Published in final edited form as:

Biom J. 2006 August ; 48(5): 860–875.

Maximum Likelihood Estimation of Marginal Pair-wise Associations with Multiple Source Predictors

Liam M. O'Brien¹, Garrett M. Fitzmaurice^{2,3}, and Nicholas J. Horton⁴

1Department of Mathematics, Colby College 5838 Mayflower Hill, Waterville, ME 04901, USA

2Department of Biostatistics, Harvard School of Public Health 655 Huntington Avenue, Boston, MA 02115, USA

3Division of General Medicine, Brigham and Women's Hospital 1620 Tremont Street, Boston, MA 02120, USA

4Department of Mathematics, Smith College Northampton, MA 01063, USA

Summary

Researchers interested in the association of a predictor with an outcome will often collect information about that predictor from more than one source. Standard multiple regression methods allow estimation of the effect of each predictor on the outcome while controlling for the remaining predictors. The resulting regression coefficient for each predictor has an interpretation that is conditional on all other predictors. In settings in which interest is in comparison of the marginal pairwise relationships between each predictor and the outcome separately (e.g., studies in psychiatry with multiple informants or comparison of the predictive values of diagnostic tests), standard regression methods are not appropriate. Instead, the generalized estimating equations (GEE) approach can be used to simultaneously estimate, and make comparisons among, the separate pairwise marginal associations. In this paper, we consider maximum likelihood (ML) estimation of these marginal relationships when the outcome is binary. ML enjoys benefits over GEE methods in that it is asymptotically efficient, can accommodate missing data that are ignorable, and allows likelihood-based inferences about the pairwise marginal relationships. We also explore the asymptotic relative efficiency of ML and GEE methods in this setting.

Keywords

Log-linear models; mixed parameter transform; multiple informants; multivariate logistic transform

1 Introduction

Statistical data analysis often involves analyzing data obtained from several covariates, or predictors, and how these covariates relate to an outcome. Standard multiple regression models typically produce estimates of coefficients for the covariates that have conditional interpretations. That is, the interpretation of each regression coefficient can only be made by holding the values of all other covariates fixed. This is the standard, and preferred, method of analysis when the partial pairwise relationships between the outcome and each predictor, conditional on the remaining predictors, is of interest.

While each covariate may provide some unique information about changes in the outcome, there are many settings where the covariates provide overlapping information. This is common in psychiatry where reports on a subject's psychopathology are gathered from multiple informants. Multiple informant data can be particularly useful in studies of child psychopathology since gathering information from parents, teachers, and mental health workers may give a more accurate assessment of the underlying state of the child (Achenbach

et al., 1987). In this setting, it may be of interest to estimate and compare regression parameters that are specific to each informant (or source). The use of multiple sources is also common in the field of diagnostic testing, where more than one type of test may be available to detect the presence of a risk factor, symptom, or disease. (Leisenring *et al.*, 2000). In the latter setting, it will be of interest to determine the predictive value of each test unconditional on the results of the other tests under consideration. This will be the analytic method of choice when the goal is to choose a single diagnostic or screening instrument from a group of candidate instruments. A similar type of example arises in studies of obesity. In this setting the goal may be to determine the best marginal predictor of obesity in adulthood (see, for example, Whitaker *et al.*, 1997). Thus interest is not in prediction of obesity in adulthood as a function of many childhood risk factors, but rather in determining its best *single* predictor early in life. We explore the latter two applications in more detail in the following paragraphs and return to a multiple informant application in Section 4.

To better illustrate this non-standard application of regression in the diagnostic testing setting, consider the following study reported in Leisenring *et al.* (2000). This study was designed to compare two binary screening measures for coronary artery disease (CAD). The true presence or absence of CAD for each subject was known from the results of an angiogram—the gold standard. While the angiogram is the gold standard, the procedure is invasive and expensive, making screening procedures valuable tools for assessing the need to administer more rigorous testing. The two screening measures under consideration were presence or absence of chest pain (CP), and the passing or failing of an exercise stress test (EST). Note that if these data were jointly analyzed in a standard regression modelling framework, then CAD status would be regarded as the outcome, with CP and EST being the predictors in, for example, a multiple logistic regression model. The results would provide probabilities or risk of CAD conditional on values for both CP and EST. However, when the goal is to choose a single screening measure, the regression parameter relating CP (or EST) to CAD has a somewhat unappealing interpretation due to conditioning on EST (or CP), and vice versa.

Here, the question of interest is which of the two screening measures is more strongly associated with the presence of CAD unconditional of the value of the screening measure not under consideration. While this comparison could be made in terms of the misclassification rates, Pepe *et al.* (1999) recommend obtaining and comparing estimates of the association (expressed in terms of an odds ratio) of each of the screening measures with the outcome (unconditional on the results of the other screening measure). Interestingly, when the marginal probabilities of the two screening measures are very similar, the apparent misclassification rate can be shown to be a monotonic function of the odds ratio between the screening measure and CAD status. Note that interest is not in prediction of CAD as a function of CP and EST, but rather in comparing the unconditional associations of each of the predictors with CAD. While the screening measures used in this particular example may be relatively inexpensive to obtain, in general it would not be desirable to gather information from more than one diagnostic test or screen. Thus, the goal is to select the screening measure with the strongest (unconditional) association with the outcome. The comparison of the regression coefficients from the two separate analyses required to assess this is a non-standard problem because the estimates are correlated, and there is no widely accepted method for jointly obtaining and comparing estimates of such marginal associations.

