WILEY

I B S

Multivariate Methods in Ophthalmology with Application to Other Paired-Data Situations
Author(s): Bernard Rosner
Source: *Biometrics*, Vol. 40, No. 4 (Dec., 1984), pp. 1025-1035
Published by: International Biometric Society
Stable URL: http://www.jstor.org/stable/2531153
Accessed: 11-02-2016 13:02 UTC

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at http://www.jstor.org/page/info/about/policies/terms.jsp

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.

Wiley and International Biometric Society are collaborating with JSTOR to digitize, preserve and extend access to *Biometrics*.

http://www.jstor.org

This content downloaded from 132.77.150.148 on Thu, 11 Feb 2016 13:02:14 UTC
All use subject to JSTOR Terms and Conditions

www.manaraa.com

BIOMETRICS 40, 1025–1035
December 1984

Multivariate Methods in Ophthalmology with Application to Other Paired-Data Situations

Bernard Rosner

Channing Laboratory, Department of Preventive Medicine and Clinical Epidemiology, Harvard Medical School and Brigham and Women's Hospital, 180 Longwood Avenue, Boston, Massachusetts 02115, U.S.A.

SUMMARY

Methods are presented for performing multiple regression analyses and multiple logistic regression analyses on ophthalmologic data with normally and binomially distributed outcome variables, while accounting for the intraclass correlation between eyes. These methods are extended to more general nested data structures where a variable number of subunits are available for each primary unit of analysis, as in familial data. These methods can also be applied to other types of paired data, as in matched studies with a variable matching ratio, where one has a continuous outcome variable and wishes to control for other confounding variables while maintaining the matching. Examples are given of these methods with a group of over 400 patients with retinitis pigmentosa, in which spherical refractive error and visual acuity are related to genetic type after the effects of age, sex and the presence of cataract, have been controlled.

1. Introduction

Methods have been previously described for the analysis of ophthalmologic data, where one has either a single normally distributed or binomially distributed outcome variable and the eye is the basic unit of analysis (Rosner, 1982). It is often the case that one wishes to control for the effects of other covariates while performing such analyses. In ophthalmology, these covariates may be either person-specific or eye-specific in nature. Multiple regression analysis and multiple logistic regression analysis (Cox, 1970) are multivariate methods for normally and binomially distributed outcome variables, respectively, which are commonly employed to achieve these objectives. However, these analytical methods generally require statistical independence for individual sample points. In ophthalmologic work, this condition is frequently not satisfied, since one often uses the eye as the fundamental unit of analysis, and outcome data on individual eyes of the same person are generally highly correlated (Rosner, 1982; Ederer, 1973).

In this paper, we present appropriate extensions of these methods for ophthalmologic data, which allow for either person-specific or eye-specific covariates in the same model, and account for the intraclass correlation between eyes of the same person. These methods have also been extended to the situations (i) where k (>2) units of analysis are provided by each person, as in dental data, and (ii) where data are available for a variable number of subunits for each major unit of analysis, as in the analysis of familial data. Finally, these results have implications for more general paired-data situations, as in the analysis of case–control studies with normally distributed outcome variables, where a variable number of cases and/or controls are available for each pairing, and one wishes to control for other covariates while maintaining the matching.

Key words: Ophthalmology; Multiple regression; Multiple logistic regression; Polychotomous logistic regression; Nested design; Intraclass correlation; Matched pair data; Case–control studies; Familial data; Retinitis pigmentosa; Twin data.

This content downloaded from 132.77.150.148 on Thu, 11 Feb 2016 13:02:14 UTC
All use subject to JSTOR Terms and Conditions

2. Normally Distributed Outcome Variable

2.1 *Multiple Regression Model*

We assume a nested data structure with n primary units of analysis, where within the ith primary unit of analysis there are t_i secondary units of analysis (or subunits), $i = 1, \cdots, n$. For example, for ophthalmologic data, the primary unit of analysis is the person, the secondary units are the eyes, and if there are no missing data, then $t_i = 2$, $i = 1, \ldots, n$. We wish to relate the value of a normally distributed outcome variable y_{ij} for the jth subunit of the ith primary unit ($i = 1, \ldots, n$; $j = 1, \ldots, t_i$) to the values of K independent variables $x_{ij1}, \ldots, x_{ijK}$, where x_{ijk} denotes the kth independent variable for the jth subunit of the ith primary unit ($i = 1, \ldots, n$; $j = 1, \ldots, t_i$; $k = 1, \ldots, K$). We consider the following multiple regression model:

$$y_{ij} = \beta_0 + \sum_{k=1}^{K} \beta_k x_{ijk} + e_{ij}, \qquad i = 1, \ldots, n, \quad j = 1, \ldots, t_i, \tag{2.1}$$

where

$$\text{var}(e_{ij}) = \sigma^2, \quad \rho(e_{ij_1}, e_{ij_2}) = \rho, \qquad i = 1, \ldots, n, \quad j_1 \neq j_2 = 1, \ldots, t_i.$$

