70-19,506

CAMPBELL, Mary Anne, 1944-
STUDIES ON QUASI-CONTINUITY.

University of Hawaii, Ph.D., 1969
Biology-Genetics

University Microfilms, A XEROX Company, Ann Arbor, Michigan

STUDIES ON QUASI-CONTINUITY

A DISSERTATION SUBMITTED TO THE GRADUATE DIVISION OF THE
UNIVERSITY OF HAWAII IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

IN GENETICS

DECEMBER 1969

By

Mary Anne Campbell

Dissertation Committee:

Newton E. Morton, Chairman
C. S. Chung
N. Rashad
L. Dickenson
Walton K. T. Shim

ACKNOWLEDGMENT

My gratitude to the staff of the Population Genetics Laboratory for their assistance in programming and data processing and to the private physicians, hospitals and families whose cooperation made this study possible. I would like to thank Dr. Stanley W. Wright for his understanding, guidance and constant encouragement providing some elements essential for the successful completion of this undertaking.

ABSTRACT

This study was an attempt to fit a model of quasi-continuous variation to three sets of data. The model is based on the assumption that the trait under study is completely additive and wholly polygenic. Conditions due to a major gene, in the sense of classical Mendelian inheritance, would give unreasonable parameter estimates and a poor fit on such a model. Information on population prevalence and recurrence in sibs is used to estimate the parameters and predict risk for other degrees of relationship.

Records on all surgically corrected cases of pyloric stenosis occurring in the period 1942 to 1966 in Hawaii provided the first set of data. Nearly 200 families were interviewed to obtain a pedigree and family history. Birth certificate numbers were matched with a file of all births for that time period to provide additional information on race and sociological variables. The sex ratio was similar to all earlier reports, approximately 4:1. The segregation analysis was compatible with a major gene hypothesis but the estimates had very large standard errors. The regression analysis of interracial crosses indicated a depression in the frequency in F_1, indicative of recessivity. The fit to

the model of quasi-continuity was neither good nor consistent. Such a theory cannot be eliminated as the possible genetic mechanism for pyloric stenosis, but the current model is inadequate to explain the data.

Males with serum cholesterol levels greater than two standard deviations above the mean of a random sample of 7,000 forty to sixty year old Japanese males were selected for a study of hypercholesterolemia. Sibs were contacted, interviewed and serum cholesterol determinations made. The 219 sibships were analyzed by segregation analysis, providing an estimate of the segregation frequency of 0.52 and the proportion of sporadics as zero, evidence for a dominant gene. Estimates of heritability on the additive model exceeded 1.5. This condition appears to be a dominant gene with nearly complete penetrance. This finding is in agreement with other studies.

Information from the Bureau of Identification provided a sample of individuals with a dermal ridge count of zero, an arch pattern on all ten fingers. These individuals were defined as "affected." Ridge counts were made for all available relatives. A control sample was selected from the Bureau files.

The frequency of patterns and mean ridge counts for the Caucasian controls agreed well with published studies from England. The Japanese had higher mean counts and a

significantly higher frequency of whorls. Correlation coefficients computed in the control sample for sib-sib and parent-child relationships agreed well with complete heritability.

Segregation analysis in both the affected families and the controls failed to detect segregation of a major gene. Parameters were estimated for the quasi-continuous model using data from first degree relatives. Fitting these estimates to data from cousins and second degree relatives gave a non-significant deviation.

Evidence from earlier studies has shown fairly conclusively that dermal ridge count is additive and multifactorial. The good fit of our data to this model of quasi-continuity was reassuring. Hypercholesterolemia, on the other hand, seems clearly due to a major gene. The failure to fit the data on pyloric stenosis may only reflect inefficient parameter estimates and lack of sufficient sophistication or accuracy in the current model.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1. INTRODUCTION

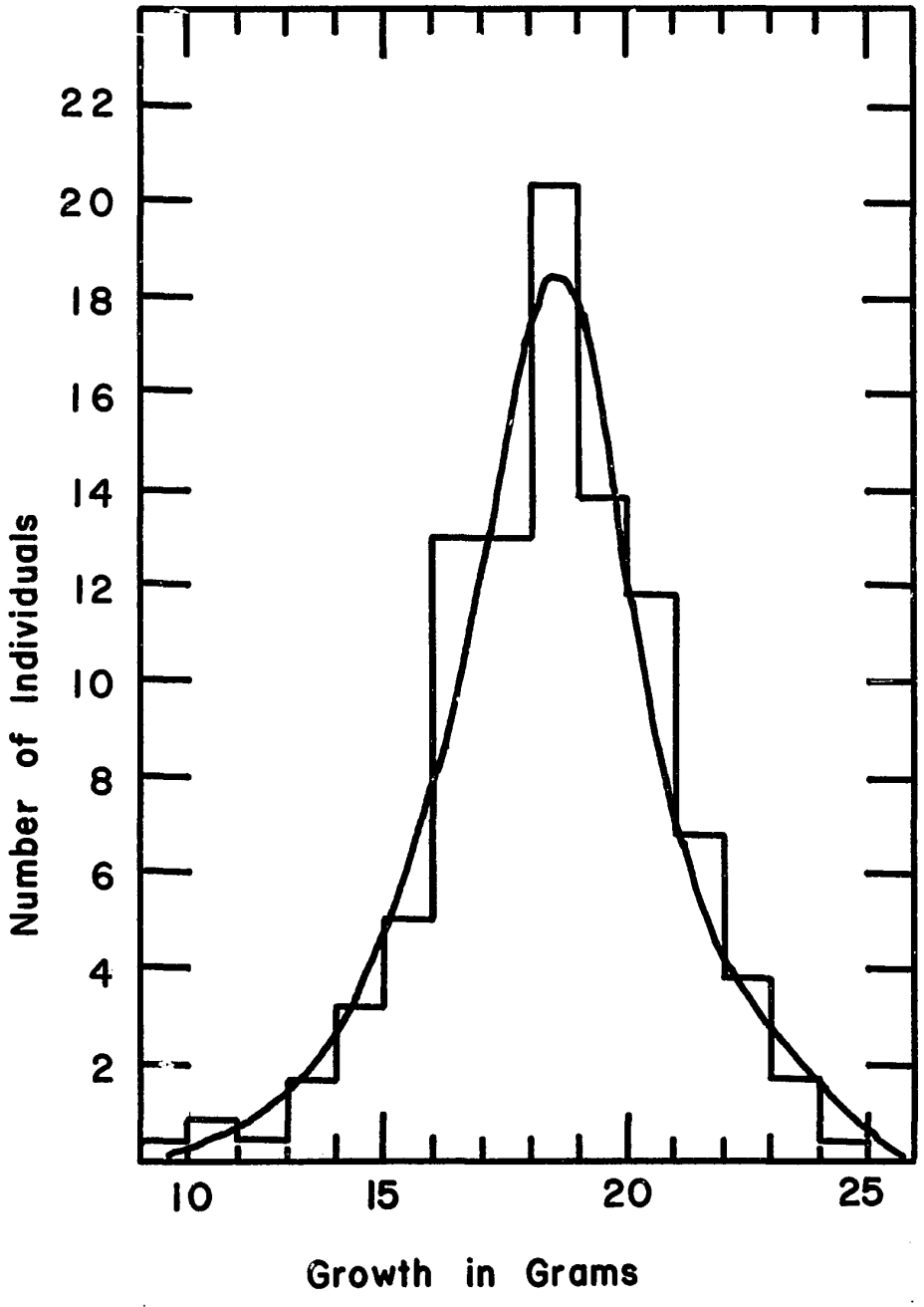
1.1 Background

The classical form of inheritance described by Mendel is that of a single locus with complete dominance. The distinct expression of the dominant gene permits a simple classification of phenotypes. The trait under study may also be the one determined by the recessive allele at the locus. These situations give different ratios for the segregation of the trait in offspring. There are more complex segregation patterns where a large environmental influence obscures the well-defined Mendelain patterns.

Some of the traits which have been subjected to genetic analysis do not exhibit these classical patterns of inheritance. Instead of a few distinct classes the phenotypic expression may assume any of the possible values in a continuous distribution. Such quantitative traits as height, longevity and weight are thought to be determined by many genes at different loci, referred to as multifactorial inheritance (Figure I). Environmental factors may also contribute to the variation in the expression of the phenotype.

There are traits which exhibit an all-or-none or discontinuous phenotypic expression which may still be multifactorially inherited. Although the phenotypic expression of the trait is discontinuous, the trait possesses a

Figure I.



From Falconer (1964), p.107.

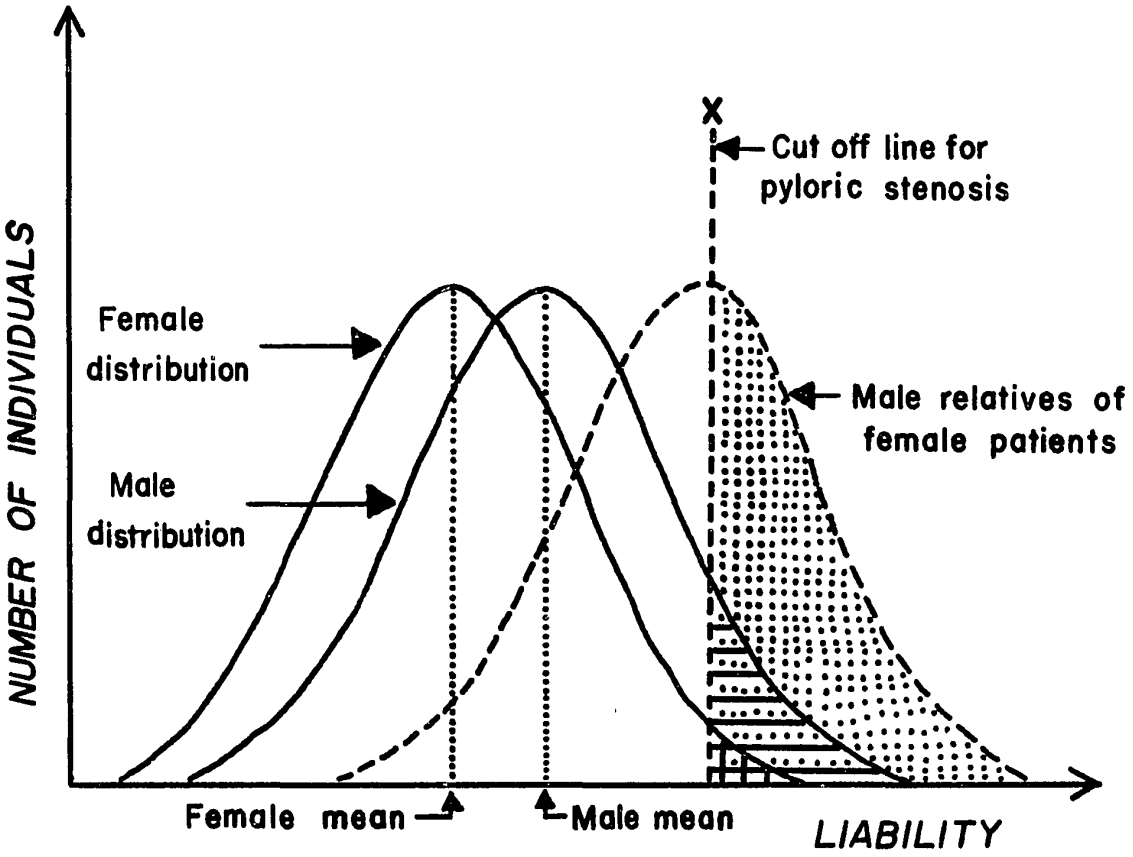
continuous aspect, which, if measured on a suitable scale, will assume a normal distribution. If the trait under study is a disease condition, this continuous aspect can be thought of as the "liability," or the combined genetic and environmental susceptibility to the disease. The continuous character has a threshold which imposes the discontinuity on the visible expression of the genotype. Individuals with a liability value greater than the threshold will appear in the visibly expressed group of affected patients, while individuals with values below the threshold will appear in the unaffected group (Figure II). Grüneberg (1952) referred to this pattern of inheritance as quasi-continuous variation.

The proportion of affected individuals is represented in Figure II by the shaded area to the right of the threshold, x. In the example given for pyloric stenosis two curves are used to represent the different incidences for males and females.

The degree of genetic determination of a trait is the total genetic variance as a proportion of the total phenotypic variance. This cannot be estimated from human data. However, the related parameter of heritability may be estimated. Heritability is defined as the ratio of the additive genetic variance to the total phenotypic variance. The additive variance is the proportion when the effects of the genes are considered singly; i.e., total genetic

Figure II.

HYPOTHETICAL DISTRIBUTION OF MULTIFACTORIAL GENOTYPE CONTRIBUTING TO PYLORIC STENOSIS IN MALES, FEMALES AND MALE RELATIVES OF FEMALE INDEX PATIENTS



From Carter (1961)

variance minus the effect due to dominance and interaction among genes. The estimate of heritability gives information about the relative importance of the genetic component in the etiology because, by definition, the total genetic determination cannot be less than the heritability. Once the heritability is known then, in theory, the incidence can be predicted for any degree of relationship.

Quasi-continuity provides a model for analysis of the genetic factors in a discontinuous trait. The basic assumption that the variation is continuous excludes the possibility of a major gene acting. Such a situation would be detected if estimates from the data were unreasonable; e.g., $h^2 >> 1$. A major gene is used here to mean one whose effect is large relative to the total variance of the trait.

1.2 Falconer's Model for Quasi-continuity

Falconer (1965) applied the concepts of quantitative genetics and quasi-continuity to the type of data available from the study of human diseases. The suggestion that this pattern of inheritance might explain the genetic component in the etiology of some diseases was first proposed by Carter (1961) for pyloric stenosis.

The necessary calculations are those of mean liability of the affected group, A, the mean liability of relatives of affected, R, and the mean liability of the general population, G. The regression of relatives on propositi

in liability gives the regression coefficient for prediction of liability of a relative of a proband as

$$b = \frac{R - G}{A - G} \qquad (1)$$

In human data the liability of an individual cannot be measured directly. By utilizing the properties of the normal distribution the mean for each group is given in standard deviation units from the threshold. Equation (1) becomes

$$b = \frac{x_g - x_r}{a} \qquad (2)$$

where x_g and x_r are the normal deviates for the general population and relatives respectively. The deviation of the mean liability of affected from the population mean is given by a.

It can also be shown from quantitative genetics (Falconer, 1964) that given the phenotypic value of any individual, P, the phenotypic value of a relative, R, and the coefficient of relationship, r, the regression coefficient becomes

$$b_{RP} = \frac{cov_{RP}}{V_P} = \frac{rV_A}{V_P} = rh^2 \qquad (3)$$

where V_A represents the additive and V_P the phenotypic variance defined above, and h^2 is the heritability.

Data on the incidence of a disease in relatives of affected lead directly to estimates of the regression

coefficient. The correlation coefficient r is defined for any specified degree of relationship, permitting a direct estimate of heritability. Heritability in turn estimates the relative importance of the genetic component in the etiology of the condition under analysis.

A complication is introduced into the calculations if the incidence varies by sex, age or time. One apparently puzzling finding was a higher incidence in relatives of females affected with pyloric stenosis when the male to female ratio among affected for this disease is 4:1. However, on this model affected females would be greater deviates from their mean than affected males (Figure II). If liability is inherited to any degree the relatives of affected females would have a higher mean liability resulting in a higher incidence.

The model leads to mathematical difficulties when attempting to derive recurrence risks in relatives (Falconer, 1965; Morton, 1967). This feature alone would make the model unattractive but in addition there is the difficulty of justifying an abrupt threshold on the continuous variate, even if the trait is arbitrarily so defined on a secondary scale. An example would be a threshold imposed on the test score called the Intelligence Quotient where severe mental defect would be defined as an I.Q. below 50. The test score serves as a crude approximation for the underlying

variate, $\underline{x}$, and some individuals below the threshold would be more severely deficient if checked on some other criteria than individuals above the threshold (Morton, unpublished).

Morton (1967) presented a hypothesis for testing for the presence of a major gene segregating against a continuous additive genetic background (Figure III). This approach was presented as an alternative to quasi-continuity which is extremely difficult to disprove. Detection of a major gene would disprove a multifactorial model but failure to detect one in a single analysis would never be sufficient evidence to assert that the model of additivity holds.

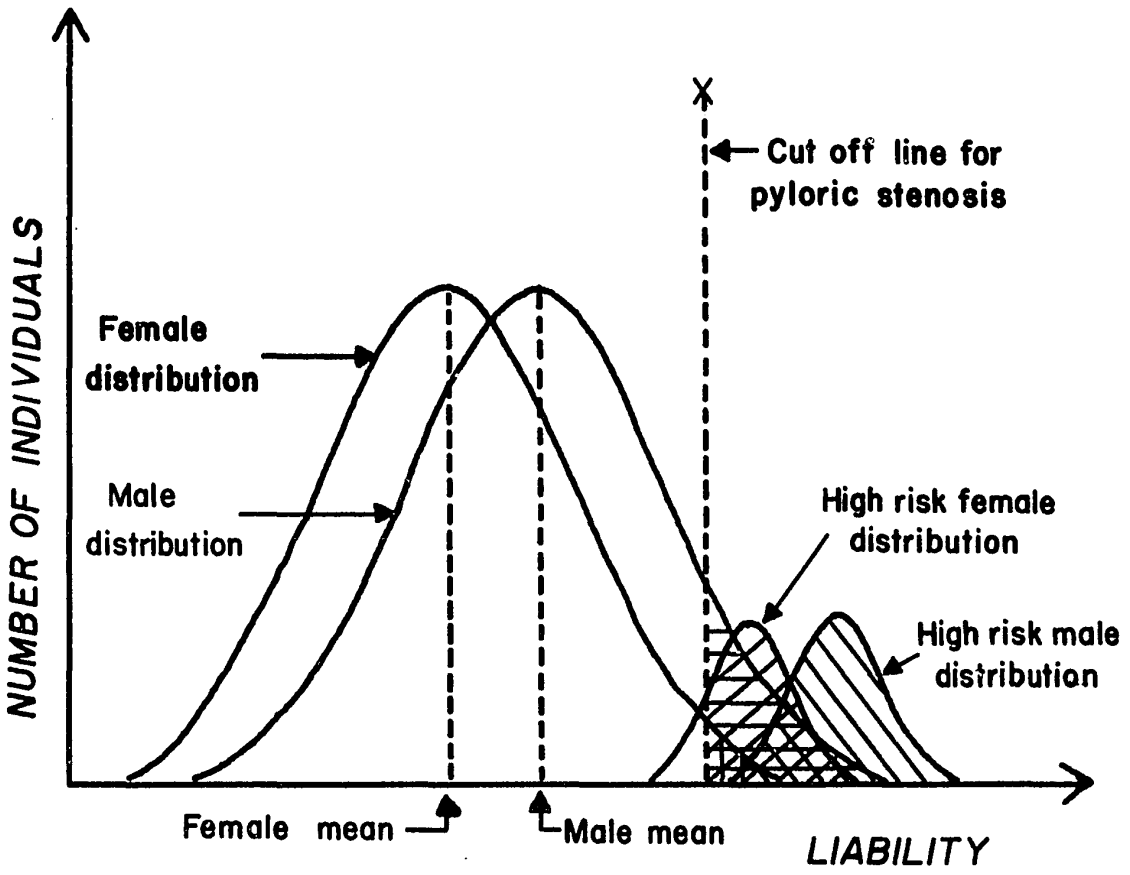
1.3 Edwards' Model for Quasi-Continuous Variation

Edwards (1967) suggested an alternative hypothesis for quasi-continuity. The underlying liability would be a measure of only the genetic variability. Then, under environmental influence the threshold would not be an abrupt cutoff but instead, a function of the liability. A sigmoid curve (Edwards, private communication) would be the most reasonable function for risk. The sigmoid curve, however, presents mathematical difficulties and Edwards chose the exponential form to represent the risk; i.e., $g(x) = ae^{bx}$ (Figure IV). Then in the population the proportion of affected would be $ae^{bx} \cdot k \cdot e^{-1/2\ x^2}$ where again liability, $\underline{x}$, is assumed to be normally distributed.

The extentions of Edwards' model presented here were derived by Elston, Campbell and Morton (in preparation).

Figure III.