Another example where information is obtained from multiple sources with the goal of relating each to the outcome marginally occurs in obesity research. A study conducted by Whitaker *et al.* (1997) sought to determine the best childhood predictors of adult obesity. Data were collected retrospectively about each subject from three sources: a measure of obesity obtained on the subject during the her or his childhood, a measure of obesity on the subject's mother, and a measure of obesity on the subject's father. These three predictors were measured at five

occasions during the subject's childhood. It is important to reiterate that interest was not in prediction of obesity in adulthood as a function of these three childhood predictors, but was rather to determine which of the three was the most strongly associated with obesity in adulthood. As noted by Pepe *et al.* (1999), such marginal associations are generally of primary interest in clinical applications due to their interpretability. In this example, obtaining such associations allows one to assess whether a child's risk of obesity is most strongly influenced by her or his own obesity, or by his or her parents' obesity status. Standard regression analyses that include all sources do not allow a direct assessment of this, however, since the associations obtained are conditional on the other sources.

Much attention has recently been given to the statistical analysis of data arising from multiple sources (Horton and Fitzmaurice, 2004). One method for jointly obtaining estimates of the separate marginal associations relating each predictor to the response was simultaneously developed by Horton *et al.* (1999) and Pepe *et al.* (1999). Both used a non-standard application of generalized estimating equations (GEEs), under a so-called "working independence" assumption. In this approach, the univariate outcome is duplicated as many times as there are numbers of predictors, producing a cluster of identical responses for each subject. A separate marginal model is specified for each predictor but all regression parameters are jointly estimated. Note that the empirical (or "sandwich") variance estimator (Huber, 1967) must then be used to obtain valid standard errors of these coefficients. While this GEE method produces valid estimates of the regression parameters, it may potentially lose efficiency in certain circumstances. Although in general, the efficiency losses associated with GEE methods are often not substantial (Liang and Zeger, 1986), the potential loss of efficiency has not been investigated in this setting. In this paper, we present a method for obtaining estimates of the separate pairwise associations with a binary outcome via maximum likelihood (ML) estimation. We compare the asymptotic relative efficiency of the working independence GEE-based method of Horton *et al.* (1999) and Pepe *et al.* (1999) to ML—particularly when the marginal associations can be assumed to be of similar magnitude.

The joint models for the marginal associations considered in this paper are a subset of those proposed by Glonek (1996) and later extended and generalized by Bergsma and Rudas (2002) as well as Rudas and Bergsma (2004). In particular, we consider a multinomial model for the binary response and binary predictors whose parameters are a mixture of marginal parameters that describe the marginal logits and pairwise log odds ratios among the response-predictor pairs separately, and conditional parameters for all other pairwise and higher-order associations. Thus, the model described here is a hybrid of log-linear models (e.g., Bishop *et al.*, 1975) and the multivariate logistic models described by Glonek and McCullagh (1995) and Bergsma and Rudas (2002). A similar type of "mixed parameter" model was developed by Fitzmaurice and Laird (1993) for longitudinal binary outcomes. Such models are desirable due to the marginal interpretation of the parameters of primary interest. In addition, ML estimation of these parameters is robust to misspecification of the conditional pairwise and higher order associations when the data are complete (Fitzmaurice and Laird, 1993; Glonek, 1996). That is, even if the model for the complementary set of conditional parameters for all other pairwise and higher-order associations is misspecified, the marginal parameter estimates remain consistent provided the model for these marginal associations has been correctly specified.

The model and estimation method are described in Section 2. We consider the efficiency losses associated with using working independence GEE-based methods over ML in Section 3. In Section 4, the proposed method is applied to multiple informant data from a psychiatric study in which there are two predictors (two informants) providing information on psychiatric distress, and it is of interest to examine and compare their marginal relationships with mortality. We conclude in Section 5 with a discussion of the potential benefits and drawbacks of the likelihood-based method described in this paper.

2 A Mixed Parameter Model for the Marginal Pair-wise Associations

Consider a binary response, Y_i , obtained from $i = 1, 2, \dots, N$ subjects. Associated with each response is a set of K binary predictors given by, X_{ik} ($k = 1, 2, \dots, K$). Given that the distribution of Y_i is Bernoulli, generalized linear models for such data only require specification of a link function, $g(\cdot)$. Note that we are not interested in $g[E(Y_i | X_{i1}, X_{i2}, \dots, X_{iK})]$, but rather in the marginal associations for each predictor, $g[E(Y_i | X_{i1})]$, $g[E(Y_i | X_{i2})]$, ..., $g[E(Y_i | X_{iK})]$. Thus, the goal of the analysis is to obtain and compare estimates of the marginal pairwise associations between the response and each of the predictors separately, i.e., the YX_k associations for $k = 1, \dots, K$. Note that we will suppress the i subscript and refer to these relationships as the marginal pairwise associations with the outcome.

The resulting data can be summarized in a 2^{K+1} contingency table for the response and the K binary predictors. For each subject a set of joint probabilities, π , describes the probability of belonging to any given cell of the contingency table. We denote the vector of all first-, second-, and higher-order products among $Y, X_1, \dots, X_K$ by

$$\mathbf{A}' = (Y, X_1, \dots, X_K, YX_1, \dots, YX_K, X_1X_2, \dots, X_{K-1}X_K, \dots, YX_1 \cdots X_K)'.$$

Standard log-linear models for the multinomial probabilities, π , are based on a transformation from π to a set of conditional parameters - through the expression,

$$\boldsymbol{\Omega} = \mathbf{C}_1' \log \boldsymbol{\pi}, \quad (1)$$

where $\mathbf{C}_1$ is a matrix that takes appropriate contrasts of the log transformed probabilities. Log-linear models enjoy a number of advantages over other model formulations. Foremost among these is that estimation methods are computationally simple and are available in standard generalized linear model software packages (e.g., PROC GENMOD in SAS; SAS Institute, Cary, NC). However, the interpretation of parameters in such models are conditional in nature. That is, the parameters describe conditional logits, conditional log odds ratios, etc. Thus, log-linear models are most useful when there is interest in conditional independence restrictions among the variables (Glonek, 1996). In the setting considered here though, we are primarily interested in the *marginal* associations relating the outcome to each predictor separately, and in making comparisons among these associations. Thus, the conditional associations, especially those for higher-order associations, are considered nuisance characteristics of the data.

