145-154

Merelender J, Rywlin J (1940) Heredity of acquired vitiligo (vitiligo in three generations). Acta Dermatovenerelogica 21: 583-585

Metzker A, Zamir R, Gazit E, David M, Feuerman EL (1980) Vitiligo and the HLA system. Dermatologica 160: 100-105

Millikan LE, Hook RR, Manning PJ (1973) Gross and ultra-structural studies in a new melanoma model: The Sinclair swine. Yeal J Biol Med 46: 631-645

Mohr J (1951) Vitiligo in a pair of monovular twins. Acta Genet Stat Med 2: 252-255

Morton NE (1982) Outline of genetic epidemiology. S Kargar, Basel

Morton NE (1993) Recent developments in genetic epidemiology. In: Majumder PP (ed) Human population genetics: A centennial tribute to J.B.S. Haldane. Plenum Press, New York

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

Moyer FH (1967) Electron microscopic observation on the origin, development and genetic control on melanine structure. The structure of eye, Acadmic Press, New York. pp 469

Naughton GK, Reggiardo D, Bystryn JC (1986a) Correlation between vitiligo antibodies and extent of depigmentation in vitiligo. J Am Acad Dermatol 15: 978-981

Naughton GK, Mahaffey M, Bystryn JC (1986b) Antibodies to surface antigens of pigmented cells in animals with vitiligo. Proc Soc Exp Biol Med 181: 423-426

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

Nordlund JJ, Lerner AB, Miller LH, Kenney JA (1978) Proceedings of the First International Workshop on Vitiligo. J Invest Dermatol 71: 165-166

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

Pardue SL, Fite KV, Bengston L, Lamont SJ, Boyle ML, Smyth JR Jr (1987) Enhanced integumental and ocular amelanosis following the termination of cyclosporine administration. J Invest Dermatol 88: 758-761

Passarge E (1993) Wither polygenic inheritance: Mapping Hirschsprung desease. Nature Genet 4: 325-326

Pawelek P, Korner A, Bergstorm A, Bologna J (1980a) Regulation of melanogenesis and precursor cytotoxicity at points distal to tyrosinase activity. J Invest Dermatol 74: 250

Pawelek P, Korner A, Bergstorm A, Bolognia J (1980b) New regulators of melanine synthesis and the autodestruction of melanoma cells. Nature 286 : 617-169

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

Poloy A, Tibor L, Kramer J, Anh-Taun N, Kraszits E, Medgyessy I, Fust G, Stenszky V, Farid NR (1991) Immunology Letters 27: 59-62

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, 3rd ed. McGraw-Hill, New York, pp 224-251

Ramaiah A, Puri N, Mojomdar M (1989) Etiology of vitiligo: A New hypothesis. Acta Derm Venereol (Stockh) 69: 323-367

Riccardi VM, Eichner JE (1986) Neurofibromatosis: phenotype, natural history and pathogenesis. Johns Hopkins University Press, Baltimore

Rice JP, Neuman RJ, Burroughs TE, Hampe CL, Daw EW, Suarez BK (1993) Linkage analysis for oligogenic traits. Am J Hum Genet 53: A66

Riordan JR, Rammens JM, Kerem B, *et al.* (1989) Identification of the cysticfibrosis gene: Cloning and characterization of complementary DNA. Science 245: 1066-1-1073

Risch N (1990) Linkage strategies for genetically complex traits. I: Multilocus models. Am J Hum Genet 46: 222-228

Risch N (1990) Linkage strategies for genetically complex traits. II: The power of affected relative pairs. Amm J Hum Genet 46: 229-241

Saha N, Ahmed MA, Wasfi AI, El-Munshid HA (1982) Distribution of serum proteins, red cell enzymes and haemoglobins in vitiligo. Hum Hered 32: 46-48

Schellreuter KU, Wood JM, Berger J (1991) Low catalase levels in epidermis of patients with vitiligo. J Invest Dermatol 97: 1081-1085

Schork NJ, Boehnke K, 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

Seiji M, Fukuzawa H, Shimao K, Itakura H (1973) On the melanization process of the melanosemes. Pigment Cell 1: 39-46

Siemens HW (1953) Die zwillingspathologie der vitiligo. Acta Genet Med Gemellol 2: 118-124

Seizinger B, Ozelius LJ, Lane AH, George-Hyslop P, Huson S, Gusella JF, Martuza RL (1987) Common pathogenic mechanism for three tumor types in bilateral acoustic neurofibrometosis. Science 136: 317-319

Silvers WK (1979) The Coat Colors of Mice. Springer-verlag, Basel

Skolnick MH, Cannon-Allbright LA, Goldgar DE, Ward Jh *et al.* (1990 Inheritance of profiferative breast disease in breast cancer kindreds. Science 250: 1715-1720.

