Am. J. Hum. Genet. 51:178–190, 1992

A Method for Assessing Patterns of Familial Resemblance in Complex Human Pedigrees, with an Application to the Nevus-Count Data in Utah Kindreds

Lue Ping Zhao,*† John Grove,† and Filemon Quiaoit*

*Program of Epidemiology, Cancer Research Center and †Biostatistics Program, School of Public Health, University of Hawaii, Honolulu

Summary

An analytic method is described for estimating phenotypic correlations between pairs of members of specific relationships in pedigrees. In estimating correlations, this new method allows simultaneous adjustment for available covariates such as age, gender, environmental factors, and variables reflecting ascertainment mode, through mean- and variance-regression models. The estimated correlations and regression coefficients corresponding to covariates are consistent and asymptotically normally distributed. Differing from a full-likelihood approach, this new method does not require the assumption of a particular joint distribution of phenotypes from a pedigree, such as the multivariate normal distribution, but instead only requires correct specification of mean- and variance-regression models. Within this framework, missing data, if they are *missing completely at random,* can be ignored without biasing estimates. The method is illustrated by an application using nevus-count data from 28 Utah kinships. The results from the analysis are that covariate-adjusted nevus counts are correlated between parents and children (correlation .22; $P < .001$) and between siblings (correlation .32; $P < .001$), while the correlation of $-.04$ between husband and wife is not significantly different ($P = .31$) from 0. This result is consistent with a genetic etiology of nevus count.

Introduction

Human pedigrees, defined as clusters of blood relatives and their spouses, are frequently collected in human genetic studies. Such studies are often concerned with the estimation of the degree of the familial resemblance of a particular phenotype, between paired relatives of a given relation (e.g., mother and daughter, brother and sister) (Falconer 1989, 148–162). A commonly used measure of familial resemblance is the correlation coefficient; it is particularly useful for quantitative phenotypes, since it has a simple interpretation and has been widely used in many other biometric applications. Recently, it has been argued that, despite its boundary, the correlation is also useful for binary phenotypes (Zhao and Prentice 1990).

Received October 31, 1991; revision received February 25, 1992.

Address for correspondence and reprints: Lue Ping Zhao, Ph.D., Program of Epidemiology, Cancer Research Center, University of Hawaii, 1236 Lauhala Street, Honolulu, HI 96813.

0002-9297/92/5101-0019$02.00

Despite its apparent simplicity, the development of analytic methods for estimating correlations in complex human pedigrees has been difficult. To avoid complex structures of general human pedigrees, many researchers, including Rosner et al. (1977), Konish and Khatri (1991), Karlin et al. (1991), and Srivastava et al. (1988), have considered a special family structure where each family pedigree has only one parent and several children. Their proposed estimates of the correlations between parent and child and between siblings have been reviewed by Eliasziw and Donner (1990; Donner and Eliasziw 1991). Other researchers, including Donner and Koval (1981), Rao et al. (1985), and Shoukri and Ward (1989), considering a nuclear family structure with two parents and several children, proposed maximum-likelihood estimates of correlations derived under the assumption of a multivariate normal distribution.

With no distributional assumption, an intuitive and consistent estimate of a correlation between a specific pair of relatives is the estimate of the Pearson product-moment correlation. To obtain this estimate, one lists all possible pairs who have the same specific relation-

ship in each pedigree and then computes the Pearson correlation of these paired relatives' phenotypes. This estimate has been criticized for being disproportionally influenced by large pedigrees (Smith 1980); and its statistical properties are not well understood, because the resulting pairs from the same pedigree are not independent (Rosner et al. 1977).

Statistical methods for the analysis of dependent data had not been well developed until the recent introduction of the estimating equations that evolved from the generalized linear model (GLM) (Nelder and Wedderburn 1972) and quasi-likelihood (Wedderburn 1974). Both the GLM and quasi-likelihood were proposed to unify apparently diverse regression techniques. Their properties and applications have been summarized by McCullagh and Nelder (1989). The GLM approach requires two assumptions: (1) the mean of a response variable is a known function of a linear combination of covariates with a vector of regression coefficients to be estimated; and (2) the distribution of the response variable, given the covariates, arises from a known member of the exponential family of distributions. Under the GLM, the score-estimating equations are completely specified by the mean and variance of the response. The estimated coefficients from solving the score-estimating equations are consistent and have an asymptotic normal distribution with an easily estimable variance matrix when all assumptions are true. It is found that, even if the response arises outside the exponential family of distributions, the estimates from the score-estimating equations continue to have these desirable properties. This robustness property led to the conceptual development of quasi-likelihood: by assuming only the function for the mean and variance of the response, one will obtain consistent and normally distributed estimates of regression coefficients, from the score-estimating equations. Recently, Liang and Zeger (1986; Zeger and Liang 1986) extended the GLM and quasi-likelihood approaches to the estimating-equations approach for multivariate analysis of longitudinal responses. Let $y' = (y_1, \ldots, y_n)$ denote a vector of n longitudinal responses. The estimating equations approach assumes that the marginal distribution of each response, y_j, follows the GLM with a correctly specified mean, μ_j, conditional on observed covariates, and variance, σ_j^2, of the jth response. The mean is indexed by a vector of p regression coefficients, $\alpha^t = (\alpha_1, \ldots, \alpha_p)$, and is of particular interest. Let $\mu' = (\mu_1, \ldots \mu_n)$ and $\sigma' = (\sigma_1, \ldots, \sigma_n)$, where σ_j is the square root of the variance. These longitudinal responses are correlated. Let $R = \|r_{ij}\|_{n \times n}$ denote a "working correlation matrix" that need not be correctly specified, where $\gamma_{ii} = 1$ if $i = j$ and where γ_{ij} equals the "working correlation" between y_i and y_j if $i \neq j$. The estimated regression coefficients in the mean vector satisfy the estimating equations

$$\sum \frac{\partial \mu'}{\partial \alpha} W^{-1}(y - \mu) = 0 , \qquad (1)$$

where the summation Σ is over all independent samples, $\partial \mu' / \partial \alpha$ is a $p \times n$ matrix of derivatives of the mean vector with respect to the regression coefficients, and the weight matrix $W = \text{diag}(\sigma) \times R \times \text{diag}(\sigma)$ and $\text{diag}(\sigma)$ is a diagonal matrix with the diagonal elements σ. The degenerated estimating equations (1) for a single response are the score-estimating equations under the GLM. Prentice (1988) formalized Liang and Zeger's idea of estimating R by an additional set of estimating equations. In summary, the essence of the estimating-equations approach is that, without making distributional assumptions about the correlated responses, the estimated regression coefficients from the estimating equations (1) are consistent and have an asymptotic normal distribution, provided that only the model for the mean is correctly specified. Following the same idea, Zhao and Prentice (1990; Prentice and Zhao 1991) extended Liang and Zeger's estimating equations to a set of estimating equations for jointly estimating parameters in mean-, variance-, and correlation-regression models, since the second-order moments, especially the correlation, are of particular interest in many scientific disciplines, such as genetics (Zhao and Prentice 1991).