The marginal parameters of interest could be estimated by fitting a multivariate logistic model of the type considered by McCullagh and Nelder (1989). Such models are based on a transformation of π to a set of marginal parameters, ψ , described by,

$$\boldsymbol{\Psi} = \mathbf{C}_2' \log(\mathbf{L}\boldsymbol{\pi}), \quad (2)$$

where $\mathbf{L}$ and $\mathbf{C}_2$ are matrices that appropriately marginalize, and take contrasts of the first-, second-, and higher-order marginal probabilities, respectively. These parameters describe the marginal logits, marginal log odds ratios, etc. Thus, the key advantage to using such models in the multiple source predictor setting of interest in this paper is that the interpretation of the pairwise outcome-predictor parameters is not conditional on the remaining predictors. However, these models are computationally more difficult to employ since the inverse transformation from η_2 to π is not as straightforward as it is in the case of the log-linear models described by (1). Methods that avoid the inverse transformation, e.g., Lang and Agresti (1994) and Bergsma and Rapsák (2005), can reduce some of the computational burden. Also, consistent estimates of the outcome-predictor pair-wise associations are obtained *only* if the

entire joint distribution is correctly specified, i.e., the model for all higher-order associations must be correctly specified.

Fitzmaurice and Laird (1993) noted this conundrum and developed a class of models that includes a mixture of log-linear and multivariate logistic parameters in such a way as to realize advantageous features of both. In their formulation, where interest is in the marginal means, the first-order logits were marginal while all two- and higher-order parameters were conditional. We, however, consider an alternative mapping,

$$\boldsymbol{\pi}' \Rightarrow \left(\psi_Y, \psi_{X_1}, \dots, \psi_{X_K}, \psi_{YX_1}, \dots, \psi_{YX_K}, \omega_{X_1X_2}, \dots, \omega_{X_{K-1}X_K}, \dots, \omega_{YX_1 \dots X_K} \right)',$$

where the ψ parameters are functions of marginal probabilities and the ω parameters are functions of conditional probabilities. Glonek (1996) described such models in a general way, illustrating a closely related model in which the first-order logits and second-order log odds ratios were marginal, while all third- and higher-order parameters were conditional. Note that in our mapping, only the subset of pairwise associations between the outcome and the predictors are marginal (i.e., $\psi_{YX_1}, \dots, \psi_{YX_K}$); the remaining pairwise associations (among the predictors), and the third- and higher-order associations are conditional.

Given our mapping, we define a vector of marginal parameters, $\boldsymbol{\psi}$, for the first-order logits and the marginal pairwise outcome-predictor associations defined by the subset of $\mathbf{A}$ given by, $\mathbf{A}'_M = (Y, X_1, \dots, X_K, YX_1, \dots, YX_K)'$. The remaining second-order predictor-predictor associations and higher-order association parameters are described by a set of parameters, $\boldsymbol{\omega}$ (defined by the complementary subset of $\mathbf{A}$, $\mathbf{A}_M^c$). Given this notation, we can stack the vectors $\boldsymbol{\psi}$ and $\boldsymbol{\omega}$ into a single vector, $\boldsymbol{\eta}' = (\boldsymbol{\psi}', \boldsymbol{\omega}')$. We use the parameterization of Grizzle *et al.* (1969) to relate $\boldsymbol{\eta}$ to $\boldsymbol{\pi}$ through,

$$\boldsymbol{\eta} = \mathbf{C}' \log(\mathbf{L}\boldsymbol{\pi}), \quad (3)$$

for appropriate choices of $\mathbf{C}$ and $\mathbf{L}$. Details on the construction of $\mathbf{C}$ and $\mathbf{L}$ for the example presented in Section 4, can be found in Appendix B.

ML Estimation

ML estimation of the parameters in the model given by (3) can be implemented using a variety of techniques. However, given this formulation, parameter estimates can be readily obtained using a similar algorithm to the one proposed by Glonek and McCullagh (1995). In their paper, the model parameters consisted entirely of marginal parameters as given by (2) and a Fisher scoring procedure was implemented for estimation. Note that the transformation $\boldsymbol{\eta}_2 \rightarrow \boldsymbol{\pi}$ is not straightforward. Glonek and McCullagh (1995) thus used a Newton-Raphson procedure to iteratively obtain the joint probabilities, $\boldsymbol{\pi}$, from $\boldsymbol{\eta}$.

Now consider the linear model relating the mixture of marginal ($\boldsymbol{\psi}$) and conditional ($\boldsymbol{\omega}$) parameters given in (3) to a known design matrix, $\mathbf{Z}$,

$$\boldsymbol{\eta} = \begin{pmatrix} \boldsymbol{\psi} \\ \boldsymbol{\omega} \end{pmatrix} = \mathbf{Z}\boldsymbol{\beta}, \quad (4)$$

where the first p components of $\boldsymbol{\beta}$ characterize the marginal associations, $\boldsymbol{\psi}$; the remaining q components of $\boldsymbol{\beta}$ characterize the complementary conditional associations, $\boldsymbol{\omega}$. The known design matrix, $\mathbf{Z}$, is of dimension $2^{K+1} \times (p+q)$, and $\boldsymbol{\beta}$ is a $(p+q) \times 1$ vector of unknown parameters, where $p+q \leq 2^{K+1}$. The linear model will facilitate constraints on $\boldsymbol{\psi}$ and $\boldsymbol{\omega}$.