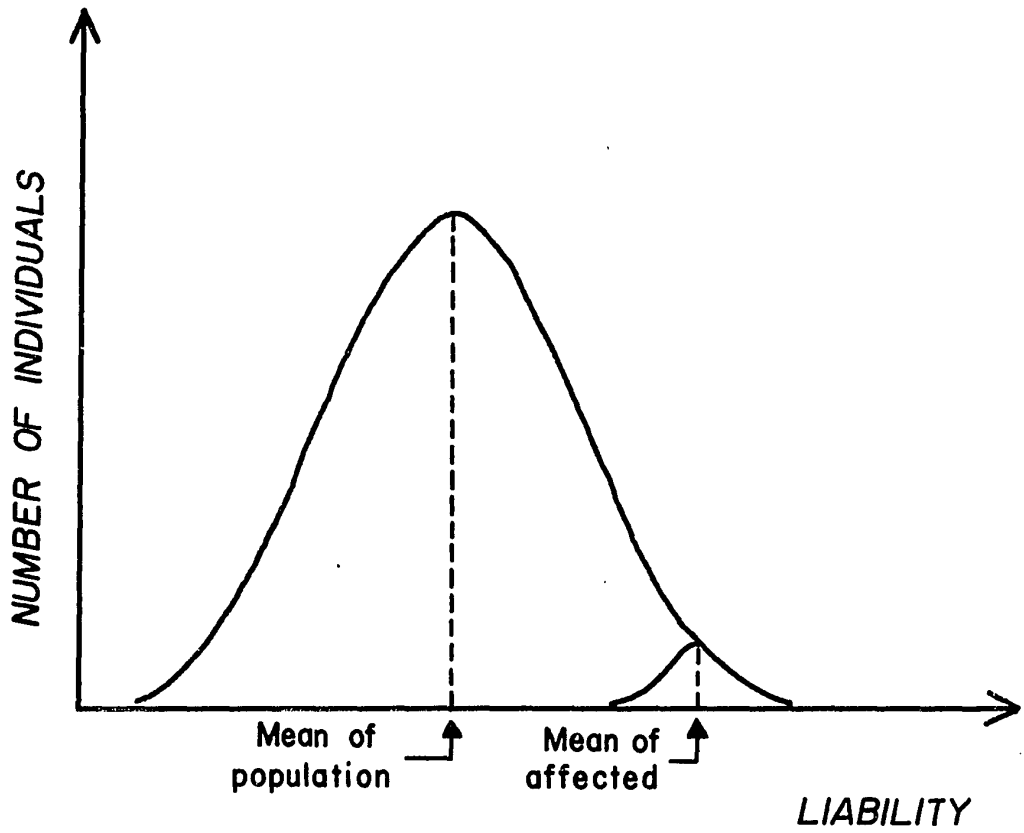
ALTERNATIVE HYPOTHESIS OF SEX-INFLUENCED INCIDENCE OF PYLORIC STENOSIS



From Morton (1966)

Figure IV.

EDWARDS' HYPOTHESIS OF EXPONENTIAL THRESHOLD FOR TRAITS DISPLAYING QUASI-CONTINUOUS VARIATION



The underlying continuous variate, $\underline{x}$, is called "genotype," and is assumed to be wholly polygenic. The genotype is analogous to Falconer's liability but here it is restricted to the genetic effects with exclusion of environmental influence.

In the general population the genotype is assumed to follow the truncate normal distribution

$$f(x) = \frac{1}{\Phi(c)} \cdot \frac{1}{\sqrt{2\pi}} e^{-x^2/2}, \quad x \leq c \qquad (4)$$

where Φ is the standard cumulative normal distribution given by

$$\Phi(c) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{c} e^{-x^2/2} \, dx$$

The truncate normal rather than the untruncated normal was used because a trait determined by a finite number of genes would have to assume a truncate distribution. If there is more than one population then the genotype mean and variance could vary from population to population, and would be denoted for the ith population by $N(\mu_i, \sigma_i)/\left(\frac{c - \mu_i}{\sigma_i}\right)$, or below as the subscripted b_i.

The probability that an individual with genotype $\underline{x}$ will manifest the trait, is designated by the exponential risk function suggested by Edwards (1967) with slightly altered notation; i.e.,

$$g(x) = e^{b(x-c)}, \quad x \leq c, \; b > o \qquad (5)$$

In this situation $\underline{x}$ is taken to be less than or equal to $\underline{c}$ so the risk will not exceed one. One could also assume

that the risk is one after the threshold point c in the form of a constant function. The expected risks would probably not differ greatly with this alteration in the model. However, the mathematics becomes more complex because of the discontinuity at the point c, and all the expressions must then be treated as the sum of two quantities. For these reasons that alternative hypothesis will not be presented here.

All estimates to date (Morton, unpublished; Campbell, unpublished) have indicated that the parameter c in equation (5) remains reasonably constant for any particular disease, and in fact the range of c has been fairly restricted, extending from 3.25 to 4.5. Therefore, in estimating the parameters for different populations; e.g., males and females for pyloric stenosis, only the b_i's will be allowed to differ.

The risk g in the entire population is truncate log normally distributed (appendix A.I.).

$$h(g) = \frac{1}{\Phi(\gamma)} \cdot \frac{\delta}{\sqrt{2\pi}\ g} \cdot e^{-1/2(\delta \ln g + \gamma)^2} \qquad (6)$$

where $\delta = 1/b$, $\gamma = c$, and $g \leq 1$. The prevalence of the trait in the general population would be given by the mean of g; i.e., (appendix A.II.)

$$P = \Phi(c-b)\ e^{1/2(b-2c)} \qquad (7)$$

by ignoring $\Phi(c)$ which is virtually unity.

In order to estimate the parameters b_i and to fit these to actual data classes as a test of the quasi-continuous hypothesis it is necessary to have expressions for the prevalence of the trait in relatives.

For two individuals, $\underline{X}$ and $\underline{Y}$ with coefficient of relationship $\underline{r}$ and class effects b_x, by, the joint distribution in the population is the bivariate truncate normal. If x, y $\leq$ c the truncation factor may be ignored and the distribution given $\underline{X}$ is affected is

$$f(x,y|X) = \frac{f(x,y)\ g\ (x)}{\int_{-\infty}^{c}\int_{-\infty}^{c} f(x,y)\ g\ (x)dxdy}$$

Here the numerator may be written as

$$\frac{1}{2\pi\ \sqrt{1-r^2}}\ e^{-\frac{1}{2(1-r^2)}\ (x-b_x)^2 - 2r\ (x-b_x)(y-b_x r)}$$

$$+ (y-b_x r)^2 \quad . \ e^{1/2\ b_x(b_x - 2c)}$$

The denominator upon integrating the last expression becomes

$$\frac{1}{\sqrt{2\pi}} \cdot \int_{-\infty}^{c} \Phi\left(\frac{c-rx}{\sqrt{1-r^2}}\right) \quad e^{1/2\ (x-b_x)^2} \quad dx$$

The term $\Phi\left(\frac{c - rx}{\sqrt{1-r^2}}\right)$ decreases as $\underline{x}$ and $\underline{r}$ increase. We are interested in situations where $r \leq 1/2$, $x \leq c$ and the smallest value of Φ would occur then at $r = 1/2$, $x = c$; i.e., $\Phi\,(0.58c)$. For $\underline{c}$ as small as 3.25 this is 0.97 and

the error in defining = 1 is negligible. With this approximation we can integrate over $\underline{x}$ from $-\infty$ to $\underline{c}$ and obtain the marginal distribution of $\underline{g}$ given $\underline{X}$ is affected as

$$f(y\ x) = \frac{\Phi\ \dfrac{(c-ry - b_y(1 - r^2)}{\sqrt{1-r^2}}}{\Phi\ (c - b_x\ )} \cdot \frac{1}{\sqrt{2\pi}}\ e^{1/2(y-b_x r)^2} \tag{8}$$

An approximation for equation (8) is $f(y\ X) = N(b_x r,1)/\Phi\ (c - b_x r)$, which certainly holds as r 0. Numerical comparisons were done and for r = 1/4 the ordinates of the two distributions differ less than .01 provided c 3.25, $c - b_x$ 1.5. For the case r = 1/2 the ordinates differed only by .012 when c 3.25 and $c - b_x$ 1.75.

The expected value of g(y) over the distribution f(x X) gives the prevalence of the trait in relatives of affected individuals.

$$P(r) = \overset{c}{} \quad \overset{c}{} \quad g(y)\ f\ (x,y\ 1\ x)\ dydx$$

$$= \frac{\Phi}{\Phi\ (c - b_y)}\ e^{b_y(b_x r - c) + b_y^2/2}$$

where

$$\Phi = \frac{1}{2\pi\sqrt{1-r^2}} \quad \overset{c}{} \quad \overset{c}{} \quad e^{-\frac{1}{2(1-r^2)}\ (x - b_y r - b_y)^2}$$

$$-2r\ (x-b_y r - b_x)(y-b_x r - b_y) +$$

$$(y - b_x r - b_x)^2 \qquad dx \tag{9}$$

or

$$= \frac{1}{2\pi\sqrt{1-r^2}} \int_{-\infty}^{c-b_y r - b_x} \int_{-\infty}^{c-b_x r - b_y} e^{-\frac{1}{2(1-r^2)}\{x^2 - 2rxy + y^2\}} dxdy \quad (10)$$

There are appropriate approximations for Φ if the parameters meet certain conditions; i.e., $c - b_y r - b_x \geq 2.25$ or $c - b_x r - b_y \geq 2.25$. Then it can be shown that the univariate normal approximations never err by greater than 1%. In the subsequent analysis only one Fortran program was available and all calculations were done using equation (10).

Most of the traits to be studied by this type of analysis are rare in the general population. With such traits the mating type of the parents is often normal x normal and the probability for an affected child will not differ greatly from equation (7). If one parent is affected the probability would not differ significantly from (9) since r = .5 for both sib and parent-child relationships. The probability that both parents would be affected for rare traits is quite small. However, the probability for a child to be affected would differ considerably from any of the previous expressions if both parents were affected.

The joint distribution of genotypes x,y for two affected parents X,Y is

$$f(x,y|X,Y) = \frac{f(x,y)\ g\ (x)\ g\ (y)}{\int_{-\infty}^{c}\int_{-\infty}^{c} f(x,y)\ g\ (x)\ g\ (y)\ dxdy} \qquad (11)$$

If X,Y have a child Z with genotype z, then the distribution of $\underline{z}$ follows from the assumption of additive effects as

$N[x+y/2, \sigma\ ]\ \ /\ \Phi\ \frac{2c\ -\ (x+y)}{\sqrt{2(1-r)}}$ The unconditional distribution of Z is $N[(1+r)(b_x + b_y)/2,\ 1]$ This leads to the approximation

$$f(z\ |X,Y) = N[(1+r)(b_x + b_y)/\ 2,1]/\Phi\ [c\ -\ (1+r)(b_x + b_y)/2] \qquad (12)$$

However, this approximation is good only if

$\Phi\ \frac{(1-r)\ (c-\mu\ )}{\sqrt{2\ -\ (1-r^2)}}$ is close to unity. This will not always be the case, and there is an alternative approximation (appendix A.III.).

The probability that out of $\underline{s}$ children $\underline{r}$ would be affected may be obtained from the log normal distribution of $\underline{g}$ in the population and the binomial distribution for affected, not affected in which the parameter is $\underline{g}$.

$$P_s\ (r) = \frac{1}{\Phi\ (\gamma)} \int_0^1 \binom{s}{r} g^r\ (1-g)^{s-r}\ \frac{\delta}{\sqrt{2\pi}\ g}$$

$$e^{-1/2(\gamma + \delta \ln g)^2}\ dg \qquad (13)$$

Upon substituting $g = e^y$ and $b_z = 1/\delta$ this may be written as

$$P_s\ (r) = \frac{1}{\Phi\ (\gamma)} \binom{s}{r} e^{-\frac{\gamma^2}{2}} \sum_{i=o}^{s-r} (-1)^i$$

$$e^{1/2\ \{b_z\ (r+i)\ -\ \gamma\}^2} \cdot \quad \Phi(\gamma\ -b_z\ (r+i) \qquad (14)$$

which was programmed in Fortran IV for data analysis.

It is also desirable to predict risk within a family; i.e., with r affected in a sibship of size s what is the probability that the next child be affected. At the present time this may only be answered if there are no class effects or, the prediction is restricted to a single class.

Given g, the probability of being affected, and $f(g|r,s)$ the distribution of g given r and s, then the required expression is the expected value of g over the distribution or:

$$\frac{1}{P_s\ (r)\ \Phi(\gamma)} \binom{s}{r} g^r\ (1-g)^{s-r}\ \frac{\delta}{\sqrt{2\pi}\ g}$$

$$e^{1/2\ (\gamma + \delta \ln g)^2}\ ,\ o \leq g \leq 1 \qquad (15)$$

Integrating (15) g times with respect to g and substituting for $P_s(r)$ from (14), the risk may be written

$$\frac{\sum_{i=1}^{s-r+1} \binom{s-r}{i-1} (-1)^{i-1} e^{1/2(r + i/\delta)^2 - (r+i/\delta)} \Phi(\gamma - \frac{r+i}{\delta})}{\sum_{j=o}^{s-r} \binom{s-r}{j} (-1)^j e^{1/2(r+j/\delta)^2 - (r+j/\delta)} \Phi(\gamma - \frac{r+j}{\delta})} \tag{16}$$

where obviously j = i + 1 for simplified programming.

Morton (unpublished) suggested a Beta distribution as an approximation for the distribution of g prior to the derivation of the log normal. Skellam (1948) showed that if g did follow the Beta distribution, then

$$f(g) = \frac{1}{\beta(p,q)} g^{p-1} (1-g)^{q-1}, \quad p, q > o$$

and,

$$P_s(r) = \frac{\binom{p+r-1}{r}\binom{q+s-r-1}{s-r}}{\binom{p+q+s-1}{s}} = \binom{s}{r} \frac{\beta(p+r, q+s-r)}{\beta(p, q)} \tag{17}$$

In view of the simplicity of this result it was numerically compared with (16) in the subsequent analysis of data on pyloric stenosis, by equating the means and variances of the log normal and the Beta distributions such that

$$\omega = e^{b_z^2/2} \quad , \quad \rho = e^{-b_z \gamma}$$

$$\frac{p}{p+q} = \omega\rho \quad , \quad \frac{pq}{(p+q+1)(p+q)^2} = \omega^2 \rho^2 (\omega^2 - 1)$$

and finally

$$q = \frac{(1 - \omega\rho)^2}{\omega\rho \ (\omega^2 - 1)} + \omega\rho - 1$$

$$p = \frac{\omega\rho}{1 - \omega\rho} \cdot q$$

The above extensions to Edwards' model were programmed in Fortran IV for use on the CDC 3100 computer. The analysis is carried out in the following manner. Given the prevalence in the general population in the form of incidence figures and the recurrence in relatives, usually from segregation analysis, initial estimates are made for the parameters b and c using maximum likelihood scoring (Morton, unpublished). These estimater are in turn used to calculate δ, γ, ω and ρ for determining the probabilities specified by (13) or (14), (16) and (17). Finally, the observed number of affected in each class of sibs is utilized in a program of searching the maximum likelihood surface to achieve estimates of b_i and c. The four classes used were male and female relatives of male probands, and male and

female relatives of female probands. Then, these estimates were applied to all classes of relatives for which data were available, usually for r = .25 and r = .125 for a test of the goodness of fit of the data to the quasi-continuous model. The results of this and related analyses are presented in the subsequent sections.

2. PYLORIC STENOSIS

2.1 Introduction

Pyloric stenosis is a disease of infancy resulting from a hypertrophied pylorus. Symptoms usually occur within the first two months of life and may be present at birth (Benson et al., 1964). The clinical signs include projectile vomiting and weight loss, which may lead to malnutrition and metabolic alkalosis. Treatment is by a surgical operation in which the muscle fibers of the pylorus are split down to the mucosal lining. Surgical therapy has reduced the mortality rate from 90% to less than 1%.

Males are more commonly affected than females with an average ratio of 4:1 (Gordon et al., 1959). A preponderance of affected firstborn has been reported (Cockayne and Penrose, 1943; Carter, 1961), although the question of birth order effect has not yet been resolved. McKeown (1952) and Gerrard (1955) suggest that this effect is present only among those patients in whom onset of symptoms occurs later than three weeks after birth.

The incidence of pyloric stenosis varies markedly among countries. Laron and Horne (1957) reported the incidence within the United States as 0.12% in Caucasians and 0.046% in Negroes. The incidence in Orientals has been reported as low (Goldschmidt, 1963; Carter, 1961).

The highest incidence noted is 0.4% in Sweden (Wallgren, 1941) and the lowest is 0.0078% in Iran (Gharib, 1964).

The present study was designed to determine the incidence of pyloric stenosis in the racial groups represented in Hawaii and to investigate the nature and extent of genetic factors in the etiology of pyloric stenosis. Clinical data collected during the study will be presented as it relates to incidence and genetic analysis (Shim, Campbell, and Wright, 1969).

In 1943 Cockayne and Penrose proposed that pyloric stenosis was recessively inherited based on an increased familial incidence and the increased consanquinity in parents of affected which were present in their data. They reported a familial incidence of 1/20, with a strong excess of affected males, 86%, and first born, 48.5%.

In 1961 Carter presented a study on pyloric stenosis based on 648 patients who had been treated at The Hospital for Sick Children, London, between 1920 and 1940. His series included 562 males and 96 females. The proportion of affected children of index patients was 6.5% and was higher than the proportion of sibs affected, 4.0%. This does not indicate a recessive gene hypothesis.

Investigators of the genetic aspects of pyloric stenosis have demonstrated familial concentration, the incidence in relatives to increase with the degree of relationship (Falconer, 1965), and an appreciable concordance for monozygotic twins (Carter and Savage, 1951;

Powell and Carter, 1951). These findings strongly suggest a genetic component in the etiology of this disease.

2.2 Methods

The data for this study were collected from the records of major hospitals in Hawaii from 1942 to 1966. This time period was chosen to coincide with the information available from census reports and the Bureau of Statistics of the Department of Public Health. For supplementary information the State of Hawaii death certificates were checked for notations of pyloric stenosis in stillbirths, fetal deaths, and neonatal deaths.

In this study surgical treatment was the criterion for diagnosis of pyloric stenosis. The 276 propositi, 228 males and 48 females, were surgically treated and the characteristic hypertrophied pylorus was confirmed at surgery. The four medically treated cases were omitted from the data. The surgery had to have taken place within the state, but the child may have been born elsewhere. For estimates of incidence the inclusion of these few patients is balanced by those patients born in Hawaii who developed the condition after leaving the state. Sibs and other relatives who had surgery elsewhere or outside the designated time period were considered affected but not propositi.

Clinical information including physical signs, results of diagnostic studies and post-operative complications was taken directly from the hospital records (Shim et al., 1969). Data were collected from the public hospitals on 148 resident families of various racial extractions. Permission was obtained from the family physician, or attending surgeon, to visit each family in order to obtain a family history, pedigree, and racial information. Families who had moved from the islands were sent a questionnaire by mail.

A total of 111 of the resident families were interviewed, 25 could not be located, and 12 were illegitimate or adopted children. There were an additional 125 predominantly Caucasian military families from Tripler General Army Hospital. Seven of these were known to be multiplex families; i.e., more than one sib affected. Only four of these families were located and interviewed. No effort was made to contact any of the other 118 Tripler families, as it was felt that too much time would be required to locate the reassigned military personnel. However, family information and history were available from the interview portions of the medical charts for 73 of these families. These were randomly distributed over the 18 year period. The remaining 45 Tripler families were included only in the incidence calculations.

The average duration from birth to surgery for the local resident families was 39.0 days, while the average duration for the military families was 43.6 days. A linear step-wise regression of resident and military families on sex and duration indicated that this difference was not significant for location ($t = 0.306$, 271 d.f.) nor for sex ($t = 0.439$, 271 d.f.). This suggests that diagnosis and treatment were homogeneous between the two groups and this permits the pooling of the data for both the clinical and genetic analysis.

A complete file of birth certificates was available for the 383,079 live births included in the survey of this study. The information on these certificates included race, mother's age, birth weight and birth order. By utilizing this file as prepared for the CDC 3100 computer, it was possible to confirm interview information and to provide additional information on affected children.

Family data were available on an additional 249 Caucasian families from a ten-year study of pyloric stenosis. These data were provided by Cedric O. Carter of the Institute of Child Health, Clinical Genetics Research Unit, London. In Carter's data medically treated cases were included if the characteristic tumor had been palpated by a pediatrician. This criterion was used in his data because medical treatment is more frequent in England than in the United States for confirmed cases of pyloric stenosis. Any

doubtful cases were omitted from the data (Carter, personal communication).

The racial structure of Hawaii has been described in detail by Morton et al. (1967). In the present study individuals were also classified using the seven major racial groups: Caucasian, Japanese, Chinese, Hawaiian, Filipino, Korean, and Puerto Rican; and one group for "others" which includes Negro, Samoan and other Pacific groups. A final small group of "part-Hawaiian, not otherwise specified" completes the racial categories. This classification permits 900 possible mating combinations of which Morton et al. were able to observe 524.

The Portuguese and Mexican groups were included in the Caucasian group for analysis. Other group variables were designed to measure Oriental and Caucasian influence, hybridity, etc. These are described in detail below in the discussion of the regression analysis.

2.3 Racial effects

2.3.1 Multiple regression analysis

The live birth certificate file for the State of Hawaii mentioned above was used in the analysis of the effects of sociological and racial factors on pyloric stenosis. The entire file was matched with the birth certificate numbers of the known pyloric stenosis cases ascertained through our study. This matching process generated a data tape which included all the live births

for the 25 year period of study, each coded zero or one to indicate not affected or affected respectively.

The data were analyzed by the method of multiple regression using the CDC 3100 fortran program MULREG. This program carries out a least squares regression analysis, providing as output the sums of squares and cross products, sums of squares due to regression and residual sum of squares, degrees of freedom, partial regression coefficient and standard error for each independent variable, and a test of significance (Burian, in press).

The analysis was carried out using four categories or classes of variables: sociological, maternal, general and hybridity factors. The analysis parallels that of Morton, Chung and Mi presented in Genetics of Interracial Crosses in Hawaii (1967).

The sociological factors represent those which are to be eliminated from the genetic analysis. This is accomplished through the use of covariance analysis. Examples of these variables are year, maternal and paternal age, father's occupation, birth order and their higher power terms.