Acknowledging the complexity of general human pedigrees, we propose an analytic method for assessing patterns of familial resemblance by estimating those specific correlations of interest, on the basis of a modification of Zhao and Prentice's estimating equations. This new method is illustrated by an application to the analysis of nevus-count data in Utah kindreds.

Data and Method

Data

Notation.—Consider m independent pedigrees. In the ith pedigree, there are n_i individuals. The phenotype of the jth individual in the ith pedigree is denoted by y_{ij}, while the ijth individual is characterized by a vector of p covariates, $x_{ij}^t = (x_{ij1}, \ldots, x_{ijp})$, including

age, sex, and other demographic or environmental variables. In the ith pedigree, the jth and kth individuals are related through a particular genealogical relationship, denoted by $G(j,k)$, where $G(j,k)$ takes values $1,2, \ldots, G$ corresponding to G exclusive types of relationships. Let S_{ig} denote the set of all pairs of individuals from the ith pedigree who have a specific relationship of type g, i.e., $S_{ig} = \{(j,k)|G(j,k) = g$ over all possible pairs in the ith pedigree$\}$, where $g = 1, 2, \ldots, G$ numerate all G-exclusive relationships. Complex relationships such as double first cousins may be treated simply as relationships of their own kind and have their own index value g. Also, let $y_i^t = (y_{i1}, \ldots, y_{in_i})$ denote a vector of n_i phenotypes, and let $x_i^t = (x_{i1}, \ldots, x_{in_i})$ denote a $p \times n_i$ matrix of covariates.

Nevus-count data in the Utah kindreds.—This set of nevus-count data from Utah kindreds was contributed to the seventh genetic analysis workshop (MacCluer et al. 1992). An updated version of this data set was made available to the authors by Dr. D. Goldgar from the University of Utah (personal communication). The updated data set includes only 28 kindreds, a subset of the pedigrees provided to the workshop. Details of the study design and results of the previous analyses have been reported elsewhere (Meyer et al. 1988; MacCluer et al. 1992). In brief, the first group of 12 kindreds was obtained by referral of families having two or more cases of melanoma or dysplastic nevus syndrome in first- or second-degree relatives. An additional 16 kindreds were identified through sequential melanoma cases. On average, each kindred has about 23 individuals, except that two particularly large kindreds, 1764 and 1771, have 165 and 126 individuals, respectively.

The total nevus count, which is believed to be associated with melanoma (Green and Swerdlow 1989), is the trait chosen for analysis here. The covariates included in the analysis are age, sex, skin type, hair color, eye color, proband status, an indicator for the two ascertainment modes, and variables for the ascertainment correction.

Accompanying this data set are the issues of ascertainment bias and missing data. In this example, the pedigree data are ascertained on the basis of the occurrences of either melanoma or dysplastic nevi syndrome (DNS), while the total nevus count is the phenotype of interest. Since the nevus count has been found to be associated with both melanoma and DNS, ignoring melanoma or DNS status might yield biased estimates of correlations. We propose below (see the Ascertainment Correction Subsection) to perform a conditional analysis that includes an indicator variable for proband status as a covariate. By including the proband indicator, we are actually estimating correlations in the subpopulation of individuals who are not the probands, and the estimated correlations are thus generalizable (at least under the modeling assumptions).

The problem of missing either outcomes or covariates is common in human genetic studies. In this data set, information on nevus count or age was missing on 128 individuals. One of the main reasons for missing data is outmigration, which is unlikely to be associated with nevus count or with other important covariates. Another reason for missing data is death. For those subjects who are deceased because of causes other than melanoma, missing data are again unlikely to be associated with the total nevus count. Overall, the missing data due to the outmigration or death of other diseases are approximately *missing completely at random,* since the missing information does not depend on either the missed data or the observed data (Little and Rubin 1987). Missing data due to death from melanoma are not *missing completely at random,* since they probably influence the nevus-count data. Fortunately, the mortality from melanoma is relatively low, and the small percentage of such missing data presumably has a limited influence on estimates. Therefore, in either situation, the missing data can be simply ignored without biasing estimates. Note that the estimates are generalizable only to those subjects who are more than 15 years old, since this study measured nevus counts in them only.

Method

Rationale.—Our primary concern is with the estimation of the correlations, of phenotypes, between paired relatives due to unexplained shared environmental and genetic factors. For example, consider a quantitative phenotype y_{ij} determined by an unexplained factor U_{ij} and an observable random covariate x_{ij} through a linear model $y_{ij} = \alpha_0 + \alpha_1 x_{ij} + U_{ij} + \varepsilon_{ij}$, where (α_0, α_1) are parameters, ε_{ij} is a random error, $(U_{ij}, \varepsilon_{ij})$ are assumed to have mean zero, and $(U_{ij}, x_{ij}, \varepsilon_{ij})$ are independent of each other. The mean and variance of the phenotype y_{ij}, given x_{ij}, are expressed as $\mu_{ij} = \alpha_0 + \alpha_1 x_{ij}$, and $\sigma_{ij}^2 = \sigma_{uij}^2 + \sigma_e^2$, respectively, where σ_{uij}^2 is the variance of U_{ij} and σ_e^2 is the error variance. The covariance between y_{ij} and y_{ik}, given x_{ij}, is given by the covariance between U_{ij} and U_{ik}, denoted by ζ_{ijk}. Thus, the correlation of interest is given by $\gamma_{ijk} = \zeta_{ijk}/\sqrt{\sigma_{ij}^2\sigma_{ik}^2}$. Note that if the jth and kth individuals share

a common factor, i.e., $U_{ij} = U_{ik}$, the correlation above is simplified to be $\gamma_{ijk} = \sigma_{ui}^2/(\sigma_{ui}^2 + \sigma_e^2)$, which is the intraclass correlation (Falconer 1989, pp. 148–162).

An appropriate estimate of the correlation γ_{ijk} requires the correct specification of the mean, given the observable covariate x_{ij}; otherwise, the estimate may be asymptotically biased. One such misspecification can be due to ignoring the observable covariate x_{ij}. Ignoring the observable covariates would make the mean and variance of the phenotype equal α_0 and $\alpha_1^2\sigma_{xi}^2 + \sigma_{ui}^2 + \sigma_e^2$, respectively. The resulting covariance between y_{ij} and y_{ik} would equal $[\zeta_{ijk} + \text{cov}(x_{ij}, x_{ik})]$. Since the correlation is influenced by the covariance between x_{ij} and x_{ik}, it is necessary to specify correctly the mean, by including all relevant x_{ij} as covariates in the regression model for means. Even if the covariate x's are independent of each other, i.e., the covariance of (x_{ij}, x_{ik}) equals zero, the correlation may still be influenced by the variance $\text{var}(x_{ij})$, since the variance is involved in the calculation of correlations above. It is therefore necessary to specify correctly the variance as well, by including x_{ij} in the regression model for variances.