Slack J, Evans KA (1966) The increased risk of death from ischemic heart disease in first degree relatives of 121 men and 96 women with ischemic heart disease. J Med Genet 3: 239-257

Slominski A, Paul R (1989) Hypothesis: possible role of melanin receptor in vitiligo: discussion paper. J Royal Soc Med 82: 539-541

Smyth JR Jr, Boissay RE, Fite KV (1981) The DAM chicken: a model for postnatal cutaneous and ocular amelanosis. J Hered 72: 150-156

Spritz RA, Giebel LB, Holms SA (1992) Dominant negative and loss of function mutations of the c-kit (mast/stem cell growth factors receptor) proto-onco gene in human piebaldism. Am J Human Genet 50: 261

Steinberg AG, Becker SW, Fitzpatric TB, Kierland RR (1951) A genetic and statistical study of psoriasis. Am J Hum Genet 3: 267-281

Thompson EA (1980) Genetic etiology and clusters in a pedigree. Hered 45: 323-334

Thomson G (1988) HLA disease associations: Model for insulin dependent diabetes mellitus and the study of complex human disorders. Ann Rev Genet 22: 31-50

Thomson G (1980) A two-locus model for juvenile diabetes Ann Hum Genet 43: 383-398

Tiwari JL, Tarasaki PI (1985) HLA and disease associations Springer-Verlag, Berlin

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

Venneker GT, Westerhof W, de Vries IJ, Drayer NM, Wolthers BG, de Waal LP, Bos JD, Asghar SS (1992) Molecular heterogeneity of the forth component (C4) and its genes in vitiligo. J Invest Dermatol 99: 853-858

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

Wasfi AI, Saha N, El-Munsid HA, El-Sheik FS, Ahmed MA (1980) Genetic association in vitiligo. Clin Genet 17: 415-417

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 Chron Dis 35: 539-553

Weeks DE, Lathrop GM, Ott J (1993) Multipoint mapping under genetic interference. Hum Hered 43: 86-97

Wendt GG, Drohm D (1972) Fortschritte der allgemeinen und klineschen humangenetik. Vol IV: Die Huntingtonsche Chorea. Stuttgart, Thieme

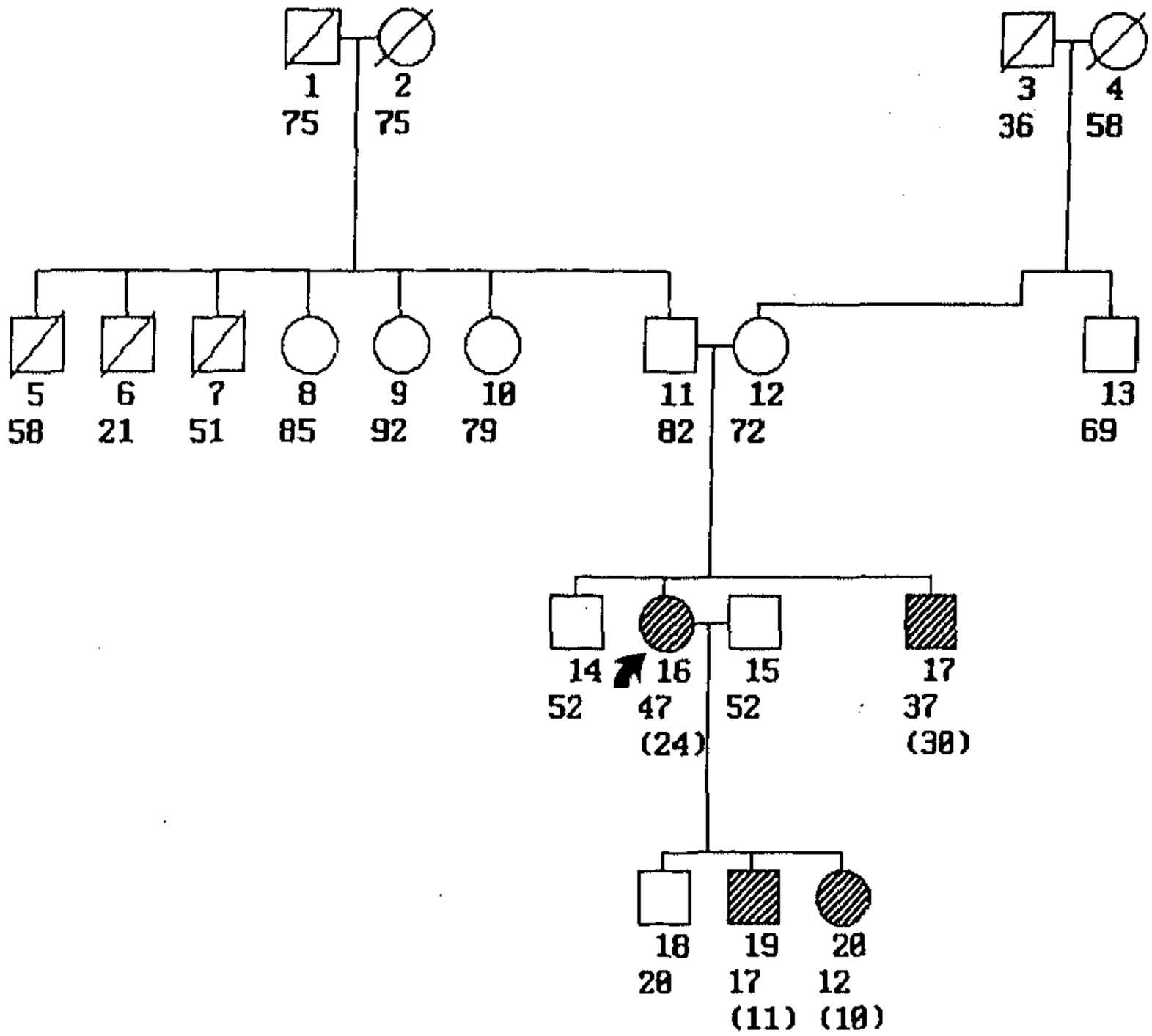
Wick MM, Byers L, Frei E (1977) Selective toxicity of L-dopa for melanoma cells. Science 197: 468-469