If ρ is known, then it follows (Rao, 1967, p. 188) that the maximum likelihood estimator (MLE) of $\boldsymbol{\beta}$ is given by

$$\hat{\boldsymbol{\beta}}(\rho) = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{Y}, \tag{2.2}$$

where $T = \sum_{i=1}^{n} t_i$, $\mathbf{X}$ is a $T \times (K+1)$ matrix $= (\mathbf{X}^{(1)\prime}\mathbf{X}^{(2)\prime}, \ldots, \mathbf{X}^{(n)\prime})'$, $\mathbf{X}^{(i)}$ is a $t_i \times (K+1)$ matrix consisting of predictor variables for the ith primary unit, whereby $X^{(i)}_{j1} = 1$, $X^{(i)}_{jk} = x_{ij,k-1}$, $j = 1, \ldots, t_i$, $k = 2, \ldots, K+1$; $\mathbf{V}^{-1} \equiv \mathbf{W}$ is a $T \times T$ matrix given by

$$\mathbf{W} = \begin{bmatrix} \mathbf{W}^{(1)} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{W}^{(2)} & \cdots & \mathbf{0} \\ \vdots & \mathbf{0} & & \vdots \\ \mathbf{0} & \mathbf{0} & & \mathbf{W}^{(n)} \end{bmatrix},$$

where $W^{(i)}_{jj} = \{1 + (t_i - 2)\rho\}/z_i$, $W^{(i)}_{jk} = -\rho/z_i$, $z_i = (1-\rho)\{1 + (t_i - 1)\rho\}$, $i = 1, \ldots, n$, $j \neq k = 1, \ldots, t_i$; $\mathbf{Y}$ is a $T \times 1$ matrix $= (\mathbf{Y}^{(1)\prime}\mathbf{Y}^{(2)\prime}, \ldots, \mathbf{Y}^{(n)\prime})'$, where $\mathbf{Y}^{(i)}$ is a $t_i \times 1$ matrix consisting of outcome variables for the ith primary unit, whereby $Y^{(i)}_{j1} = y_{ij}$, $i = 1, \ldots, n$, $j = 1, \ldots, t_i$; and $\hat{\boldsymbol{\beta}}(\rho)$ is a $(K+1) \times 1$ matrix $= [\hat{\beta}_0(\rho), \ldots, \hat{\beta}_K(\rho)]'$.

2.2 *Estimation of* $\boldsymbol{\beta}$ *and* ρ

Since ρ is not known, one must use iterative methods to obtain the MLE for $\boldsymbol{\beta}$. In particular, if $\boldsymbol{\beta}$ is known and $t_i = t$, then it follows from Donner and Koval (1980) that the MLE of ρ is given by

$$\hat{\rho}(\boldsymbol{\beta}) = \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j \neq l}}^{t} \sum_{l=1}^{t} y^*_{ij} y^*_{il} \Bigg/ \left\{ (t-1) \sum_{i=1}^{n} \sum_{j=1}^{t} y^{*2}_{ij} \right\}, \tag{2.3}$$

where $y^*_{ij} = y_{ij} - (\beta_0 + \sum_{k=1}^{K} \beta_k x_{ijk})$, $i = 1, \ldots, n$, $j = 1, \ldots, t$. Thus, one can successively alternate between (2.2) and (2.3) to obtain $\hat{\boldsymbol{\beta}}$ and $\hat{\rho}$.

If t_i is not the same for all i, then an explicit solution for the MLE of ρ as a function of $\boldsymbol{\beta}$, as presented for the balanced case in (2.3), does not exist, and Newton–Raphson methods must be used to obtain $\hat{\boldsymbol{\beta}}$ and $\hat{\rho}$.

This content downloaded from 132.77.150.148 on Thu, 11 Feb 2016 13:02:14 UTC
All use subject to JSTOR Terms and Conditions

2.3 *Hypothesis Testing*

If one substitutes the MLE for $\boldsymbol{\beta}$ and ρ into the model in (2.1), then it follows immediately that the log likelihood, ln L, is given by

$$\lambda \equiv -2 \ln L = \text{constant} + T \ln\{(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})'\mathbf{V}^{-1}(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})/T\} + (T - n)\ln(1 - \hat{\rho}) + \sum_{i=1}^{n} \ln\{1 + (t_i - 1)\hat{\rho}\}. \quad (2.4)$$

One can now use asymptotic methods to perform hypothesis tests concerning $\boldsymbol{\beta}$. In particular, if one wishes to test the hypothsis H_0: $\beta_k = 0$, all other $\beta_i \neq 0$ versus H_1: *all* $\beta_i \neq 0$, *then from (2.4) one can obtain* λ *under either* H_0 *or* H_1*; we denote these values by* $\lambda_{K-1}^{(k)}$ *and* λ_K, *respectively. We now compute* $U_k = \lambda_{K-1}^{(k)} - \lambda_K \sim \chi_1^2$ under H_0, and reject if $U_k > \chi_{1,1-\alpha}^2$, the $100(1 - \alpha)$th percentile of a χ_1^2 distribution.

2.4 *Implications for General Paired-Data Situations*

The model in (2.1) was designed primarily for nested data structures such as those naturally occurring in ophthalmology, otolaryngology, twin data, etc. However, such a model can also be applied to general paired-data situations where the outcome variable can be assumed to be normally distributed. In this case, the pairing can be considered as the primary unit of analysis and the individuals within the pairing the secondary units of analysis. Let y_{ij} denote outcome for the jth person in the ith pairing, $i = 1, \ldots, n$ $j = 1, \ldots, t_i$. In particular, if some members of the pairing can be considered as cases and others as controls, then case/control status can be included in the set of predictor variables, where $x_{ijk} = 1(0)$ for a case (control) member of the ith pairing, respectively. Then β_k measures the difference in outcome between cases and controls after adjustment for other potential confounding variables. For example, this would be an appropriate model for examining the relationship between serum cholesterol level and retinitis pigmentosa (RP) after adjustment for potential confounding variables such as age, sex and weight, in a study where the primary sampling unit is the family, and one ascertains a variable number of cases and unaffected siblings (controls) within a given family.