Correlation-, mean-, and variance-regression models.—Let a binary indicator h_{ijkg} take the value 1 if the jth and kth individuals are relatives of gth type from the ith pedigree, and let it take the value of 0 otherwise. The vector of G binary indicators $h_{ijk}^t = (h_{ijk1}, \ldots, h_{ijkG})$ takes the values $(1, 0, \ldots, 0), (0, 1, \ldots,), \ldots, (0, 0, \ldots, 1)$ corresponding to $g = 1, 2, \ldots, G$ relationships, respectively. The correlations corresponding to G types of relationships can be combined into one regression model; for example, the linear correlation regression model is $\gamma_{ijk} = h_{ijk}^t\delta$, where $\delta^t = (\delta_1, \ldots \delta_G)$ is a vector of G parameters of interest. Each δ_g in this model corresponds to the correlation for paired relatives of the gth type.

The correlation-regression model can be generally expressed as

$$\gamma_{ijk} = \gamma(h_{ijk}^t\delta), \tag{2}$$

where $\gamma(\bullet)$ is a specified function. A chosen function may be nonlinear. For example, a hyperbolic function $\gamma(\alpha) = [1 - \exp(-\alpha)]/[1 + \exp(-\alpha)]$ ensures that the correlation falls in $(-1,1)$.

The vector h_{ijk} may be extended to include other covariates for testing certain hypotheses. For example, one may wish to compare the other correlations with the first correlation: H_0: $\delta_g = \delta_1$, $g = 2, \ldots, G$ in the above formulation (2). To perform such a test, one may include a constant term in h_{ijk}, such as $h_{ijk}^t = (1, h_{ijk2}, \ldots, h_{ijkG})$. Under the model (2) indexed by parameters $\delta^t = (\delta_1, \ldots, \delta_G)$, the parameter δ_1 is an intercept and $\delta_g (g>2)$ corresponds to the differences $(\delta_g - \delta_1)$. To test the hypothesis, one may equivalently test H_0: $\delta_g = 0$. Another example arises from this study with the nevus-count data. As described earlier, the ascertainment of the first 12 kindreds is different from that of the remaining 16 kindreds. To test the consistency of estimates from the two different ascertainment modes, one may include an additional binary covariate, $h_{ijk,G+1}$, for describing ascertainment. To test the hypothesis, one introduces this variable as a covariate in the correlation-regression model.

Consider a mean-regression model

$$\mu_{ij} = \mu(\dot{x}_{ij}^t\alpha), \tag{3}$$

where $\mu(\bullet)$ is a specified function, $\dot{x}_{ij}$ is a vector of $\dot{p}$ covariates, a subset of the covariate vector x_{ij}, and $\alpha^t = (\alpha_1, \ldots, \alpha_{\dot{p}})$ is a vector of $\dot{p}$ parameters. The choice of this function $\mu(\bullet)$ is up to investigators. If scientific theory does not provide the function, investigators might wish to choose some appropriate and flexible function with limits that are equal to the range of the response variable (some examples of such functions have been given in McCullagh and Nelder 1989).

Consider a variance-regression model

$$\sigma_{ij}^2 = \sigma(\ddot{x}_{ij}^t\beta, \mu_{ij}), \tag{4}$$

where $\sigma(\bullet)$ is a specified function, $\ddot{x}_{ij}$ is a vector of $\ddot{p}$ covariates, also a subset of x_{ij}, and $\beta^t = (\beta_1, \ldots, \beta_{\ddot{p}})$ is a vector of $\ddot{p}$ parameters. The specified variance-regression model may depend not only on the covariates $\ddot{x}_{ij}$ but also on the mean μ_{ij}. Note that, for a binary phenotype, the variance is fully specified, by the mean, as $\sigma_{ij}^2 = \mu_{ij}(1 - \mu_{ij})$ and thus is not subject to any further modeling.

Regression models for nevus-count data.—The correlation-, mean-, and variance-regression models (2)–(4) are rather general and can be applied to either continuous or discrete phenotypes. In the following, we consider a set of specific models for nevus counts. The mean- and variance-regression models are chosen for consistency with the analysis by Thomas (1992).

The correlation-regression model considers the estimation of the correlations between parent and child, between siblings, and between parents and is expressed by the covariate vector $h_{ijk}^t = (h_{ijk1}, h_{ijk2}, h_{ijk3})$, which equals (1,0,0), (0,1,0), and (0,0,1), respec-

tively. The covariate vector h_{ijk} will be extended to include a binary indicator for the two ascertainment modes, as described in the previous section. A linear function $\gamma(\bullet)$ is useful for exploratory modeling and is chosen here for its simple interpretation.

The mean-regression model is used for nevus counts; we use an exponential regression model $\mu_{ij} = exp(\dot{x}_{ij}^t\alpha)$ to ensure that the estimated mean is positive. The inclusion of a covariate in the covariate vector $\dot{x}_{ij}$ depends on whether its association with the average nevus count is of interest, whether it influences the estimated correlations (by comparing two estimated correlations with and without adjusting for the covariate), and whether it is part of the ascertainment correction. For example, age and sex are included in $\dot{x}_{ij}$ as covariates, both because their association with the averaged nevus count is of interest and also because ignoring these covariates may influence the estimated correlations. Indicators for ascertainment correction are also included (and are discussed in the Ascertainment Correction Subsection).

The variance-regression model $\sigma_{ij}^2 = \mu_{ij}\exp(\ddot{x}_{ij}^t\beta)$ allows dependence of the variance on the mean and also ensures a positively estimated variance ($\sigma_{ij}^2 > 0$) by the exponential function. Count responses (such as total nevus count) are often assumed to arise from the Poisson distribution. The variance of such a count response equals the mean. Thus, the assumed variance-regression model above can be thought of as a Poisson variance with an extravariability factor, $\exp(\ddot{x}_{ij}^t\beta)$. The criterion for including a covariate in the variance-regression model is the same as that for including a covariate in the mean-regression model.

Estimation.—A recently developed extension of the method for estimating equations (Zhao and Prentice 1990, 1991; Prentice and Zhao 1991) is used to estimate the parameter vector (δ, α, β), the estimates of which are denoted by $(\hat{\delta},\hat{\alpha},\hat{\beta})$. The estimates $(\hat{\delta},\hat{\alpha},\hat{\beta})$ satisfy a set of $(G + \dot{p} + \ddot{p})$ equations

$$\sum_{i=1}^{m} X_i^t F_i = 0\,, \qquad (5)$$

where the summation $\Sigma_{i=1}^m$ is over all m independent pedigrees. The design matrix X_i^t is a block-diagonal matrix with diagonal block elements of $h_i^t = (\ldots, h_{ijk}, \ldots)$, $\dot{x}_i^t = (\dot{x}_{i1}, \ldots, \dot{x}_{in_i})$ and $\ddot{x}_i^t = (\ddot{x}_{i1}, \ldots, \ddot{x}_{in_i})$. The vector F_i is a concatenation of the three vectors $(z_i - \zeta_i)$, $(y_i - \mu_i)$, and $(s_i^2 - \sigma_i^2)$, where a vector of covariances $\zeta_i^t = (\ldots, \zeta_{ijk}, \ldots)$, a vector of means $\mu_i^t = (\mu_{i1}, \ldots, \mu_{in_i})$, and a vector of variances $(\sigma_i^2)^t = (\sigma_{i1}^2, \ldots, \sigma_{in_i}^2)$. In F_i, $s_i^2 = (y_i - \mu_i)^2$ is an $n_i \times 1$ vector of unbiased estimates of the variances σ_i^2, since $E(s_i^2) = \sigma_i^2$, given μ_i; $z_i^t = (\ldots, z_{ijk}, \ldots)$ includes all relevant pairs in S_{ig}, where each $z_{ijk} = (y_{ij} - \mu_{ij})(y_{ik} - \mu_{ik})$ is an unbiased estimate of the covariance ζ_{ijk}, since $E(z_{ijk}) = \zeta_{ijk}$, given μ_{ij} and μ_{ik}.