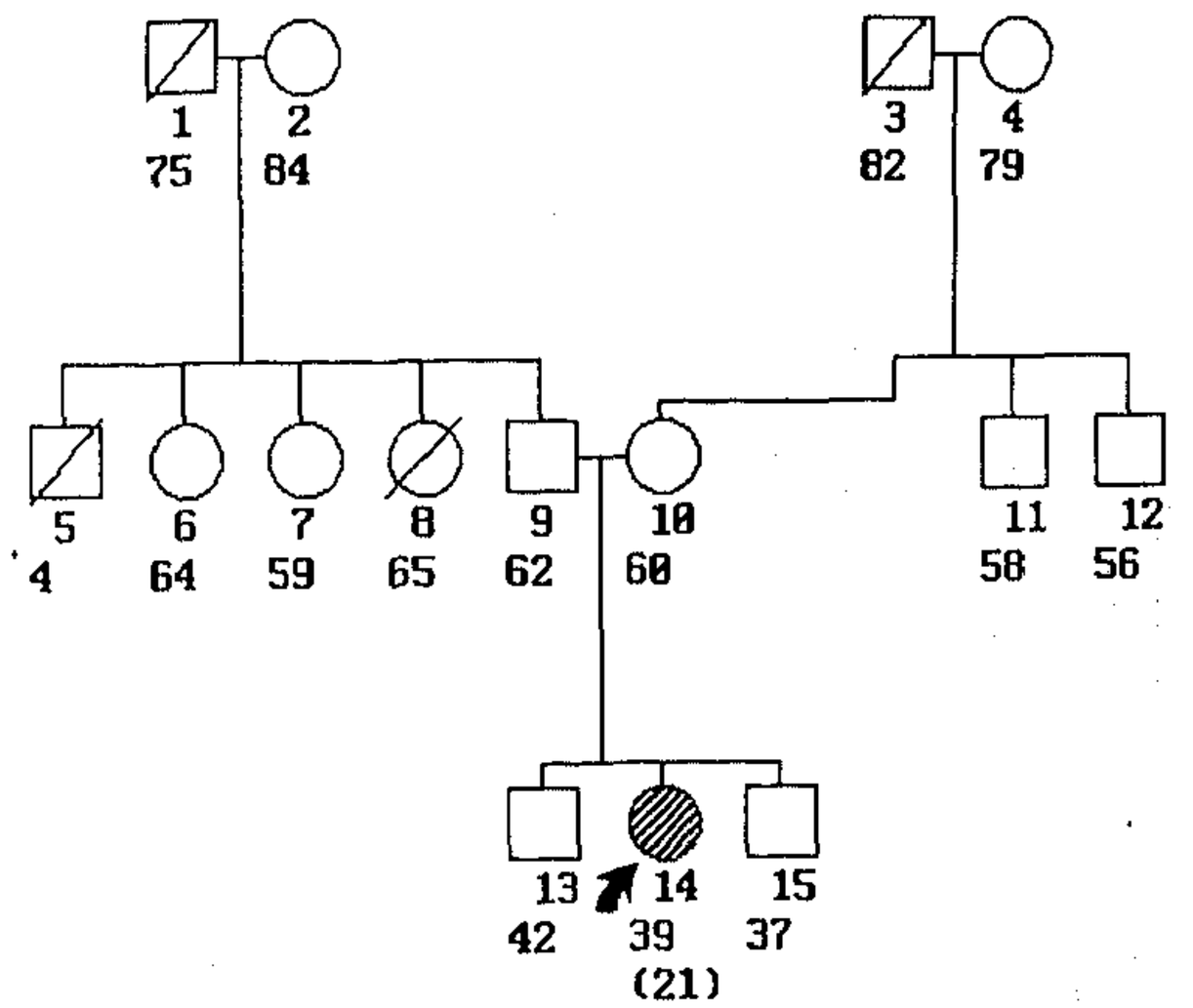
Wick MM, Hearing VJ, Rorsman H (1987) Biochemistry of melanization. In: Fitzpatrick TB, Eisen AZ, Wolff K, Freedberg IM, Austen KF (eds) Dermatology in general medicine, 3rd ed. McGraw-Hill, New York, pp 251-258