From (3) and (4), we have $\mathbf{C}' \log(\mathbf{L}\boldsymbol{\pi}) = \mathbf{Z}\boldsymbol{\beta}$; and each subject's contribution to the log-likelihood is

$$\ell_i = (\boldsymbol{\beta}; Y_i) = Y_i' \log \boldsymbol{\pi}, \quad (5)$$

where each Y_i is a 2^{K+I} vector that indicates the cell of the contingency table to which subject i belongs, and has a multinomial distribution, $Y_i \sim M(1, \boldsymbol{\pi})$. The score and information, as given in Glonek and McCullagh (1995), are

$$\mathbf{s} = \sum_{i=1}^N \mathbf{s}_i(\boldsymbol{\beta}; Y_i) = \sum_{i=1}^N \left[\left(\frac{\partial \boldsymbol{\pi}}{\partial \boldsymbol{\beta}} \right)' \text{diag}(\boldsymbol{\pi})^{-1} Y_i \right], \quad (6)$$

and

$$\mathbf{I} = \sum_{i=1}^N \mathbf{I}_i(\boldsymbol{\beta}) = \sum_{i=1}^N \left[\left(\frac{\partial \boldsymbol{\pi}}{\partial \boldsymbol{\beta}} \right)' \text{diag}(\boldsymbol{\pi}_i)^{-1} \left(\frac{\partial \boldsymbol{\pi}}{\partial \boldsymbol{\beta}} \right) \right], \quad (7)$$

respectively, where summation is over N independent subjects. Noting that

$(\partial \boldsymbol{\pi} / \partial \boldsymbol{\beta}) = \{\mathbf{C}' [\text{diag}(\mathbf{L}\boldsymbol{\pi})]^{-1} \mathbf{L}\}^{-1} \mathbf{Z}$, we can obtain estimates of $\boldsymbol{\beta}$ using a Fisher scoring algorithm.

Note that each update on the Fisher scoring algorithm requires updating $\boldsymbol{\pi}$. However, there is no analytical expression for the transformation $\boldsymbol{\eta} \rightarrow \boldsymbol{\pi}$ (or, equivalently; $\boldsymbol{\beta} \rightarrow \boldsymbol{\pi}$) since $\boldsymbol{\eta}$ is a vector of both marginal and conditional parameters. Glonek & McCullagh (1995) and Glonek (1996) used a Newton-Raphson scheme to obtain $\boldsymbol{\pi}$ from $\boldsymbol{\psi}$. We, however, follow the approach of Fitzmaurice and Laird (1993) and use an iterative proportional fitting (IPF) algorithm (Deming and Stephan, 1940) to obtain the estimated joint probabilities, $\boldsymbol{\pi}$, from $\boldsymbol{\beta}$. While we used IPF to obtain these joint probabilities, there are alternative approaches that do not require an inverse transformation. For example, Lang and Agresti (1994) provide a method, based on earlier work by Aitchison and Silvey (1958), of maximizing the likelihood subject to model constraints via Lagrange multipliers. They then solve the Lagrangian likelihood equations using a Newton-Raphson or Fisher scoring algorithm. This approach has been further modified by Bergsma and Rapcsák (2005) to handle a larger number of multinomial probabilities.

To implement the IPF, we specify a 2^{K+I} "start table," $(S(\boldsymbol{\omega}))$, constructed using the correct conditional log odds ratios (but arbitrary margins). We then use the IPF algorithm to scale $S(\boldsymbol{\omega})$ until its margins satisfy those defined by $\boldsymbol{\psi}$. Once the algorithm has converged, the scaled table contains the set of joint probabilities, $\boldsymbol{\pi}$, for the set of parameters defined by $\boldsymbol{\eta}$. This method will always result in a valid set of joint probabilities, $\boldsymbol{\pi}$, unlike the Newton-Raphson procedure implemented by Glonek & McCullagh (1995) and Glonek (1996). Utilizing the IPF method, we implement the following two-stage iterative procedure for obtaining the ML estimate, $\boldsymbol{\beta}$

1. Calculate $\hat{\boldsymbol{\beta}}^{J+1}$ from $\hat{\boldsymbol{\beta}}^J$ using the scoring step, $\hat{\boldsymbol{\beta}}^{J+1} = \hat{\boldsymbol{\beta}}^J + \mathbf{I}^{-1} \mathbf{s}$, where $\mathbf{I}$ and $\mathbf{s}$ are obtained from (7) and (6), respectively. Note that both (6) and (7) require knowledge of $\boldsymbol{\pi}$.
2. Use the IPF procedure to obtain $\hat{\boldsymbol{\pi}}^{J+1}$ from $\hat{\boldsymbol{\beta}}^{J+1}$.

A key advantage of the mixed parameter model proposed here is that even if the model for $\boldsymbol{\omega}$ has been misspecified, the ML estimate $\boldsymbol{\psi}$ is consistent provided the model for $\boldsymbol{\psi}$ is correct. That is, the p -dimensional subset of the vector $\boldsymbol{\beta}$ that characterize $\boldsymbol{\psi}$ is consistently estimated even when the model for $\boldsymbol{\omega}$ is misspecified. This has been pointed out by Fitzmaurice and Laird (1993) for the specific mixed parameter transform that they consider; Glonek (1996) proved this result for the general mixed parameter transform of which our model is a special case. Given that interest is primarily in the marginal pairwise associations between Y and each of

the predictors, X_k , $k = 1, \dots, K$, this result provides a key advantage to using the mixed parameterization over a fully marginal parameterization. Note, however, that if the model for ω is misspecified then the model-based standard errors of β will not be correct. Fitzmaurice and Laird (1993) using an empirical (or “sandwich”) covariance estimator (Huber, 1967) to protect against misspecification of ω .

