11-1-1989

Regression Models for Bivariate Binary Responses

Juni Palmgren
National Public Health Institute, Finland, juni@math.su.se

1. Introduction

Analysis of binary observations from nested designs has been the focus of much recent interest. Simple one level nesting of subunits within blocks is commonly encountered in empirical studies. Two or more observations made on the same individual serves as a simple example. The observations may be on different outcome variables or they may be repeated ordered or unordered observations on the same variable. It is natural to view the outcomes within each block as a multivariate binary response and assume independence between blocks. The nature of the outcome variables and the empirical problem guides the choice of parametrisation for the multinomial block probabilities. For modelling effects of covariates on marginal response probabilities allowance has to be made for the fact that observations within a block are likely to be more similar than observations between blocks. And a decision has to be made whether the regression parameters should reflect effects of covariates on the individual response probabilities or on the group prevalence (Zeger, Liang and Albert, 1988).

When modelling individual response probabilities the regression model is specified conditionally on observations belonging to a block, and the form for the heterogeneity between blocks is given. Williams (1982), Stiratelli, Laird and Ware (1984) and Laird (1989) capture block heterogeneity by specifying some parameters as random variables. Williams deals with repeated unordered observations and block specific covariates, which reduces the problem to that of overdispersion relative to the binomial for the block totals. For the special case where covariate values vary within blocks, but take the same values across blocks, it is appropriate to use a conditional analysis (Breslow and Day, 1980, Chapter 7). Conditioning on the block totals blots out the between block variation, which in this case does not contribute information about the regression parameters.

When modelling group prevalence, the marginal probabilities constitute averages over blocks and the similarity between units within a block is modelled by specifying a within block dependence structure. McCullagh and Nelder (1989) introduce a general multivariate logistic model specifying logistic transforms of the marginal response probabilties combined with a general log odds ratio structure for the association. Their model is flexible and does not a priori assume absense of higher order interactions. A major drawback is that although a one to one relationship may exist between the multinomial probabilites in the joint distribution and the model parameters, there does not in general exist an explicit analytic expression for the former set of parameters in terms of the latter. If such analytic expressions were available, then full maximum likelihood inference for the model parameters would be feasible, even if the relationship had a complicated nonlinear form.

Liang and Zeger (1986) and Zeger and Liang (1986) use a set of generalized estimating equations to estimate parameters in the marginal logits. Instead of specifying a full distribution for the multivariate binary response within a block, they make assumptions about means, variances and correlations. When correlations are used to describe dependence between binary variables, the value of the coefficient is constrained, in order for the probabilities in the joint distribution to remain in the unit interval. Liang and Zeger use simple correlation structures without constraints, thus setting up generalized estimating equations or quasi likelihood equations that in general do not correspond to a probability distribution. They show, however, that even for misspecified correlation structure the estimates of the marginal logit parameters are consistent and asymptotically normally distributed. They base inference about the parameters in the marginal logits on a Wald type statistic, using an estimator for the parameter covariance matrix, which is robust against misspecification of the association structure. Prentice (1988), Sharples (1989) and Zhao and Prentice (1989) discuss extensions of Liang's and Zeger's estimating equations, which allow more efficient estimation of the correlation parameters.

In this paper we focus on regression models for bivariate binary responses. When only one or two subunits are observed within each block, the multinomial response probabilities can be expressed explicitly in terms of the parameters in the marginal logits and the log odds ratio (Dale 1986, Palmgren 1989). The multinomial response probabilities can alternatively be expressed in terms of the marginal logits and the correlation coefficient. We discuss maximum likelihood estimation using the IRLS algorithm in GLIM for fitting exponential family nonlinear models (Ekholm et al, 1986). The nonlinear model formulation is flexible and the algorithm is fast, whereby the computational difficulties (McCullagh, 1989) in obtaining maximum likelihood estimates for this bivariate logistic model are largely overcome.

Our model differs from the bivariate binary regression model discussed by Rosner (1984). He deals with effects of covariates on individual response probabilities, and he accounts for block heterogeneity by assuming a beta-binomial distribution. He also uses a log linear parametrisation. McCullagh and Nelder (1989, Chapter 6) give a comparison of the interpretation of the log linear parameters and the parameters in the bivariate logistic model.

We illustrate use of the bivariate logistic model and maximum likelihood estimation on two data set. For the classical British coalminers data, marginal logits of breathlessness and wheeze as functions of age are shown to be highly insensitive to prespecified values of the log odds ratio. Corresponding stability for the marginal regression parameters is not present if the association is modelled using the correlation coefficient. In the second example we model discrete time survival of total hip arthroplasties as function of patient specific and hip specific covariates. One third of the arthroplasties were bilateral, and the bivariate logistic models provides an opportunity to formulate and compare survival for bilateral and unilateral prostheses.

In both examples we also look at inference about the regression parameters in the marginal logits when the assumption of independence within blocks is used. In Section 5 we use simulated data to evaluate the performance of estimates and standard errors under independence. Liang's and Zeger's robust version of the standard errors appear to make allowance for the misspecified association when the covariates are block specific, but not when they are subunit specific.

2. The Bivariate Logistic Model

Let $Y = (Y_1, Y_2)$ denote a bivariate binary response with distribution given by the four multinomial probabilities $\pi_{kl} = pr(Y_1 = k, Y_2 = l)$, $k, l = 0, 1$. Let $\pi_{1.} = \pi_{11} + \pi_{10}$, $\pi_{.1} = \pi_{11} + \pi_{01}$ and $\psi = \pi_{11}\pi_{00}/(\pi_{10}\pi_{01})$ denote the marginal probabilities and the odds ratio. The probability π_{11} is expressed in terms of $\pi_{1.}$, $\pi_{.1}$ and ψ in the following way:

$$\pi_{11} = \begin{cases} \frac{1}{2}(\psi - 1)^{-1}\{a - \sqrt{a^2 + b}\} & \text{if } \psi \neq 1 \\ \pi_{1.}\pi_{.1} & \text{if } \psi = 1, \end{cases} \tag{1}$$

with $a = 1 + (\pi_{1.} + \pi_{.1})(\psi - 1)$ and $b = -4\psi(\psi - 1)\pi_{1.}\pi_{.1}$ (Dale, 1986; Palmgren, 1989). The other multinomial probabilities follow from the margins. The bivariate logistic model is specified by expressing $logit\,\pi_{1.}$, $logit\,\pi_{.1}$ and $log\,\psi$ as linear predictors $\beta_1 x$, $\beta_2 x$ and γx, respectively. The covariate vector x may include block specific and subunit specific covariates. For covariate values associated with Y_1 but not with Y_2 the corresponding elements in β_2 are set to zero, and similarly for β_1.