The expression for the estimating equations (5) is rather compact. To describe this expression more explicitly, let us consider a simple pedigree with six subjects (fig. 1) and detail the vectors and matrices involved in the estimating equations (5). The size of the ith pedigree is six ($n_i = 6$). If our primary interest is in the correlations between parents, between parent and child, and between siblings, then there are only nine relevant pairs ($G_i = 9$): (1,2) and (4,5) (husband, wife); (1,3), (1,4), (2,3), (2,4), (4,6), and (5,6) (parent, child); and (3,4) (sib,sib). Suppose that the mean μ_i of the phenotypes depends on sex and age, the variance σ_i^2 depends on sex, and the covariance ζ_i depends on the familial relationship and an additional covariate that is constant within the pedigree (with value .2 for this kinship). The matrices h_i, $\dot{x}_i$, and $\ddot{x}_i$, and vectors $(z_i - \zeta_i)$, $(y_i - \mu_i)$ and $(s_i^2 - \sigma_i^2)$ for this kinship are given in the Appendix, as are the F_i and X_i. The h_i matrix is easily modified to estimate more specific correlations. For example, to estimate sex-specific parent-child correlations, one replaces the single variable h_{ijk1} for parent-child by four binary variables—

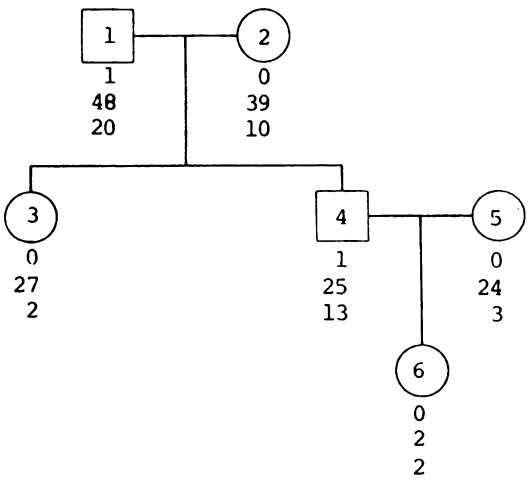

Figure 1 Pedigree used to demonstrate the configuration of vectors and matrices for the ith contribution to the estimating equations illustrated in the Appendix. The number inside the square or circle is the individual's index number; the numbers beneath are, respectively, sex, age, and number of nevi (the phenotypic trait, y_{ij}).

one each for father-son, father-daughter, mother-son, and mother-daughter.

It can be shown, in a way similar to that described by Zhao and Prentice (1990), that the estimates $(\hat{\delta},\hat{\alpha},\hat{\beta})$ are consistent and have an asymptotic normal distribution as the number of pedigrees m increases. The asymptotic variance matrix, Σ, may be estimated by

$$\hat{\sum} = A^{-1}(\sum_{i=1}^{m} X_i^t F_i F_i^t X_i) A^{-1},$$

where $A = \Sigma_{i=1}^{m} A_i$, and A_i is a derivative matrix of $X_i^t F_i$ with respect to parameters (δ,α,β). In $\hat{\Sigma}$, the diagonal elements are the estimated variances for the estimated parameters, while the off-diagonal elements yield the estimated covariances. The estimates' asymptotic normal distribution allows one to construct test statistics for testing correlations, as well as regression parameters, in the mean- and variance-regression models.

Also, note that the estimating equations (5) include only those pairs of correlations that are present in $S_{ig}, i = 1, \ldots, m; g = 1, \ldots, G$, in contrast to all possible pairs of correlations in the estimating equations proposed earlier by Zhao and Prentice (1990; Prentice and Zhao 1991). To differentiate it from other methods, we call this method a "pairwise method." By the pairwise method, one can focus on the relevant correlations of interest.

A program PFR (Patterns of Familial Resemblance) for this pairwise method has been written in GAUSS (Aptech System 1984), on a personal computer. The computing time depends on the sizes of pedigrees and the numbers of covariates. In most of our analysis on the pedigree data from 28 Utah kindreds, the computing time for fitting a model on an IBM PS2/model 70 386 is less than 20 min. The program also allows the user to specify several different functions for correlations, means, and variances.

Ascertainment correction.—Human pedigree data are frequently collected according to a particular ascertainment mode. Two modes are discussed here. One is that the ascertainment of a particular pedigree depends only on one or more relatives' phenotypes, who are known as probands. Let $(y_0, y_1, \ldots, y_m)$ denote a vector of relatives' phenotypes, where $y_1, \ldots$ and y_m correspond to nonproband relatives' phenotypes, and y_0 is a scalar phenotype corresponding to a single proband (but can also be a vector of phenotypes corresponding to multiple probands). The information from these phenotypes may be characterized by the conditional distribution function $f(y_1, \ldots, y_m|y_0)$.

The other ascertainment mode is that the ascertainment of a particular pedigree depends on one or more probands' covariates. Let $(x_0, x_1, \ldots, x_m)$ denote a vector of relatives' covariates. The quantitative information from these phenotypes can be characterized by the conditional distribution function $f(y_0,y_1, \ldots, y_m,x_1, \ldots x_m|x_0)$. The second ascertainment mode is emphasized in the following discussion, since the Utah kindreds were ascertained on the basis of the occurrence of melanoma or DNS, while the phenotype of interest is the total nevus count.

The conditional distribution function, $f(y_0,y_1, \ldots, y_m, x_1, \ldots, x_m|x_0)$, is proportional to $f(y_0, y_1, \ldots, y_m|x_0,x_1, \ldots, x_m)$, since the former function can be decomposed into the product of the latter function with the conditional distribution function, $f(x_1, \ldots, x_m|x_0)$, which is unlikely to contribute relevant information. In view of the estimating equations described earlier in this paper, the quantitative information may be characterized by the conditional moments: the conditional mean $E(y_j|x_0, \ldots, x_m)$, the conditional variance $\text{var}(y_j|x_0, \ldots, x_m)$, and the conditional covariance $\text{cov}(y_j,y_k|x_0, \ldots, x_m)$. A conditional analysis using the estimating equations would yield valid information about the familial resemblance of the phenotypes, given the covariates. Additional assumptions, however, are needed in order to appropriately model these conditional moments.