The ML estimation routine described here is not implemented in widely available commercial software. Thus, we implemented the iterative routine in R utilizing the IPF algorithm that is included as part of the *loglin()* function. In the next section, we investigate the asymptotic relative efficiency of the independence GEE as compared to ML. We illustrate the proposed method in Section 4 using data collected from two predictors (two informants) of the psychiatric status of 953 subjects from the Stirling County study.

3 Asymptotic Relative Efficiency of Independence GEE Estimator Relative to ML

In this section the asymptotic efficiency of the working independence GEE estimator of Horton *et al.* (1999) and Pepe *et al.* (1999) is compared to the ML estimator described in Section 2. The minimal asymptotic relative efficiencies (ARE) of GEE estimation with a working independence assumption, relative to ML estimation, were obtained over a wide grid of values for ψ and ω . A full description of the independence GEE method can be found in Appendix A. We define the asymptotic relative efficiency of the GEE estimator relative to the ML estimator by,

$$\frac{V(\hat{\beta}_{ML})}{V(\hat{\beta}_{GEE})}$$

where $V(\hat{\beta}_{ML})$ and $V(\hat{\beta}_{GEE})$ represent the asymptotic variances of $\hat{\beta}_{ML}$ and $\hat{\beta}_{GEE}$ respectively.

First, we consider the case where there are two predictors. The values for the first-order marginal parameters (e.g., ψ_Y , ψ_{X_1} , ψ_{X_2}) were varied between -3 to 3 in steps of 1, corresponding to marginal probabilities that range from 0.047 to 0.953. The values for the second- and higher-order conditional parameters (e.g., $\omega_{X_1 X_2}$, $\omega_{Y X_1 X_2}$) ranged from -10 to 10 in steps of 1. Note that these values are on the log scale, and thus cover a wide grid of marginal and conditional odds ratios. The minimal ARE was calculated for values of the marginal pairwise outcome-predictor log odds ratios ranging from 0 to 4 (corresponding to marginal odds ratios ranging from 1.0 to 54.6). We consider both the case where each outcome-predictor marginal association is unique ($\psi_{Y X_1} \neq \psi_{Y X_2}$), and the case where there is assumed to be a shared parameter for these associations ($\psi_{Y X_1} = \psi_{Y X_2}$). We have restricted the marginal predictor-predictor odds ratios to be greater than 1 since it is assumed that the multiple predictors provide overlapping information about the univariate outcome.

In the first scenario, where the two coefficients for the outcome-predictor marginal log odds ratios are unique, the ARE=1, and asymptotically, there is no efficiency loss associated with independence GEE estimation when compared to ML estimation. This occurs because the GEE and ML estimators for the marginal outcome-predictor associations are equivalent when the multinomial model is saturated (see Appendix A). Thus, the ARE=1 when there are no shared parameters, regardless of the strength of the marginal pairwise associations or the values of the conditional parameters considered ($\omega_{X_1 X_2}$ and $\omega_{Y X_1 X_2}$). However, when a common

outcome-predictor parameter (i.e., $\psi_{YX_1} = \psi_{YX_2}$) is assumed, the multinomial model is no longer saturated and the minimal ARE is approximately 0.90. Of note, the ARE=1 when the two conditional parameters are both set to 0, regardless of the values of the marginal parameters; however, such conditional independencies are unlikely to arise in practice in this setting.

Table 1 shows the minimum ARE across a wide range of values for ψ and ω for various strengths of the outcome-predictor log odds ratio (ψ_{YX}). The minimum ARE always occurs when the conditional associations take on their maximum values (i.e., $\omega_{X_1X_2} = \omega_{YX_1X_2} = 10$). This is illustrated in Figure 1 which shows the ARE as a function of $\omega_{X_1X_2}$ and $\omega_{YX_1X_2}$ for a fixed set of first-order marginal logits and $\psi_{YX} = 4$. Here the values, $\psi_Y = -3$, $\psi_{X_1} = 3$, and $\psi_{X_2} = 2$, were chosen since that is where the minimum ARE occurs for the value of ψ_{YX} considered. However, the pattern is similar regardless of the values of the marginal parameters. Table 1 also shows the minimum ARE obtained when values of the marginal predictor-predictor odds ratios ranged from 1 to 2, as well as those obtained for marginal predictor-predictor odds ratios that ranged from 1 to 5. This indicates that, even though extreme values of ψ and ω were needed to obtain the minimum ARE, values quite close to this minimum were seen for reasonable values of these marginal predictor-predictor odds ratios.