APPENDIX

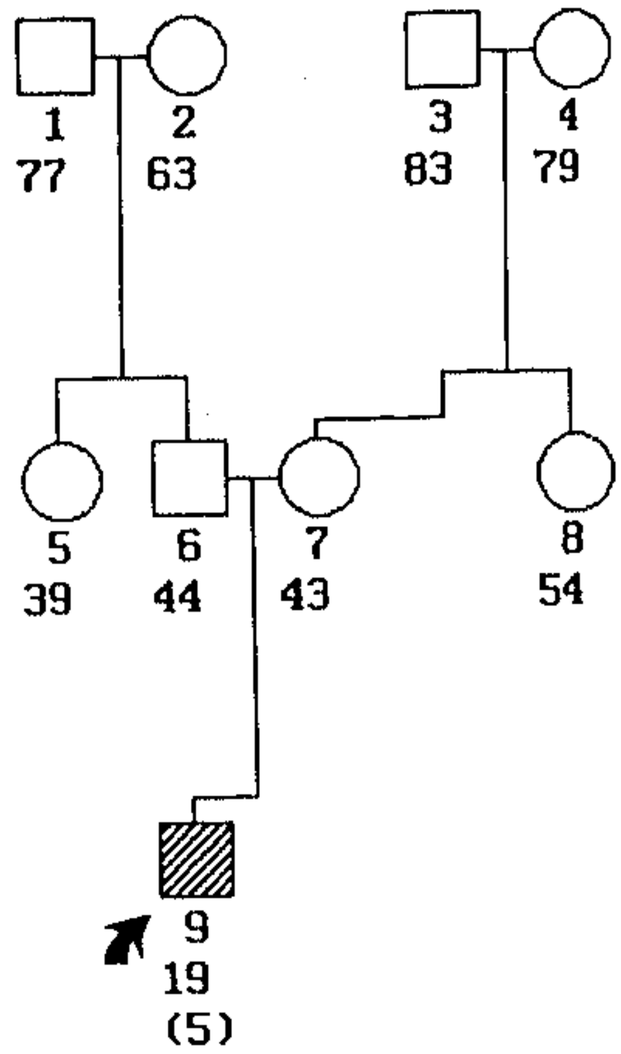
Pedigree Charts of a Selection of 10 Sampled Families.

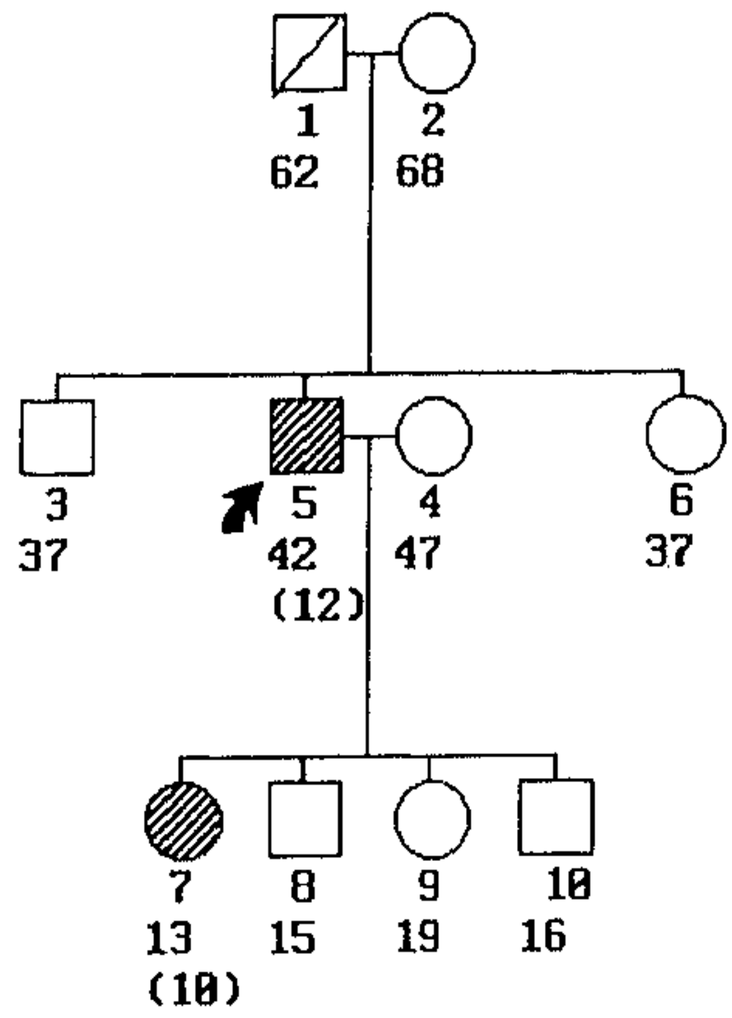
KEY : Below each symbol (circle or square), the first number represents serial number of the individual within the pedigree, the second number represents current age (in years) and the third number in parentheses represents age at onset, if the individual is affected. The proband is marked by an arrow.