To set up likelihood equations for the parameters $\beta = (\beta_1, \beta_2, \gamma)$ we use as technical device the connection between the multinomial and the Poisson distributions. Let $I_{kl} = 1$ if $Y_1 = k$ and $Y_2 = l$ and $I_{kl} = 0$ otherwise, and define I_{kl}, $k, l = 1, 0$ to be independent $Poisson(\pi_{kl})$ variates. By definition $\sum_{kl} I_{kl} = 1$ and $\sum_{kl} \pi_{kl} = 1$, yielding identical large sample likelihood inference for $\pi = (\pi_{11}, \pi_{10}, \pi_{01}, \pi_{00})$ from the multinomial and Poisson likelihoods (Birch, 1963; Palmgren, 1981).

For observations $Y_i = (Y_{i1}, Y_{i2})$ with covariate vectors x_i, $i = 1, \ldots, n$ the

Poisson based likelihood equations have the form

$$\sum_{i=1}^{n} D_i' V_i^{-1}(I_i - \pi_i) = 0, \tag{2}$$

with $I_i = (I_{11}, I_{10}, I_{01}, I_{00})_i'$, $\pi_i = (\pi_{11}, \pi_{10}, \pi_{01}, \pi_{00})_i'$, $V_i = diag(\pi_{11}, \pi_{10}, \pi_{01}, \pi_{00})_i$ and D_i a $4 \times p$ matrix with p the dimension of $\beta = (\beta_1, \beta_2, \gamma)$ and the columns of D_i given by $D_{is} = (\frac{\partial \pi_{11}}{\partial \beta_s}, \frac{\partial \pi_{10}}{\partial \beta_s}, \frac{\partial \pi_{01}}{\partial \beta_s}, \frac{\partial \pi_{00}}{\partial \beta_s})'$, for s=1,...,p. Estimates for β are obtained by starting from an arbitrary value $\hat{\beta}_0$ sufficiently close to β and iteratively computing values $\hat{\beta}_{t+1}$, $t = 0, 1, \ldots$ from the linear expressions

$$\hat{\beta}_{t+1} = \hat{\beta}_t + (\sum_i D_{it}' V_{it}^{-1} D_{it})^{-1} \sum_i D_{it}' V_{it}^{-1}(I_i - \pi_{it}), \tag{3}$$

where D_{it}, V_{it} and π_{it} are evaluated at $\hat{\beta}_t$. Provided that the likelihood function has a unique maximum the algorithm will converge to the maximum likelihood estimates (Green, 1984). The estimated asymptotic covariance matrix for the maximum likelihood estimate $\hat{\beta} = (\hat{\beta}_1, \hat{\beta}_2, \hat{\gamma})$ is given by the inverse of the Fisher information matrix

$$\Sigma(\hat{\beta}) = \sum_i (D_i' V_i^{-1} D_i)^{-1}, \tag{4}$$

evaluated at $\hat{\beta}$. In Appendix A we show that $\Sigma(\hat{\beta})$ is block diagonal with zeroes for elements corresponding to $(\hat{\beta}_1, \hat{\gamma})$ and $(\hat{\beta}_2, \hat{\gamma})$.

It could be, that for some blocks only one of the two responses Y_{i1} and Y_{i2} is observed. If only Y_{i1} is observed, then the likelihood equations (2) and the expressions in (3) and (4) are modified by replacing I_i, π_i and V_i with $I_i = (I_{11} + I_{10},\ I_{01} + I_{00})_i$, $\pi_i = (\pi_{11} + \pi_{10},\ \pi_{01} + \pi_{00})$ and $V_i = diag(\pi_{11} + \pi_{10},\ \pi_{01} + \pi_{00})$. The matrix of derivatives D_i is modified accordingly. Note that π_i now has the form

$$\pi_i = [\exp(\beta_1 x_i)\{1+\exp(\beta_1 x_i)\}^{-1}, \{1+\exp(\beta_1 x_i)\}^{-1}]$$

and that the incomplete observations contribute no information about the association. Corresponding modifications are made if only Y_{i2} is observed.

From (1) it is clear that the elements of the matrices D_i are rather complicated expressions of the parameters $\beta = (\beta_1, \beta_2, \gamma)$. Ekholm et al (1986) provide a general set of macros NLIN in GLIM which construct the D_i matrices in each iteration by forming numerical derivatives from the nonlinear expressions $\pi_i = \pi_i(\beta)$. We use the macros NLIN extensively in the sections that follow.