With any method of estimation, correct ascertainment adjustment depends on the underlying mechanism that produced the data. In practice, it is preferable to consider alternate ascertainment corrections, in order to see whether the estimated correlations have changed. One such a correction assumes that

$$E(y_j|x_0, \ldots, x_m) = E(y_j|x_j);$$

$$\text{var}(y_j|x_0, \ldots, x_m) = \text{var}(y_j|x_j);$$

$$\text{cov}(y_j,y_k|x_0, \ldots, x_m) = \text{cov}(y_j, y_k|x_j,x_k);$$

that is, these moments, dependent on only their own covariates, are independent of the other relatives' covariates. The dependence of these simplified moments on the covariates can be modeled easily by the mean-, variance-, and correlation-regression models and can be estimated by using the estimating equations (5). This conditioning constitutes an ascertainment correction for pedigree data that are ascertained by using either single or multiple probands.

An alternate model assumes that

$$E(y_j|x_0, \ldots, x_m) = E[y_j|x_j, \text{rel}(x_j,x_0)];$$

$$\text{var}(y_j|x_0, \ldots, x_m) = \text{var}[y_j|x_j,\text{rel}(x_j,x_0)];$$

$$\text{cov}(y_j,y_k|x_0, \ldots, x_m) = \text{cov}[y_j,y_k|x_j,\text{rel}(x_j,x_0), x_k, \text{rel}(x_k,x_0)];$$

where $\text{rel}(x_j,x_0)$ indicates the genealogical relation between the jth relative and one or more probands. It is expected that, if the inclusion of x_j's do not adequately account for the ascertainment, the conditional moments may, in addition, depend on the genealogical relations, $\text{rel}(x_j, x_0)$, with probands. These dependencies can be modeled and tested, and biases in estimating correlations—biases that are due to ignoring such relations—can be observed. To specify the relations as covariates in respective regressions, one may create a vector of indicators for the genealogical relationships with probands. If pedigrees are ascertained through a single proband, the creation of indicators is straightforward, by, e.g., binary indicator variables specifying whether a subject is a parent, sibling, child, or spouse of the proband. If data are ascertained by using multiple probands, additional indicators may be created.

Ascertainment correction is important in the development of methods for pedigree data analysis, since misspecifying it can invalidate parameter estimation. For example, a simple attempt at a correction might be to exclude probands, i.e., the analysis based on the likelihood function $f(y_1, \ldots, y_m)$ for the analysis of data ascertained by the first ascertainment method. The resulting estimate may be biased, since the actual likelihood function $f(y_1, \ldots, y_m|y_0)$ is conditional. Similarly, the same analysis on the data ascertained by the second ascertainment method would be based on $f(y_1, \ldots, y_m|x_1, \ldots, x_m)$, which is again different from the actual likelihood function $f(y_1, \ldots, y_m|y_0,x_0,x_1, \ldots, x_m)$; and consequently the estimates might be biased.

Results

Under the assumed regression models for correlations, means, and variances, the estimated correlations and parameters for mean and variance models are presented in table 1. The estimated correlations of .20 between parent and child and .32 between siblings

Table 1

Estimated Correlations between Parent and Child, between Siblings, and between Husband and Wife and Estimated Parameters in Mean-Regression Model $\mu_{ij} = \exp(\dot{x}'_{ij}\alpha)$, and Variance-Regression Model, $\sigma^2_{ij} = \mu_{ij}\exp(\ddot{x}'_{ij}\beta)$

Covariate	Coding	Coefficient (standard error)	*P*
Correlations:			
Between parent and child	1 = Yes; 0 = no	.20 (.04)	<.001
Between sibling and sibling	1 = Yes; 0 = no	.32 (.06)	<.001
Between husband and wife	1 = Yes; 0 = no	−.08 (.05)	.10
Means:			
Intercept	1	3.27 (.15)	<.001
Ascertainment	1 = Multiples; 0 = single	.12 (.09)	.21
Proband	1 = Yes; 0 = no	.50 (.08)	<.001
Age	Age/10	.39 (.08)	<.001
Age^2	(age/10)^2	−.07 (.01)	<.001
Sex	1 = Male; 0 = Female	.01 (.06)	.93
Variances:			
Intercept	1	2.36 (.25)	<.001
Ascertainment	1 = Multiples; 0 = single	.40 (.23)	.08
Proband	1 = Yes; 0 = no	−.03 (.41)	.95
Age	Age/10	.43 (.21)	.04
Age^2	(age/10)^2	−.05 (.04)	.22
Sex	1 = Male; 0 = female	−.27 (.21)	.19

are significantly different from zero, suggesting that familial aggregation of nevus count is due to shared environment or genes. The latter correlation is significantly greater than the former correlation (not shown). The −.08 correlation between husband and wife is not significantly different from zero.

In the mean-regression model, both estimated parameters .39 and −.07, corresponding to the linear and quadratic terms of age, respectively, are significantly different from zero. The combination of linear and quadratic terms can be approximated as $-.07(\text{age}-30)^2$ plus a constant; taken at face value, it suggests that the nevus count increases with age up to 30 years and decreases after that. This is consistent with the known growth pattern of nevi (Green and Swerdlow 1989). The average nevus count does not differ between males and females ($P = .93$), nor does it differ across the two ascertainment processes ($P = .21$). The estimate of .50 for the effect of proband status is significantly different from zero ($P < .001$), implying an association of nevus count with melanoma or DNS status. Ignoring proband status might therefore affect the estimates of the correlations. It would be preferable to include the occurrences of melanoma and DNS, instead of the proband indicator, in the regression model, since the occurrences of these two diseases were the ascertaining criteria and since their association with the nevus count would be of interest. Unfortunately, a large proportion of participants were not examined for the occurrence of DNS.

In the variance-regression model, it appears that the variability of nevus counts increases with age ($P = .04$) but not with the quadratic term of age ($P = .22$). The variability from pedigrees ascertained through multiple cases is marginally significantly greater than that ascertained through a single case ($P = .08$). The variability of nevus counts is not significantly associated with either gender or proband status ($P = .19$ and .95, respectively).

After excluding insignificant covariates from the mean- and variance-regression models, we next investigated the associations of means and variances with skin types, hair color, and eye color separately (in table 2), as well as the effects on estimating the correlations by the inclusion of these covariates in the mean- and variance-regression models (not shown). In the first four rows of table 2, the association with skin type is investigated, with skin type 1 as a reference. The estimated parameters .34 and .39, of skin types 2 and 3, respectively, are significantly different from that of the skin type 1 ($P = .01$), while skin type 4 or 5 is not significantly different from skin type 1 ($P = .62$). The variability of nevus count with skin type 4 or 5 is significantly different from that of skin type 1 ($P = .01$), while those of skin types 2 and 3 are not significantly different from that of skin type 1 ($P = .84$ and .96, respectively). In the fifth to eighth rows, the association with hair color is investigated, with black hair as a reference. The coefficient −.46 for red hair is significantly different from that for black hair ($P = .02$), suggesting that individuals with red hair tend to have a lower nevus count. The rest of the estimated parameters in either means or variances are not significantly different from zero. In the last four rows, eye color is investigated, with blue eyes as a reference. None of the estimated parameters in mean- and variance-regression models are significantly different from zero. Finally, the estimated correlations after adjustment for skin type, hair color, and eye color individually are not much different from those without any adjustment.