The remaining categories are summarized in Table 2.1. Caucasian mothers are given the value zero for all six maternal variables, designated M_i, i = 1 to 6. In this way the mean for Caucasian mothers is the regression intercept and the effects are measured as deviations from the

Table 2.1 Coding for variables for interracial crosses analysis

Atlantic = Caucasian + Puerto Rican; Pacific = Hawaiian + Chinese + Filipino + Japanese + Korean; 20 = part-Hawaiian not otherwise specified

Class of effect	Variable	Effect	Definition
Maternal	M_1	Pacific mother	0 for Atlantic, 1/4 for 20, 1/2 for other part-Pacific, 1 for Pacific
	M_2	Hawaiian mother	0 for non-Hawaiian, 1/2 for part-Hawaiian, 1 for Hawaiian
	M_3	Chinese mother	0 for non-Chinese, 1/4 for 20, 1/2 for part-Chinese, 1 for Chinese
	M_4	Filipino mother	0 for non-Filipino, 1/2 for part-Filipino, 1 for Filipino
	M_5	Puerto Rican mother	0 for non-Puerto Rican, 1/2 for part-Puerto Rican, 1 for Puerto Rican
	M_6	Korean mother	0 for non-Korean, 1/2 for part-Korean, 1 for Korean
General	G_i---G	General combining effects	half the sum of M_i and the corresponding paternal effect

Table 2.1 Coding for variables for interracial crosses analysis (continued)

Class of effect	Variable	Effect	Definition
Hybridity			Let p_i, m_i = proportion of ith Atlantic race in father, mother, respectively ($i = 1,2$) q_i, n_i = proportion of ith Pacific race in father, mother, respectively ($i = 1,..,5$)
	H_1	Major maternal hybridity	$4 \sum_{i,j} m_i n_j$
	H_2	Minor maternal hybridity	$4(\Sigma m_i m_j + \Sigma n_i n_j)$, $i \neq j$
	H_3	Major recombination	$2(\sum_{i,j} m_i n_j + \Sigma p_i q_j)$
	H_4	Minor recombination	$2(\Sigma m_i m_j + \Sigma n_i n_j + \Sigma p_i p_j + \Sigma q_i q_j)$, $i \neq j$
	H_5	Major hybridity of child	$(\Sigma p_i)(\Sigma n_i) + (\Sigma m_i)(\Sigma q_i)$
	H_6	Minor hybridity of child	$\Sigma p_i m_j + \Sigma q_i n_j$, $i \neq j$

From Morton, N., Chung, C., and Mi, M.-P. 1967. pp. 40-41.

Caucasian mean as standard. The Japanese mothers are placed in a Pacific variable and the Hawaiian, Chinese, Filipino and Korean effects measured as deviation from the Pacific mean. The contribution of the race to the mother determines the value of the variable; i.e., the value 1 if she is entirely of the race, 1/2 if biracial, etc. Part Hawaiian mothers not otherwise specified are treated as 1/2 Hawaiian and 1/4 Chinese since the data from Morton, et al. (1967) indicate that this group is 1/2 Hawaiian, 1/4 Caucasian, and 1/4 Oriental, mostly Chinese.

The general variables, G_i, are defined as half the sum of the paternal and maternal effects. These measure the additive effects of genes received from both parents and the average environmental effect of the parental ethnic group. This excludes the environmental effects controlled by the covariance analysis of the sociological variables.

Major hybridity is defined as between Atlantic and Pacific races while minor hybridity refers to within the major groups. The six hybridity variables, indicated by H_i, measure the effects of maternal hybridity, recombination, and heterozygosity of the child.

The analysis is carried out in a stepwise manner within mating types, the data having been previously sorted into mating types by parental racial group. The first step in the process is a regression on the preforced sociological

variables, a pre-forced variable being one that is included in the analysis before regression regardless of the statistical significance. With environmental variability thus controlled the regression continues on the unforced or independent racial variables. The hybridity variables are then included in the analysis even if they were not found to be significant in the step-wise procedure. This insures estimates of the hybridity effects.

The significance is tested by comparison with the residual mean square. The sociological effects are compared with the residual mean squares within mating type while the racial effects are compared with the residual mean squares among the mating types.

2.3.2 Results

The data were processed forming the matrices of sums of squares and cross products and providing means and group totals. The mean for the pyloric stenosis variable provided the incidence estimate for each racial group as listed below. The sample size and small number of cases make some of the estimates unreliable, particularly for the mixed racial groups. The incidence for Caucasians and Puerto Ricans compare favorably with published figures. The Oriental groups have a lower incidence but not as low as reported for Negroes in the United States (Laron and Horne, 1957).

Table 2.2

Incidence of Pyloric Stenosis in Hawaii
1942-1966

Racial Group	Incidence Per 1000 Live Births	No. of Births
Caucasian	1.8496 ± 0.15	83,257
Chinese	0.0	11,274
Filipino	0.0884 ± 0.06	22,630
Japanese	0.5200 ± 0.08	73,831
Puerto Rican	1.0389 ± 0.52	3,850
Korean	0.8264 ± 0.82	1,120

The step-wise regression was performed on the sociological variables. Those which were significantly related to pyloric stenosis are listed in Table 2.3 with their partial regression coefficients, means, and significance test values.

The strong association with year is not surprising since the recent hospital records proved more accessible and more accurate with the introduction of coding systems and full time medical librarians.

The associations with birth order and sex were expected from their consistent appearance in the literature (McKeown *et al.*, 1951; Gordon *et al.*, 1959). An overall sex ratio of 4.8 to 1 males to females or 82.8% males also compared favorably with other studies (81.7% males, MacMahon *et al.*, 1951a). The Caucasian ratio was 4.8:1

Table 2.3

Variables with a Significant Effect on Pyloric Stenosis

Variable	Regression Coefficient ($\times 10^{-3}$)	Mean	t Test Value (∞ d.f.)
SOCIAL:			
Year of birth	.05763	55.998	7.036
Sex	.98939	1.514	10.049
First born[1]	.22944	-0.465	2.043
Armed Forces[2]	.42851	0.246	3.027
Birth weight	.12638	7.031	3.013
RACIAL:			
G1 Pacific	-.91056	.15910	-5.723
G2 Hawaiian	-.59164	.29822	-1.982
G4 Filipino	-.52998	.19304	-2.746
G5 Puerto Rican	-.77338	.39312	-1.967
G3 Chinese	-.28874	.24813	-1.164
G6 Korean	.23924	.59709	.401
HYBRIDITY:			
H5 Major hybridity of child	-.60578	0.1649	-3.271

1. coded (-1) if first born, (+1) for all other birth orders.

2. coded (1) if father in Armed Forces, (0) for all other occupations.

and the Japanese was 3.3:1. This difference was not significant with a chi square of 0.8173 and $p \sim 0.4$.

McKeown (1951b) reported that the sex ratio was not significantly different between first born and later births. This was tested in our data by comparing first births with all birth orders greater than one pooled. These were summed over all families for which birth order information was available and gave the following:

Table 2.4

Birth Order by Sex for Pyloric Stenosis

	Male	Female	Total
First Births	340	84	424
Later Births	349	73	422
	689	157	846

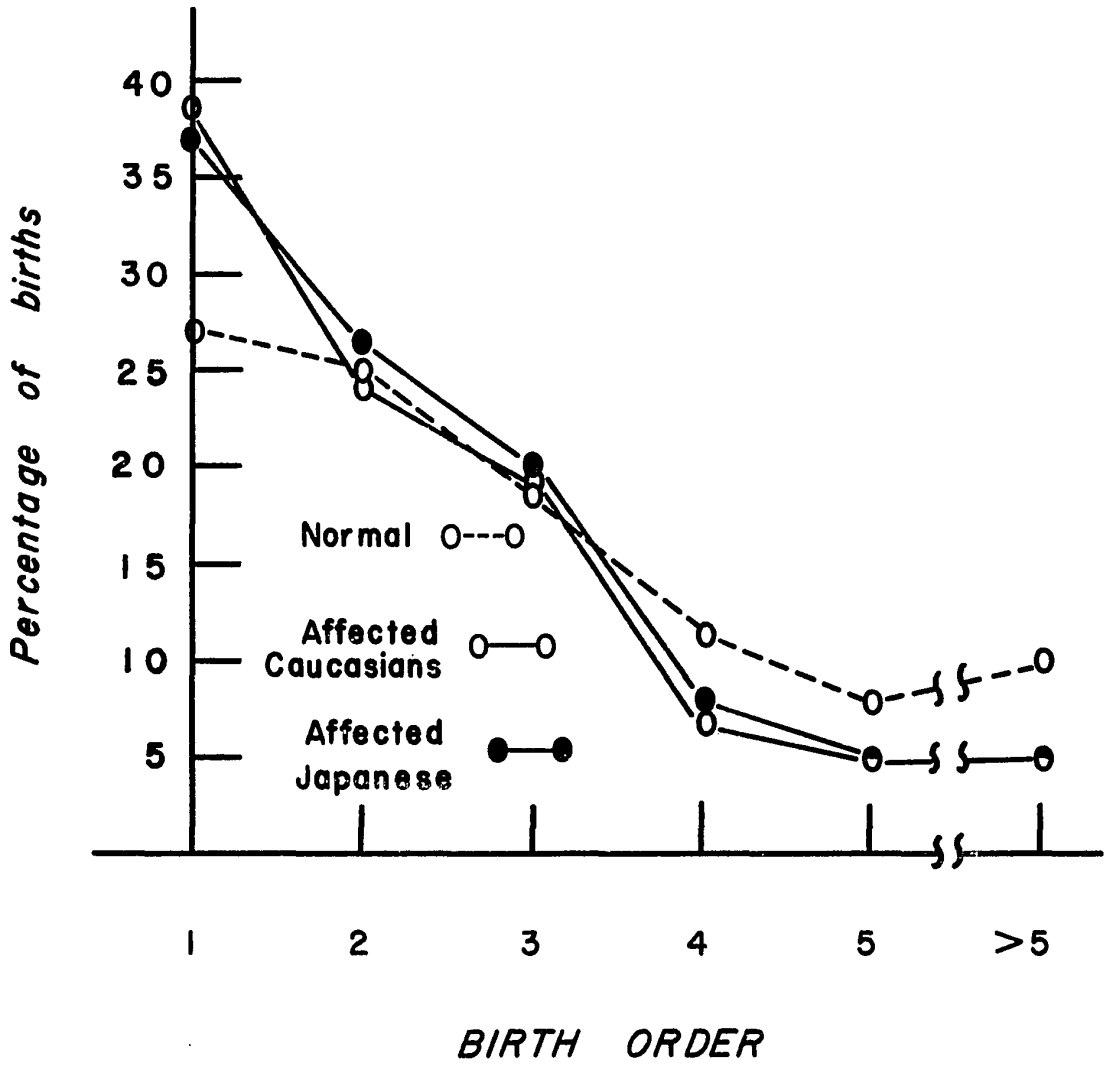
$X_1^2 = 0.8833$, $P \sim 0.4$

This is not significant and there is no greater distortion of the sex ratio for first births in our data.

The birth order effect is presented in Figure V. The analysis revealed a definite increase in first and second births. The deviation from total births was significant, chi square of 10.6 and $P < 0.01$ but the difference between Caucasian and Japanese was not significant, chi square of 1.7011 and $P \sim 0.15$. The variable coded zero or one to indicate a first birth was significant in the multiple

Figure V.

Distribution of birth order in affected Caucasian, affected Japanese and total normal live births



regression, $\underline{t}$ = 2.043 with ∞d.f., but the variable indicating birth order was not significant.

The correlation with birth weight was initially felt to be a spurious result of the highly significant relationship with first birth and sex. An additional regression was prepared to determine all sociological variables significantly related to birth weight. Year of birth, sex, birth order and parental age were significant. These variables were pre-forced into the regression model and birth weight was regressed on pyloric stenosis within racial groups. The resulting regression had a positive slope indicating affected babies were heavier and was highly significant, t = 2.9956, $\underline{P}$ ~ 0.01. The mean birth weights for the two largest racial groups, Caucasian and Japanese, are compared below as well as the mean for the part Caucasian part Japanese group. The mean for each group is expressed in pounds.

Table 2.5

Mean Birth Weight in Pounds

Race	Normal	Affected
Caucasian	7.222	7.565
Cauc/Japanese	7.073	7.359
Japanese	6.909	7.129

Not only is the birth weight effect constant within groups but appears to be additive among groups.

The variable for father in the armed services was significant. This resulted from ascertainment of almost all affected Caucasians at Tripler Army Hospital, and the Caucasian group constituted the majority of cases ascertained.

The significant sociological variables were pre-forced and the step-wise regression executed on the racial variables. The variables whose effects were significant are also listed in Table 2.3.

None of the six maternal variables were significant in the regression analysis. In a homogeneous environment these variables are designed to measure the direct effect of the mother's genotype. The environment under study may not be homogeneous enough to permit a completely genetic interpretation of these variables, but the analysis does show that the maternal influence measured does not have a significant effect on the occurrence of pyloric stenosis.

The general Pacific variable, G_1, was the most strikingly significant, demonstrating their greatly reduced incidence as compared with the Caucasians. The remaining five general variables measure the effects of the other groups in addition to the Japanese effect given by G_1. The only positive coefficient was for the Korean group, G_6. This is indicative of an incidence greater than that for the Japanese group, but still below the Caucasian

estimate. The Korean incidence estimate has a large standard error and in fact this group does not differ significantly from the Japanese, $\underline{t} = 0.40067$.

The variable measuring minor hybridity, H_6, or crosses within groups was slightly negative but not significant. However, the major hybridity variable H_5 was highly significant, with $p < .01$. In racially mixed mating types the incidence of affected children tends to resemble that of the parent with the lower incidence, an effect suggestive of recessivity.

The regression analysis did detect significant sociological effects but failed to show any maternal effect. The group and hybridity variables demonstrate what is clear from the incidence calculations with additional information on the recessive effect in F_1 progeny.

2.4 Segregation analysis

The methods for segregation analysis were primarily developed by N. E. Morton (1958, 1959, 1962, 1965, 1967). A segregation analysis computer program, SEGRAN, prepared for the CDC 3100 computer was used for the analysis (Yee, *et al.*, 1969).

Three parameters were estimated in the segregation analysis and used to describe the genetic trait under study:

(i) $\underline{p}$ is the segregation frequency; e.g., for recessive genes in normal X normal matings, $p = 1/4$.

(ii) $\underline{x}$ is the proportion of cases that are sporadics; i.e., due to mutation, phenocopies, rare heterozygote expression, et cetera.

(iii) $\underline{\pi}$ is the ascertainment probability; i.e., the probability that an affected individual in the population will be a proband.

The data were grouped by family size and the distribution of affected and probands within sibship size was analyzed. Both parents and children were classified as either affected (with pyloric stenosis) or normal. The affected sibs, $\underline{r}$, plus the normal sibs, $\underline{c}$, sums to the sibship size, $\underline{s}$, $(r + c = s)$. The number of probands within the family was designated $\underline{a}$, where $\underline{a}$ must be less than or equal to $\underline{r}$.

The probability distribution of $\underline{r}$ affected among $\underline{s}$ sibs was used to provide information about the segregation frequency $\underline{p}$, where

$$P(r = 1 | a > 0) = \frac{s\, p\, \pi\, [\, x + (1-x)(1-p)^{s-1}\,]}{xsp\pi + (1-x)\, [\, 1-(1-p\,\pi)s\,]} \qquad (18)$$

$$P(r, r > 0 \mid a > 0) = \frac{(1-x)\, \binom{s}{r}\, p^{r}\, (1-p)^{s-r}\, [\, 1-(1-\pi)^{r}]}{sp\pi\, x + (1-x)\, [\, 1-(1-p\ \ )^{s}\pi]} \qquad (19)$$

The probability for the distribution of a probands among r affected

$$P(a|a>o) = \frac{\binom{r}{a}\pi^{a}(1-\pi)^{r-a}}{1-(1-\pi)^{r}} \qquad (20)$$

gives information about π, free of assumptions concerning p and x, when the probands are independently ascertained.

The ascertainment probability is often estimated by the distribution of t independent ascertainments per proband. When more than one independent ascertainment is possible; e.g., birth certificates, death certificates, hospital or institutional records and private physicians, etc., then t indicates the number of these sources through which each proband was ascertained. In the present study the only independent ascertainment was the hospital surgical record. Therefore, the distribution of t could not provide information about the ascertainment probability. This situation was approached in a slightly different manner in each of the four sets of data subjected to segregation analysis.

The first set of data was the 111 resident families provided by the present study. The method of ascertainment was incomplete selection; i.e., sampling through the affected children. This type of selection, also referred to as multiple selection, means that both simplex (one child affected) and multiplex families would be included

in the sample. One proband would be sufficient to insure selection of a family but a family could contain as many as r probands. Pedigrees of the multiplex families ascertained in this study are given as Appendix B. A preliminary calculation of the Caucasian incidence compared favorably with the known mainland incidence and provided support for the prediction of a high ascertainment probability. The probability model (20) was used in estimating π .

Two additional sets of similar data were provided by other workers; the first being Carter's sample of 249 families from England. The method of study paralleled the one employed in our data collection. However, this survey was restricted to only one hospital over a specified ten-year time period.

McKeown *et al*. (1951c) published an appendix of their family information. The completeness of this appendix made possible the addition of their 476 families for the segregation analysis. These data included information on each parent, birth order and sex for each child. It was indicated whether or not the affected individuals were probands. The method of ascertainment was similar to the other sets of data with incomplete selection. Unlike Carter's sample this study included all the hospitals for the city of Birmingham over a ten-year time period.

In these third and fourth groups of data the ascertainment probability, $\underline{\pi}$, was also based on the distribution of probands among affected, model (20). No information could be provided by the distribution of $\underline{t}$ ascertainments. In Carter's data there were no families with more than one proband; i.e., $a > 1$. It was felt that the ascertainment probability would be lower for this body of data.

The second group of data, the Tripler Army Hospital families, differed in selection model and could not be pooled with the other sets of data. It was omitted from the initial segregation analysis estimates.

The remaining three groups were pooled and tested for heterogeneity. The comparison was made using the uncorrected chi square, here represented by $\underline{\tau} \cdot \underline{\tau} = X^2 - (UK^{-1}U')$ where the X^2 and $(UK^{-1}U')$ scores were obtained from the fit of the data to a given hypothesis and the values provided by the program SEGRAN. The number of degrees of freedom was calculated from the sum of the observed classes of sibship size minus the number of parameters estimated. The difference in the sum of the corrected chi square for each of the three groups and the corrected chi square taken over all groups provided the following value:

ra Table Uncorrected chi square $\underline{\tau} = 14.51$,
12 d.f., $P \sim 0.3$

On the hypothesis that the data were homogeneous this chi

square value gives a P value which is not significant. The groups, therefore, were homogeneous and could be pooled for the analysis.

The same calculation for P over the sr table involves acceptance of large sample theory applied to observations of size one for large sibship size, which is not necessarily reliable. However, the results confirmed those from the ra table and on the hypothesis of homogeneity gave a P value of approximately 0.5.