3. Correlation versus Odds Ratio

The correlation coefficient $\rho = corr(Y_1, Y_2)$ could be an alternative to the odds ratio as measure of association between the two responses in a block. Instead of expression (1) we have

$$\pi_{11} = \pi_{1.}\pi_{.1} + \rho\sqrt{\pi_{1.}(1-\pi_{1.})\pi_{.1}(1-\pi_{.1})}, \quad (5)$$

and if ρ is restricted, so that $0 \leq \pi_{11}(\pi_{1.}, \pi_{.1}, \rho) \leq \min(\pi_{1.}, \pi_{.1})$ then maximum likelihood estimation is carried out as in Section 2. Note that if $\pi_{1.} = \pi_{.1} = \pi$ and $\pi \geq 0.5$, then ρ may take any value in the interval [-1, 1]. If $\pi < 0.5$, then the lower bound is $-\pi(1-\pi)^{-1}$, the upper bound remaining at 1. However, if $\beta_1 \neq \beta_2$ or if there are subunit specific covariates, then $\pi_{1.}$ and $\pi_{.1}$ may be very different in some blocks. For $\pi_{1.} = \pi$ and $\pi_{.1} = k\pi$ the correlation ρ is restricted to the interval

$$-\pi\sqrt{k(1-\pi)^{-1}(1-k\pi)^{-1}} \leq \rho \leq (1-\pi)\sqrt{k(1-\pi)^{-1}(1-k\pi)^{-1}}. \qquad (6)$$

For $\pi = 0.8$ and $k = 0.2$ the upper bound in (6) is as low as 0.21.

We compare use of the odds ratio and the correlation coefficient as measures of association by analysing the data in Table 1. These data were first presented by Ashford and Sowden (1970) and concern selfreported symptoms of breathlessness and wheeze among working coalminers in Britain, classified into nine five year age groups. They have subsequently been used as illustration in several text books on categorical data analysis. McCullagh and Nelder (1989) give a thorough motivation for applying the bivariate logistic model to these data, and they report the fit for a model with linear age effect for the marginal logits and the log odds ratio. McCullagh (1989) states that it does not appear possible to fit the bivariate logistic model with currently available computer packages. We do, however, in Appendix B fit the same model as McCullagh, using the GLIM procedure given in Section 2. All standard output in GLIM is available. Note in particular the block diagonal structure for the parameter correlation matrix. We further fit a sequence of models assuming fixed constant log odds ratio, log ψ, in the range [0, 6]. For log odds ratio values far from the maximum likelihood estimate, log $\hat{\psi} = 2.8$ ($s.e.$(log $\hat{\psi}$) $= 0.06$), the overall fit of the model is very bad. However, Figure 1a shows the stability of the slope parameter in relation to its standard error, even for grossly misspecified values for the odds ratio. The stability of the intercept parameters, although not reported, was equally striking. Another feature of Figure 1a is that the standard errors for the slope estimates do not vary nearly at all with ψ. Age is here a block specific covariate with different intercepts and slopes for the two margins. If, however, the parameters were restricted to be the same for the margins, then the standard errors would decrease with decreasing association. The degree of association determines the amount of information available in the data about the joint parameters. We return to this question in Section 5.

Figure 1b shows the estimates for the corresponding slope parameters when the correlations coefficient ρ is used as measure of association, and ρ is given fixed values in the range [0, 0.5]. For $\rho > 0.52$ the algorithm converged to negative estimated cell frequencies and stopped. The maximum likelihood estimate $\hat{\rho} = 0.49$ ($s.e.(\hat{\rho}) = 0.009$) is very close to this upper bound. The most striking feature of Figure 1b is, however, how much more unstable the slope parameter estimates are when the correlation is used instead of the odds ratio.

Cox and Reid (1987) state that an important consequence of two parameters being orthogonal is that the maximum likelihood estimate of one parameter varies only *slowly* with given values of the other parameter. This consequence of ψ being orthogonal to β_1 and β_2 (cf. Appendix A) is, indeed, exemplified in Figure 1a. In contrast, the covariance matrix for the estimates $\hat{\beta}_1$, $\hat{\beta}_2$ and $\hat{\rho}$ is not block diagonal and the slope estimates are sensitive to values for ρ.

4. Survival of hip prostheses

We illustrate use of the bivariate logistic model by analysing discrete time survival of total hip prostheses as function of patient specific and hip specific covariates. The data, which are described in detail in Turula (1989) consist of 2681 primary hip arthroplasties performed between 1967 and 1986 at the Helsinki Invalid Foundation Hospital. For 432 patients the arthroplasty was performed bilaterally, involving 864 hips. Of the unilaterally performed arthroplasties, 1072 concerned the right hip and 745 the left hip. In all cases considered here primary osteoarthrosis was the reason for the operation. The patients were followed till May 1988 and failure of the prosthesis was defined as the replacement or removal of one or more components for any reason other than infection. During follow-up a total of 22% of the prostheses failed. There is some indication that different failure mechanisms operate for the left and the right hip and that unilateral and bilateral arthroplasties differ (Turula, 1989). For the bilaterally treated patients the two hips were not operated on the same occasion, but the operations were close in time.

We analyse the effect on survival of gender, age and weight at the time of operation, payclass and operation score. The payclass covariate is an indicator for private or non private patients. Private patients were usually operated by highly skilled staff surgeons. The operation score is computed as the sum on the log scale of the length of the operation and the amount of blood lost during the operation. The score is an indication of the difficulty of the operation (Turula, 1989). For bilateral arthroplasties, age and weight at the time of the left and right operation are not identical, but very similar. The operations were usually both performed in the same payclass. We may thus view age, gender, weight and payclass as patient specific covariates and operation score as a hip specific covariate.

The first twelve years of follow-up are divided into three year intervals. The hazard is assumed constant within each interval and also after twelve years of follow-up. We first describe the model for bilateral prostheses. Let n_{1j} and n_{2j} denote the number of right and left hips, respectively at risk in the beginning of interval j, $j = 1, \ldots, 5$. In general $n_{1j} \neq n_{2j}$, since the two hips in a pair may have different follow up times. We write n_j for the number of *pairs* at risk at the beginning of interval j, and $n_{1j} - n_j$ and $n_{2j} - n_j$ for the number of *single* hips at risk in the beginning of interval j. Let (Y_{i1j}, Y_{i2j}), $i = 1, \ldots, n_j$ denote a bivariate binary response for patient i with both hips at risk at the beginning of interval j. We set $Y_{i1j} = 1$ if the right hip failed during that interval and zero otherwise, and similarly for Y_{i2j} and the left hip. The joint multinomial probabilities are defined in terms of the conditional marginal failure probabilities π_{i1j} and π_{i2j} and the conditional odds ratio ψ_i using the expression in (1). Note that although the same hip may contribute to the denominator in several risk sets, conditioning on the status at the beginning of each interval induces independence between responses in the different intervals.