This problem has been considered previously (Rosner and Hennekens, 1978) for paired-data situations with a continuous outcome variable, in the special case of a 1–1 matching ratio. By the use of multiple regression, differences in outcome are modelled as a linear function of differences in potential confounding variables between case and control members of a pair. The regression intercept provides an estimate of the difference in outcome between the case and control members of a pair, after adjustment for potential confounding variables. The model in (2.1) is more general in that (i) outcome is considered explicitly for individual members of a pairing, rather than as a difference score between case and control members of a pair, and (ii) variable numbers of cases and controls are allowed per pairing.

3. Binomially Distributed Outcome Variables

3.1 *Polychotomous Logistic Regression Model for Ophthalmologic Data* ($t_i = 2$)

We restrict our attention in this section to the case where each primary unit of analysis has exactly two subunits, as in the situation encountered for ophthalmologic data with no

This content downloaded from 132.77.150.148 on Thu, 11 Feb 2016 13:02:14 UTC
All use subject to JSTOR Terms and Conditions

missing values. Henceforth, for ease of presentation, we will refer to the primary units of analysis as persons, the subunits as eyes, and will arbitrarily label the first (second) eye as the right (left) eye, respectively.

Let the outcome for each eye be binary, where + (−) represents the diseased (nondiseased) condition, respectively. There are then four possible disease states for each person, ++, +−, −+, and −−, where the first (second) symbol refers to the condition of the right (left) eye, respectively. Let the $(N_P + 2N_E) \times 1$ matrix of independent variables for the ith person be

$$\mathbf{x}_i = (\mathbf{x}_i^{(0)\prime}, \mathbf{x}_i^{(1)\prime}, \mathbf{x}_i^{(2)\prime})', \tag{3.1}$$

where $\mathbf{x}_i^{(0)}$ is an $N_P \times 1$ matrix of person-specific variables, and $\mathbf{x}_i^{(1)}$ ($\mathbf{x}_i^{(2)}$) are $N_E \times 1$ matrices of right (left) eye-specific variables, respectively. We also assume that the components of the right and left eye-specific matrices correspond to the same type of variables; for example $x_{i1}^{(1)}$ and $x_{i1}^{(2)}$ might correspond to visual acuity for the right and left eyes, $x_{i2}^{(1)}$ and $x_{i2}^{(2)}$ to intraocular pressure, and so on.

Let us now consider an individual with $\mathbf{x} = \mathbf{x}_0 \equiv (\mathbf{0}_{N_P}', \mathbf{0}_{N_E}', \mathbf{0}_{N_E}')$, where $\mathbf{0}_c$ is a $c \times 1$ vector of zeros. We use the beta-binomial distribution to model the probability of the four disease states (++, +−, −+, −−) for such an individual. Specifically, we assume that this individual has probability p of having any particular eye affected, where p follows a beta distribution with parameters a and b over the class of all individuals with covariates $\mathbf{x}_0$. It follows immediately that the probability distribution of disease states for such an individual is given by

$$\begin{aligned} P(++ \mid \mathbf{x}_0) &= (a + 1)a/\{(a + b + 1)(a + b)\}, \\ P(+- \mid \mathbf{x}_0) &= P(-+ \mid \mathbf{x}_0) = ab/\{(a + b + 1)(a + b)\}, \\ P(-- \mid \mathbf{x}_0) &= (b + 1)b/\{(a + b + 1)(a + b)\}. \end{aligned}$$

We now wish to specify the probability distribution of disease states for an individual with arbitrary covariates $\mathbf{x} = (\mathbf{x}^{(0)\prime}, \mathbf{x}^{(1)\prime}, \mathbf{x}^{(2)\prime})'$. Let $\alpha_1 = \ln\{a/(b + 1)\}$, $\alpha_2 = \ln[(a + 1)a/\{(b + 1)b\}]$, z_R (z_L) be indicator variables set to 1 if the right (left) eye is affected and 0 otherwise, and p_R (p_L) be the conditional probability that the right (left) eye is affected given the outcome status of the left (right) eye. We use a logistic model to relate p_R (p_L) to $\mathbf{x}$, z_L (z_R):

$$\begin{aligned} \ln\{p_R/(1 - p_R)\} &= \alpha_1 + (\alpha_2 - 2\alpha_1)z_L + \boldsymbol{\beta}\mathbf{x}^{(0)} + \boldsymbol{\gamma}\mathbf{x}^{(1)}, \\ \ln\{p_L/(1 - p_L)\} &= \alpha_1 + (\alpha_2 - 2\alpha_1)z_R + \boldsymbol{\beta}\mathbf{x}^{(0)} + \boldsymbol{\gamma}\mathbf{x}^{(2)}. \end{aligned} \tag{3.2}$$