Families with one or more parent affected were deleted and the remaining 828 families compiled as shown in Tables 2.6 and 2.7. The iteration of p, x, and π using the SEGRAN program over the combined groups gave the following maximum likelihood estimates:

$$\hat{p} = 0.19630 \pm 0.11749$$

$$\hat{x} = 0.78941 \pm 0.10472$$

$$\hat{\pi} = 0.59696 \pm 0.08420$$

When x was set equal to zero and held constant during the iteration for p and π the following estimates were obtained:

$$\hat{p} = 0.03592 \pm 0.00616$$

$$\hat{\pi} = 0.59402 \pm 0.08450$$

A comparison of these two sets of estimates was used to indicate if there were a division of the data into a group of families with a high risk of recurrence and a group with

Table 2.6 Segregation Analysis, Pyloric Stenosis
Normal x Normal Matings (sr Tables)

sr Table $p = 0.19630$, $x = 0.78941$, $\pi = 0.59696$

Children s	Affected r	CAMPBELL No.	U_p	U_x	U_π	CARTER No.	U_p	U_x	U_π
1	1	17	0	0	0	35	0	0	0
2	1	30	-4.68086	4.36323	0.62783	87	-13.57450	12.65338	1.82070
2	2	1	5.15789	-4.80789	-0.69181	3	15.47366	-14.42366	-2.07543
3	1	22	-5.43549	5.87955	0.85871	60	-14.82407	16.03515	2.34210
3	2	-				5	19.84353	-24.31952	-3.36852
4	1	18	-5.18361	6.60321	0.98305	36	-10.36721	13.20642	1.96610
4	2	1	2.77184	-4.91676	-0.65812	2	5.54367	- 9.83353	-1.31625
5	1	10	-2.91888	4.46249	0.67933	8	- 2.33510	3.56999	0.54346
5	2	1	1.56810	-4.96661	-0.64481	-			
5	3	-				-			
6	1	7	-1.88898	3.55472	0.55467	4	- 1.07942	2.03127	0.31696
6	2	-				-			
7	1	3	-0.69245	1.66199	0.26627	-			
8	1	-				-			
9	1	1	-0.12782	0.60952	0.10323	-			
9	3	-				1	3.03694	- 5.13586	-1.05046
10	1	-				-			
13	1	-				-			
TOTAL		111	-11.43026	12.44345	2.07841	241	1.1750	- 6.21936	-0.82133

$$K = \begin{bmatrix} 41.84973 & -52.97345 & -7.41946 \\ & 81.67461 & 11.24119 \\ & & 1.55224 \end{bmatrix} \qquad \begin{bmatrix} 192.89263 & -225.10538 & -32.58519 \\ & 275.98725 & 40.24789 \\ & & 5.97384 \end{bmatrix}$$

Table 2.6 Segregation Analysis (continued)

sr Table $p = 0.19630$, $x = 0.78941$ $\pi = 0.59696$

Children s	Affected r	No.	U_p	U_x	U_π	No.	U_p	U_x	U_π
1	1	150	0	0	0	202	0	0	0
2	1	152	-23.71637	22.10705	3.18100	269	-41.97174	39.12366	5.62953
2	2	9	46.42097	-43.27098	-6.22629	13	67.05251	-62.50253	-8.99354
3	1	82	-20.25956	21.91470	3.20087	164	-40.51913	43.82940	6.40174
3	2	3	11.90612	-14.59171	-2.02111	8	31.74965	-38.91124	-5.38962
4	1	38	-10.94317	13.94011	2.07533	92	-24.49399	33.74975	5.02447
4	2	2	5.54367	- 9.83353	-1.3.625	5	13.85918	-24.58381	-3.29062
5	1	21	- 6.12964	9.37122	1.42659	39	-11.38362	17.40370	2.64939
5	2	1	1.56810	- 4.96661	-0.64481	2	3.13619	- 9.93322	-1.28961
5	3	3	7.90658	- 4.96661	-1.08576	1	7.90658	- 4.96661	-1.08576
6	1	3	- 0.80956	1.52345	0.23772	14	- 3.77796	7.10944	1.10935
6	2	1	0.35823	- 5.01360	-0.63350	1	0.35823	- 5.01360	-0.63350
7	1	6	- 1.38489	3.32398	0.33254	9	- 2.07734	4.98597	0.79881
8	1	3	- 0.54519	1.76149	0.29008	3	- 0.54519	1.76149	0.29008
9	1	1	- 0.12782	0.60952	0.10323	2	- 0.25563	1.21904	0.20646
9	3	-				1	3.03694	- 5.13886	-1.05046
10	1	2	- 0.14570	1.24609	0.21707	2	- 0.14570	1.24609	0.21707
13	1	1	0.07661	0.62657	0.11852	1	0.07661	0.62657	0.11852
TOTAL		476	9.71839	- 6.21886	-0.54477	828	0.00562	0.00523	0.71231

$$K = \begin{bmatrix} 381.46993 & -374.43969 & -56.16773 \\ & 425.39207 & 61.58314 \\ & & 9.07545 \end{bmatrix} \quad \begin{bmatrix} 616.21229 & -652.51853 & -96.17239 \\ & 783.05393 & 113.07221 \\ & & 16.60153 \end{bmatrix}$$

Table 2.7 Segregation Analysis

Pyloric Stenosis, normal x normal matings
(ra table)

ra Table $\pi = 0.59696$

Affected r	Probands a	CAMPBELL No.	CAMPBELL U_π	CARTER No.	CARTER U_π	McKEOWN No.	McKEOWN U_π	TOTAL SAMPLE No.	TOTAL SAMPLE U_π
1	1	108	-1.76841	230	0	459	0	797	0
2	1	1	-1.76841	10	-17.68405	6	-10.61043	17	-30.06289
2	2	2	4.77578	0	0	10	23.87892	12	28.65471
3	1	-	-	1	- 3.80860	0	0	1	- 3.80860
3	2	-	-	0	0	0	0	0	0
3	3	-	-	0	0	1	4.50400	1	4.50400
TOTAL		111	3.00738	241	-21.49265	476	17.77249	828	- 0.71278
		K	=12.66828	K	= 51.41663	K	= 76.75320	K	= 140.83811

a low risk. Such a division would result if $\underline{x}$ differed significantly from zero (Morton, 1965, 1967). There are three chi square tests available, one from the $UK^{-1}U'$ score, one from the L score and one from the differences of the chi squares for goodness of fit. By large sample theory these are equivalent tests. With the null hypothesis that $\underline{x} = 0$, they gave the following:

TEST	CHI SQUARE VALUE	P
$UK^{-1}U'$	11.92	0.005
L	5.68	0.02
X_1^2	17.96	0.0001

All three tests were significant and the null hypothesis can be rejected: $\underline{x}$ differs significantly from zero. There is a low risk group and a high risk group in this data, or at least the risk varies markedly among families.

An additional test for internal consistency of the data was made by comparing the Caucasian and Oriental portions of the Hawaii data to the parameter estimates from the entire sample. Neither sample deviated significantly from the pooled sample nor from the other racial group. The values for the parameters were homogeneous between racial groups.

Families with an affected parent were analysed separately (Table 2.8). These are families which were ascertained through an affected child. There were no

Table 2.8 Segregation Analysis

Pyloric Stenosis Normal x Affected Matings

sr Table $p = 0.19630$, $\chi = 0.41993$, $\pi = 0.59696$

Children	Affected		ALL THREE GROUPS		
s	r	No.	U_p	U_x	U_π
1	1	2	0	0	0
2	1	2	-0.95075	0.32174	0.11787
2	2	1	5.27347	-1.78458	-0.65380
3	3	3	-2.49559	0.97532	0.33692
4	1	1	-1.07109	0.48753	0.16040
TOTAL		9	0.75603	0.00001	-0.03860

$$K = \begin{bmatrix} 31.48474 & -10.89744 & -3.95592 \\ & 3.79127 & 1.37346 \\ & & 0.49797 \end{bmatrix}$$

ra Table $\pi = 0.59696$

Affected	Probands		
r	a	No.	U_π
1	1	8	0
2	1	1	-1.76841
TOTAL		9	-1.76841

$K_{\pi\pi} = 4.22276$

families in which the parent was the proband, nor were there any families with more than one affected parent.

A proportion $\underline{x}$ of the affected parents would be sporadic and a proportion (1 - x) would not be sporadic. For the affected parents who were sporadic the probability of having an affected child is just the incidence in the general population, $\underline{I}$, provided that dominant mutations do not account for some sporadics. For the affected parents who were not sporadic the probability of having an affected child is (p + I - pI), which is approximately p + I. The proportion of sporadic cases $\underline{x}$ in families with an affected parent is shown in Appendix A.V. to be:

$$\tilde{x} = \frac{I}{I + (1 - x)p}$$

where $\underline{I}$ is the incidence in the population, $\underline{x}$ is the proportion of sporadic cases in families without affected parents and $\underline{p}$ is the segregation frequency. With $\underline{I} \sim 0.002$, $\underline{x} \sim 0.8$ and $\underline{p} \sim 0.2$ then

$$\tilde{x} = \frac{.002}{.002 + .04} \sim 0.05 \qquad (21)$$

This value for $\underline{\tilde{x}}$ was taken as the trial value in the iteration for $\underline{\tilde{x}}$ in the data with one parent affected. Values for $\underline{p}$ and $\underline{\pi}$ were taken from the previous results where $\underline{p} = 0.19630$ and $\underline{\pi} = 0.59696$. The iterative solution for $\underline{\tilde{x}}$ was 0.41993 ± 0.51358. The large standard error is attributible to the small sample size and the resultant

small amount of information. The chi square (L score) for goodness of fit was not significantly altered by fitting $\underline{x}$. Therefore, the data were compatible with the hypothesis put forth above, but a wide range of values would fit about as well.

The second set of data was the families from Tripler Army Hospital. Treatment was different in this case to avoid bias which could be introduced because the entire sample of 125 families could not be utilized in the analysis. The 80 families used were treated as a random sample of the 125 family group, and sibs of the proband were analysed as for complete selection.

Under complete selection $\underline{p}$ may be estimated from

$$P(r = o) = h + (1 - h)\ (1 - p)^2 \qquad (22)$$

$$P(r \mid r > o) = (1 - h)\ \binom{s}{r}\ p^r\ (1 - p)^{s-r} \qquad (23)$$

The parameter $\underline{h}$ represents the proportion of families that cannot segregate because of sporadicity of the proband.

Alternatively (and more conveniently with SEGRAN) it is possible to employ the incomplete selection model used above by setting $\underline{\pi}$ very nearly equal to zero and including the proband. For, as $\underline{\pi}$ approaches zero and therefore is omitted from the model, the distribution approaches that for complete selection. Under these conditions $\underline{\pi}$ is fixed

at 0.01, $\underline{p}$ again represents the segregation frequency but $\underline{x}$ now assumes the role of $\underline{h}$ as defined above.

Iteration was attempted over the 80 families for $\underline{p}$ and $\underline{x}$ (see Table 2.9). The sample size combined with the small segregation frequency precluded simultaneous convergence for the two parameters. The $\underline{p}$ value taken alone would not converge but instead oscillated about a value slightly greater than 0.14. This oscillation indicated that the maximum likelihood curve did not give a sharp peak and therefore one estimate, but rather consisted of a flattened curve with the estimate extending over a range of values. Several attempts at achieving a convergence within this range failed. The boundaries for the range were taken from the two values with the smallest U score of opposite sign. These were 0.14038 (U score = +1.79075) and 0.16373 (U score = -1.14733). The average of these values was then used as the estimate of $\underline{p}$, such that $\hat{\underline{p}} = 0.15205$.

The values for the parameters from the Tripler study could then be given as the following:

$$\hat{\underline{p}} = 0.15025 \pm 0.12101$$

$$\hat{\underline{x}} = 0.70627 \pm 0.16165$$

$$\underline{\pi} = 0.01$$

As with the analysis of the other groups $\underline{x}$ was then set equal to zero and an estimate was obtained for $\underline{p}$ giving the values:

Table 2.9 Segregation Analysis

Pyloric Stenosis

(Tripler Data - Complete Selection Model)

sr Table $p = 0.15025$ $x = 0.70674$ $\pi = 0.59696$

Children Affected

s	r	No.	U_p	U_x	U_π
1	1	40	0	0	0
2	1	16	-4.89316	2.50297	0.35314
3	1	14	-7.56886	4.21579	0.61752
3	2	1	5.48169	-3.40599	-0.45840
4	1	3	-2.14005	1.30092	1.9833
4	2	1	4.30634	-3.40674	-0.43640
5	1	4	-3.33190	2.21558	0.35232
5	2	1	3.13099	-3.40749	-0.41443

$$K = \begin{bmatrix} 68.28701 & -49.82816 & -6.56651 \\ & 38.27000 & 4.98255 \\ & & 0.65151 \end{bmatrix}$$

$$K_{\pi\pi} = 153.04247$$

$\underline{x} = 0.0$

$\hat{\underline{p}} = 0.03865 \pm 0.02308$

$\underline{\pi} = 0.01$

The sr table provided the test of the null hypothesis $\underline{x} = 0$ for this group. The three values were

TEST	CHI SQUARE VALUE	P
$UK^{-1}U'$	0.84	> 0.5
L	1.01	> 0.5
X_1^2	39.62	$\ll 0.01$

These tests were not in close agreement as were the values for the larger set of data. However, on the basis of the highly significant chi square value it can be stated the $\underline{x}$ does differ significantly from zero in this sample.

Despite the difficulties with convergence in this smaller sample and the use of a different ascertainment model the values for the parameters are in close agreement with those determined by the first analysis. This necessarily gives support to both sets of estimates.

2.5 Quasi-continuous Analysis

The initial estimates for the quasi-continuous parameters were obtained from the NUQUAC program (Morton, unpublished). Then, these estimates were used to predict the probability of selecting from the general population a sibship of size $\underline{s}$ with $\underline{r}$ affected. With mean and

variances equated the log normal and Beta distributions should not give very different answers. For families from size 2 to 9 the probability of selecting a family with $r = 0$ was never less than 0.9 with either distribution. The probability of selecting a family with $r > 0$ was very small, often approaching zero. This is the expected result for very rare traits.

Estimates were also made of the probability that the $s + 1$ child be affected given $\underline{r}$ affected in the sibship. These are listed in Table 2.10 for the Beta approximation. The Beta is found by fitting two points representing population prevalence and the segregation frequency. It should be noted that the probability increases rapidly with increasing $\underline{r}$ in a sibship of fixed $\underline{s}$ while it decreases fairly slowly with fixed $\underline{r}$ and increasing $\underline{s}$.

Next an attempt was made to estimate the parameter b_i and c from the data of Carter (1965) on members of affected for first degree relatives of male and female probands. Since on the quasi-continuous model the estimates for $r = .5$ would apply to both sibs and children, these classes were pooled for the estimates. There was some difficulty in obtaining the estimates as the effect of sex is quite strong in pyloric stenosis, section 2.3.2. The final estimate of $b_1 = 1.611$, $b_2 = 2.827$ and $c = 3.005$ were fitted to the same data of Carter for different degrees of relationship. The estimates gave a very poor fit, not only

Table 2.10 P(rs) Table for Pyloric Stenosis, Beta Approximation

S/R	00	01	02	03	04	05	06
0	.001850						
1	.001787	.035920					
2	.001728	.034734	.067741				
3	.001673	.033625	.065577	.097529			
4	.001621	.032583	.063546	.094509	.125472		
5	.001572	.031605	.061638	.091671	.121703	.151736	
6	.001526	.030683	.059841	.088998	.118155	.147312	.176469
7	.001483	.029814	.058145	.086476	.114807	.143138	.171470
8	.001442	.028993	.056543	.084094	.111644	.139195	.166745
9	.001403	.028215	.055027	.081839	.108651	.135463	.162275
10	.001367	.027479	.053590	.079702	.105814	.131926	.158037
11	.001332	.026779	.052227	.077674	.103121	.128569	.154016
12	.001299	.026115	.050931	.075746	.100562	.125378	.150194

Table 2.10 (continued)

S/R	07	08	09	10	11	12
0						
1						
2						
3						
4						
5						
6						
7	.199801					
8	.194296	.221847				
9	.189087	.215898	.242710			
10	.184149	.210261	.236373	.262484		
11	.179463	.204910	.230358	.255805	.281252	
12	.175010	.199825	.224641	.249457	.274273	.299089

to all the data but also to the set of the classes that was used for the estimation (Table 2.11). Additional attempts to get convergence always led back to these estimates. The fit to the small sample of affected sibs available through the Hawaii study was not significantly bad, $p \doteq .1$, but the sample is very small and not terribly informative (Table 2.12).

There are several possible explanations for the highly significant chi square. First, the estimation based on only four classes is not fully efficient, and the largest contribution to chi square comes from the fifth class which does not contribute to the iteration. In all attempts to fit estimates this class of male cousins of male probands gave the worst fit. Secondly, this theory of quasi-continuity is only one of several possible, and involves some approximation which can be improved. Finally, each class is the sum of many sibships, and so introduces non-binomial variation. This could be allowed for by replacing sample size with an equivalence based on segregation analysis in data where each sibship is reported separately. There is considerable opportunity for improvement in the theory of quasi-continuity, its application to data, and the completeness with which the data are reported.

In the future it will be of interest to obtain reasonable estimates of the parameters from one set of data on

Table 2.11 Pyloric Stenosis

Parameters estimated from quasi-continuity

$\beta_1 = 1.611, \quad \beta_2 = 2.827, \quad C = 3.005$

r	Sex of Proband	Sex of Relative	Affected Observed	Affected Expected	χ^2	Normal Observed
.500	1	1	24	35.2002	3.8194	502
.500	1	2	12	11.8668	.0015	504
.500	2	1	18	10.0870	6.8742	86
.500	2	2	9	3.9727	6.6122	96
.125	1	1	6	36.1701	26.0536	1055
.125	1	2	3	9.5106	4.4977	1046
.125	2	1	2	9.1054	5.7653	236
.125	2	2	0	2.3618	2.3873	221
.250	1	1	1	9.8143	8.2740	226
.250	1	2	1	2.7018	1.0857	212
.250	2	1	1	2.3551	.8238	43
.250	2	2	0	.8869	.9023	52
			Sum χ^2 =		67.0970	

1 = male
2 = female

Table 2.12 Campbell's Data Pyloric Stenosis

Fit of parameters estimated from Carter's Data

$\beta_1 = 1.611, \quad \beta_2 = 2.827, \quad C = 3.005$

r	Sex of Proband	Sex of Relative	Affected Observed	Affected Expected	χ^2	Normal Observed
.500	1	1	2	6.4913	3.3304	95
.500	1	2	0	2.1618	2.2127	94
.500	2	1	1	1.6488	.2828	16
.500	2	2	0	1.0972	1.1404	29
				Sum χ^2_3 =	6.9662	

1 = male
2 = female

pyloric stenosis. Then, as more complete sets of data become available from other sources the same set of estimates could be fitted for comparison.

2.6 Summary

The estimates for the quasi-continuous model presented in the previous section and the lack of a good fit to the model present a major problem in discussing the genetics of pyloric stenosis, which is often considered the classical example of a quasi-continuous trait.

Several other pieces of evidence, however, come to bear on this particular question. The regression analysis reported in section 2.3.2 on the racial variables indicated a reduction in the frequency of affected in F_1, a finding indicative of recessivity. From segregation analysis we know that the evidence for sporadic cases is strong, with an estimate of nearly 80%. The segregation frequency, $\underline{p}$, estimated in this analysis actually conforms with a major gene hypothesis, but the standard error is large. Additional data from other mating types would be required for a stringent test of reduced penetrance.

A strong influence of several sociological variables including birth order and birth weight were found by regression analysis. The significant deviation in sex ratio is an acknowledged characteristic of pyloric stenosis.

The difficulty of this situation then is clear. The results from the quasi-continuous analysis are inconclusive

enough to suggest that such a model should remain an open possibility for at least a role in the etiology of this disease. Some evidence, however points in the direction of real heterogeneity in this condition. Perhaps this represents a confounding of environmentally caused phenocopies and one or more genetic conditions. The parameters of the quasi-continuous model presented in section 1.3 are designated to represent the environmental variation. The fact that not a sufficient amount of the variation can be accounted for to permit an overall fit to the data by this model could mean that additional refinements must be made.

This study contributes some additional information on the occurrence of pyloric stenosis. However, the question of the genetic mechanism/mechanisms in the etiology of this condition continues at the present time to defy precise resolution. Pyloric stenosis should not be considered a good example of a quasi-continuous trait until a theory of quasi-continuity is developed which will give a good fit to the data.

3. HYPERCHOLESTEROLEMIA

3.1 Introduction

One of the most vital questions in medical research today concerns the relationship between atherosclerosis, ischematic heart disease, and elevated serum cholesterol levels. The current study affords the opportunity to test one possible mode of inheritance for these conditions.

Stanbury, Wyngaarden and Fredrickson (1966) divide familial hyperlipoproteinemia into five types with the characteristics listed in Table 3.1. All of these are referred to as familial hyperlipemia or essential familial hyperlipemia. Type I is a severe form of the disease usually detected in childhood and will not be included in this study.

Types II to Type V are separated on the basis of the induction, if present, and the affected form of lipoproteins. One form of lipoprotein appears as small concentrations normally present in the serum. As these particles increase in size with ingestion of dietary fat they contain an increasing proportion of triglyceride. When the particles reach a density of approximately 0.9 they are referred to as chylomicrons and contain over 90% triglyceride. The pre- β -lipoproteins, or very low density lipoproteins, have a glyceride level of 20 to 60% and a higher cholesterol content. They are primarily responsible for transporting

Table 3.1 Some clinical features of different types of familial hyperlipoproteinemia

Type	Usual Age of Detection	Xanthomas: Tendinous	Xanthomas: Tuberous	Xanthomas: Eruptive	Vascular disease	Diabetes in family	Abnormal glucose tolerance	Hyperuricemia
I	< 10 years	0	0	+	0	0	0	0
II	30 years	+	+	0	+	0	0	+
III	Adulthood	+	+	0 +	+	+	+	+
IV	Adulthood	0	0	0 +	+	+	+	0
V	Adulthood	0	0	+	0 +	+	+	0

+ = frequently present, 0 = usually absent

From Stanbury, Wyngaarten and Fredrickson

glyceride in the plasma. Cholesterol (Δ^5 -cholestene-3- β -ol) is the second commonest lipid in plasma exceeded only by the phosopholipids. Approximately 70% of the cholesterol present in the serum is esterfied with long-chain fatty acids, primarily of the unsaturated type. The plasma cholesterol is transported predominantly by the β -lipoproteins (S_f 0 to 12, density 1.019 to 1.063).

Type II, which represents an increase in β -lipoproteins, is the most common form of hypercholesteremia, usually with normal glyceride levels. Type III is similar to II except the glyceride level is elevated either endogenously or through carbohydrate induction. With this condition a glucose intolerance is common. Unlike Types II and III, Type IV has a normal cholesterol level but the level of pre- β -lipoproteins is greatly increased. Again, a glucose intolerance is common; the condition may be carbohydrate induced. Type V displays a complex pattern of defects which usually includes excesses of chyclomicrons and pre- β -lipoproteins. It may be related to Type I or Type IV.

Wilkinson et al. (1948) was among the first to study the genetic variations among individuals with elevated total cholesterol levels. His study extended over four generations and based on his finding of 1/2 of the children affected if one parent was affected, he proposed a dominant gene as the mode of genetic action. This study first

proposed the idea that the presence of xanthomatosis in association with elevated cholesterol levels represented the homozygous state and the absence of xanthomatosis represented the heterozygous state. Although other studies have supported this idea it has been fairly conclusively shown that this is not the case (Khachdurian, 1964).

Aldersberg et al. (1952) also found that a recessive gene could not account for the high percentage of affected in sibs and children. They proposed a dominant gene with reduced penetrance.

By selecting probands for both hypercholesteremia and xanthomatosis, Wheeler (1957) found that 67.3% of the probands over 40 years of age had a history of heart disease. He could find no relation between any of these conditions and birth order or sex. He also favored a dominant genetic hypothesis.

Epstein et al. (1959) were the first to distinguish a bimodal distribution of cholesterol levels, but only for males. The authors were puzzled by the degree of variability in the expression of one gene and proposed the idea that two genes might be responsible with one determining the less severe manifestation. In response to criticism to their earlier work Hirschorn and Wilkinson (1957) also proposed the idea of multiple abnormal alleles to explain the variability in expression.