For bilaterally treated patients with a single hip at risk in the beginning of interval j, we observe only one of the responses Y_{k1j} and Y_{m2j}, with respective failure probabilities π_{k1j} and π_{m2j}, $k = 1, \ldots, n_{1j} - n_j$, $m = 1, \ldots, n_{2j} - n_j$. The marginal failure probabilities are assumed not to depend on whether the other hip in the pair was at risk or not. For unilateral prostheses the responses and the failure probabilites are defined in a way analogous to those for single bilateral prostheses at risk.

We thus define failure at time t, $t = 0, 3, 6, 9$ years as the conditional probability of failing in the next three years, given that one is currently at risk. At time $t = 12$ the failure probability refers to the rest of the follow up period. Correspondingly, survival is defined as the probability of surviving for at least another three years (till the end of the follow-up period), given that one is currently at risk. We assume the covariates to have constant effect on the logit of the failure probabilities. An alternative would be to use the complementary log -log link (McCullagh, 1980).

In Table 2 we report the analysis of deviance from imposing equality constraints on parameters for the right and left hips. For unilateral prostheses no difference is seen neither in the baseline probabilities nor in the effects of covariates. Baseline is here defined as a 62 year old non privately operated male patient, who weighs 71 kg and has operation score 11.8. For bilateral prostheses there is a significant drop in the maximized likelihood when the baseline probabilities are set equal. No difference is seen for effects of covariates.

Table 3 gives the conditional survival probabilities at baseline. For bilateral prostheses, but not for unilateral prostheses, the survival probability is decreasing. The significant difference in baseline survival for the bilateral right and left prostheses apparently stems from a faster decrease for the right hip. Since the model fit was worse when restricting the parameters to be equal, we do not report the joint survival probabilities for the bilateral prostheses. The odds ratio for the two bilaterally treated hips is positive and close to significant.

In Tables 4 and 5 we report the covariate effects for the unilateral and bilateral prostheses, respectively. The pattern is very similar in the two tables and there is no difference between right and left hips. Survival is significantly higher for women, and high operation score has a significant negative effect on survival.

In Table 5 we also show the estimates and the robust standard errors from assuming independence between the left and the right hip. The form for the robust covariance matrix is given in expression (8) in the next Section. Inference from the independence model is virtually identical to that from the bivariate logistic model. When solving Zeger and Liang's generalized estimating equations for unspecified correlation between the right and left hip, the correlation estimate was -0.02. This implies a slightly weeker dependence than what was indicated by the odds ratio estimate in Table 3. Apparently, however, little is lost for these data if one assumes independence between outcomes on the right and the left hip. In the next section we take a closer look at properties of the bivariate logistic model and the independence model, when the association is stronger than in these data.

5. Independence and Robust Standard Errors

The orthogonality of the marginal logits and the log odds ratio suggests that if the bivariate logistic model is true, then nearly unbiased and normally distributed estimates of the marginal logit parameters are obtained, even if we wrongly assume independence within blocks.

The likelihood equations under independence are

$$\sum_{i=1}^{n} D_i' V_i^{-1}(Y_i - \pi_i) = 0, \tag{7}$$

with $Y_i = (Y_1, Y_2)_i'$, $\pi_i = (\pi_{1.}, \pi_{.1})_i'$, $V_i = diag\{\pi_{1.}(1-\pi_{1.}), \pi_{.1}(1-\pi_{.1})\}_i$ and $D_i = diag\{\frac{\partial \pi_{1.}}{\partial \beta_1}, \frac{\partial \pi_{.1}}{\partial \beta_2}\}_i$. Equations identical to (7) are obtained from (1) and (2) when $\pi_{11} = \pi_{1.}\pi_{.1}$.

Although solving equations (7) gives nearly the correct parameter estimates, the model based covariance matrix for the estimates, $\Sigma_I(\hat{\beta}) = (\sum_{i=1}^{n} D_i' V_i^{-1} D_i)^{-1}$ does not in general reflect the true variation. Zeger and Liang (1986) suggest the following modification, expected to be robust against misspecification of the dependence

$$\Sigma_R(\hat{\beta}) = n(\sum_{i=1}^{n} D_i' V_i^{-1} D_i)^{-1} \{\sum_{i=1}^{n} D_i' V_i^{-1} cov(Y) V_i^{-1} D_i\} (\sum_{i=1}^{n} D_i' V_i^{-1} D_i)^{-1}, \tag{8}$$

where $cov(Y) = E(Y_i - \pi_i)(Y_i - \pi_i)'$ is the empirical covariance matrix. If the independence assumption is true, then $\Sigma_R(\hat{\beta})$ reduces to $\Sigma_I(\hat{\beta})$.

To evaluate the performance of the independence assumption and the use of robust standard errors we performed a small simulation study using the model

$$\begin{aligned} logit\, \pi_{1.} &= \beta_1 x_1 \\ logit\, \pi_{.1} &= \beta_2 x_2 \\ log\, \psi &= \gamma. \end{aligned}$$

We set up two experiments. In experiment (a) the covariate was block specific and in experiment (b) it was subunit specific. The covariate values were generated as

(a) $x_1 \sim N(0,1), \quad x_2 = x_1$.

(b) $x_1 \sim N(0,1), \quad x_2 \sim N(0,1), \quad x_1$ and x_2 independent.

For given x-values, data sets of 250 bivariate binary observations were generated from the above model with $\beta_1 = \beta_2 = 1$ and $\gamma = 6$. In each of the experiments data was replicated 400 times and the following models, all having zero intercepts were fitted each time:

(i) The bivariate logistic model in (1) and (2) with different slope parameters for the marginal logits and with constant odds ratio.