It follows directly from (3.2) that

$$\begin{aligned} P(++ \mid \mathbf{x})/P(-+ \mid \mathbf{x}) &= \exp(\alpha_2 - \alpha_1 + \boldsymbol{\beta}\mathbf{x}^{(0)} + \boldsymbol{\gamma}\mathbf{x}^{(1)}), \\ P(+- \mid \mathbf{x})/P(-- \mid \mathbf{x}) &= \exp(\alpha_1 + \boldsymbol{\beta}\mathbf{x}^{(0)} + \boldsymbol{\gamma}\mathbf{x}^{(1)}), \\ P(-+ \mid \mathbf{x})/P(-- \mid \mathbf{x}) &= \exp(\alpha_1 + \boldsymbol{\beta}\mathbf{x}^{(0)} + \boldsymbol{\gamma}\mathbf{x}^{(2)}), \\ P(++ \mid \mathbf{x})/P(-- \mid \mathbf{x}) &= \{P(++ \mid \mathbf{x})/P(-+ \mid \mathbf{x})\}\{P(-+ \mid \mathbf{x})/P(-- \mid \mathbf{x})\} \\ &= \exp(\alpha_2 + 2\boldsymbol{\beta}\mathbf{x}^{(0)} + \boldsymbol{\gamma}\mathbf{x}^{(1)} + \boldsymbol{\gamma}\mathbf{x}^{(2)}). \end{aligned}$$

In summary, one has the following polychotomous logistic regression model

$$P(j \mid \mathbf{x}_i) = \exp(\boldsymbol{\theta}_j\mathbf{x}_i^*) \Big/ \sum_{k=1}^{4} \exp(\boldsymbol{\theta}_k\mathbf{x}_i^*), \tag{3.3}$$

This content downloaded from 132.77.150.148 on Thu, 11 Feb 2016 13:02:14 UTC
All use subject to JSTOR Terms and Conditions

where $\mathbf{x}_i^* = (1, \mathbf{x}_i^{(0)\prime}, \mathbf{x}_i^{(1)\prime}, \mathbf{x}_i^{(2)\prime})'$, $\boldsymbol{\theta}$ is a $4 \times (N_P + 2N_E + 1)$ matrix such that

$$\boldsymbol{\theta} = \begin{bmatrix} \boldsymbol{\theta}_1 \\ \boldsymbol{\theta}_2 \\ \boldsymbol{\theta}_3 \\ \boldsymbol{\theta}_4 \end{bmatrix} \equiv \begin{bmatrix} \alpha_2 & 2\boldsymbol{\beta} & \boldsymbol{\gamma} & \boldsymbol{\gamma} \\ \alpha_1 & \boldsymbol{\beta} & \boldsymbol{\gamma} & \mathbf{0}'_{N_E} \\ \alpha_1 & \boldsymbol{\beta} & \mathbf{0}'_{N_E} & \boldsymbol{\gamma} \\ 0 & \mathbf{0}'_{N_P} & \mathbf{0}'_{N_E} & \mathbf{0}'_{N_E} \end{bmatrix},$$

$\boldsymbol{\beta}$ is a row vector of length N_P of person-specific regression coefficients, $\boldsymbol{\gamma}$ is a row vector of length N_E of eye-specific regression coefficients which are assumed to be the same for the right and left eye, $\mathbf{0}_c$ is a $c \times 1$ vector of zeros, and $j = 1, \ldots, 4$ correspond to the disease states ++, +−, −+, and −−.

We note from (3.2) that β_p has an odds-ratio interpretation similar to that given by ordinary logistic regression. In particular, from (3.2), if $x_p^{(0)}$ is a binary exposure variable such that $x_p^{(0)} = 1$ (0) for exposed (unexposed), then $\exp(\beta_p)$ is the ratio of odds in favor of disease for a particular eye for the exposed versus the unexposed, all other variables held constant, including the disease status of the fellow eye. Furthermore, γ_q has a similar interpretation, with either $x_q^{(1)}$ or $x_q^{(2)}$ replacing $x_p^{(0)}$. Finally, we define the 'pairwise odds ratio' (OR) as

$$\text{OR} = \{P(++ \mid \mathbf{x})/P(-+ \mid \mathbf{x})\}/\{P(+- \mid \mathbf{x})/P(-- \mid \mathbf{x})\}.$$

It can be easily seen from (3.2) that $\text{OR} = \exp(\alpha_2 - 2\alpha_1) = \{(a+1)(b+1)\}/(ab)$. Thus, OR provides a measure of the dependence between eyes for binary outcome variables. We estimate the parameters in (3.3) by the method of maximum likelihood where the likelihood $L = \prod_{i=1}^{n} P(j \mid \mathbf{x}_i)$. No closed-form expressions exist for the parameter estimates, and thus an iterative procedure based on the Newton–Raphson method may be employed.

We note from (3.2) that OR is assumed to be the same for all primary sampling units, and in particular, is independent of $\mathbf{x}^{(0)}$. This assumption can be relaxed by adding interaction terms of the form $\boldsymbol{\beta}^*\mathbf{x}^{(0)}z_L$ ($\boldsymbol{\beta}^*\mathbf{x}^{(0)}z_R$) to the right-hand side of the first and second equations in (3.2), respectively. It then follows that $\boldsymbol{\theta}_1 = (\alpha_2, 2\boldsymbol{\beta} + \boldsymbol{\beta}^*, \boldsymbol{\gamma}, \boldsymbol{\gamma})$, while $\boldsymbol{\theta}_i$ ($i = 2, 3, 4$) remain unchanged in (3.3). One can then test H_0: $\boldsymbol{\beta}^* = \mathbf{0}'_{N_P}$ versus H_1: $\boldsymbol{\beta}^* \neq \mathbf{0}'_{N_P}$, after controlling for other variables in the model, by estimating the parameters of the augmented model by the method of maximum likelihood and comparing likelihoods under H_0 and H_1 using asymptotic methods.