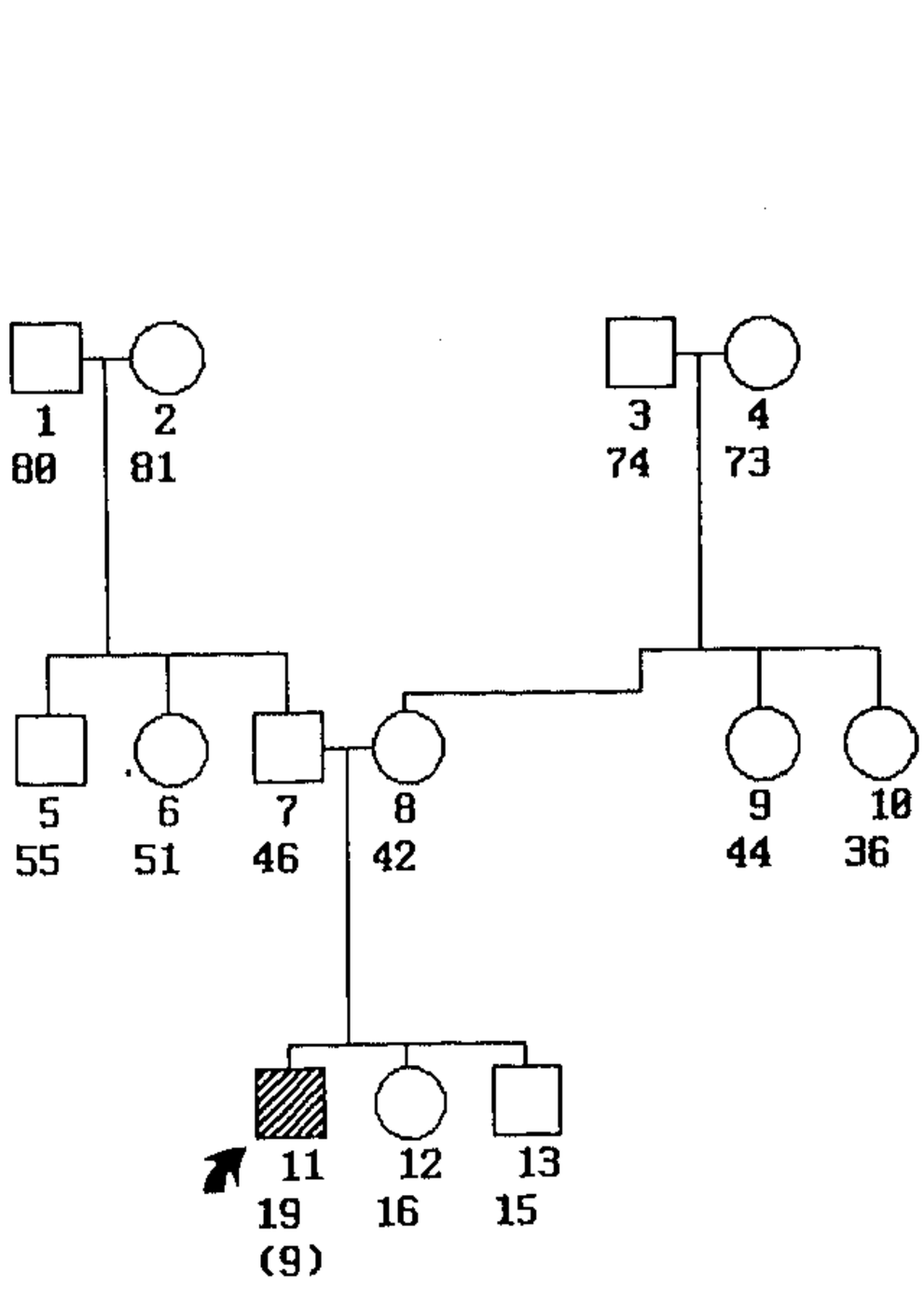

1
75
2
75
3
36
4
58
5
58
6
21
7
51
8
85
9
92
10
79
11
82
12
72
13
69
14
52
16
47
(24)
15
52
17
37
(30)
18
20
19
17
(11)
20
12
(10)