After including these significant terms in the mean- and variance-regression models, we estimated the correlations again; the results are listed in the second to fourth rows of table 3. The estimated correlation .22 between parent and child is slightly increased over the previous value of .20 in table 1, while the estimated correlation of .32 between siblings is the same as the

Table 2

Estimated Parameters Corresponding to Skin Type, Hair Color, and Eye Color, in Mean-Regression Model $\mu_{ij} = \exp(\dot{x}_{ij}\alpha)$ and Variance-Regression Model $\sigma^2_{ij} = \mu_{ij}\exp(\ddot{x}_{ij}\beta)$

Covariate Coding	Mean: Coefficient (standard error)	Mean: P	Variance: Coefficient (standard error)	Variance: P
Skin:				
Type 1 (reference) ...	0		0	
Type 2	.34 (.12)	.01	.06 (.29)	.84
Type 3	.39 (.15)	.01	−.02 (.35)	.96
Type 4 or 5	.04 (.19)	.62	−.93 (.38)	.01
Hair color:				
Black (reference)	0		0	
Brown	.04 (.14)	.77	.12 (.28)	.67
Blond	.04 (.13)	.75	.12 (.25)	.63
Red	−.46 (.20)	.02	.08 (.22)	.71
Eye color:				
Blue (reference)	0		0	
Green	−.12 (.15)	.40	.51 (.31)	.10
Hazel	.02 (.10)	.83	.33 (.20)	.10
Brown or grey	−.07 (.10)	.49	−.16 (.26)	.53

Table 3

Estimated Correlations between Parent and Child, between Siblings, and between Husband and Wife

Covariate	Coefficient (standard error)	*P*
Correlations:[a]		
Parent and child	.22 (.05)	<.001
Sibling and sibling	.32 (.06)	<.001
Husband and wife	−.04 (.04)	.31
Correlations without ascertainment correction:[a]		
Parent and child	.22 (.05)	<.001
Sibling and sibling	.32 (.06)	<.001
Husband and wife	−.07 (.05)	.15
Correlations with misspecified mean:[b]		
Parent and child	.10 (.04)	.01
Sibling and sibling	.36 (.06)	<.001
Husband and wife	−.07 (.07)	.28
Correlations with misspecified variance:[c]		
Parent and child	.23 (.05)	<.001
Sibling and sibling	.31 (.06)	<.001
Husband and wife	−.05 (.05)	.31

[a] Estimated under specified mean $\mu_{ij} = \exp(\dot{x}_{ij}^t\alpha)$ and variance $\sigma_{ij}^2 = \mu_{ij}\exp(\ddot{x}_{ij}^t\beta)$.
[b] Estimated under misspecified mean $\mu_{ij} = \exp(\alpha)$.
[c] Estimated under misspecified variance $\sigma_{ij}^2 = \mu_{ij}\exp(\beta)$.

previous value. Both estimated coefficients are significantly different from zero, but the difference between them is no longer significantly different from zero. The correlation −.04 between husband and wife is not significantly different from zero ($P = .31$).

To quantify the effects of the ascertainment correction on the correlations, the binary-indicator variable for proband was excluded from the mean-regression model. The fourth to sixth rows of table 3 give the estimated correlations. The estimated correlations between parent and child and between siblings have changed only slightly. The estimated correlation between husband and wife, however, is biased away from the null, while the *P* value decreases to .15. Overall, the adjustment for ascertainment in this data set has not substantially altered the estimated correlations of nevus counts. To further check the validity of this ascertainment correction, we fitted another model including four indicators (for parent, sibling, child, and spouse) into the mean- and variance-regression models (not shown). Each indicator variable takes the value 1 if a proband has the specified relation with the subject and takes the value 0 otherwise. The variability among the spouses of probands is significantly lower than that among other subjects, while the rest of the indicators affect neither the variability nor the mean. After adjustment, the estimated correlations between parent and child and between husband and wife decreased from .32 and −.04 to .27 and −.03, respectively, while that between parent and child remained unchanged. Such modest changes suggest that simply including the proband indicator as a covariate was an adequate adjustment for these data, and further modeling on ascertainment correction was not attempted.

To demonstrate the effect of misspecifying the mean and variance functions, the correlations were estimated without adjusting for covariates in the mean- and variance-regression models. The 9th–11th rows of table 3 give the estimated correlations under the assumption of a constant mean $\mu_{ij} = \exp(\alpha)$. The correlation between parent and child is biased toward zero, while that between siblings is biased away from zero; the correlation between husband and wife is hardly affected. In the next three rows of table 3, the correlations are estimated by using the variance function $\sigma_{ij}^2 = \mu_{ij}\exp(\beta)$. It appears that the estimated correlations are only slightly affected.

An approximate test for heterogeneity due to the two different ascertainment processes was made by including a binary-indicator variable and relevant cross-products with relationship variables in the

correlation-regression model; the .30 correlation between parent and child in the first 12 pedigrees apparently is significantly different from the −.02 correlation between parent and child in the other 16 pedigrees ($P < .01$). The correlations between siblings and between spouses are not much different between these two groups ($P = .10$ and .43, respectively).

Discussion

We are proposing a new analytic method for estimating correlations between members of specified relationships in complex human pedigrees. This method requires correct specification of the mean-, variance-, and correlation-regression models but does not otherwise require assumptions concerning the joint distribution of correlated phenotypes. The estimates of parameters in these regression models have an asymptotic multivariate normal distribution that can be used for statistical inferences.

The estimating equations proposed by Zhao and Prentice (1990) and Prentice and Zhao (1991), which are presumably more efficient than the pairwise method, can be expressed as

$$\sum_{i=1}^{m} X_i' D_i W_i^{-1} F_i = 0 , \qquad (6)$$

where D_i is a derivative matrix of covariances, means, and variances with respect to linear combinations $h_{ijk}'\delta, \dot{x}_{ij}'\alpha$ and $\ddot{x}_{ij}'\beta$ and where W_i is a chosen weight matrix. The choice of this weight matrix does not, in general, alter the consistency of the estimates ($\hat{\delta}, \hat{\alpha}, \hat{\beta}$) but determines the efficiency of estimation. Indeed, the estimates are fully efficient if the weight matrix is proportional to the variance matrix of F_i, which involves knowing the third- and fourth-order moments in addition to the first- and second-order moments (Prentice and Zhao 1991). In other words, for fully efficient estimates the specified third and fourth moments in W_i should equal their true values, which may require unattainable knowledge about the exact joint distributions of relatives' phenotypes. The applications of the more efficient estimating equations (6) to arbitrary and large human pedigrees, however, may be limited. In the estimating equations (6), both D_i and W_i are $(2n_i + p_i) \times (2n_i + p_i)$ size matrices, where $p_i = .5n_i(n_i - 1)$ is the total number of possible pairs in the ith pedigree. Thus, the size of the matrices rapidly increases with the pedigree size n_i. For example, a pedigree (Utah kindred 1764) with 165 individuals may involve a 13,860 × 13,860 matrix W_i. Matrix inversion or even multiplications of matrices can be a formidable task. This limitation, however, does not prohibit the application of equations (6) to pedigree analysis of many but small pedigrees. In fact, in such circumstances the estimating equations (6) may be preferred to equations (5), for the gain of statistical efficiency.