3.2 *Extension to Unbalanced Designs*

We now consider the general case, where the ith primary unit of analysis has t_i subunits, $i = 1, \ldots, n$. Let $y_{ij} = 1$ if the jth subunit of the ith primary unit is affected, and 0 otherwise. Let the $(N_P + t_i N_E) \times 1$ matrix of independent variables for the ith primary unit be $\mathbf{x}_i = (\mathbf{x}_i^{(0)\prime}, \mathbf{x}_i^{(1)\prime}, \ldots, \mathbf{x}_i^{(t_i)\prime})'$, where $\mathbf{x}_i^{(0)}$ is an $N_P \times 1$ matrix of person-specific variables and $\mathbf{x}_i^{(1)}, \ldots, \mathbf{x}_i^{(t_i)}$ are $N_E \times 1$ matrices of subunit-specific variables. Suppose that each subunit in the ith primary unit has probability p of being affected, where p is assumed to follow a beta distribution.

We again first consider an individual with $\mathbf{x} = \mathbf{x}_0 \equiv (\mathbf{0}'_P, \mathbf{0}'_E, \ldots, \mathbf{0}'_E)'$. Define $a_0 = b_0 = 1$, $a_k = \prod_{j=0}^{k-1}(a+j)$, $b_k = \prod_{j=0}^{k-1}(b+j)$, $k \geq 1$, $y_{i\cdot} = \sum y_{ij}$. It follows immediately from the beta-binomial distribution that $P(y_{i\cdot} = k \mid \mathbf{x}_0)/P(y_{i\cdot} = 0 \mid \mathbf{x}_0) = a_k b_{t_i-k}/b_{t_i}$. We now wish to specify the probability distribution of disease states for an individual with arbitrary covariates $\mathbf{x}_i$. Suppose there are s_i affected subunits in the ith primary unit ($s_i > 0$) and let $J = \{j_1, \ldots, j_{s_i}\}$ denote the ordered set of affected subunits. Let $\mathbf{y}_i^{(0)}, \ldots, \mathbf{y}_i^{(s_i)}$ be column vectors of length t_i such that $\mathbf{y}_i^{(0)} = \mathbf{y}_i$, $\mathbf{y}_{ij}^{(m)} = \mathbf{y}_{ij}^{(m-1)}$ for all $j \neq j_m$, $\mathbf{y}_{ij_m}^{(m)} = 0$, $m = 1, \ldots, s_i$. Thus, $\mathbf{y}_i^{(1)}$ is formed by changing the status of the lowest-order affected subunit in $\mathbf{y}_i^{(0)}$ to

This content downloaded from 132.77.150.148 on Thu, 11 Feb 2016 13:02:14 UTC
All use subject to JSTOR Terms and Conditions

unaffected, and similarly for $\mathbf{y}_i^{(2)}, \ldots, \mathbf{y}_i^{(s_i)}$, where $\mathbf{y}_i^{(s_i)} = \mathbf{0}_{t_i}$. We now express $P(\mathbf{y}_i \mid \mathbf{x}_i)/P(\mathbf{0}_{t_i} \mid \mathbf{x}_i)$ as $\prod_{m=0}^{s_i-1} P(\mathbf{y}_i^{(m)} \mid \mathbf{x}_i)/P(\mathbf{y}_i^{(m+1)} \mid \mathbf{x}_i)$ and assume a logistic model similar to that given in (3.2) for the conditional distribution of the j_{m+1}th subunit given the response status of the other $t_i - 1$ subunits, whereby

$$P(\mathbf{y}_i^{(m)} \mid \mathbf{x}_i)/P(\mathbf{y}_i^{(m+1)} \mid \mathbf{x}_i) = \exp(\alpha^{(m)} + \boldsymbol{\beta}\mathbf{x}_i^{(0)} + \boldsymbol{\gamma}\mathbf{x}_i^{(j_{m+1})}), \qquad m = 0, \ldots, s_i - 1,$$

and

$$\alpha^{(m)} = \ln\{(a + s_i - m - 1)/(b + t_i - s_i + m)\}$$

from the beta-binomial distribution. Upon collecting terms we then have

$$\frac{P(\mathbf{y}_i \mid \mathbf{x}_i)}{P(\mathbf{0}_{t_i} \mid \mathbf{x}_i)} = \left(\frac{a_{s_i} b_{t_i - s_i}}{b_{t_i}}\right) \exp\left\{\sum_{j=1}^{t_i} y_{ij}(\boldsymbol{\beta}\mathbf{x}_i^{(0)} + \boldsymbol{\gamma}\mathbf{x}_i^{(j)})\right\}, \tag{3.4}$$