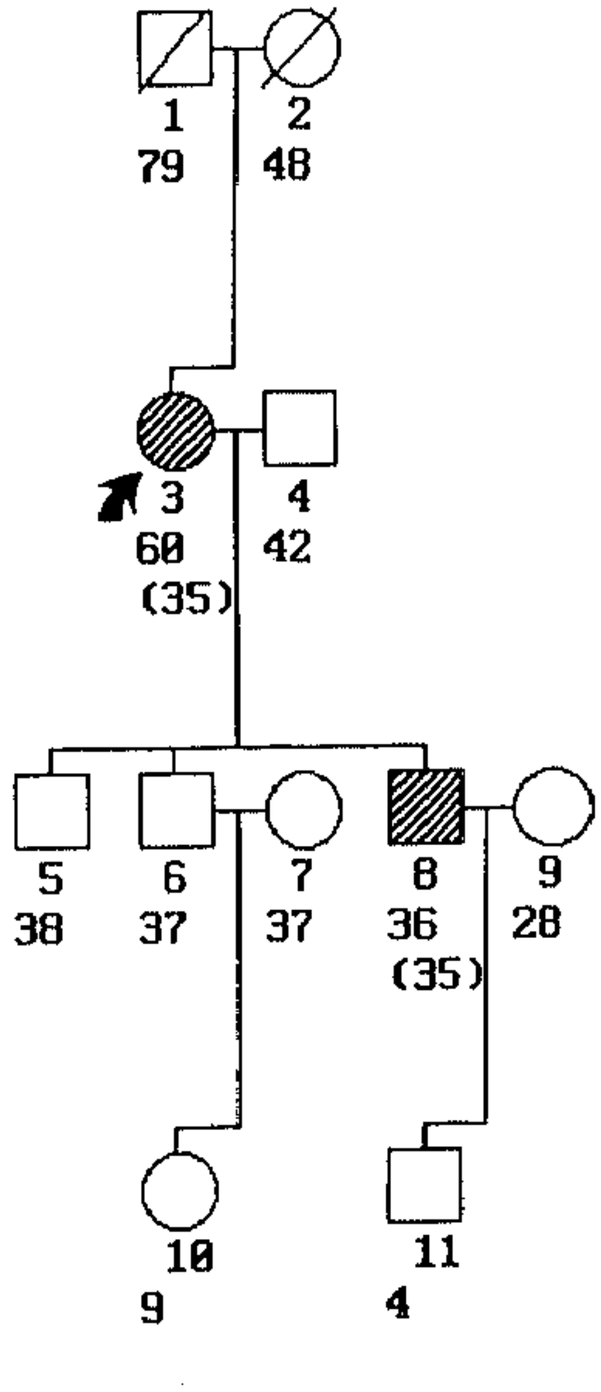

1
75
2
84
3
82
4
79
5
4
6
64
7
59
8
65
9
62
10
60
11
58
12
56
13
42
14
39
(21)
15
37