Estimated correlations from the estimating equations are comparable with the Pearson product-moment correlations. In the simplest case with no covariates and with linear correlation and mean- and variance-regression models, an estimated correlation of a type g relationship, from the estimating equations (5), is identical to the corresponding Pearson correlation and is given by

$$\hat{\gamma}_g = \sum_{i=1}^{m} \sum_{(j,k)\in S_{ig}} (y_{ij} - \hat{\mu})(y_{ik} - \hat{\mu}) / \hat{\sigma}^2 ,$$

where

$$\hat{\mu} = \frac{1}{N} \sum_{i=1}^{m} \sum_{j=1}^{n_i} y_{ij};$$

$$N = \sum_{i=1}^{m} n_i;$$

$$\hat{\sigma}^2 = \frac{1}{N} \sum_{i=1}^{m} \sum_{j=1}^{n_i} (y_{ij} - \hat{\mu})^2 .$$

To account for covariates in the means of the responses, the estimated correlation from the estimating equations with covariates in the mean-regression model and with a constant variance is similar to the Pearson correlation of the regression residuals, $(y_i - \hat{\mu}_i)$, where $\hat{\mu}_i$ are the fitted values from the usual regression of responses on the covariates. However, the use of the estimating equations (5) allows one to adjust both the means and variances for covariates and to perform a regression analysis on correlations. This approach yields estimates of asymptotic variances that can be used for statistical inference.

The estimating-equations technique differs from maximum-likelihood estimation in several ways. Maximum-likelihood estimation requires restrictive assumptions about joint distributions. Our estimating-equations approach requires models only for first- and second-order moments. For the use of maximum likelihood, even if one were to assume a distribution which is determined by its first- and second-order moments

(e.g., the multivariate normal distribution), one implicitly would be specifying all higher-order moments. These moment assumptions may be unrealistic and often are not checked in practice. (We note that the almost universal assumption of multivariate normality in correlation analyses stems from the mathematical simplicity that accompanies it.) If the assumptions of these nuisance components fail to hold, the maximum-likelihood estimates may not be consistent. Even if they are consistent in some cases, their standard errors may be quite inaccurate, and the likelihood ratio test can be misleading. Moreover, the computation of maximizing a complicated likelihood function for complex human pedigrees can be a formidable task. In contrast, the pairwise method is straightforward in implementation, can be used for the analysis of pedigrees with either continuous or discrete phenotypes, and does not require making distributional assumptions. The results are easily interpreted, since they are simply estimates of parameters in mean-, variance-, and correlation-regression models. Alternative models can be used to assess the robustness of the correlation estimates.

In exchange for the simpler computations and the robustness in inference is a potential loss in statistical efficiency. First, the method of unweighted estimating equations (5) may not be as efficient as the method of maximum likelihood. For example, an assumed multivariate normal distribution with specified means, variances, and correlations implies all higher-order moments, which are not assumed in the estimating equations. These higher-order moments may carry some relevant information about the correlations, which is captured by the likelihood function but not by the estimating equations. To gain efficiency, one can extend the estimating equations to include those implicitly specified higher-order moments, at the expense of greatly increased computations.

Second, this pairwise method as presented here estimates correlation only for specific pairs of relationships, ignoring those relationships that are not explicitly included in the model. This is both a strength and a weakness. At the cost of assumptions concerning the exact functional relations of different pairwise correlations, information from additional pairs might be pooled into a single estimate. Consider the correlation between grandparents and grandchildren. A variable might be included in the matrix h_i to estimate this correlation without using assumptions concerning its relation to the parent-child correlation; estimated separately, it gives no information about the parent-child correlation. If one were willing to assume some functional relation—e.g., that is equal to the square of the parent-child correlation—the h_i matrix could be suitably modified to allow pooling of information across both pairs of relationships, to estimate one correlation. The degree of efficiency, then, depends on the assumptions that the investigator is willing to bear, and each assumption must be specified explicitly. This additional flexibility to the pairwise method is currently under investigation, and the associated procedure for estimating such parameters will be developed.

This new method was applied to the analysis of nevus-count data in 28 Utah kindreds. The results suggests that nevus counts are correlated between parent and child and between siblings, with correlations .22 and .32, respectively (P's $< .001$). These nonzero correlations could be due to shared genes as well as due to other common environmental factors which have not been accounted for. It appears that the shared adult environment may not contribute much to this familial resemblance: the correlation between husband and wife, who share many aspects of adult environment, is actually negative, although not significantly different from zero.

There are reasons to be cautious in interpreting these results. First of all, the estimates of correlations may be influenced by the method of sampling. As described earlier, the first 12 pedigrees were ascertained through multiple cases of either melanoma or DNS, while the other 16 pedigrees were ascertained through single cases of melanoma. Of the three correlations estimated, it was found that the parent-child correlation from the first 12 pedigrees apparently was significantly greater than that from the remaining 16 pedigrees.

Second, the estimated correlations can also be affected by seriously misspecifying the mean and variance functions. We found that ignoring the covariates related to the mean number of nevus counts biased the parent-child correlation toward zero and biased the sib-sib correlation away from zero. Deliberately misspecifying the variance function had little effect on the estimates of the correlations, although a general conclusion of such is not warranted.

In summary, this newly proposed method allows one to assess patterns of familial resemblance, by estimating specific correlations of interest. The relatively simple computations for estimating correlations allow one to thoroughly explore patterns of familial aggregation in complex human pedigrees. Since the exact joint distribution of phenotypes need not be specified, the investigator is free to concentrate on the more rele-

vant scientific aspects of the problem: the dependence of familial correlations, means, and variances on covariates.

Acknowledgments

The authors would like to thank Dr. David Goldgar for providing the updated version of the 28 Utah kindreds that are used for the illustration, and they would like to thank Dr. Dave Curtis for providing the program PEDRAW, which draws pedigrees; these pedigrees helped us to interpret the data. They also thank the two reviewers for their helpful comments. This work was supported in part by NIH grants CA 55670-01 and PO1 CA 33619.