or the equivalent polychotomous logistic regression model

$$P(\mathbf{y}_i \mid \mathbf{x}_i) = \frac{a_{s_i} b_{t_i - s_i} \exp\left\{\sum_{j=1}^{t_i} y_{ij}(\boldsymbol{\beta}\mathbf{x}_i^{(0)} + \boldsymbol{\gamma}\mathbf{x}_i^{(j)})\right\}}{\sum_{\mathbf{z}_i} a_{\sum z_{ij}} b_{t_i - \sum z_{ij}} \exp\left\{\sum_{j=1}^{t_i} z_{ij}(\boldsymbol{\beta}\mathbf{x}_i^{(0)} + \boldsymbol{\gamma}\mathbf{x}_i^{(j)})\right\}}, \tag{3.5}$$

where the summation in the denominator is over all possible permutations $\mathbf{z}_i = (z_{i1}, \ldots, z_{it_i})$ of 0s and 1s. The likelihood of the entire sample is then given by $L = \prod_{i=1}^{n} P(\mathbf{y}_i \mid \mathbf{x}_i)$ and standard numerical methods can be used to obtain maximum likelihood estimates for a, b, $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$, although the computations become cumbersome for large t_i. However, if only person-specific covariates are considered in the model, then the computations can be reduced considerably since, in this case, the summation in the denominator of (3.5) is with respect to $\sum z_{ij}$ rather than $\mathbf{z}_i$. Finally, we can use asymptotic methods similar to those considered in §2.3 to perform hypothesis tests concerning $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ for the models considered in (3.3) and (3.5).

4. Examples

We now present examples of the use of these methods on a data set obtained from an outpatient population of 456 persons with retinitis pigmentosa (RP), aged 6 to 80, who were seen at the Massachusetts Eye and Ear Infirmary from 1970 to 1979. The patients were classified on the basis of a detailed family history into the genetic types of autosomal dominant RP (DOM), autosomal recessive RP (AR), sex-linked RP (SL), and isolate RP (ISO) for a study of differences between these four groups on certain measurements made during a routine ocular examination. In order to simplify the analysis, only one person was selected from each family, and if more than one affected person was available for analysis, then a randomly selected affected person was chosen. Thus, the 456 persons were from 456 unique families. The details of the design of the study and the procedures for genetic classification are given by Berson, Rosner and Simonoff (1980).

We first present an analysis of the differences in spherical refractive error between genetic types using the multiple regression model in (2.1). For this analysis, genetic type was represented in the form of three indicator variables for the groups DOM, AR and SL, respectively. Two other person-specific variables (age and sex), and two eye-specific variables (the presence of a posterior subcapsular cataract on slit-lamp examination separately for each eye, and an indicator variable distinguishing between the right and left eye), were also

This content downloaded from 132.77.150.148 on Thu, 11 Feb 2016 13:02:14 UTC
All use subject to JSTOR Terms and Conditions

considered in the analysis as potential confounding variables. The sample used for this analysis consisted of the subgroup of 427 persons who had complete information on spherical refractive error and all the above predictor variables. Data were collected on spherical refractive errors based on retinoscopy following dilation and cycloplegia with 10% phenylephrine HCL and 1% cyclopentolate HCL. The results of the multiple regression analysis are presented in Table 1.

We see that there are significant differences between the refractive error of eyes of persons from the four genetic types ($P < .001$) after controlling for the effects of age, sex and the presence of cataracts. In addition, males ($P = .050$) and eyes with posterior subcapsular

Table 1

Results of multiple regression model comparing spherical refractive error (diopters) in different genetic types after controlling for age, sex, the presence of a posterior subcapsular cataract, and the right versus left eye

(a) Estimates of regression parameters and tests of significance of overall effects

Variable	Regression coefficient	Standard error	Chi square*	df	*P*-value	Intraclass correlation	Residual variance
						0.943	9.966
Constant	0.447	0.433					
Genetic type			24.44	3	<.001		
DOM†	0.875	0.531					
AR†	0.319	0.442					
SL†	−2.898	0.646					
Age	−0.011	0.010	1.23	1	NS		
Sex‡	−0.623	0.317	3.84	1	.050		
Presence of posterior subcapsular cataract§	−0.410	0.182	5.40	1	.020		
Right eye**	−0.012	0.049	0.05	1	NS		

(b) Comparison of specific genetic types: mean difference †† ± SE (with *P*-value)

Group, i_1	Comparison group, i_2			
	DOM	AR	SL	ISO
DOM	—	0.556 ± 0.642 (NS)	3.773 ± 0.799 ($P < .001$)	0.875 ± 0.531 (NS)
AR		—	3.217 ± 0.750 ($P < .001$)	0.319 ± 0.442 (NS)
SL			—	−2.898 ± 0.646 ($P < .001$)

* Chi square statistics are given by $-2 \log L_2 + 2 \log L_1$, where L_1 (L_2) are the likelihoods under, respectively, (i) the full model and (ii) the reduced model obtained by deleting the variable(s) corresponding to the effect of interest from the full model.

† Coded as 1 if person has this genetic type and 0 otherwise.

‡ Coded as 1 if male and 0 if female.

§ Coded for each eye as 1 if yes and 0 if no.

** Coded as 1 if right eye and 0 if left eye.

†† Mean difference = $\beta_{i_1} - \beta_{i_2}$; SE($\beta_{i_1} - \beta_{i_2}$) computed from the asymptotic variance–covariance matrix of the regression parameters obtained from the Fisher information matrix; *P*-value obtained by comparing $(\beta_{i_1} - \beta_{i_2})/\text{SE}(\beta_{i_1} - \beta_{i_2})$ to an N(0, 1) distribution.