A1

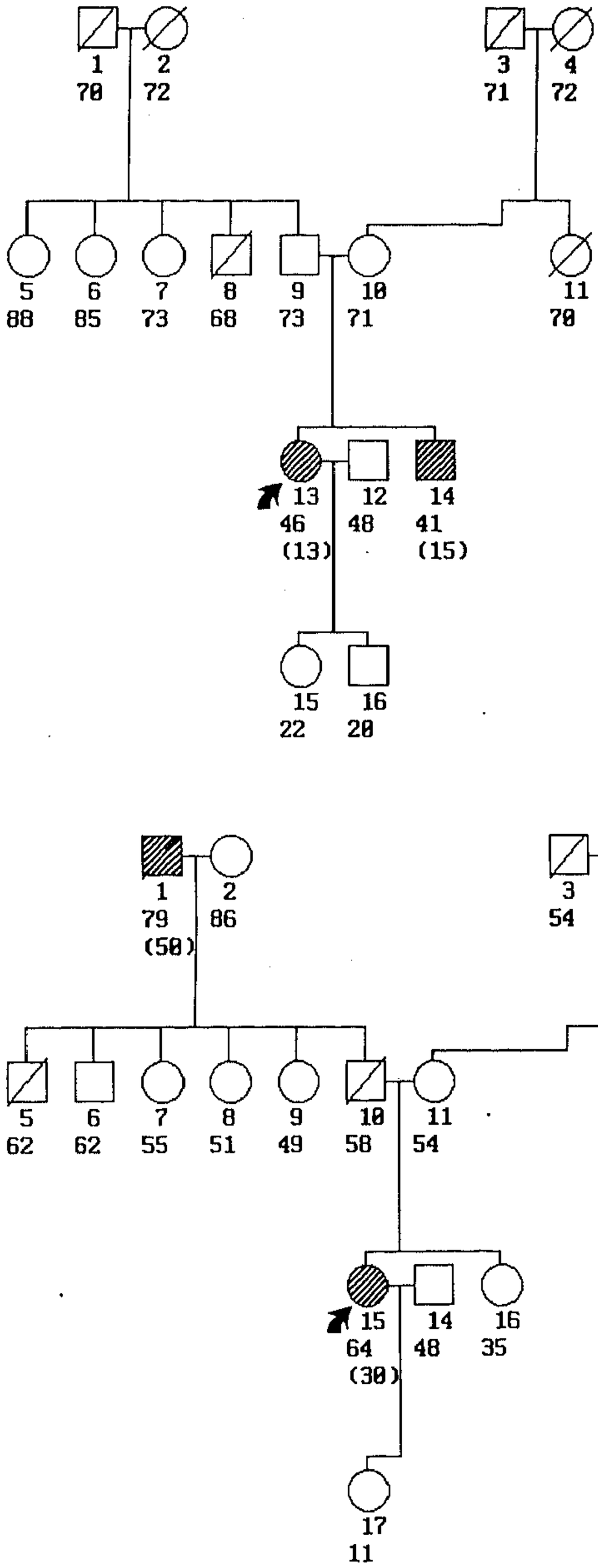

1
77
2
63
3
83
4
79
5
39
6
44
7
43
8
54
9
19
(5)