One real difficulty in interpreting these papers is in not knowing if one and only one trait was under study. Few if any of these workers measured triglyceride levels, using only elevated cholesterol and/or xanthomatosis as criteria for ascertainment. If the divisions of Stanbury et al. (1966) are not totally arbitrary, then several clinical and genetically distinct conditions may have been pooled for analysis in these earlier papers. However, it could be that the changes in glyceride levels only represents one portion of a continuum of symptoms comprising one genetic condition.

Nevin and Slack (1968; Slack and Nevin, 1968) studied both cholesterol and triglyceride levels in their small sample. Patients were selected for hyperlipidemia with associated xanthomatosis without determining the history of heart disease. Upon analysis they found a 6 to 9 fold increased risk of heart disease in relatives of patients with hypercholesteremia. The relatives of patients with hypercholesteremia and hypertriglyceremia showed no significant increase in the risk for heart disease above the general population. The authors admit that their sample is too small to permit drawing definite conclusions. They did not consider the possibility that their subjects with hypercholesteremia and hypertriglyceridemia were non-fasting normals.

A single large kindred of nine generations provided four living generations and 659 individuals for study by Harlon, et al. (1966). Seventy-nine individuals were defined as affected with hypercholesterolemia using an age and sex adjusted discriminant. Clinical manifestations were elevated serum cholesterol with normal triglyceride values. The disorder was found to be segregating as a simple autosomal dominant with high penetrance. In this kindred "'coronary heart disease' at an unusually early age is not epidemic;" a finding not compatible with the hypothesis of Nevin and Slack (1968).

The findings presented in these investigations, although suggestive of major gene acting, leave doubts about the mode of inheritance for hypercholesterolemia. The questions of the interaction among the lipoprotein forms and the number of genetic entities under study remain unanswered.

3.2 Methods

The Honolulu Heart Program (HHP) under the auspices of the National Institutes of Health (NIH) and National Heart Institute (NHI) surveyed over 8,000 Japanese males born between 1900 and 1919, inclusive. Names were drawn from the records of the selective service registry for men registering during World War II. These men were sent a questionnaire by mail and all respondents still residing on the island of Oahu were contacted by phone and scheduled for

an interview. Information collected on social background, occupation and diet supplemented a complete physical examination which included an electrocardiogram and serum lipoprotein determinations.

A 20 m.l. sample of blood was drawn with vacutainers and standard venepuncture technique. The samples were allowed to stand at room temperature for 10 minutes. After centrifuging at 2400 r.p.m. for 15 minutes a 10 m.l. sample of serum was placed in a vial and frozen. All samples were sent directly to Milton Nichaman, M.D. at the NIH Laboratory, San Francisco for determination of serum cholesterol, non-fasting triglyceride and uric acid.

The level of serum cholesterol is affected by many factors and even the problem of distinguishing between normal and elevated levels is complex. As age increases the cholesterol level rises, reaches a plateau, and then declines. The changes differ in men and women with the increase occurring earlier in males, about ages 20 to 30, and the decrease in later years lessened (Aldersberg, et al., 1956). Racial differences have been suggested but the reports in this area are still fragmentary (Seiver, 1968; Schaefer et al., 1953).

In our study the initial selection of probands was made dividing the sample available from the HHP into two groups, with triglyceride level known and without. The number of individuals who had been examined by the HHP and

were available to us at this time was 7,406. Regression analysis showed no significant correlation of cholesterol level with age in this group, but significant correlation with time since last food intake and triglyceride level. Correcting for the significant factors, 223 men were selected as probands whose uncorrected and corrected cholesterol levels exceeded 290 mg.% or approximately two standard deviations above the samples mean value of 218. 09 mg. %. By selecting men through elevated serum cholesterol levels Types II and III would be included in the study with the exclusion of Types IV and V.

An attempt was made to locate and contact all full sibs of probands residing on Oahu. No attempt was made to interview the parents who tended to be quite elderly if still living. Many of the children were no longer living at home and it was felt that the additional effort required to locate these individuals would not prove worthwhile. All contacted individuals were asked to come to the HHP Laboratory in order to obtain a pedigree and records on personal health and a family history. A blood sample was taken at the time of the interview, processed like those described above, and sent to the same laboratory for analysis. The complete set of data consisted of 219 sub-ships with 223 probands, 185 brothers and 151 sisters.

3.3 Regression Analysis

A second multiple regression analysis was undertaken to examine the effects of age and sex as well as fasting time and triglyceride levels on the serum cholesterol level of the sibs.

The effect of age differs in the two sexes as shown by the inclusion of the interaction term 'age x sex' in the regression, significant at the .01 level, "where sex was coded 1, -1." Harlon et al. (1966) also reported a different age effect in the two sexes.

In the sibs the effects of fasting time and triglyceride levels were not significant. One possible explanation of the lack of a fasting time effect is that the sibs were seen at only restricted times of day, not from 8:00 a.m. to 8:00 p.m. as was done for the probands. This restriction may have limited the possible variation in fasting time sufficiently to make its effect negligible.

The lack of control data for females prevented direct comparison by regression of sibs and controls. In the final analysis the sibs' cholesterol levels were adjusted for age, sex, age x sex, age squared, and age squared by sex. These adjustments would force the sibs to resemble the control population sufficiently so that again all values over 290 mg. % could be designated affected.

A regression was then performed on the control data to determine the mean, standard deviation and significant

variables related to triglyceride levels. The unadjusted mean for the non-fasting triglyceride values was 511.23 with a standard deviation of 258.62. Even with adjustments the deviation about the mean was so great that two deviations above the mean fell very close to the value of 1000. Unfortunately only three columns had been set aside for coding triglycerides by the HHP and any value greater than 1000 was coded 998. Transformation of the triglyceride level did not seem to appreciably improve the situation. Therefore, elevated triglyceride levels were defined as those above 950 mg. % after adjustments for age, fasting time, age squared, fasting time squared and cholesterol levels.

These criteria were applied to the family data from our study. Fourteen individuals had initial triglyceride levels equal to or greater than 990 while all the others fell below 950. The adjusted values for all fourteen were still above 950 and all of these individuals were designated as having abnormally high triglyceride.

The sibships were divided into two groups based on the criterion of elevated triglyceride. All but 34 individuals in ten families were included in the group of normal triglycerides, a disappointingly small proportion of the sample.

Regression analysis of a with/without high triglyceride variable on the occurrence of heart attacks,

hypertension and diabetes revealed very little. The difference in frequency of heart attacks was not significant, but it is of interest to note that all eleven reported attacks did occur in the group with normal triglyceride levels. Only the variable for high blood pressure was significant and then only barely at the .05 level. Also, this was probably the most inaccurately recorded information despite an effort to determine if the individual was actually under doctor's care. We are therefore uncertain whether elevated triglyceride is of clinical significance.

3.4 Segregation Analysis

The segregation analysis was done with the sibs of probands defined as affected or not affected by the methods outlined above. Under incomplete selection the parameters $\underline{p}$, $\underline{x}$ and $\underline{\pi}$ were estimated (section 2.4).

All estimates of $\underline{\pi}$ in this analysis were approximately 0.1 or less. The study is one approaching single selection; i.e., the probability of ascertainment is so low that there is virtually no probability of having two probands in one family. In this situation the probability that a family be ascertained is approximately proportional to the number of affected sibs. In this particular instance single selection resulted directly from the HHP study sampling only a limited number of individuals, often only one, from a family and effectively restricting the number of possible probands to very nearly one.

There are two complications which make single selection somewhat less desirable than it might seem (Morton, unpublished). As $\pi \to 0$ it becomes difficult if not impossible to get an estimate of the prevalence in the population. Secondly, there would be a poor representation of families with isolated cases and therefore a low probability of ascertainment of sporadic cases. The estimate of $\underline{x}$ remains unbiased but the power of any test of hypothesis on $\underline{x}$ is lowered.

From the total sample the estimates of the parameters were:

Table 3.2

Segregation Analysis for Cholesterol

$\hat{x} = 0.04155 \pm .04060$	$x = 0$
$\hat{p} = 0.54278 \pm .03356$	$\hat{p} = 0.52455 \pm .02716$
$\hat{\pi} = 0.04347 \pm .02101$	$\hat{\pi} = 0.04385 \pm .02109$

The U and K scores are given in Tables 3.3 and 3.4.

The test of the null hypothesis $x = 0$ showed that $\underline{x}$ did not differ significantly from zero in this data. The heterogeneity chi square values being 2.258, 0.590 and 1.264 with one degree of freedom.

The estimate for $\underline{p}$ of nearly one half suggests a dominant gene, one locus hypothesis. The heterogeneity chi squares for the null hypothesis $p = .50$ were 2.456,

Table 3.3 Cholesterol sr Table

U and K Score from Segregation Analysis

sr Table $r = 0.54278$ $x = 0.04155$ $\pi = 0.04347$

Children s	Affected r	No.	U_p	U_x	U_π
2	1	30	-59.3002	33.5822	7.8927
2	2	16	29.8149	-16.8844	-3.9683
3	1	10	-35.8109	32.4568	5.2380
3	2	20	-6.0565	-21.3458	.2538
3	3	16	59.6268	-17.0767	-7.8509
4	1	6	-26.7095	40.5370	4.6928
4	2	9	-22.2235	-9.7145	2.4392
4	3	8	12.4818	-8.6351	-1.8588
4	4	5	27.9486	-5.3969	-3.6400
5	2	2	-9.2718	-2.1831	1.0540
5	3	7	-4.2449	-7.6409	.1652
5	4	3	10.2693	-3.2747	-1.4161
5	5	1	7.4526	-1.0916	-0.9600
6	2	1	-6.8027	-1.1038	.7806
6	3	1	-2.7732	-1.1038	.2772
6	4	4	5.0250	-4.4153	-0.8738
6	6	1	9.3153	-1.1038	-1.1866
7	3	1	-4.9403	-1.1162	.5284
7	5	1	3.1187	-1.1162	-0.4552
7	6	2	14.2965	-2.2323	-1.8708
9	5	1	-1.2159	-1.1411	.0401
Totals		145	.0000	.0000	.0217

Table 3.4 Cholesterol ra Table

U and K Score from Segregation Analysis

ra Table $\pi = 0.04347$

Affected r	Probands a	No.	U_π
2	1	44	-23.5108
2	2	4	94.0634
3	1	33	-35.5215
4	1	12	-19.5145
5	1	3	- 6.5512
6	1	3	- 8.2469
Totals		99	.7184

$K_\pi = 2266.1$

0.81 and 1.188 respectively, and the null hypothesis could not be rejected. It would be reasonable to expect a segregation of one half even though the mating types of the parents are unknown if the trait is assumed to be rare. Under such an assumption nearly all matings producing affected children would be of the type normal x heterozygote affected with very little contribution from intercrosses of homozygous x normal matings. If you assume the increase in p above .50 to be due to the contributions of the latter matings the gene frequency estimates would be .05 with a large standard error.

The data were divided into two groups based on the criterion of normal or elevated triglyceride levels. The small group of ten families with high serum triglycerides gave the following estimates with $x = 0$; $\hat{p} = 0.49870 \pm .10134$, $\hat{\pi} = .14942 \pm .13113$. The large standard errors result from the small sample size. The estimates do not differ significantly from those obtained from the overall samples nor from the estimates from the families defined as having normal triglyceride levels.

These studies would indicate a major dominant gene as the genetic mechanism responsible for elevated serum cholesterol, a finding in agreement with several other authors including Harlon, et al. (1966) and Aldersberg, et al. (1952). There is either no genetic heterogeneity in hypercholesterolemia between those families with normal and those with elevated triglycerides, or both types of elevated cholesterol are due to dominant genes at different loci or different alleles at the same loci. One other reasonable alternative is that the sample with elevated triglycerides was too small a proportion of the total sample to permit detection of heterogeneity.

3.5 Quasi-Continuous Analysis

The segregation analysis of the cholesterol material has indicated that the data did not differ significantly from a major dominant gene hypothesis. Therefore, the

initial analysis of this data on the quasi-continuous model was the estimation of heritability by the method of Falconer. The Fortran program QUAC had been prepared to estimate heritability and, given heritability to estimate the recurrence in sibs using the approximation derived from Falconer and Edwards (Morton, 1967).

For data the estimate of the gene frequency at .05 was used to estimate a population prevalence on the order of magnitude of .10. With the recurrence risk in sibs estimated to be 0.5 from segregation analysis the estimate of heritability was never less than 1.50. Setting the heritability at the maximum; i.e., 0.99, and using the same program, Edwards' model predicted a value slightly greater than 0.3 while Falconer's model predicted a recurrence in sibs slightly less than 0.3. Neither estimate approaches the value obtained through segregation analysis, even though the heritability was set at the maximum.

This information alone would indicate a major gene and consequently a poor fit to the quasi-continuity model for the cholesterol data. However, the data were submitted to additional analysis.

NUQUAC provided the initial estimates of b and c and these were used to predict the probability that the s + 1 child be affected given r out of s are affected. The probabilities are given in Table 3.5.

Table 3.5 Hypercholesterolemia

Probability s+1 child be affected given r and s

S/R	00	01	02	03	04	05	06
0	.058340						
1	.029466	.524400					
2	.019710	.350785	.681859				
3	.014808	.263535	.512262	.760989			
4	.011858	.211043	.410227	.609412	.808596		
5	.009889	.175989	.342089	.508189	.674289	.840389	
6	.008480	.150921	.293361	.435802	.578243	.720683	.863124
7	.007423	.132104	.256785	.381466	.506147	.630828	.755509
8	.006600	.117459	.228318	.339177	.450036	.560895	.671754
9	.005941	.105737	.205533	.305328	.405124	.504920	.604716
10	.005402	.096142	.186883	.277623	.368363	.459103	.549844
11	.004953	.088144	.171336	.254527	.337718	.420910	.504101
12	.004572	.081375	.158177	.234979	.311781	.388583	.465385

Table 3.5 (continued)

S/R	07	08	09	10	11	12
0						
1						
2						
3						
4						
5						
6						
7	.880190					
8	.782613	.893472				
9	.704511	.804307	.904103			
10	.640584	.731324	.822064	.912804		
11	.587293	.670484	.753676	.836867	.920058	
12	.542187	.618990	.695792	.772594	.849396	.926198

The remaining estimates of the parameters, b_i, and the test for the model of quasi-continuous variation were not done for two reasons. The evidence from all other analyses including the estimate of heritability strongly indicated a major gene and not an additive multifactorial model was most appropriate for hypercholesterolemia. Secondly, there were only two classes in these data, male and female relatives of male probands and only for r = 0.5. Therefore, three parameters could not be estimated and tested for goodness of fit because of the lack of sufficient degrees of freedom.

3.6 Summary

The estimate from segregation analysis, p = .5, x = .0, and the estimate of heritability of $h^2 > 1$, would indicate a major gene active in the etiology of hypercholesterolemia. This is a finding concurrent with several studies (Harlon et al., 1966, Alderberg, et al., 1952).

The inability to detect any differences in segregation frequency or sociological variables between families with normal and families with elevated triglyceride levels may well have been a result of small sample size. The contradictory reports of Nevin and Slack (1968) and Harlon et al. (1966) and the inability of this author to resolve the question, leaves unanswered a major question on the

heterogeneity, genetical or clinical, of the conditions defined as hypercholesterolemia.

This study was done only on Japanese subjects but the results appear compatible with the published studies mentioned above in Caucasian populations. The major difficulties with this study were the treatment of female sibs without adequate control data, the estimation of a population prevalence, and the lack of sufficient classes for estimates of the quasi-continuous parameters to permit ample tests for goodness of fit to that hypothesis, which, however appears to be excluded by the data.

4. DERMATOGLYPHICS

4.1 Introduction

There are three basic patterns in the dermal ridges of finger tips: arches, loops and whorls. Loops are further defined as radial or ulnar depending on which side of the hand they open toward. These dermal ridge patterns begin to form during the third and fourth month of fetal life and are complete by the seventh month (Hale, 1952). The patterns once formed remain unchanged for life barring physical disruption; e.g., scarring.

The patterns vary in frequency among various fingers as well as between the sexes and among racial and ethnic groups. Females tend to have fewer arches, radial loops and whorls than men (Holt, 1968).

Early studies by Galton and Wilder as reported by Holt (1968) indicated a hereditary basis for these patterns. There are many variations on each of the basic patterns and classification of intermediary types becomes extremely difficult. To avoid this complication another quantity, total ridge count, was defined and studied. The ridge count for each finger is determined by the number of ridges bisected by a line drawn from the center of the pattern to the tri-radii. For arches the count is zero as there are no tri-radii. For whorls the sum of the count to

both tri-radii is used (Cummins and Midlo, 1943). The sum for all ten fingers gives the total finger ridge count.

Total ridge count is one of the best examples of a quantitative trait giving good agreement between observed and expected on a quantitative model. The table below gives correlation coefficients calculated from the total ridge count data (McKusick, 1964).

Table 4.1

Correlation Coefficient

Relationship	Observed	Theoretical
Parent - Child	0.48	0.50
Parent - Parent	0.05 ± 0.07	0
Mid-Parent - Child	0.66 ± 0.03	0.70
Sib - Sib	0.50 ± 0.04	0.50
Monozygotic Twins	0.95 ± 0.01	1.00
Dizygotic Twins	0.49 ± 0.08	0.50

Additional evidence from Holt (1968) supports the theory of additive gene effects with an absence of dominance in this trait. The correlation between child and mid-parental value is higher than the correlation with either parent. Further, the regression of ridge count of child on mid-parental count shows no significant deviation from linearity, as expected in the absence of dominance (Penrose, 1949).

4.2 Methods

For our study a threshold was imposed on this trait by selecting individuals with a zero ridge count as probands. The individuals with ten arches were ascertained by Ian Shine and William S. Pollitzer while visiting professors in Hawaii. The ridge count was determined on all first degree relatives of probands for which information was available. The data include 232 probands and information on parents, grandparents, sibs, children, aunts, uncles and cousins. There were additional probands for whom no family information was available. These were included only in the incidence calculation.

A control for each proband was selected by Pollitzer as the first record following the proband in the file who was not a relative and for which family information was available. Data on the 963 individuals of the control group consist of ridge count, race, sex and family relationship. All records were taken from the Bureau of Identification of the Civilian Records Division of the City of Honolulu. This file contains 684,502 records of persons living in Hawaii collected since 1941.

The standard convention for recording ridge counts was observed. Two entries were made for each finger, one for the radial and one for the ulnar side. For patterns other than whorls one or both entries were coded zero.

A discriminant function was determined through regression of degree of relationship on ridge count and number of arches as well as interaction and higher order terms. The function giving the most distinct discrimination (as an interclass correlation) between probands and non-affected was

$$D = 2.73505 - 0.009666\ (\text{ridge count}) + 0.000015\ (\text{ridge count})^2 + 0.00154\ (\text{ridge count} \times \text{no. of arches})$$

All individuals with a discriminant function value equal to or greater than that of the probands were defined as secondary affected cases. These 90 individuals appeared in both the affected and control families (3 cases).

Records with only family number, sex, and type of parent coded were added to the data file whenever information on one or both parents was missing. The discriminant function was left blank and this "unknown" phenotype was coded and included as a possible mating type when compiling family records for segregation analysis but excluded from all other calculations.

4.3 Preliminary Analysis

4.3.1 Pattern Frequencies

The dermal ridge pattern frequencies for a British sample of 500 men and 500 women are given below (Holt, 1964):

Table 4.2

Frequency of Dermal Ridge Patterns

Pattern	Males	Females
whorls	28.3	23.9
ulnar loops	61.5	65.6
radial loops	5.9	4.8
arches	4.3	5.7

These frequencies vary among racial groups and between the sexes. Within each racial group, however, the frequencies are similar for the two sexes. The mean value of Holt's frequencies were used to compute expected values for our data. The comparison, shown in Table 4.3 indicates that the Caucasian control group did not deviate significantly from the English group, $\chi^2_3 = 5.75$. However, the deviation of the Japanese group was highly significant, particularly the high frequency of whorls.

The frequency distribution for the affected group is also given in Table 4.3. As expected, the frequency of arches is much higher than any other group. The deviation from Holt's data or our control group is highly significant. The probands were removed from the affected group for this analysis.

Differences due to race and sex also occur in total ridge count. Mean ridge counts and standard deviations

Table 4.3 Frequency Distribution for Dermal Ridge Patterns

Population:	Control		Affected		Holt	
	Caucasian	Japanese	Caucasian	Japanese	Male	Female
Arches	5.89	2.27	27.88	17.60	4.3	5.7
Radial Loops	4.66	3.41	5.00	4.32	5.9	4.8
Ulnar Loops	60.32	49.43	54.09	60.89	61.5	65.6
Whorls	29.13	44.88	13.03	17.19	28.3	23.9

$$\chi^2_3 = 5.75$$

were calculated by Holt (1955) using a sample of 825 Caucasian males and 825 Caucasian females. Only the higher of the two counts for a whorl was used in this study. The overall mean for males was 145.18, standard deviation 50.49. The females had consistently lower means for each of the ten fingers with an overall mean of 126.97 standard deviation 52.33.

Mean ridge counts for the control sample are given in Table 4.4. Our sample was small and the standard deviations for the mean ridge counts quite large. As expected from the higher frequency of whorls the Japanese mean exceeded the Caucasian mean for both males and females. The high mean ridge count for Caucasian females is undoubtedly due to sampling error. The means are generally greater than those from Holt because both tri-radii counts were used when a whorl was recorded.

The families ascertained through individuals with low arches have strikingly lower mean ridge counts, Table 4.4. Probands were omitted for this analysis.