This content downloaded from 132.77.150.148 on Thu, 11 Feb 2016 13:02:14 UTC
All use subject to JSTOR Terms and Conditions

cataracts ($P = .020$) were significantly more likely to have myopic refractive errors, while no significant effect of age on refractive error was apparent in this series. Finally, spherical refractive error was not significantly different between the right and left eye. We now look in more detail for differences in refractive error between the specific genetic types, using the information matrix to obtain asymptotic standard errors of linear contrasts of the regression coefficients. We find that there are significant differences between the sex-linked group and each of the other groups ($P < .001$), while there are no significant differences between the DOM, AR and ISO groups. We note that the estimated correlation between spherical refractive error of the two eyes of an individual is .943.

Table 2

Results of polychotomous logistic regression model comparing visual acuity in different genetic types after controlling for age, sex, the presence of a posterior subcapsular cataract, and the right versus left eye*

(a) Estimates of regression parameters and tests of significance of overall effects

	Variable	Regression coefficient	Standard error	Chi square†	df	*P*-value	Odds ratio between eyes
							26.4
	α_1	−2.164	0.230				
	α_2	−1.056	0.389				
β	Genetic type			12.38	3	.006	
	DOM‡	−0.334	0.186				
	AR‡	0.207	0.157				
	SL‡	0.601	0.252				
	Age	0.0098	0.0037	7.26	1	.007	
	Sex§	0.041	0.111	0.14	1	NS	
γ	Presence of posterior subcapsular cataract**	0.277	0.118	5.57	1	.018	
	Right eye††	0.055	0.239	0.05	1	NS	

(b) Comparison of specific genetic types: odds ratio‡‡ (with *P*-value)

Group, i_1	Comparison group, i_2			
	DOM	AR	SL	ISO
DOM	—	0.58 ($P = .017$)	0.39 ($P = .002$)	0.72 (NS)
AR		—	0.67 (NS)	1.23 (NS)
SL			—	1.82 ($P = .017$)

* An eye is defined as affected if visual acuity is 20/50 or worse and normal if 20/40 or better.

† Chi square statistics ar given by $-2 \log L_2 + 2 \log L_1$, where L_1 (L_2) are the likelihoods under, respectively, (i) the full model and (ii) the reduced model obtained by deleting the variable(s) corresponding to the effect of interest from the full model.

‡ Coded as 1 if person has this genetic type and 0 otherwise.

§ Coded as 1 if male and 0 if female.

** Coded for each eye as 1 if yes and 0 if no.

†† Coded as 1 if right eye and 0 if left eye.

‡‡ Ratio of odds in favor of an eye being affected for a person in Group i_1 as compared with a person in Group i_2, all other factors being the same (including the acuity status of the opposite eye).

This content downloaded from 132.77.150.148 on Thu, 11 Feb 2016 13:02:14 UTC
All use subject to JSTOR Terms and Conditions

We next present an analysis of the differences in best-corrected Snellen visual acuity (VA) between genetic types, using the polychotomous logistic regression model in (3.3) with the same set of predictor variables as in Table 1. For this purpose, an eye was considered affected if VA was 20/50 or worse and normal if VA was 20/40 or better. A total of 444 persons with complete data on visual acuity and all predictor variables were used in the analysis; the results are presented in Table 2.

We see that there are significant differences between the VA of eyes of persons from the four genetic types ($P = .006$) after controlling for the effects of age, sex and the presence of cataracts. In addition, age ($P = .007$) and the presence of a posterior subcapsular cataract ($P = .018$) both significantly increased the probability of an eye having VA of 20/50 or worse, while sex had no significant effect on acuity. Finally, VA was not significantly different for the right and left eye. In order to assess more specific differences in VA, for each pair of genetic types, we computed the ratio of odds in favor of reduced acuity in a particular eye, for one genetic type versus the other, after controlling for other risk factors and the acuity status of the fellow eye. It follows immediately from (3.2) that, for a comparison of the pth genetic type with ISO ($p \leqslant 3$), the odds ratio is given by $\exp(\beta_p)$, while for a comparison of the p_1th and p_2th genetic type, $1 \leqslant p_1, p_2 \leqslant 3$, it is given by $\exp(\beta_{p_1} - \beta_{p_2})$ where the standard error of the latter odds ratio is obtained from the information matrix. In each case, asymptotic normality of ln(odds ratio)/SE{ln(odds ratio)} is used to provide a basis for significance testing. We see that eyes of sex-linked persons were significantly more likely to have reduced acuity than eyes of DOM and ISO persons (DOM versus SL, odds ratio = .39, $P = .002$; SL versus ISO, odds ratio = 1.82, $P = .017$). The only other significant difference was found between eyes from DOM and AR persons (DOM versus AR, odds ratio = .58, $P = .017$). As was the case for spherical refractive error, a strong association was found between the presence of reduced acuity in two eyes of the same person. The estimates of the parameters of the beta distribution were $a = .162$, $b = .373$, with pairwise odds ratio, $\text{OR} = (a + 1)(b + 1)/(ab) = 26.4$ after adjusting for the other variables considered in Table 2.