We next consider the case when there are three predictors. Given that the set of conditional parameters, ω , is 8-dimensional, we place some simplifying constraints on the values investigated. Specifically, all second-order predictor-predictor associations were constrained to be equal ($\omega_{X_1X_2} = \omega_{X_1X_3} = \omega_{X_2X_3}$), as well as all third-order associations among the outcome and pairs of predictors ($\omega_{YX_1X_2} = \omega_{YX_1X_3} = \omega_{YX_2X_3}$). The remaining three-way ($\omega_{X_1X_2X_3}$) and four-way ($\omega_{YX_1X_2X_3}$) associations were allowed to vary independently. While these conditional parameters are not required to be constrained, and such equality constraints are undesirable in practical applications, we do so here solely to reduce the dimensionality of the study of ARE in the three-predictor variable case. The ARE was calculated for a grid of values for the marginal logits ranging from -3 to 3, and values of the conditional parameters ranging from -5 to 5, again at steps of 1. The range of the conditional parameters was more restricted than in the 2-predictor case due to extremely small joint probabilities encountered. The values for the ψ_{YX} associations again ranged from 0 to 4. As described in the two-predictor case, the GEE and ML estimators for the marginal outcome-predictor associations are equivalent when the model is saturated. This is reflected by the ARE being equal to 1 when the coefficients for the marginal outcome-predictor associations are unique ($\psi_{YX_1} \neq \psi_{YX_2} \neq \psi_{YX_3}$).

The ARE also equals 1 when $\omega = 0$ (even for a shared ψ_{YX} association). However, when a common parameter for all three marginal outcome-predictor associations was considered ($\psi_{YX_1} = \psi_{YX_2} = \psi_{YX_3}$) with $\omega \neq 0$, the minimum ARE ranged from approximately 0.78 to 0.82. The largest losses in efficiency of the independence GEE estimator were observed for large positive values of $\omega_{X_1X_2X_3}$ and $\omega_{YX_1X_2X_3}$ and large absolute values of ψ_{YX} . Table 2 lists the minimum relative efficiencies seen for varying values of a shared outcome-predictor association coefficient. It also shows the minimum ARE obtained for values of all marginal predictor-predictor odds ratios ($OR_{X_1X_2}$, $OR_{X_1X_3}$, and $OR_{X_2X_3}$) ranging from 1 to 2, as well as those obtained for all marginal predictor-predictor odds ratios ranging from 1 to 5. Again, while the minima occurred for large marginal predictor-predictor odds ratios (on the order of 10^2 to 10^4), most of the loss occurs when these marginal odds ratios take on less extreme values. When the minima were obtained, one or more cell probabilities approach zero. In order for all the cell probabilities to be of modest size ARE values of 0.90 or larger were seen.

We see that, in general, the efficiency loss of the independence GEE increases as the magnitude of the marginal outcome-predictor association (ψ_{YX}) increases. In the case of two predictors this loss is modest (approximately 10 percent or less). However, in the case of three predictors, we see losses approaching 25 percent, even for reasonably small values of the predictor-predictor association. Thus, these results indicate that the efficiency losses associated with using the working independence GEE method of Horton *et al.* (1999) and Pepe *et al.* (1999) relative to ML are dependent on the number of predictors in the model. We see that there may be some value to using ML estimation when there are three predictors, but perhaps not when there are only two.

4 Example

In this section, we utilize the multiple predictor model described in Section 2 to analyze data arising from the Stirling County depression study. The data come from a large cohort study consisting of 953 subjects in Eastern Canada. More information about the Stirling County study can be found in Leighton (1959), Murphy (1980), and Murphy *et al.* (1985).

The data were obtained prospectively from 1952 to 1968, with one outcome of interest being mortality during this 16-year period. Information on psychiatric distress was obtained from two predictors, and it is of interest to relate each of these predictors to the outcome, marginally. One predictor was a self-report measure called the DPAX (depression-anxiety scale) and was processed via computer algorithm (see Murphy *et al.*, 1985, for more information about the DPAX). This measure is designed to detect the presence of anxiety and/or depression through a self-report questionnaire. The second predictor (denoted GP) contained information about the presence of psychiatric distress as determined by a general physician. Each physician diagnosis was validated by a psychiatrist. Note that while the DPAX detects the presence of depression and/or anxiety, the GP data indicates any general mental disturbance deemed relevant by the physician. These data are summarized in Table 3.

Note that since we do not have covariates in addition to the two binary predictors (X_1 for DPAX; X_2 for GP), we can calculate the joint probabilities directly. However, it is of interest to determine whether X_1 and X_2 (DPAX and GP) have marginal pair-wise associations with the outcome (mortality) that are significantly different. Standard regression techniques would not readily allow such a comparison since the coefficients from a regression analysis including both X_1 and X_2 (DPAX and GP) have conditional interpretations. In order to assess this, we first fit the saturated model,

$$\begin{pmatrix} \psi_Y \\ \psi_{X_1} \\ \psi_{X_2} \\ \psi_{YX_1} \\ \psi_{YX_2} \\ \omega_{X_1X_2} \\ \omega_{YX_1X_2} \end{pmatrix} = \mathbf{I}_{7 \times 7} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \\ \beta_5 \\ \beta_6 \\ \beta_7 \end{pmatrix}. \quad (8)$$

The matrix, $\mathbf{I}$, is the identity matrix of indicated dimension. Note that $\beta_1, \dots, \beta_5$ correspond directly to the marginal parameters ($p = 5$) with primary interest in β_4 and β_5 , while β_6 and β_7 correspond directly to the conditional parameters ($q = 2$). The construction of $\mathbf{C}$ and $\mathbf{L}$ are shown in Appendix B for this two-predictor example.

We next fit a reduced model in which we assume a shared association parameter, with $\beta_4 = \beta_5$ (describing the marginal log odds ratio between the outcome with each of the two predictors separately). The ML estimates of β and their model-based standard errors are shown in Table 4 for the saturated and reduced models. In order to test the hypothesis that $\beta_4 = \beta_5$, we performed a likelihood ratio test (LRT) of $H_0: \beta_4 = \beta_5$ and found that the model with a shared parameter for the marginal pairwise associations between mortality and DPAX, and mortality and GP, is tenable (LRT = 0.03; $p > 0.85$). We caution the reader that such a comparison is meaningful only when the predictors have the same underlying scales of measurement (as they do in this example).