Initially the validity of the control group was questioned. However, the consistently good fit of the Caucasians to the expected and the reasonable differences exhibited by the Japanese indicate a suitable control. Further support was provided by the reasonable estimates of correlation coefficients from this data.

Table 4.4 Mean Total Ridge Count

Sample	Sex	Number of Individuals	Mean	Standard Deviation
Japanese (controls)	M	145	198.9	86.6
	F	154	188.5	75.5
Caucasian (controls)	M	33	155.1	70.9
	F	40	176.6	75.2
Japanese (affected)	M	83	100.3	94.7
	F	93	72.6	75.3
Caucasion (affected)	M	44	76.6	78.7
	F	55	54.2	70.7
British*	M	825	145.2	50.5
	F	825	126.9	52.3

*from Holt, 1955

4.3.2 Correlation Coefficients

The correlation coefficient, r, was designed by Galton in the late 19th century for use as a measure of the inherited resemblance in human families. When phenotypic values are known on a quantitative trait for two sets of relatives they can be regressed on each other. The correlation coefficient was defined as the geometric mean of these two regression coefficients; i.e.

$$r = \frac{\text{covariance } XY}{(\text{variance } X \cdot \text{variance } Y)^{1/2}} = \frac{\text{covariance } XY}{\sigma_X \quad \sigma_Y}$$

where x and y represent the phenotypic values of the two groups and σ is the standard deviation.

It was R. A. Fisher (1918) who showed that the regression coefficient of child on parent is equal to the proportion of genes they have in common. This is true if the trait is continuously distributed, there is no dominance, assortative mating nor effect of common environment. Under these conditions the interpretation holds for any relative on relative regression; e.g., sib on sib or parent on parent. In regular relationships; i.e., without inbreeding the variances are equal in theory and the correlation coefficient is equal to the regression coefficient. Predictions of theoretical values for the correlation coefficients given in section 4.1, are based on the proportion of genes in common for the two groups of relatives studied.

Fisher also showed that dominance would reduce the likeness between relatives and depress the correlation coefficient, particularly the parent - child correlation. Penrose (1949) extended this principle to the regression line test for dominance. The child's phenotypic value is regressed on the mean value of the two parents, Galton's mid-parental value. The regression coefficient will be $\sqrt{1/2}$ and the regression linear if the genes are perfectly additive and environmental effects are negligible. Dominance induces both departure from linearity and depression of the regression coefficient.

For parent - child and parent - parent correlations the observations are treated as independent and the standard interclass correlation coefficient computed (Holt, 1956; Snedecor and Cochran, 1966). In the calculation of the sib - sib correlations, however, the data are grouped by sibship resulting in an unequal number of observations per group. Two methods for correcting for unequal numbers have been suggested. Smith (1957) used analysis of variance techniques and weighted the among group sum squares. Earlier Cochran (1939) suggested an uncorrected sum squares within and among groups but weighted in the following manner:

$$r = \frac{1 - \frac{msq_W}{msq_B}}{1 + \left(\frac{N^2 - \Sigma n_i^2}{N(k - 1)} - 1 \right) \frac{msq_W}{msq_B}}$$

where msq_W = mean square within, msq_B = mean square between, n_i = number of sibs in the ith sibship, $N = \Sigma_i n_i$ and k = number of sibships. Both approaches were programmed and the estimates agreed consistently to two decimal places. Only results from the semi-weighted analysis of variance are presented as Smith provided the necessary formula for computation of the standard error.

The estimated values for the correlation coefficients are given below. Holt also reported a higher value for parent - child estimates when using only the mother than when using only the father. All estimates closely

approximate the theoretical values. A few show a slightly significant deviation which may well be due to the fact that this analysis was done over racial groups, so that there is a significant marital correlation.

Table 4.5 Correlation Coefficients
Estimated from Hawaii Data

Relationship	Observed r	Theoretical r	
Parent - Child	.4896 ± .0194	0.50	NS
Father - Child	.4427 ± .0269		NS
Mother - Child	.5609 ± .0225		*
Sib - Sib	.5544 ± .0009	0.50	*
Mid-Parent - Child	.6701 ± .0198	0.71	NS
Parent - Parent	.0752 ± .0358	0.00	*

NS = not significant deviation from expected; * = significant deviation from expected by Z transform (Fisher, 1950).

Additional estimates were made using the group of affected families. All correlation coefficients fell below the expected values. Because by selecting a group from one tail of distribution the variance among groups was reduced. From the standard formula for correlation coefficient $r = \sigma_B^2 / (\sigma_w^2 + \sigma_B^2)$, it can readily be seen why the results were obtained. The correlation coefficient

loses any interpretive value when estimated from such a restricted sample of the population.

4.4 Segregation Analysis

The families for the dermatoglyphics study were ascertained under incomplete selection as defined in section 2.4. The parameters to be estimated in the analysis are the segregation frequency $\underline{p}$, the proportion of sporadics $\underline{x}$, and the ascertainment probability, $\underline{\pi}$.

Unlike the cholesterol studies the parental mating types were known. Therefore, the two mating types, normal x normal and normal x affected were analyzed separately. No affected x affected matings were observed. The sample of normal x unknown mating type gave estimates not significantly different from normal x normal matings and the two samples were pooled in the final analysis. Reciprocal matings were pooled throughout the study.

In the first analysis families selected through a proband were analyzed. The U and K scores are given in Tables 4.6 and 4.7.

Table 4.8 Dermatoglyphics - Normal x Normal Matings
Estimates of Parameters from Segregation Analysis

$x = 0$	$\hat{x} = 0.2689 \pm 0.2987$
$\hat{p} = 0.1049 \pm 0.1059$	$\hat{p} = 0.1419 \pm 0.0658$
$\hat{\pi} = 0.1169 \pm 0.0253$	$\hat{\pi} = 0.4088 \pm 0.1053$

Table 4.6 Segregation Analysis

Dermatoglyphics

sr Table $p = .1419$ $x = .2689$ $\pi = .1149$ $p = .3429$ $x = .1372$ $\pi = .4088$

Children s	Affected r	No.	Normal x Normal Matings U_p	U_x	U_π	No.	Normal x Affected Matings U_p	U_x	U_π
2	1	9	-6.9610	1.3506	0.4695	3	-4.8351	1.4580	.6356
2	2	1	7.0915	-1.3760	-0.4783	3	9.5060	-2.8665	-1.2497
3	1	11	-16.1724	3.4112	1.1420	4	-14.9101	5.7948	1.6676
3	2	0	-	-	-	2	3.7823	-2.0738	-0.4231
4	1	9	-18.7586	4.3104	1.3946	2	-14.4235	7.2383	1.2305
4	2	1	4.8442	-1.3924	-0.3755	2	1.2114	-2.2477	-0.0264
4	3	1	13.0587	-1.3924	-0.8832	0	-	-	-
5	1	2	-5.2219	1.3101	.4111	0	-	-	-
5	2	1	3.7199	-1.4006	-0.3249	2	-1.3748	-2.4331	.3574
5	3	1	11.9344	-1.4006	-0.8326	1	3.7506	-1.2165	-0.3174
5	4	0	-	-	-	1	8.1886	-1.2165	-0.6975
6	1	3	-9.1420	2.5099	.7670	0	-	-	-
6	2	2	5.1904	-2.8177	-0.5496	0	-	-	-
6	3	1	10.8097	-1.4089	-0.7825	1	2.4501	-1.3153	-0.1316
6	4	0	-	-	-	1	6.8881	-1.3153	-0.5117
7	1	1	-3.3925	1.0216	.3053				
7	2	1	1.4701	-1.4171	-0.2252				
7	3	1	9.6846	-1.4171	-0.7329				
8	1	0	-	-	-				
8	2	1	.3446	-1.4253	-0.1761				
9	1	1	-3.8158	1.3930	.4029				
9	2	1	-0.7814	-1.4336	-0.1276				
10	1	1	-3.9025	1.5749	.4510				
Totals		49	-0.0000	-0.0000	- .1449	22	.2336	- .1935	.5336

Table 4.7 Segregation Analysis

Dematoglyphics

ra Table		π = .1149			π = .4088
Affected $\underline{r}$	Probands $\underline{a}$	No.	Normal x Normal Matings U_π	No.	Normal x Affected Matings U_π
2	1	7	-4.1959	8	-8.5036
2	2	1	9.2277	1	3.0748
3	1	4	-4.8868	1	-2.2583
3	2	0	-	1	1.8795
3	3	0	-	0	-
4	1	0	-	0	-
4	2	0	-	1	.5682
4	3	0	-	1	4.7059
4	4	0	-	0	-
Totals		12	0.14493	13	-0.5336
			K_π = 90.131		K_π = 60.994

On the null hypothesis that x=0 the three chi squares available from the sr tables and explained in section 2.4 were 1.555, 0.59, and 0.0551 respectively. In this material the proportion of sporadics did not differ significantly from zero. For a trait that develops in utero and is subject to very little if any environmental influence post partem this is a reasonable finding.

The null hypothesis x=0 was also tested in the normal x affected matings. The results were non-significant with χ^2 = 0.643, 0.37 and 0.1363. The parameter estimates are given below, the U and K scores are given in Tables 4.6 and 4.7.

Table 4.9 Dermatoglyphics Normal x Affected Matings Estimates of Parameters from Segregation Analysis

$x = 0$	$\hat{x} = -0.1372 \pm 0.1122$
$\hat{p} = 0.3876 \pm 0.0684$	$\hat{p} = 0.3429 \pm 0.0872$
$\hat{\pi} = 0.4021 \pm 0.1223$	$\hat{\pi} = 0.4088 \pm 0.1285$

The possibility that the trait defined by ten arches could be due to a recessive gene was tested in the following manner. Under incomplete selection only doubly heterozygous normal x normal matings would be ascertained giving an expected segregation frequency of 1/4. The observed frequency of 0.1049 could be explained by a gene that was not fully penetrant. The penetrance would be the

ratio of observed to expected or 0.5674. On the recessive hypothesis the class of normal x affected matings would consist of heterozygous backcrosses, with a theoretical segregation frequency of 1/2. The only testable model is to assume that the penetrance of the gene in the two mating types is equal. Then, the expected segregation frequency for the normal x affected would be 0.5 times 0.5674 or 0.2837. This value for p and the estimated value of 0.3876 were fitted to the data given x=0 and π =0.4021. The chi square tests from the sr tables for heterogeneity were again computed. The values with one degree of freedom of 6.336, 2.63 and 3.0423 represent both significant and non-significant deviation. These results could not be used to completely reject a major recessive gene hypothesis but neither are they indicative of one.

An additional attempt was made to detect a major gene segregating within the continuous distribution of ridge count (Figure III). Individuals defined as affected in the first analysis were omitted and the remaining individuals including controls were divided into affected and non-affected on the basis of their discriminant function value. The opposite tail of the distribution, high ridge count, constituted the affected group with the criterion being all values less than the mean value of the discriminant for the control group.

The segregation analysis was done separately for each of the mating types. This analysis was done under complete selection with parameters $\underline{p}$, $\underline{h}$ and $\underline{y}$ as defined in section 2.4, equations (22) and (23).

The estimates of the parameters from the normal x normal, normal x affected and affected x affected matings respectively, are given below. U and K scores are given in Tables 4.8 and 4.9.

Table 4.10 Dermatoglyphics - Complete Selection

Estimates of Parameters by Mating Type

N x N	N x A	A x A
$\hat{p} = 0.1538 \pm 0.0289$	$\hat{p} = 0.5919$	$\hat{p} = 0.8356 \pm .0250$
$\hat{h} = 0.0065 \pm .3271$	$\hat{h} = 0.1106 \pm .0391$	$\hat{h} = 0.0000 \pm .0004$
$\hat{y} = 0.0000 \pm .0002$	$\hat{y} = 0.1298 \pm .0318$	$\hat{y} = 0.1554 \pm .1316$

The estimates for $\underline{h}$ are consistent among mating types and not significantly different from zero.

The estimates of $\underline{p}$ are not compatible with a major gene hypothesis. Affected x affected matings have an expected segregation of 1.0 while the observed is only 0.8356. The segregation in normal x affected exceeds 0.5, indicating complete penetrance, while again the normal x normal matings show reduced penetrance. However, the penetrance values would not be the same if estimated from the two mating types which exhibit a reduction in expression.

Although a theory of variable penetrance is a possibility such a theory would be impossible to test and of little or no value.

Neither segregation analysis has detected the presence of a major gene in the distribution of ridge counts. These findings, of course, cannot completely rule out the possibility of the existence of such a gene.

After this thesis was written, a paper delivered by Frota-Pessoa at a Werner-Gren symposium at Burg Warterstein called attention to evidence from Brazil that blood group 0 may be associated with a high frequency of whorls, and he suggested hererozygous effect of the 0 gene on this trait. The possibility of racial stratification was discussed and discounted. Hawaii provides an excellent opportunity to test this hypothesis in a racially-stratified sample within which spurious associations should be randomized. Large samples of finger prints and ABO groups are available and need only linked with races (Morton, personal communication).

4.5 Quasi-continuous Analysis

This analysis was also done by the method previously delineated. After the initial estimates from NUQUAC, the probability for a sibship of size s with r affected was calculated as well as the risk for the s + 1 child given r affected. The latter is given in Table 4.11.

Table 4.11 Dermatoglyphics

Probability for s+1 child given s and r

(Beta approximation)

S/R	00	01	02	03	04	05	06
0	.000732						
1	.000653	.109007					
2	.000589	.098350	.196112				
3	.000536	.089592	.178647	.267703			
4	.000493	.082266	.164039	.245812	.327585		
5	.000455	.076047	.151639	.227230	.302822	.378414	
6	.000423	.070703	.140982	.211261	.281540	.351819	.422098
7	.000396	.066060	.131724	.197389	.263053	.328717	.394382
8	.000371	.061989	.123608	.185226	.246844	.308462	.370080
9	.000350	.058391	.116433	.174475	.232517	.290559	.348600
10	.000330	.055188	.110046	.164904	.219761	.274619	.329477
11	.000313	.052318	.104323	.156328	.208333	.260338	.312342
12	.000298	.049732	.099166	.148600	.198034	.247468	.296902

Table 4.11 (continued)

S/R	07	08	09	10	11	12
0						
1						
2						
3						
4						
5						
6						
7	.46046					
8	.431699	.493317				
9	.406642	.464684	.522726			
10	.384335	.439192	.494050	.548908		
11	.364347	.416352	.468357	.520362	.572367	
12	.346336	.395770	.445204	.494638	.544072	.593507

The data on all first degree relatives divided by sex of proband and sex of relatives were used to estimate b_i and c, where b_1 represents the estimated parameter for males and b_2 the estimate for females. The estimates are $b_1 = 1.959$, $b_2 = 1.192$, $c = 2.490$.

A sample for r = 0.125 and 0.25 was obtained by tabulating affected individuals, probands excluded, by sex and by the sex of the proband through which that family was ascertained. The estimate from the first degree relatives were fitted to all ten classes of relatives for which there was information (Table 4.12). The fit was actually quite good, $\chi^2_9 = 4.1151$, $p \sim 0.9$.

The c estimate of approximately 2.5 was somewhat lower than earlier estimates from other data. In some of the equations such a low value would result in a Φ that was not negligible. However, the estimates were obtained using equation (10), an exact expression subject only to numerical rounding error.

4.6 Summary

Several attempts to discern a major gene through segregation analysis of this body of data failed. Estimates were provided from more standard methods of analysis, e.g. correlation coefficients, which tended to indicate that total ridge count followed closely the predictions for an additive trait. Some slight discrepancies such as the slight though significant correlation between

Table 4.12 Dermatoglyphics

Parameters estimated from quasi-continuity

$B_1 = 1.959$, $B_2 = 1.192$, $c = 2.490$

r	Proband Sex	Relative Sex	Affected Observed	Affected Expected	χ^2	Normal Observed
.500	1	1	11	8.2172	1.0338	82
.500	1	2	19	18.5176	.0153	85
.500	2	1	14	15.5918	.1744	214
.500	2	2	31	32.2788	.0595	187
.250	1	1	1	.3512	1.2729	5
.250	1	2	1	1.1877	.0342	8
.250	2	1	1	1.5199	.1873	29
.250	2	2	3	4.0513	.3097	31
.125	2	1	0	.3014	.3149	7
.125	2	2	0	.6374	.7131	6

Sum $\chi^2_9 =$ 4.1151

1 = male
2 = female

parents are probably resulted from racial endogamy since they did not appear in intra-racial samples (Holt, 1952).

The estimates from the quasi-continuity analysis are reasonable, even though c is somewhat lower than the earlier observations from other data. The fit of the data to the hypothesis for all tested degrees of relationship was quite good.

The evidence from this study would indicate that total ridge count is an additive trait. Imposing a threshold effect by defining certain ridge counts, or discriminant function values as affected did not destroy the additive properties. This analysis would also seem to indicate that this model of quasi-continuous variation does have accurate and reasonable predictive value when applied to a trait which fulfills the basic assumptions of the model.

5. SUMMARY

The application of the theory of quasi-continuous variation to data on human diseases has been presented. The early model of Falconer predicts heritability which serves as an estimate of the importance of the genetic constituent in the etiology of a condition. However, as pointed out by Falconer (1965) and Morton (1967) the model requires extensions to provide recurrence risks in certain relatives of affected individuals.

The model suggested by Edwards' (1967) and presented here as derived by Elston, Campbell and Morton (in preparation) does permit prediction of recurrence risks in relatives. Also, in theory, it is possible to allow for effects of sex, age or other environmental factors by designating groups as sub-populations and estimating parameters for each population. Such estimates permit estimates of the type of recurrence risks demanded in genetic counseling.

It is possible to test the goodness of fit of the model by estimating prevalence in several classes of relatives of varying degrees of relationship. However, as was shown with the cholesterol study, a stringent test of the hypothesis will often require much more extensive data than is usually collected in a family study of a particular trait.

The model, as derived, assumes the trait under study to be solely polygenic and additive. If the trait in fact meets these criteria the fit should be good and the model of predictive value. A larger, faster computer, or improved numerical estimations should smooth out the difficulties encountered in estimating the parameters. It is encouraging that the fit in the dermatoglyphics study, an acknowledged additive trait, was quite good.

Genetic load theory (Morton, Crow and Muller, 1956) can be extended to predict risk in relatives on a multifactorial, non-synergistic model (Morton, personal communication). Such a theory would assume that many loci could bring about manifestation of the trait in question, each with a specified and usually small probability. However, the effects would not be additive. By fitting both this model and the additive multifactorial model presented here one could test for additivity. This would serve as a major addition to the analysis of traits assumed to be additive and quasi-continuous.

A summary of the analysis for each of the three traits studied is presented at the end of each section. In brief, hypercholesterolemia appears due to a dominant gene, very low ridge count is typically quasi-continuous, and pyloric stenosis (sometimes considered the example <u>par excellence</u> of quasi-continuity) is consistent with a mixture of megaphenic and microphenic effects but not with the current

models of quasi-continuity. The overall poor fit to the data on pyloric stenosis was somewhat disappointing but at the same time the good fit to the data on dermatoglyphics was reassuring. At least some resolution of the genetic mechanisms for these conditions was achieved through the combined use of segregation analysis, regression analysis and the expanded model of quasi-continuous variation.

The problems of detecting major gene effects while genetic heterogeneity, occurrence of sporadics, phenocopies, etc. obscure basic genetic segregation is difficult if not impossible. The tendency has been to suggest that traits which lacked clear resolution were multifactorial or quasi-continuous without acknowledging the difficulties in analysis. A more precise method of testing for additivity and quasi-continuity may help to prevent a general inclusion of unresolved traits in this particular category.

APPENDIX A:

Sections I-IV from Elston, Campbell and Morton in preparation.

I. Suppose x follows the truncate normal distribution

$N(\mu, \sigma^2)/\Phi(\frac{c-\mu}{\sigma})$, i.e.

$$f(x) = \frac{1}{\Phi(\frac{c-\mu}{\sigma})} \cdot \frac{1}{\sqrt{2\pi}\sigma} e^{-(1/2)(\frac{x-\mu}{\sigma})^2}$$

$x \leq c$

and suppose further $g(x) = e^{b(x-c)}$, $b > 0$. Then $x = (\ln g)/b + c$, $dx/dg = 1/bg$ and the distribution of g is

$$h(g) = f(x) \left| \frac{dx}{dg} \right|$$

$$= \frac{1}{\Phi(\frac{c-\mu}{\sigma})} \cdot \frac{1}{\sqrt{2\pi}\sigma bg} e^{-1/2(\frac{x-\mu}{\sigma})^2}$$

$$= \frac{1}{\Phi(\gamma)} \cdot \frac{\delta}{\sqrt{2\pi}\, g} e^{-(1/2)(\delta \ln g + \gamma)^2}$$

$g \leq 1$,

where $\delta = 1/b\sigma$ and $y = (c - \mu)/\sigma$.