5. Discussion

We have presented multivariate models for normally and binomially distributed outcome variables for ophthalmologic data which allow one to look simultaneously at the effects of person- and eye-specific predictor variables while accounting for the intraclass correlation between eyes. These methods are also directly applicable to other special areas, such as for twin data or otolaryngological data. Furthermore, these methods have also been extended to the situation where a variable number of subunits are available for each primary unit of analysis, as in familial data. In the binomial case, these methods are extensions of the beta-binomial model used in the analysis of binary response data from toxicological experiments (Williams, 1975; Haseman and Kupper, 1979) in that other covariates of both a person-specific and eye-specific nature are controlled for explicitly in the analysis.

Other possible methods of multivariate analysis for the designs discussed in this paper include (a) treating each eye as an independent random variable and performing analyses over all eyes, (b) performing separate analyses for the left and right eye, and comparing results (Ederer, 1973), and (c) representing all variables on a person-specific basis, and performing standard analyses. The disadvantage of the first two methods have been discussed previously (Rosner, 1982). In particular, Method (a) is generally invalid, resulting in greatly exaggerated levels of significance. This point is emphasized in the examples considered in the present paper, where (i) the intraclass correlation between eyes for spherical refractive error was estimated as .943, and (ii) the pairwise odds ratio between

This content downloaded from 132.77.150.148 on Thu, 11 Feb 2016 13:02:14 UTC
All use subject to JSTOR Terms and Conditions

eyes for reduced visual acuity (VA 20/50 or worse) was estimated as 26.4. Method (b) is valid, but may be inefficient, particularly for outcome variables which have lower correlations between eyes than the variables considered in this paper. In particular, it is possible that analyses of separate eyes would both yield nonsignificant results (or one eye-specific analysis would yield significant results while the other would not), while an analysis appropriately combining evidence over individual eyes would result in overall significant results. Method (c) is also likely to be inefficient, particularly for eye-specific predictor variables, where it is desirable to relate findings for outcome and predictor variables on the same eye.

The findings in this paper, while intended mainly for nested data structures as in ophthalmologic data, also have implications for more general paired-data situations. In particular, for matched-pair studies with variable numbers of cases and controls per pairing, the model in (2.1) can be used to compare cases and controls on continuous outcome variables in the presence of other confounding variables, while maintaining the matching. This is a generalization of previous work (Rosner and Hennekens, 1978) which only permitted such analyses for matched studies with a 1–1 matching ratio.

Computer programs to implement the methods in this paper have been written in FORTRAN IV and are available upon request from the author.

Acknowledgements

The author wises to thank Dr Eliot L. Berson, Department of Ophthalmology, Harvard Medical School, Massachusetts Eye and Ear Infirmary, Boston, for permission to use information from his retinitis pigmentosa outpatient population data base for the examples in this paper. I also wish to acknowledge Mr Martin Van Denburgh for programming assistance in implementing the algorithms described in this paper.

Résumé

Cet article présente des méthodes de régression multiple et de régression multiple logistique dans le contexte de variables ophthalmologiques aussi bien normales que binomiales avec prise en compte du coefficient de corrélation intra-classes entre les deux yeux. Ces méthodes sont étendues au cas plus général de schémas emboités où il existe un nombre variable de sous unités pour chaque unité primaire (exemple des données familiales). Elles peuvent être appliquées à d'autres types de données appariées telles que celles où la réponse est continue et on veut contrôler d'autres facteurs de confusion tout en maintenant l'appariement. Des exemples de ces méthodes sont donnés dans un groupe de plus de 400 malades avec pigmentation rétinienne pour lesquels l'erreur de réfraction sphérique et l'acuité visuelle sont reliés à des marqueurs génétiques après contrôle de l'âge, du sexe, et de la présence de cataracte.

References

Berson, E. L., Rosner, B. and Simonoff, E. (1980). Risk factors for genetic typing and detection in retinitis pigmentosa. *American Journal of Ophthalmology* **89,** 763–775.

Cox, D. R. (1970). *Analysis of Binary Data.* London: Methuen.

Donner, A. and Koval, J. J. (1980). The estimation of intraclass correlation in the analysis of familial data. *Biometrics* **36,** 19–25.

Ederer, F. (1973). Shall we count numbers of eyes or numbers of subjects? *Archives of Ophthalmology* **89,** 1–2.

Haseman, J. K. and Kupper, L. L. (1979). Analysis of dichotomous response data from certain toxocological experiments. *Biometrics* **35,** 281–293.

Rao, C. R. (1967). *Linear Statistical Inference and its Applications.* Wiley: New York.

Rosner, B. and Hennekens, C. H. (1978). Analytic methods in matched pair epidemiologic studies. *International Journal of Epidemiology* **7,** 367–372.

This content downloaded from 132.77.150.148 on Thu, 11 Feb 2016 13:02:14 UTC
All use subject to JSTOR Terms and Conditions

Rosner, B. (1982). Statistical methods in ophthalmology: an adjustment for the intraclass correlation between eyes. *Biometrics* **38,** 105–114.
Williams, D. A. (1975). The analysis of binary responses from toxicological experiments involving reproduction and teratogenicity. *Biometrics* **31,** 949–952.

Received July 1983; *revised December* 1983 *and April* 1984

This content downloaded from 132.77.150.148 on Thu, 11 Feb 2016 13:02:14 UTC
All use subject to JSTOR Terms and Conditions