(ii) The independence model in (7) with different slope parameters. Here the model based and the robust standard errors coincide.

(iii) The bivariate logistic model in (1) and (2) with equal slope parameters for the marginal logits and with constant odds ratio.

(iv) The independence model in (7) with equal slope parameters for the marginal logits. Both model based and robust standard errors were computed.

For Model (ii) the model based covariance matrix $\Sigma_I(\hat{\beta})$ is diagonal, and if we assume binomial variance for the observations, then $diag\{\Sigma_R(\hat{\beta})\} = diag\{\Sigma_I(\hat{\beta})\}$. For Model (iv) the robust standard errors were computed as the square root of

$$var_R(\hat{\beta}) = \frac{\sum_i \pi_{i1.}(1-\pi_{1i.})x_{i1}^2 + \sum_i 2(\pi_{i11} - \pi_{i1.}\pi_{i.1})x_{i1}x_{i2} + \sum_i \pi_{i.1}(1-\pi_{i.1})x_{i2}^2}{\{\sum_i \pi_{i1.}(1-\pi_{1i.})x_{i1}^2 + \sum_i \pi_{i.1}(1-\pi_{i.1})x_{i2}^2\}^2}, \tag{9}$$

with $\pi_{i1.} = \pi_{i1.}(\hat{\beta})$, $\pi_{i.1} = \pi_{i.1}(\hat{\beta})$ and $\pi_{i11} = \pi_{i11}(\hat{\beta}, \hat{\gamma})$. Expression (1) was used for $\pi_{i11}(\hat{\beta}, \hat{\gamma})$ with $\hat{\gamma}$ obtained from the fit of model (iii).

In both experiments and in each data set the estimated slope parameters were virtually identical for all four models. The results of the simulations in terms of averages over the 400 replications are given in Table 6 for models (i) and (ii) and in Table 7 for models (iii) and (iv). We focus on the behavior of the estimated standard errors.

A feature of both Table 6 and Table 7 is that in experiment (b), where the covariate is subunit specific the standard errors are consistently smaller than in experiment (a). In the absense of measurement error and for finite odds ratios, different covariate values for the two responses adds information about the variation that is not attributed to a 'common cause'. This is coupled with less knowledge about the magnitude of the 'common cause', i.e. in larger standard error for the odds ratio itself.

Table 6 shows that if the parameters are not restricted to be equal in the two margins and the covariate is block specific, then there is no need to account for possible dependence within blocks. The analysis separates into the estimation of ordinary logistic models for each of the two responses. This supports the finding in Section 4 where, regardless of the value for the odds ratio, the same inference is obtained for the effect of age on the prevalence of wheeze and breathlessness. If the covariate is subunit specific, then model (ii) is inefficient.

We turn to Table 7, where the parameters in the two marginal logits are restricted to be equal. In experiment (a), where the covariate is block specific the model based standard error from model (iv) is on average considerably smaller and more variable than the standard error from the true model (iii). This is to be expected, since under independence we wrongly assume information about the joint parameter from two independent observations in each block. If the observations are dependent their joint "information value" lies between that of one and two independent observations. However, in experiment (a) the mean and the dispersion of the *robust* standard error from model (iv) corresponds very well to the true standard error. Comparison between model (iii) and (iv) in experiment (b), where the covariate is subunit specific, suggests that inference based on the independence assumption is again inefficient, regardless of whether the model based or the robust version of the standard error is used. The model based standard error appears to be more labile than the robust standard error.

This simulation experiment is limited in scope. With som caution one could, however, summarize as follows: For block specific covariates the assumption of independence, coupled with use of robust standard errors seems to produce correct inference. This procedure is simple, since standard statistical software can be used to obtain the estimates. Computation of the standard errors in (8) involves matrix manipulation, but conceptually the procedure is straight forward. If subunit specific covariates are the focus of interest and the parameters in the margins are restricted to be equal, one might be willing to set up the more complicated model described in Section 2 in order to get efficient estimates. Alternatively one could use the generalized estimating equations with nonzero correlation. The efficiency of this latter procedure is not known.

APPENDIX A.

Orthogonality of $cov(\hat{\beta}_1, \hat{\beta}_2\,\hat{\gamma})$.

We show that reparametrisation of the multinomial probabilities π_{11}, π_{10}, π_{01} and π_{00}, with $\sum_{ij}\pi_{ij} = 0$ using monotone transformations f, g and h of respectively the marginal probabilites $\pi_{1.} = \pi_{11} + \pi_{10}$, $\pi_{.1} = \pi_{11} + \pi_{01}$ and the odds ratio $\psi = \pi_{11}\pi_{00}/(\pi_{10}\pi_{01})$, implies that the elements corresponding to $\{f(\pi_{1.}), h(\psi)\}$ and $\{g(\pi_{.1}), h(\psi)\}$ in the Fisher information matrix are zero. The property of parameter orthogonality between $(logit\,\pi_{.1}, log\,\psi)$ and $(logit\,\pi_{1.}, log\,\psi)$ was noted by McCullagh (1989), but it was not formally motivated.