We note that the independence GEE (using empirical standard errors) and ML methods both result in estimates and standard errors for the regression coefficients that are nearly identical. This is not surprising given the ARE results for the 2-predictor case reported in Section 3. We also used the joint probabilities estimated from the Stirling County data to determine the corresponding ARE. We obtained an ARE of approximately 1 which agrees very closely with the comparison of the standard errors from the GEE and ML analyses.

Thus, in this example, the DPAX and GP predictors have similar relationships to mortality. This indicates that receiving a positive DPAX or GP assessment is associated with 1.76 times the odds of mortality relative to someone without a positive assessment. Thus having a diagnosis of psychiatric distress is associated with mortality irrespective of source. Note that even though the predictors are exchangeable, we obtain substantially smaller standard errors (0.158 versus 0.212 and 0.206) by using all the available data to estimate the common association. In this example, we report model-based standard errors for the ML estimator, however they did not differ substantively from those obtained via the empirical variance estimator since the model for the remaining conditional associations is saturated.

5 Discussion

While obtaining information from multiple predictors is common in many diverse applications, methods for simultaneously obtaining and comparing the marginal pairwise associations between the outcome and each predictor have seen relatively little work. The approach developed by Horton *et al.* (1999) and Pepe *et al.* (1999) provides a method for obtaining and comparing estimates of such marginal relationships using a GEE strategy in which the relationship between the outcome and each predictor can be modeled separately (but estimated simultaneously). This is advantageous in that no distributional assumptions about Y need to be made in order to obtain valid estimates of β (Liang and Zeger, 1986); the GEE approach also does not require assumptions about the joint distribution of $X_1, \dots, X_K$. This method is also advantageous in that it can be easily implemented using widely available commercial software (e.g., PROC GENMOD in SAS, or xtgee in Stata).

In this paper, we have developed a likelihood-based framework in which estimates of the marginal pairwise associations can be obtained and compared. The model presented here is similar to models considered by Fitzmaurice & Laird (1993) and Glonek (1996). While such models require specification of the joint distribution for ML estimation, only the model for the marginal parameters needs to be correctly specified for consistent estimates of the pairwise marginal associations with the outcome to be obtained. ML estimation is generally advantageous relative to GEE-based estimation in that it is asymptotically efficient. However, we have shown that there is no loss in efficiency associated with independence GEE methods for estimation of β when there are no shared parameters among the marginal outcome-predictor associations. That is, when the model is saturated, the GEE and ML estimators for the marginal outcome-predictor associations are equivalent (see Appendix A). While the ARE results indicate there is some loss of efficiency when there is a shared parameter among these

associations, this loss may be modest relative to the increased computational difficulty in implementing the ML estimation routine. We note that we have not investigated the efficiency loss when there are more than three predictors. It may be the case that this loss is greater; however, the computational difficulty in implementing the ML estimation procedure would also be greater. One would thus need to weigh the relative advantage of the efficiency gain when using ML estimation against its added computational burden.

The lack of substantial efficiency gain is consistent with previous work. For example, Fitzmaurice (1995) investigated the ARE of GEE versus ML estimation for the analysis of multivariate binary data. He showed that there is some efficiency loss associated with GEE versus ML methods using a working independence assumption for covariates that vary within a subject, but very little loss in efficiency when there are no covariates that vary within a subject. While he considered multiple outcomes for any set of predictors, recall that an appealing feature of logistic regression models is that the odds ratio can be estimated from a prospective or retrospective study design. One implication of this result when X is binary is that if X is treated as the response in the logistic regression, and Y is treated as the predictor, the estimate of the pairwise log odds ratio remains the same. Considering this fact, the ARE results reported in Fitzmaurice (1995) have implications in our setting. If we were to regard $X_1, \dots, X_K$ as multiple responses, and Y as a predictor, then the ARE results from Fitzmaurice (1995) suggest that the independence GEE is almost efficient when compared to ML. That is, Y is a between-subject “predictor” in the logistic regression model for the multiple correlated “responses”, $X_1, \dots, X_K$.

Aside from efficiency, ML estimation is advantageous over GEE methods since it can incorporate missing data that are either missing completely at random (MCAR; Little and Rubin, 1987) or missing at random (MAR; Little and Rubin, 1987), but at the cost of requiring assumptions about the joint distribution of the outcomes. Given that missing data are common in survey data, this method will allow consistent estimates of the separate pairwise associations to be obtained provided missingness is ignorable. However, incorporating missing data is not straightforward since only the multivariate logistic component (ψ) of the model is reproducible (i.e., it will require the use of the EM algorithm). Also, by utilizing ML estimation, it is possible to conduct likelihood-based inference. This is advantageous in the discrete data setting since likelihood ratio tests (LRT) are purported to have better finite-sample properties than Wald tests when the outcome is binary (Hauck and Donner, 1977). ML estimation also has the advantage of allowing modelling of the joint distribution in a variety of ways. In contrast, by focusing only on the marginal probability of the outcome, GEEs do not allow modelling of the joint distribution at all.

While the method presented in this paper enjoys the benefits of any ML-based estimation procedure, for discrete multivariate data it becomes cumbersome as the number of predictors increases. That is, as the number of second- and higher-order nuisance parameters (which number $2^{K+1} - 2K - 1$) proliferates it becomes computationally demanding. When the data are complete, our ARE results suggest that using ML estimation over the GEE method of Horton *et al.* (1999) and Pepe *et al.* (1999) may yield modest gains from an efficiency perspective alone.