II. The mean value of g is easily obtained as

$$\int_{-\infty}^{c} g(s)\ f(x)\ dx$$

$$= \frac{1}{\Phi(\frac{c-\mu}{\sigma})} \cdot \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{c} e^{b(x-c)\ -(1/2)(\frac{x-\mu}{\sigma})^2}\ dx$$

$$= \frac{1}{\Phi(\frac{c-\mu}{\sigma})}\ e^{b(\mu - c) + 1/2\,\sigma^2\, b^2}$$

$$\frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{c} e^{-1/2(\frac{x-\mu-\sigma^2 b}{\sigma})^2}$$

$$= \frac{\Phi(\frac{c-\mu-\sigma^2 b}{\sigma})}{\Phi(\frac{c-\mu}{\sigma})}\ e^{1/2b(\sigma^2 b + 2(\mu - c))}$$

III. The mean of the truncate normal distribution $N(\mu, \sigma^2)/\Phi(\frac{c-\mu}{\sigma})$ can be obtained using the identity

$$\int_{-\infty}^{c} te^{-(1/2)\ t^2}\ dt = -e^{-(1/2)\ c^2}$$

(This results immediately on substituting $y = t^2$). Putting $t = (x - \mu)/\sigma$ in this identity we obtain

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{c} (x - \mu)\ e^{-(1/2)(\frac{x-\mu}{\sigma})^2} =$$

$$-\frac{\sigma^2}{\sqrt{2\pi}}\ e^{-(1/2)c^2}$$

But

$$\mu\left(\frac{1}{\sqrt{2\pi}}\right)\int_{-\infty}^{c} e^{-(1/2)\left(\frac{x-\mu}{\sigma}\right)^2} =$$

$$\mu\,\Phi\left(\frac{c-\mu}{\sigma}\right)$$

therefore

$$\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{c} xe^{-(1/2)\left(\frac{x-\mu}{\sigma}\right)^2} =$$

$$\mu\,\Phi\left(\frac{c-\mu}{\sigma}\right) - \frac{\sigma^2}{\sqrt{2\pi}}\,e^{-(1/2)c^2}$$

Dividing through by $\Phi\left(\frac{c-\mu}{\sigma}\right)$ we obtain the mean of the truncated distribution as

$$\mu - \frac{\sigma^2\, e^{-(1/2)c^2}}{\sqrt{2\pi}\;\Phi\left(\frac{c-\mu}{\sigma}\right)}$$

Thus for $\sigma = 1$ and $c > 3$ the mean is virtually the same as that of the untruncated distribution, regardless of μ ($< c$).

IV. Suppose x and y follow the truncate bivariate normal distribution

$$f(x, y) = N\left(\begin{bmatrix}\mu_x\\ \mu_y\end{bmatrix}, \begin{bmatrix}1 & r\\ r & 1\end{bmatrix}\right)/\Phi$$

where Φ is the cumulative bivariate normal integral $\Phi\,(c - \mu_x,\ c - \mu_y,\ r)$. The distribution of $t = (x + y)/2$ is then given by

$$f(t) = \frac{1}{\Phi\, 2\pi\sqrt{1-r^2}} \frac{d}{dt} \int_{-\infty}^{c} \int_{-\infty}^{2t-x} f(x,y)\,dy\,dx$$

$$= \frac{1}{\Phi\, \pi\, \sqrt{1-r^2}} \int_{-\infty}^{c} e^{-\frac{1}{2(1-r^2)}}$$

$$\{(x-\mu_x)^2 - 2r(x-\mu_x)(2t-x-\mu_x)$$

$$+ (2t-x-\mu_y)^2\}\, dx$$

$$= \frac{1}{\Phi\, \pi\, \sqrt{1-r^2}} \int_{-\infty}^{c} e^{-\frac{1}{1-r}\{x-\mu+t\}^2}$$

$$e^{-\frac{1}{1+r}\{t-\mu\}^2}\, dx$$

where $\mu = (\mu_x + \mu_y)/2$.

Thus

$$f(t) = \frac{1}{\Phi\, \sqrt{\pi(1+r)}} \Phi\left(\frac{c-\mu+t}{\sqrt{(1+r)/2}}\right)$$

$$e^{-\frac{1}{1+r}\{t-\mu\}^2}$$

$$= \Phi\left(\frac{c-\mu+t}{\sqrt{(1-r)/2}}\right) N(\mu, (1+r)/2)/\Phi$$

Now for $r \leq 1/8$ the Φ that has t in its argument will be virtually unity for $c - \mu + t > 2$; i.e., $t > 2 - c + \mu$. Conservatively taking $c = 3.25$ and the mean of the distribution to be μ, this means that Φ is always virtually unity for all t greater than the mean minus two standard deviations. Thus in a region of small probability in the left hand tail of the distribution f(t) will be less than the approximation:

$N(\mu, (1 + r)/2)/\Phi\left(\frac{c - \mu}{\sqrt{(1 + r)/2}}\right)$, but everywhere

else it will be greater by at most a factor

$\Phi\left(\frac{c - \mu}{\sqrt{(1+r)/2}}\right)/\Phi$. In the extreme case $c = 3.25$,

$\mu = 1.5$ and $r = 1/8$, this factor is 1.07. Thus in such an extreme case the approximation may not be too good; however in most practical situations this factor is much closer to unity, and then the approximation is adequate.

Using this approximation, and taking the distribution of z (given t) to be $N(t, (1 - r)/2)/$

$\Phi\left(\frac{2c - t}{\sqrt{2(1-r)}}\right)$, the distribution of z is

$$f(z) = \frac{1}{\Phi\,\dfrac{c-\mu}{\sqrt{(1+r)/2}}\,\pi\,\sqrt{1-r^2}} \int_{-\infty}^{c} \frac{1}{\Phi\left(\dfrac{2c-t}{\sqrt{2\,(1-r)}}\right)} e^{-\frac{1}{1-r}\{z-t\}^2 - \frac{1}{1+r}\{t-\mu\}^2}\,dt$$

$$= \frac{1}{\Phi\,\dfrac{c-\mu}{\sqrt{(1+r)/2}}\,\pi\,\sqrt{1-r^2}} \int_{-\infty}^{c} \frac{1}{\Phi\,\dfrac{2c-t}{\sqrt{2(1-r)}}} e^{-\frac{2}{1-r^2}\{t-1/2[(1+r)z+(1-r)\mu]\}^2 - 1/2\{z-\mu\}^2}\,dt$$

Here the Φ in the integrand is smallest when c is small, t = c and r = 0, and for the extreme case c = 3.25 this becomes $\Phi(2.3) = 0.99$. We therefore ignore it in the integration and find

$$f(z) = \frac{\Phi\left(\dfrac{2c-(1+r)z-(1-r)\mu}{\sqrt{2\,(1-r^2)}}\right)}{\Phi\,\dfrac{c-\mu}{\sqrt{(1+r)/2}}\,\sqrt{2\pi}}\; N(\mu, 1)$$

Provided the Φ here that has z in its argument is very nearly a constant, this distribution will be

well approximated by $f(z) = N(\mu, 1)/\Phi(c - \mu)$, in which it will be recalled that $\mu = (\mu_x + \mu_x)/2$. The Φ will be nearly a constant provided (substituting c for z in the argument) $\Phi\left(\frac{(1 - r)(c - \mu)}{\sqrt{2(1 - r^2)}}\right)$ is near unity. It should be noted that if this condition is fulfilled the approximation used for f(t) is certainly very good, and so in practice this condition should be checked first. The particular values $c - \mu = 3$ and $r = 0$, for example, lead to $\Phi(2.12) = 0.98$, which is certainly close enough to unity for the approximation to be good.

If $c - \mu < 3$ the approximation may not be good enough, but we can nevertheless obtain the expected value of g(z) approximately as follows. The true distribution of z is

$$f(z) = \frac{1}{\Phi(\sqrt{\pi}(1 + r)} \int_{-\infty}^{c} \Phi \frac{c - \mu + t}{\sqrt{(1 - r)/2}}$$

$$e^{-\frac{1}{1 + r}\{t - \mu\}^2} \Phi\left(\frac{\frac{1}{2c - t}}{2(1 - r)}\right) e^{\frac{1}{1-r}\{z-t\}^2} dt$$

and so the expected value of g(z) is (using I, above)

$$\frac{e^{b_Z^2 (1 - r) - b_Z c}}{\Phi \sqrt{\pi} (1 + r)} \int_{-\infty}^{c} \frac{\Phi\left(\frac{c - \mu + t}{\sqrt{(1-r)/2}}\right)}{\Phi\left(\frac{2c - t}{\sqrt{2}(1 - r)}\right)}$$

$$\Phi\left(\frac{2c - 2t - (1 - r) b_Z}{\sqrt{2}(1 - r)}\right) e^{b_Z t - \frac{1}{1+r}\{t - \mu\}^2} dt$$

$$= \frac{e^{(5/4)b_Z^2 (1-r) + b_Z (\mu - c)}}{\Phi \sqrt{\pi} (1 + r)} \int_{-\infty}^{c} \frac{\Phi\left(\frac{c - \mu + t}{\sqrt{(1-r)/2}}\right)}{\Phi\left(\frac{2c - t}{\sqrt{2}(1-r)}\right)}$$

$$\Phi\left(\frac{2c - 2t - (1-r)b_Z}{\sqrt{2}(1 - r)}\right) e^{-\frac{1}{1-r}\{t - (\mu + \frac{(1-r)}{2} b_Z\}^2} dt$$

Letting $K = e^{5/4\, b_Z(1-r) + b_Z (\mu - c)}$.

$$\Phi\left(\frac{2c - 2\mu - (1 + r)b_Z}{\sqrt{2}(1 - r)}\right)$$

and $\varphi(t) \dfrac{\Phi\left(\frac{(c - \mu + t)}{\sqrt{(1 - r)/2}}\right)}{\Phi\left(\frac{2c - t}{\sqrt{2}(1 - r)}\right)} \Phi\left(\frac{2c - 2t - (1-r) b_Z}{\sqrt{2}(1 - r)}\right)$

We see that this is K times the expected value of $\varphi(t)$ over the truncate normal distribution with mean $\mu + (1 - r)b_Z/2$ and variance $(1 - r)/2$. The expected value of t over this distribution is (III, above)

$$\mu_t = \mu + \left(\frac{1-r}{2}\right) b_Z - \frac{(1 - r)e^{-1/2c^2}}{2\sqrt{2}\,\Phi\left(\frac{2c - 2\mu - (1-r)b_Z}{\sqrt{2}\,(1 - r)}\right)}$$

and as a first approximation to the expected value of φ (t) we can take $\varphi(\mu_t)$; this will slightly overestimate the true value.

V. Derivation of $\tilde{\underline{x}}$, the proportion of sporadic cases among sibs in families with at least one parent affected.

In this discussion we are considering two distributions of individuals each with a different probability of being affected with pyloric stenosis (Figure III). The low-risk group consisting of the greater portion of the general population has a probability of recurrence similar to the incidence. The high-risk group consists of families with a probability of recurrence greater than the incidence. This probability is designated $\underline{p}$, the segregation frequency.

I = total probability of being affected, the incidence in the general population.

x = probability of sporadic cases among affected

p = probability of being affected in the high-risk group, the segregation frequency

1. A proportion $\underline{x}$ of the affected parents will be sporadic cases.
2. The risk of having an affected child is not increased in the case of an affected parent who is sporadic if dominant mutations are negligible. Therefore, the probability that a parent who is affected and who is sporadic will have an affected child is approximately $\underline{I}$.
3. A proportion $(1 - x)$ of the affected parents will not be sporadic cases.
4. The probability of the non-sporadic affected parent having an affected child is increased above the general incidence by the amount of the segregation frequency. The probability of an affected child becomes:

$$p + I - pI$$

approximated by $(p + I)$, where $\underline{pI}$ is small and may be neglected.

5. The total probability of being affected given that one parent is affected would be the probability contributed by the sporadic affected parents plus the probability contributed by the non-sporadic affected parents:

$$x(I + (1 - x)(p + I)$$
$$= xI + (1-x)p + I - xI$$
$$= I + (1 - x)p$$

6. Given one parent affected the probability of being affected and being a sporadic case is the total probability minus the probability of being affected but not sporadic. That is, the total minus the portion of p which represents non-sporadic cases, $(p - xp)$, or:

$$
\begin{aligned}
& I + (1 - x)p - (p - xp) \\
= \; & I + p - xp - p + xp \\
= \; & I
\end{aligned}
$$

7. From this the proportion of sporadics given at least one parent is affected is the probability of sporadic cases over the total probability of being affected:

$$\tilde{x} \doteq \frac{I}{I + (1 - x)p}$$

APPENDIX B

Pedigrees for all families of the current study of pyloric stenosis with more than one affected individual.

Symbols Used:

Propositus

Affected (not propositus)

Deceased

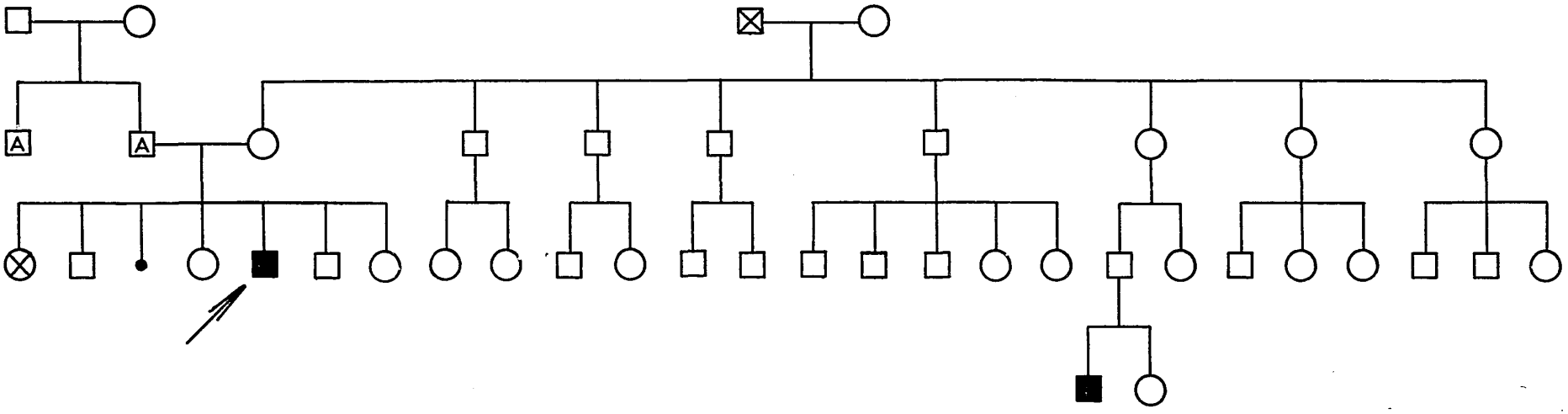
Adopted

Surgery for ulcer

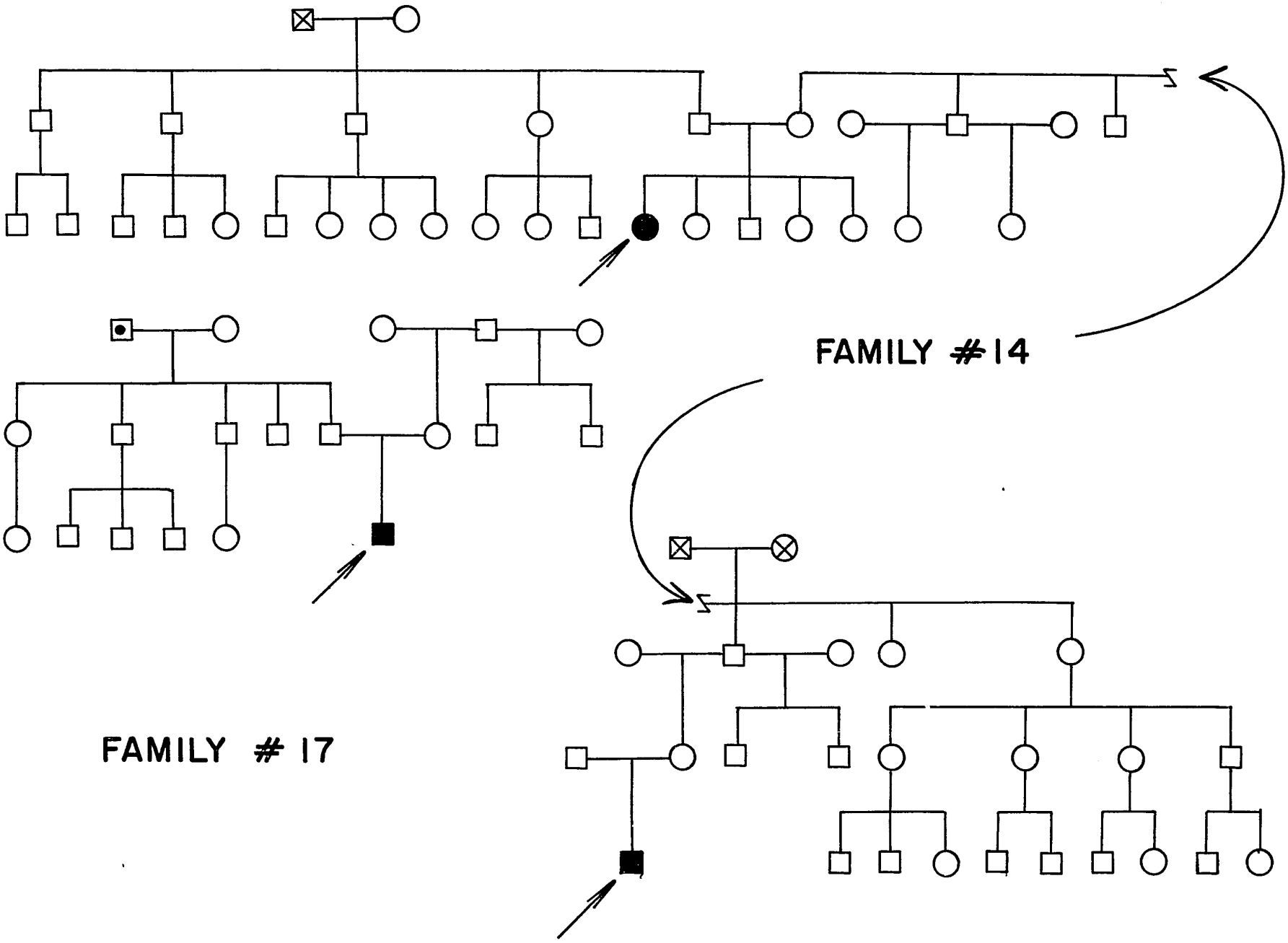
Miscarriage (or abortion)

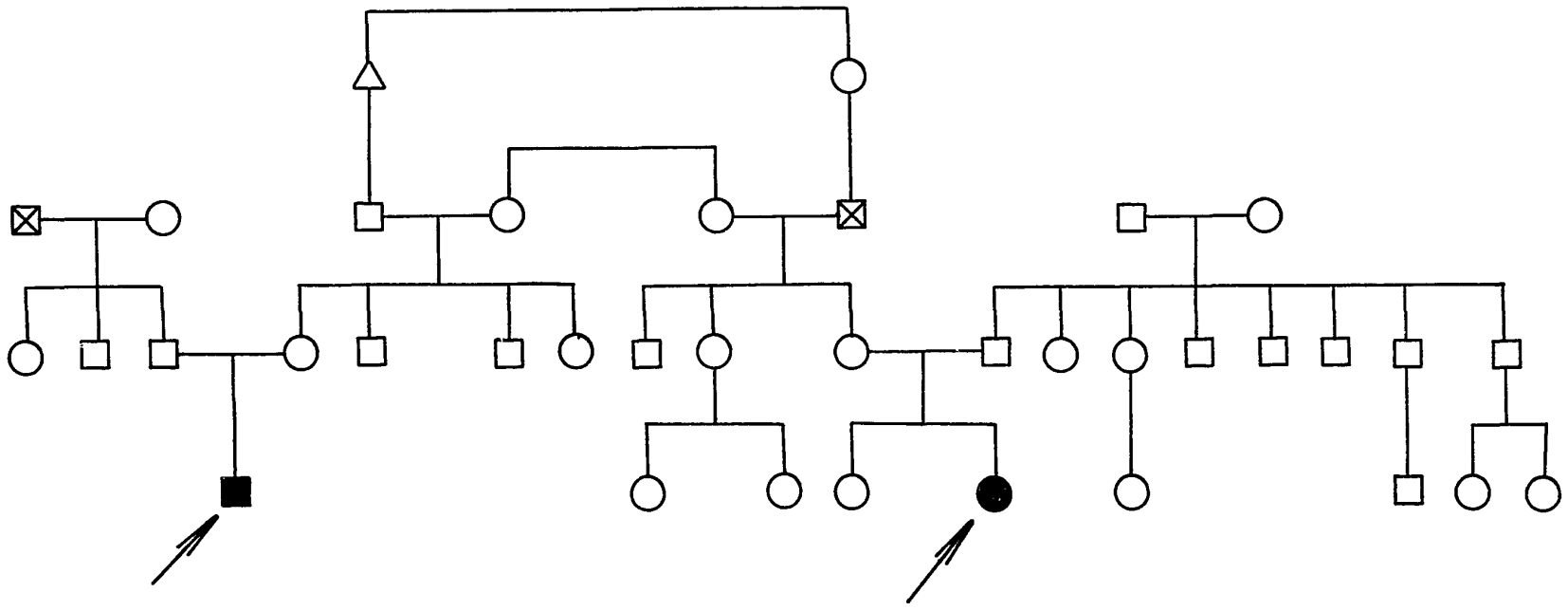
Sex unknown

Note: For some pedigrees where information was obtained by mail form it was not possible to divide the cousins into sibships.