The log likelihood for a multinomial observation $Y = (Y_{11}, Y_{10}, Y_{01}, Y_{00})'$ with $E(Y) = (\pi_{11}, \pi_{10}, \pi_{01}, \pi_{00})'$ is $logL = \sum_{ij} Y_{ij}\log\pi_{ij}$. For $f = f(\pi_{1.})$ and $h = h(\psi)$ the corresponding element in the Fisher information matrix is

$$-E(\frac{\partial^2 \log L}{\partial f \partial h}) = \frac{\partial\pi_{11}}{\partial h}\{\frac{\partial\pi_{11}}{\partial f}(\frac{1}{\pi_{11}}+\frac{1}{\pi_{10}}+\frac{1}{\pi_{01}}+\frac{1}{\pi_{00}}) - \frac{\partial\pi_{1.}}{\partial f}(\frac{1}{\pi_{10}}+\frac{1}{\pi_{00}})\}. \quad \text{(A1)}$$

Using the expression in (1) for $\psi \neq 1$ we have

$$\frac{\partial\pi_{11}}{\partial f} = \frac{1}{2}(\psi-1)^{-1}\{\frac{\partial a}{\partial f} - \frac{1}{2}(a^2+b)^{-\frac{1}{2}}(2a\frac{\partial a}{\partial f} + \frac{\partial b}{\partial f})\}, \quad \text{(A2)}$$

with

$$\frac{\partial a}{\partial f} = (\psi-1)\frac{\partial\pi_{1.}}{\partial f}\,,\quad \frac{\partial b}{\partial f} = -4\psi(\psi-1)\pi_{.1}\frac{\partial\pi_{1.}}{\partial f}.$$

After some manipulation we write (A2) as

$$\frac{\partial\pi_{11}}{\partial f} = \frac{\partial\pi_{1.}}{\partial f}\{\frac{\pi_{11}+\pi_{01}\psi}{1+(\pi_{10}+\pi_{01})(\psi-1)}\} = \frac{\partial\pi_{1.}}{\partial f}\{\frac{\frac{1}{\pi_{10}}+\frac{1}{\pi_{00}}}{\frac{1}{\pi_{11}}+\frac{1}{\pi_{01}}+\frac{1}{\pi_{10}}+\frac{1}{\pi_{00}}}\},$$

from which it is clear that the right hand side of (A1) is zero. The derivation is analogous for $g(\pi_{.1})$ and $h(\gamma)$.

If h is defined as $h = h(\rho)$, with ρ the correlation coefficient, and if (5) is used instead of (1), then the right hand side of (A1) is not zero.

USE OF NLIN MACROS TO FIT THE BIVARIATE LOGISTIC MODEL FOR THE DATA IN TABLE 1.

```
[o] GLIM 3.77 update 0 (copyright)1985 Royal Statistical Society, London
[o]
[i] $C READ IN MACROS NLIN
[i]
[i] $C READ IN DATA
[i]                 $units 36
[i]                 $data y $read
[i]                   9   7  95 1841     23   9 105 1654     54  19 177 1863
[i]                 121  48 257 2357    169  54 273 1778    269  88 324 1712
[i]                 404 117 245 1324    406 152 225  967    372 106 132  526
[i]                 $var 9 x nn
[i]                 $assign x=-4,-3,-2,-1,0,1,2,3,4
[i]                 $assign nn=1952,1791,2113,2783,2274,2393,2090,1750,1136
[i]
[i] $C CONSTRUCT INDICES AND MAKE N INTO A LONG VECTOR
[i]                 $var 9 i i1 i2 i3 $cal i=%gl(9,1) : i1=4*i-3 : i2=i1+1
[i]                 $cal i3=i1+2 : i4=i1+3 : j=%gl(9,4) : n=nn(j)
[i]
[i] $C DEFINE THE POISSON DISTRIBUTION
[i]                 $mac m3 $cal %va=%fv+0.0001 $$endmac!
[i]                 $mac m4 $cal %di=2*(%yv*%log(%yv/%fv)+%fv-%yv) $$endmac
[i]
[i] $C EXPRESS FITTED VALUES AS FUNCTIONS OF PARAMETERS
[i]                 $mac fv!
[i]                 $var 9 pia pib psi k1 k2 k3 s11 s12 s21 s22
[i]                 $cal pia=%exp(p(1)+p(2)*x) : pia=pia/(1+pia)
[i]                 $cal pib=%exp(p(3)+p(4)*x) : pib=pib/(1+pib)
[i]                 $cal psi=%exp(p(5)+p(6)*x)
[i]                 $cal k1=(0.5/(psi-1)) : k2=1+(pia+pib)*(psi-1)
[i]                 $cal k3=4*psi*(1-psi)*pia*pib : k4=k1*(k2-%sqrt(k2*k2+k3))
[i]                 $cal s11=%ne(psi,1)*k4+%eq(psi,1)*pia*pib!
[i]                 $cal s12=pia-s11 : s21=pib-s11 : s22=1-s11-s12-s21!
[i]                 $var 36 ss $cal ss(i1)=s11 : ss(i2)=s12 : ss(i3)=s21 : ss(i4)=s:
[i]                 $cal %1=n*ss!
[i]                 $$endmac!
[i]
[i] $C GIVE INITIAL VALUES, DEFINE Y-VARIATE AND CALL MACRO NLIN
[i]                 $assign p=-2,0.5,-1.5,0.3,3,-0.1
[i]                 $yvar y
[i]                 $use nlin $scale 1.0
[i]                 $use fit $disp e c $
[o] scaled deviance = 30.394 at cycle 3
[o]            d.f. = 30
[o]
[o]           estimate       s.e.     parameter
[o]      1     -2.262     0.02989     P1
[o]      2     0.5145     0.01207     P2
[o]      3     -1.488     0.02056     P3
[o]      4     0.3254    0.008868     P4
[o]      5      3.022     0.06973     P5
[o]      6    -0.1314     0.02844     P6
[o]      scale parameter taken as  1.000
[o]
[o] Correlations of parameter estimates
[o]    1   1.0000
[o]    2  -0.6202   1.0000
[o]    3   0.4219  -0.1648   1.0000
[o]    4  -0.1478   0.4297  -0.3420   1.0000
[o]    5   0.0000   0.0000   0.0000   0.0000   1.0000
[o]    6   0.0001  -0.0001  -0.0000  -0.0000  -0.5966   1.0000
[o]          1         2        3        4        5        6
[c]
```

References

Ashford JR and Sowden RR (1970). Multivariate probit analysis. Biometrics 26, 535-46.

Birch MW (1963). Maximum likelihood in three-way contingency tables. J. R. Statist. Soc. B 27, 111-24.

Breslow NE and Day NE (1980). Statistical Methods in Cancer Research I: The Analysis of Case-Control Studies. I.A.R.C, Lyon.