Appendix

Matrices and Vectors Involved in *i*th Contribution to Estimating Equations (5) from Sample Pedigree in Figure 1

$\dot{x}_i$ is the 6 × 2 ($n_i = 6$ and $\dot{p} = 2$) matrix for response means $E(y_i)$, while $\ddot{x}_i$ is the 6 × 1 ($n_i = 6$ and $\ddot{p} = 1$) matrix for response variances $E(s_i)$:

$$\dot{x}_i = \begin{array}{c} \begin{matrix} \text{Sex} & \text{Age} \end{matrix} \\ \begin{pmatrix} 1 & 48 \\ 0 & 39 \\ 0 & 27 \\ 1 & 25 \\ 0 & 24 \\ 0 & 2 \end{pmatrix} \end{array} \quad \begin{array}{c} \text{Member Index} \\ \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \end{pmatrix} \end{array} \quad \begin{array}{c} \text{Sex} \\ \begin{pmatrix} 1 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \end{pmatrix} \end{array} = \ddot{x}_i .$$

h_i for covariances is a 9 × 4 ($P_i = 9$ and $G_i + 1 = 4$) matrix of pair relationships plus one pedigree covariate:

$$h_i = \begin{array}{c} \begin{matrix} \text{Parent–Child} & \text{Sib–Sib} & \text{Husband–Wife} & \text{Covariate} & \text{Pair Indices} \end{matrix} \\ \begin{pmatrix} 0 & 0 & 1 & 0.2 \\ 1 & 0 & 0 & 0.2 \\ 1 & 0 & 0 & 0.2 \\ 1 & 0 & 0 & 0.2 \\ 1 & 0 & 0 & 0.2 \\ 0 & 1 & 0 & 0.2 \\ 0 & 0 & 1 & 0.2 \\ 1 & 0 & 0 & 0.2 \\ 1 & 0 & 0 & 0.2 \end{pmatrix} \end{array} \begin{array}{l} 1,2 \\ 1,3 \\ 1,4 \\ 2,3 \\ 2,4 \\ 3,4 \\ 4,5 \\ 4,6 \\ 5,6 . \end{array}$$

The vector F_i is a concatenation of a 9 × 1 vector $(z_i - \zeta_i)$, a 6 × 1 vector $(y_i - \mu_i)$, and another 6 × 1 vector $(s_i^2 - \sigma_i^2)$, in which z_i, y_i, and s_i may be expressed as

$$y_i^t = (20,10,2,13,3,2);$$

$$(s_i^2)^t = [(20 - \mu_{i1})^2, (10 - \mu_{i2})^2, (2 - \mu_{i3})^2, (13 - \mu_{i4})^2, (3 - \mu_{i5})^2, (2 - \mu_{i6})^2];$$

$$z_i^t = [(20 - \mu_{i1})(10 - \mu_{i2}), (20 - \mu_{i1})(2 - \mu_{i3}), (20 - \mu_{i1})(13 - \mu_{i4}), (10 - \mu_{i2})(2 - \mu_{i3}), (10 - \mu_{i2})(13 - \mu_{i4}), (2 - \mu_{i3})(13 - \mu_{i4}), (13 - \mu_{i4})(3 - \mu_{i5}), (13 - \mu_{i4})(2 - \mu_{i6}), (3 - \mu_{i5})(2 - \mu_{i6})].$$

The resulting X_i and F_i are a $(G_i + 1 + \dot{p} + \ddot{p}) \times (P_i + 2n_i)$ block-diagonal matrix and a $(P_i + 2n_i) \times 1$ vector, respectively, and may be expressed by

$$X_i = \begin{pmatrix} h_i & 0 & 0 \\ 0 & \dot{x}_i & 0 \\ 0 & 0 & \ddot{x}_i \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} z_i - \zeta_i \\ y_i - \mu_i \\ s_i^2 - \sigma_i^2 \end{pmatrix}.$$

References

Aptech System (1984) The GAUSS system, version 2.0. Aptech System, Kent, WA

Donner A, Eliasziw M (1991) Methodology for inferences concerning familial correlations: a review. J Clin Epidemiol 44:449–455

Donner A, Koval JJ (1981) A multivariate analysis of family data. Am J Epidem 114:149–154

Eliasziw M, Donner A (1990) Comparison of recent estimators of interclass correlation from familial data. Biometrics 46:391–398

Falconer DS (1989) Introduction to quantitative genetics. Ronald, New York

Green A, Swerdlow AJ (1989) Epidemiology of melanocytic nevi. Epidemiol Rev 11:204–221

Karlin S, Cameron EC, Williams PT (1981) Sibling and parent-offspring correlation estimation with variable family size. Proc Natl Acad Sci USA 78:2664–2668

Konish S, Khatri CG (1991) Inferences on multivariate measures of interclass and intraclass correlations in familial data. J R Stat Soc [B] 53:649–659

Liang KY, Zeger SL (1986) Longitudinal data analysis using generalized linear models. Biometrika 73:13–22

Little RJ, Rubin DB (1987) Statistical analysis with missing data. John Wiley, New York

MacCluer JW, Chakravarti A, Cox D, Bishop DT, Bale SJ, Skolnick MH (eds) (1992) Genetic Analysis Workshop 7: "Issues in Gene Mapping and Detection of Major Genes." Cytogenet Cell Genet 59:65–240

McCullagh P, Nelder JA (1989) Generalized linear models, 2d ed. Chapman & Hall, London

Meyer LJ, Piepkorn MW, Seuchter SA, Cannon-Albright LA, Bishop DT, Zone JJ, Skolnick MH (1988) Genetic and epidemiologic evaluation of dysplastic nevi. Pigment Cell Res [Suppl 1]: 144–151

Nelder JA, Wedderburn RWM (1972) Generalized linear models. J R Stat Soc [A] 135:370–384

Prentice RL (1988) Correlated binary regression with covariates specific to each binary observation. Biometrics 44, 1033–1048

Prentice RL, Zhao LP (1991) Estimating equations for parameters in means and covariances of multivariate discrete and continuous responses. Biometrics 47:825–839

Rao DC, Vogler GP, McGue M, Russell JM (1985) Maximum likelihood estimation of familial correlations from quantitative data on pedigrees: a general method and examples. Technique report. Division of Biostatistics, Department of Preventive Medicine and Public Health, Washington University School of Medicine, St Louis

Rosner B, Donner A, Hennekens CH (1977) Estimation of interclass correlation from familial data. Appl Stat 26: 179–187

Shoukri MM, Ward RH (1989) Use of regression models to estimate genetic parameters and measures of familial resemblance in man. Appl Stat 38:467–479

Smith CAB (1980) Estimating genetic correlations. Ann Hum Genet 43:265–284

Srivastava MS, Keen KJ, Katapa RS (1988) Estimation of interclass and intraclass correlations in multivariate familial data. Biometrics 44:141–150

Thomas DC, Fitting genetic data using Gibbs sampling: an application to nevus counts in 38 Utah kindreds. In: MacCluer JW, Chakravarti A, Cox D, Bishop DT, Bale SJ, Skolnick MH (eds) (1992) Genetic Analysis Workshop 7: "Issues in Gene Mapping and Detection of Major Genes." Cytogenet Cell Genet 59:228–230

Wedderburn RWM (1974) Quasilikelihood functions, generalized linear models and the Gauss-Newton method. Biometrika 61:439–447

Zeger SL, Liang KY (1986) Longitudinal data analysis for discrete and continuous outcomes. Biometrics 42:121–130

Zhao LP, Prentice RL (1990) Correlated binary regression using a quadratic exponential model. Biometrika 77:642–648

——— (1991) Use of a quadratic exponential model to generate estimating equations for means, variances, and covariances. In: Godambe VP (ed) Estimating functions. Oxford University Press, London, pp 103–117