1
62
2
68
3
37
5
42
(12)
4
47
6
37
7
13
(10)
8
15
9
19
10
16

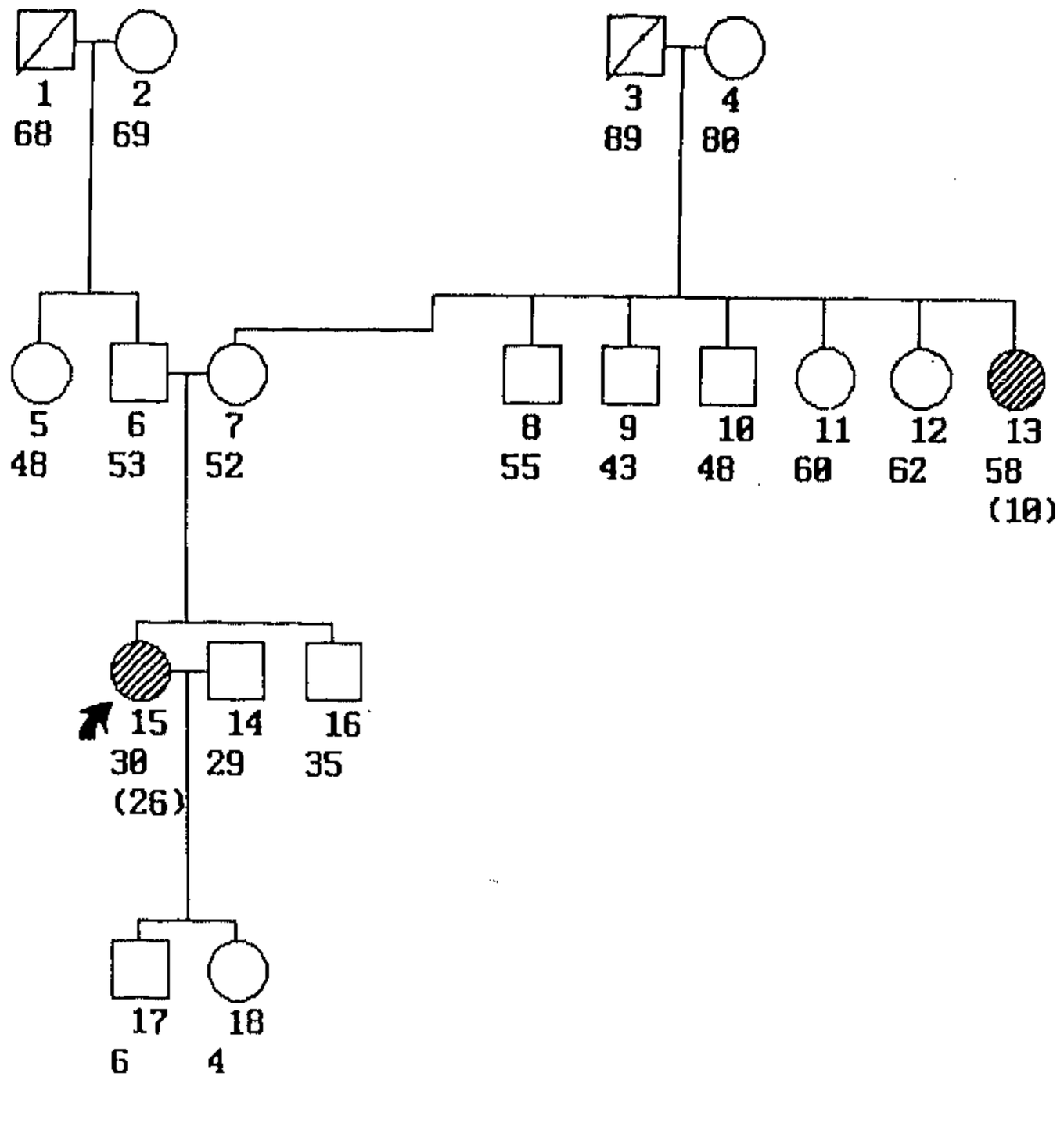

1
80
2
81
3
74
4
73
5
55
6
51
7
46
8
42
9
44
10
36
11
19
(9)
12
16
13
15

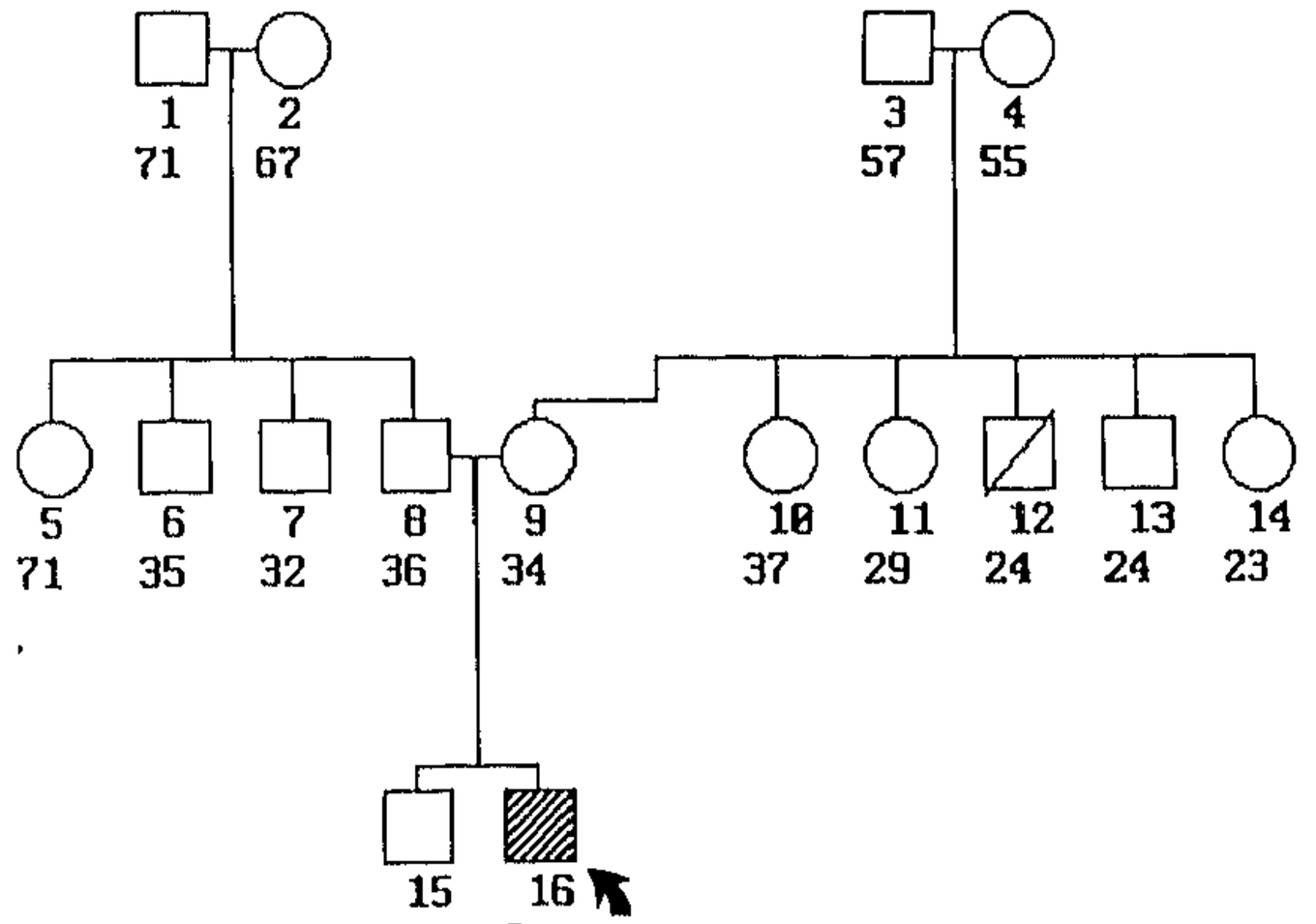

1
79
2
48
3
60
(35)
4
42
5
38
6
37
7
37
8
36
(35)
9
28
10
9
11
4

A2


1
70
2
72
3
71
4
72
5
88
6
85
7
73
8
68
9
73
10
71
11
70
13
46
(13)
12
48
14
41
(15)
15
22
16
20
1
79
(50)
2
86
3
54
4
49
5
62
6
62
7
55
8
51
9
49
10
58
11
54
12
68
13
55
15
64
(30)
14
48
16
35
17
11

A3


1
68
2
69
3
89
4
80
5
48
6
53
7
52
8
55
9
43
10
48
11
60
12
62
13
58
(10)
15
30
(26)
14
29
16
35
17
6
18
4


1
71
2
67
3
57
4
55
5
71
6
35
7
32
8
36
9
34
10
37
11
29
12
24
13
24
14
23
15
10
16
6
(5)