Cox DR and Reid N (1987). Parameter orthogonality and approximate conditional inference, (with discussion). J. R. Statist. Soc. B 49, 1-39.

Dale JR (1984). Global cross-ratio models for bivariate discrete ordered responses. Biometrics 42, 909-17.

Ekholm A, Green M and Palmgren J (1986). Fitting exponential family nonlinear models in GLIM 3.77. The Glim Newsletter 13, 4-13.

Green PJ (1984). Iteratively reweighted least squares for maximum likelihood estimation, and some robust and resistant alternatives (with discussion). J. R. Statist. B 46, 149-92.

Laird MN (1989). Topics in likelihood-based methods for longitudinal data analysis. Technical Report. Harvard School of Public Health.

Liang KY and Zeger SL (1986). Longitudinal data analysis using generalized linear models. Biometrika 73, 12-22.

McCullagh P (1980). Regression models for ordinal data (with discussion). J. R. Statist. Soc. B 42, 109-42.

McCullagh P (1989). Models for discrete multivariate responses. Proceedings for the 47th Session of the International Statistical Institute. In Press.

McCullagh P and Nelder JA (1989). Generalized Linear Models. Second Edition. Chapman and Hall, London.

Palmgren J (1981). The Fisher information matrix for log linear models arguing condionally on the observed explanatory variables. Biometrika 68, 563-6.

Palmgren J (1989). Models for discrete multivariate responses. Discussion of paper by P. McCullagh. Proceedings for the 47th Session of the International Statistical Institute. In Press.

Prentice RL (1988). Correlated binary regression with covariates specific to each binary observation. Biometrics 44, 1033-48.

Rosner B (1984). Multivariate methods in opthalmology with application to other paired data situations. Biometrics 40, 1025-35.

Sharples K. (1989). Regresson analysis of correlated binary data. PhD Dissertation. Dept of Biostatistics, University of Washington.

Stiratelli R, Laird NM and Ware JH (1984). Random effects models for serial observations with binary response. Biometrics 40, 961-71.

Turula K (1989). Factors associated with the survival of total hip prostheses. MD Dissertation. Dept of Public Health Science. University of Helsinki.

Williams DA (1982). Extra-binomial variation in logistic linear models. Appl. Statist. 31, 144-8.

ZEGER SL AND LIANG KY (1986). Longitudinal data analysis for discrete and continuous outcomes. Biometrics 44, 1019-31.

ZEGER SL, LIANG KY AND ALBERT PS (1988) Models for longitudinal data: a generalized estimating equation approach. Biometrics 44, 1049-60.

ZHAO LP AND PRENTICE RL (1989). Correlated binary regression using a generalized quadratic model. Technical Report. Division of Public Health Sciences, Fred Hutchinson Cancer Research Center.

Table 1: Working coalminers in UK collieries classified by age and self reported symptoms of breathlessness and wheeze.

Age-group	Breathlessness		No Breathlessness		
in years	Wheeze	No Wheeze	Wheeze	No Wheeze	Total
20-24	9	7	95	1841	1952
25-29	23	9	105	1654	1791
30-34	54	19	177	1863	2113
35-39	121	48	257	2357	2783
40-44	169	54	273	1778	2274
45-49	269	88	324	1712	2393
50-54	404	117	245	1324	2090
55-59	406	152	225	967	1750
60-64	372	106	132	526	1136

Table 2: Analysis of Deviance. Comparison is made to a model which has different baseline probabilities and different effects of covariates for the right and the left hips.

Unilateral Prostheses

Restrictions	Δ Deviance	Δ d.f.	p-value
Equal baseline probabilities	4.6	5	0.47
Equal covariate effects	3.5	5	0.62
Equal baseline and covariate effects	10.8	10	0.37

Bilateral Prostheses

Restrictions	Δ Deviance	Δ d.f.	p-value
Equal baseline probabilities	12.7	5	0.03
Equal covariate effects	5.9	5	0.39
Equal baseline and covariate effects	21.3	10	0.001

Table 5: Bilateral prostheses. Effects of covariates on the logit probability of survival for at least another three years. Standard errors in parentheses.

	Bivariate Logistic Model			Independence Model[2]
	Right hip	Left hip	Jointly	Jointly
Age	-0.017	-0.002	-0.011	-0.010
	(0.013)	(0.015)	(0.010)	(0.010)
Sex	0.48	0.65[1]	0.55[1]	0.52[1]
	(0.27)	(0.31)	(0.21)	(0.20)
Weight	0.017	-0.007	0.007	0.007
	(0.013)	(0.015)	(0.010)	(0.010)
Payclass	0.37	0.17	0.28	0.26
	(0.24)	(0.29)	(0.17)	(0.18)
Operation Score	-0.24	-0.53[1]	-0.34[1]	-0.34[1]
	(0.15)	(0.20)	(0.12)	(0.12)

[1] $p < 0.05$

[2] Robust s.e.

Table 6: Simulation results for bivariate logistic model (i) and independence model (ii), when $\hat{\beta}_1$ and $\hat{\beta}_2$ are estimated separately. Means and S.D.'s taken over 400 replicated data sets of 250 bivariate binary observations. True parameter values: $\beta_1 = \beta_2 = 1$ and $\gamma = 6$. In experiment (a) x is block specific, in experiment (b) subunit specific.