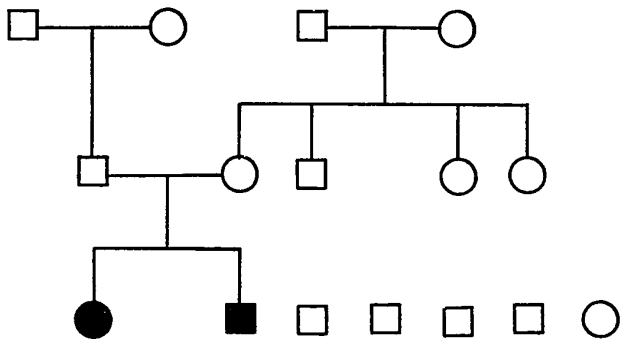
FAMILY # 008


FAMILY #14
FAMILY # 17

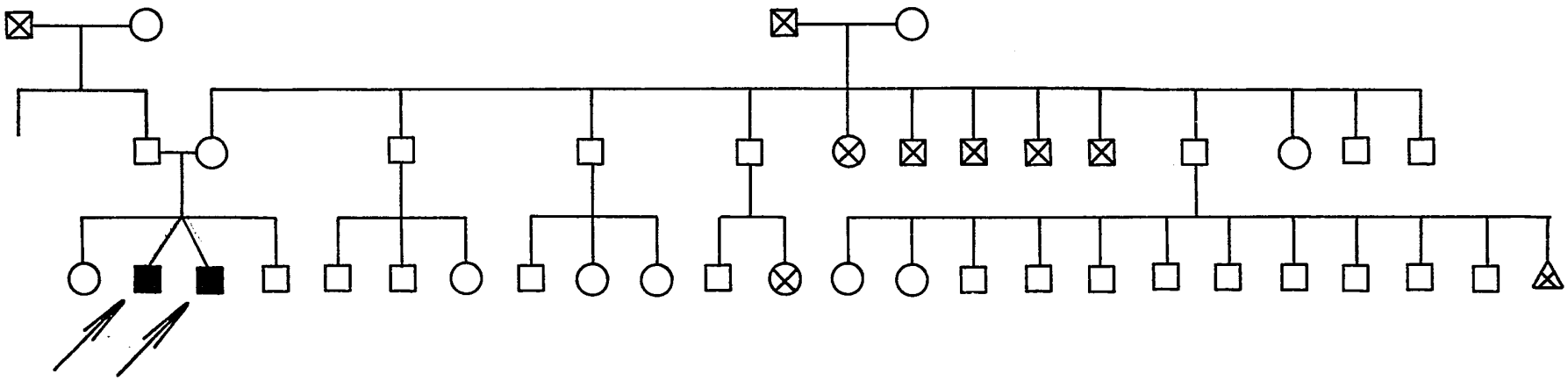


FAMILY # 112

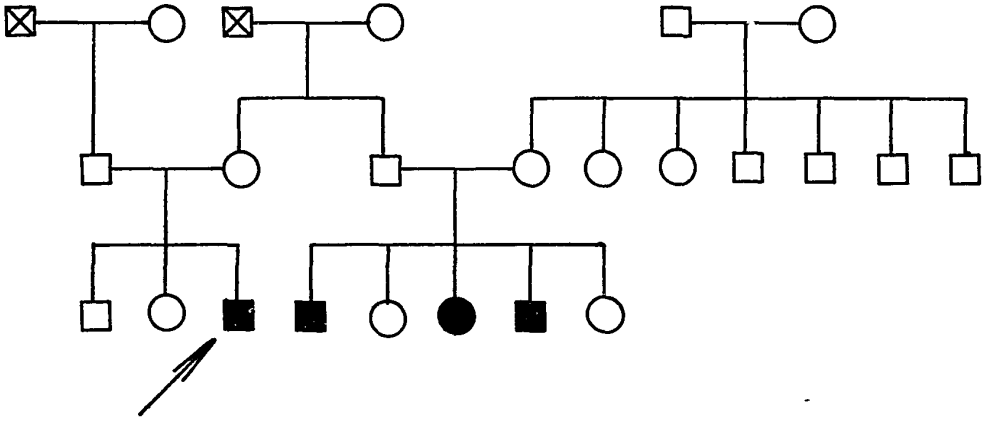
FAMILY # 025



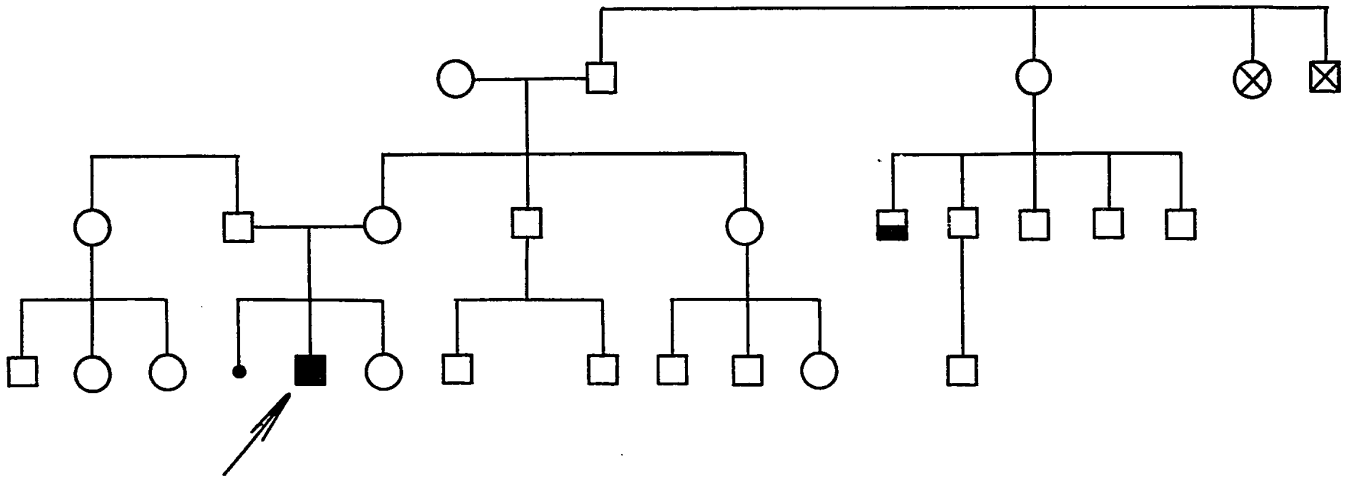
FAMILY # 31



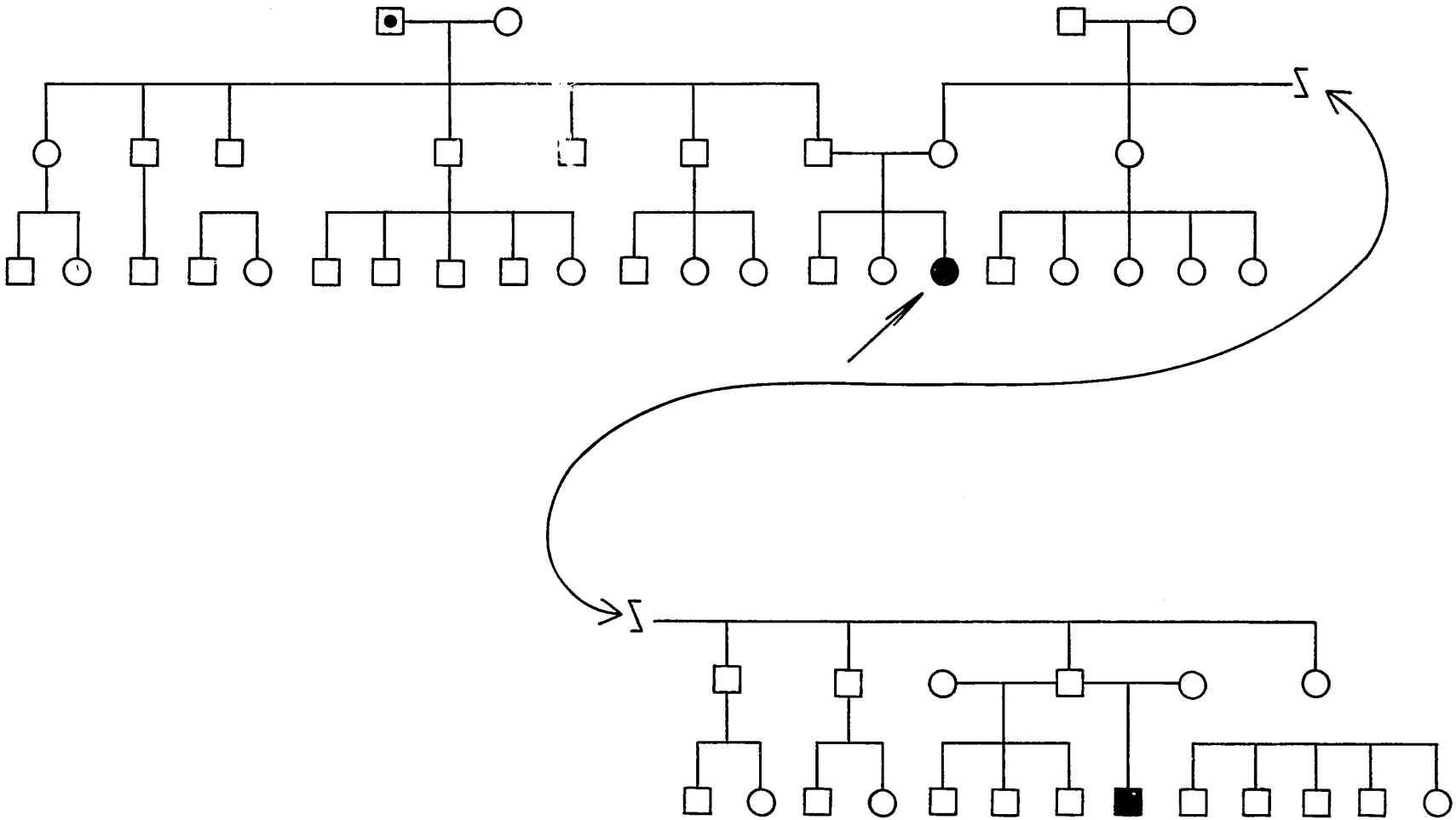
FAMILY # 41



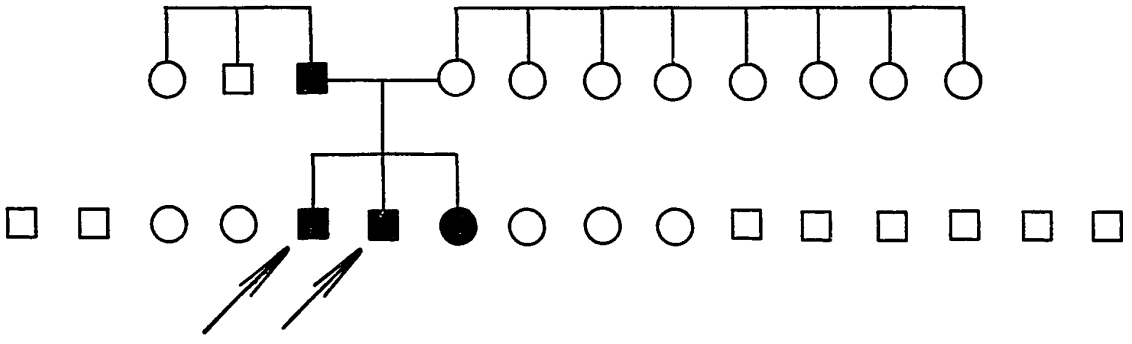
FAMILY # 47



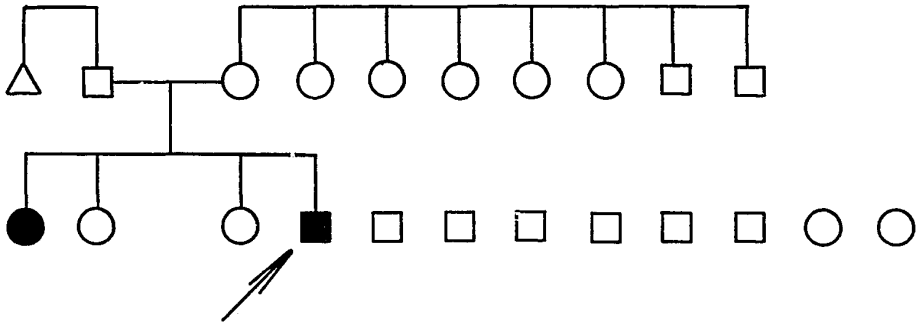
FAMILY # 66



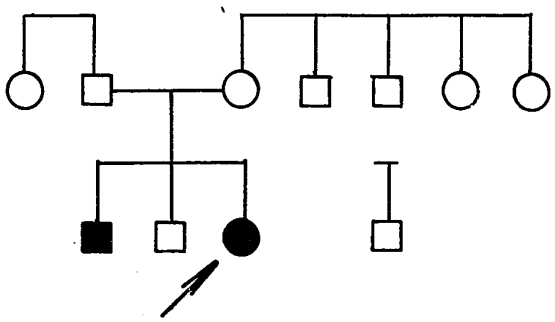
FAMILY # 94



FAMILY # 150



FAMILY # 158



FAMILY # 256

LITERATURE CITED

Aldersberg, D., L. Schaefer, J. Drachman, and A. Steinberg. 1952. Genetic Studies on Idiopathic Hypercholesteremia. J. Am. Med. Assoc. 162: 619.

Benson, D., W. Mustard, M. Ravitch, W. Snyder and K. Welch. 1964. Review of 1,120 cases of pyloric stenosis. Am. J. Surg. 107: 429.

Burian, C. Multiple Regression Analysis. Computer Applications in Human Genetics. Vol. II. ed. N. E. Morton. In press.

Campbell, M. Unpublished.

Carter, C. 1961. Inheritance of Congenital Pyloric Stenosis. Brit. Med. Bull. 17: 251.

Carter, C. 1963. The Genetics of Common Malformations. Second International Conference on Congenital Malformations. 306.

Carter, C. 1965. The Inheritance of Common Congenital Malformations. Progress in Medical Genetics. Vol. IV: 59.

Carter, C. and T. Savage. 1951. Pyloric stenosis in four first cousins. Arch. Dis. Child. 26: 50.

Cochran, W. 1939. The use of the analysis of variance in enumeration by sampling. J. Amer. Stat. Assoc. 34: 492.

Cockayne, E. and L. Penrose. 1943. The Genetics of Congenital Pyloric Stenosis. Ohio Journal of Science. 43: 1.

Cummins, H. and C. Midlo. 1943. Fingerprints, Palms and Soles. Philadelphia. Blakiston.

Edwards, J. 1960. The Simulation of Mendalism. ACTA genet. 10: 63.

Edwards, J. 1967. Linkage Studies of Whole Populations. Proc. of the III International Congress of Human Genetics. ed. J. Crow and J. Neel. 479.

Elston, R., M. Campbell, and N. Morton. Quasi-continuity. In preparation.

Epstein, F., W. Black, E. Hand, T. Frances. 1959. Familial Hypercholesteremia, Xanthomas and Coronary Heart Disease. Amer. J. Med. 26: 39.

Falconer, D. 1964. Introduction to Quantitative Genetics. Edinburgh. Oliver and Boyd.

Falconer, D. 1965. The inheritance of liability to certain diseases, estimated from the incidence among relatives. Ann. Hum. Genet. 29: 51.

Fisher, R. 1918. The correlation between relatives on the supposition of Mendalian inheritance. Trans. Roy. Soc. Edinb. 52: 399.

Fisher, R. 1950. Statistical Methods for Research Workers. 11th ed. New York. Hafner.

Gerrard, J., J Waterhouse, and D. Maurice. 1955. Infantile Pyloric Stenosis. Arch. Dis. Child. 30: 493.

Gharib, R. 1964. The incidence of infantile pyloric stenosis. J. Ped. 65: 622.

Goldschmidt, E. (ed.) 1963. The Genetics of Migrant and Isolate Populations. New York. Williams and Wilkins.

Gordon, H., W. Pollock, W. Norris, and J. Hart. 1959. Hypertrophic Pyloric Stenosis: Experience with 1573 Cases at Los Angeles Children's Hospital. West. J. Surg. 67: 139.

Gruneburg, H. 1952. Genetical studies on the skelton of the mouse: IV. Quasi-continuous variations. J. Genet. 51: 95.

Hale, A. 1952. Morphogenesis of volar skin in the human fetus. Amer. J. Anat. 91: 147.

Harlon, W., J. Graham, and E. Estes. 1966. Familial Hypercholesterolemia: A Genetic and Metabolic Study. Medicine. 45: 77.

Hirschorn, K., and D. Wilkenson. 1957. Mode of Inheritance in Essential Familial Hypercholesterolemia. Circulation 16: 509.

Holt, S. Genetics of dermal ridges: inheritance of total finger ridge-count. Ann. Eugen. 17: 140.

Holt, S. Genetics of dermal ridges: frequency of distributions of total finger ridge-count. Ann. Hum. Genet. 20: 159.

Holt, S. Genetics of dermal ridges: parent-child correlations for total finger ridge-count. Ann. Hum. Genet. 20: 270.

Holt, S. Genetics of dermal ridges: the relation between total ridge count and variability of counts from finger to finger. Ann. Hum. Genet. 22: 323.

Holt, S. Fingerprint patterns in Mongolism. Ann. Hum. Genet. 27: 279.

Holt, S. The Genetics of Dermal Ridges. Springfield, Illinois, Thomas.

Khachodurian, A. 1964. The inheritance of Essential Familial Hypercholesterolemia. Amer. J. Med. 37: 402.

Laron, Z. and L. Horne. 1957. Incidence of Infantile Pyloric Stenosis. AMA J. Dis Child. 94: 151.

McKeown, T., B. MacMahon, and R. Record. 1951a. The Incidence of Congenital Pyloric Stenosis Related to Birth Rank and Maternal Age. Ann. Hum. Genet. 16: 249.

McKeown, T., B. MacMahon, and R. Record. 1951b. The Familial Incidence of Congenital Pyloric Stenosis. Ann. Hum. Genet. 16: 260.

McKeown, T., B. MacMahon, and R. Record. 1951c. Congenital Pyloric Stenosis. An investigation of 578 cases. Brit. J. Soc. Med. 5: 185.

McKeown, T., B. MacMahon, and R. Record. 1952. Evidence of post-natal environment influences in the aetioloty of infantile pyloric stenosis. Arch. Dis. Child. 27: 386.

McKusick, V. 1964. Human Genetics. Englewood Cliffs, New Jersey. Prentice Hall.

Morton, N. E. 1958. Segregation analysis in human genetics. Science. 127: 79.

Morton, N. E. 1959a. Genetic tests under incomplete ascertainment. Amer. J. Hum. Genet. 11: 1.

Morton, N. E. 1959b. Methods of study in human genetics. In Genetics and Cancer. Austin: University of Texas Press, pp. 391-407.

Morton, N. E. 1962. Segregation and linkage. In Methodology in Human Genetics, W. J. Burdette, ed. San Francisco: Holden-Day, pp. 17-52.

Morton, N. E. 1965. Models and evidence in human population genetics. In Genetics Today, vol. 3 (Proceedings of the XI International Congress of Genetics, The Hague, The Netherlands, September, 1963), S. J. Geerts, ed. Oxford: Pergamon Press, pp. 935-951.

Morton, N. E. 1967. The Detection of Major Genes Under Additive Continuous Variation. Amer. J. Hum. Genet. 19: 23.

Morton, N. E. Quasi-continuity, unpublished.

Morton, N. E., C. Chung, and M. P. Mi, 1967. Genetics of Interracial Crosses in Hawaii. Basel. S. Karger.

Morton, N., J. Crow, and H. Muller. 1956. An estimate of the mutational damage in man from data on consanguineous marriages. Proc. Nat. Acad. Sc. 42: 855.

Nevin, N. and J. Slack. 1968. Hyperlipidaemic xanthomatosis. II: Mode of Inheritance in 55 Families with Essential Hyperlipidaemia and Xanthomatosis. J. Med. Genet. 5: 9.

Penrose, L. 1949. The Biology of Mental Defect. London. Sidgwick & Jackson.

Powell, B. and C. Carter. 1951. Pyloric Stenosis in Twins. Arch. Dis. Child. 26: 45.

Seives, M. 1968. Serum cholesterol levels in Southwestern American Indians. S. Chron. Dis. 21: 107.

Schaefer, L., S. Drachman, A. Steinberg, and D. Aldersberg. 1953. Genetics Studies of Hypercholesteremia: Frequency in a Hospital Population and in Families of Hypercholesteremic Patients. Amer. Heart J. 46(1): 99.

Shim, W., A. Campbell and S. Wright. 1968. A twenty-five year survey of pyloric stenosis in Hawaii I. Clinical Aspects.

Skellam, J. 1948. A Probability Distribution Derived From the Binomial Distribution by Regarding the Probability of Success as Variable Between the Sets of Trials. J. Royal Slat-Soc Series B. 10: 257.

Slack, J. and N. Nevin. 1968. Hyperlipidaemic Xanthomatosis I: Increased Risk of Death from Ischaemic Heart Disease in First Degree Relatives of 53 patients with Essential Hyperlipidaemic Xanthomatosis. J. Med. Genet. 5: 4.

Smith, C. 1957. On the estimate of intraclass correlations. Ann Hum. Genet. 21(4): 362.

Snedecor, G. and W. G. Cochran. 1967. Statistical Methodology. 6th ed. Iowa State University Press.

Stanbury, J., J. Wyngaarden, and D. Frederickson. 1966. The Metabolic Basis of Inherited Disease. New York: McGraw-Hill. 2nd Edition. pp. 429-485.

Wallgren, A. 1941. Incidence of Hypertrophic Pyloric Stenosis. Amer. J. Dis. Child. 62: 751.

Wheeler, E. 1957. Genetic Aspects of Atherosclerosis. Amer. J. Med. 23: 653.

Wilkenson, C., E. Hand, M. Flieglemon. 1948. Essential Familial Hypercholerterolemia. Ann. Int. Med. 29: 671.

Yee, S. and N. E. Morton. Segregation Analysis. Applications of Computers in Human Genetics. Vol. II. In press.