	Experiment (a):		Experiment (b:)	
	Model (i):	Model(ii):	Model (i):	Model (ii):
$\hat{\beta}_1$:				
Mean	1.000	1.003	1.003	1.000
S.D.	0.174	0.175	0.146	0.175
$Model\,Based\,S.E.(\hat{\beta}_1)$:				
Mean	0.171	0.173	0.138	0.172
S.D.	0.012	0.013	0.013	0.013
$\hat{\beta}_2$:				
Mean	1.006	1.008	1.004	1.011
S.D.	0.168	0.170	0.132	0.175
$Model\,Based\,S.E.(\hat{\beta}_2)$:				
Mean	0.172	0.173	0.132	0.167
S.D.	0.012	0.013	0.011	0.012
$\hat{\gamma}_1$:				
Mean	6.18		6.98	
S.D.	0.77		1.94	
$S.E.(\hat{\gamma}_1)$:				
Mean	0.66		1.81	
S.D.	0.12		1.15	

Table 7: Simulation results for bivariate logistic model (iii) and independence model (iv), when $\hat{\beta}_1 = \hat{\beta}_2 = \hat{\beta}$. Means and S.D.'s taken over 400 replicated data sets of 250 bivariate binary observations. True parameter values: $\beta_1 = \beta_2 = 1$ and $\gamma = 6$. In experiment (a) x is block specific, in experiment (b) subunit specific.

	Experiment (a):		Experiment (b:)	
	Model (iii):	Model(iv):	Model (iii):	Model (iv):
$\hat{\beta}$:				
Mean	1.003	1.005	1.001	1.000
S.D.	0.167	0.168	0.110	0.127
Model Based S.E.$(\hat{\beta})$:				
Mean	0.167	0.122	0.108	0.120
S.D.	0.011	0.071	0.007	0.070
Robust S.E$(\hat{\beta})$:				
Mean		0.167		0.125
S.D.		0.012		0.006
$\hat{\gamma}_1$:				
Mean	6.08		6.87	
S.D.	0.68		1.91	
$S.E.(\hat{\gamma}_1)$:				
Mean	0.64		1.76	
S.D.	0.09		1.18	

Figure 1a: Maximum likelihood estimates for the effect of a five year increase in age on the logit of breathlessness ($\hat{\beta}_1$) and wheeze ($\hat{\beta}_2$) for prespecified values of the log odds ratio $log\,\psi$.

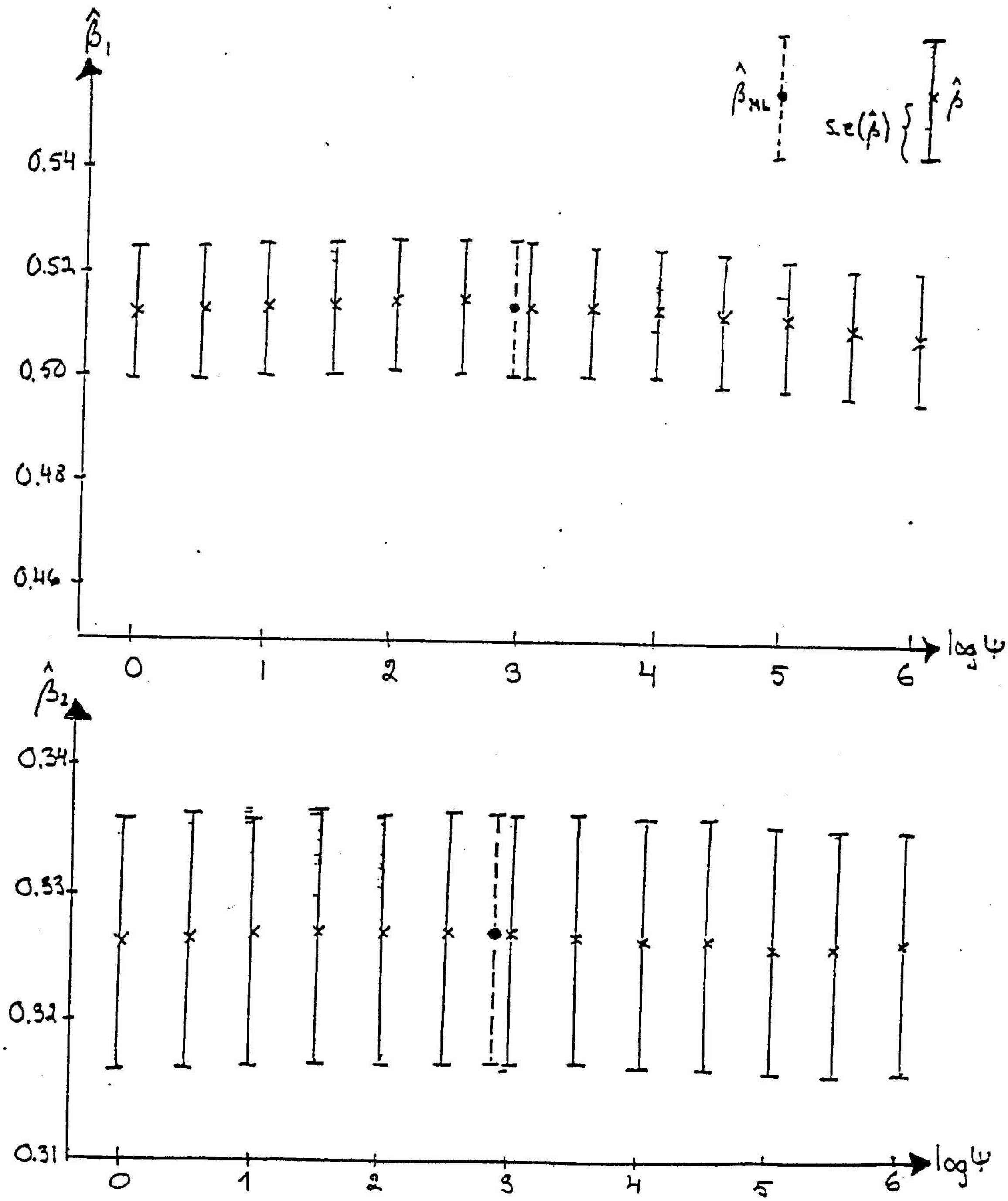

Figure 1b: Maximum likelihood estimates for the effect of a five year increase in age on the logit for breathlessness ($\hat{\beta}_1$) and for wheeze ($\hat{\beta}_2$) for prespecified values of the correlation coefficient ρ.