Investigating the asymptotic relative efficiency of GEE-based methods compared to the ML methods described in this paper when there are missing data would be a valuable extension to this work. While the complete data results show modest losses, greater losses may result when there are missing data. Also, extensions of this method when there are additional covariates in the model would be of interest in many practical settings.

Appendix A: Equivalency of Independence GEE and ML Estimators in the Saturated Model

Let Y_i be the binary response from subject i , with X_{i1} and X_{i2} being two binary predictors. These three binary variables define a $2 \times 2 \times 2$ contingency table. The number of subjects falling into a given cell is denoted by n_{uvw} where we order the subscripts denoting presence (1) or absence (0) of the outcome, first predictor, and second predictor, respectively. Thus, the number of subjects satisfying the condition $Y_i = X_{i1} = X_{i2} = 0$ is given by n_{000} . We use the "+" notation to denote summation over an index. For example, the count of subjects satisfying $Y_i = X_{i1} = 0$, but having no restriction on X_{i2} , is given by n_{00+} . We abbreviate the total count of subjects, n_{+++} , by N as in Section 2.

Given that the 8 cell counts follow a multinomial distribution, the ML estimate of the joint probability for a given cell of the $2 \times 2 \times 2$ contingency table is given by the count in that cell, divided by the total number of subjects (e.g., $\hat{\pi}_{000} = n_{000} / N$; cf. Agresti, 1991). We now consider the marginal 2×2 contingency table between Y and X_1 , marginalized over X_2 , as given in Table 5. The ML estimates of these joint probabilities are given by $\hat{\pi}_{uv} = n_{uv+} / N$ for the joint probability associated with the cell in row u and column v . We denote the log odds ratio describing the marginal association between Y and X_1 by ψ_{YX_1} ; the ML estimate of this parameter in a saturated model (e.g., the model given by (8)) is given by, $\hat{\psi}_{YX_1} = \log[(n_{00+}n_{11+}) / (n_{10+}n_{01+})]$. This follows from the invariance property of ML estimation since the saturated model given by (8) is simply a one-to-one transformation of the 7 non-redundant multinomial probabilities.

Now consider the GEE estimation method of Horton *et al.* (1999) and Pepe *et al.* (1999) for the two predictor case. Although we have a univariate response from each subject, Horton *et al.* (1999) and Pepe *et al.* (1999) derive the estimating equations by considering a 2×1 vector of responses, $Y_i = (Y_i, Y_i)'$. This vector is then related to the two predictors via a multivariate logistic regression model,

$$\text{logit}(\mathbf{Y}_i) = \mathbf{X}_i \boldsymbol{\beta}, \quad (9)$$

where,

$$\mathbf{X}_i = \begin{pmatrix} 1 & X_{i1} & 0 & 0 \\ 0 & 0 & 1 & X_{i2} \end{pmatrix}; \quad \text{and} \quad \boldsymbol{\beta} = \begin{pmatrix} \beta_{01} \\ \beta_{11} \\ \beta_{02} \\ \beta_{12} \end{pmatrix}.$$

We use a working independence assumption when deriving the estimating equations for β . The independence GEE estimates β are found as the solution to,

$$\sum_{i=1}^N \mathbf{D}_i' \mathbf{A}_i^{-1} [\mathbf{Y}_i - E(\mathbf{Y}_i | \mathbf{X}_i)] = 0;$$

where

$$\mathbf{D}_i = \left(\frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\beta}} \right); \quad \text{and} \quad \mu_{ik} = E[Y_i | X_{ik}] = \frac{\exp(\beta_{0k} + \beta_{1k} X_{ik})}{1 + \exp(\beta_{0k} + \beta_{1k} X_{ik})}; \quad k = 1, 2.$$

Since we are utilizing a working independence assumption the weight matrix is given as, $\mathbf{A}_i = \text{diag}[\mu_{i1}(1 - \mu_{i1}), \mu_{i2}(1 - \mu_{i2})]$. Note that a diagonal weight matrix is generally required

to ensure consistency of the point estimates (Pepe, *et al.*, 1999). In this setting, the matrix, $\mathbf{D}'_i$ is given by $\mathbf{X}'_i \mathbf{A}_p$, resulting in the following estimating equations,

$$\sum_{i=1}^N \mathbf{X}'_i [\mathbf{Y}_i - E(\mathbf{Y}_i | \mathbf{X}_i)] = 0.$$

Without loss of generality, we focus on the marginal $Y X_1$ log odds ratio, $\psi_{YX_1} = \beta_{11}$. We note that the estimating equations for the parameters relating the first predictor to the mean are orthogonal to those relating the second predictor to the mean. Thus, we focus on the first two rows of the estimating equations and obtain,

$$\sum_{i=1}^N \sum_{X_{i2} \in (0,1)} [Y_i - E(Y_i | X_{i1})] = 0$$

and

$$\sum_{i=1}^N \sum_{X_{i2} \in (0,1)} X_{i1} [Y_i - E(Y_i | X_{i1})] = 0.$$

Using the notation in Table 5, these equations can be expressed as,

$$n_{1++} - \left[n_{+0+} \frac{\exp(\beta_{01})}{1 + \exp(\beta_{01})} + n_{+1+} \frac{\exp(\beta_{01} + \beta_{11})}{1 + \exp(\beta_{01} + \beta_{11})} \right] = 0; \quad (10)$$

and

$$n_{11+} - \left[n_{+1+} \frac{\exp(\beta_{01} + \beta_{11})}{1 + \exp(\beta_{01} + \beta_{11})} \right] = 0. \quad (11)$$

Subtracting (11) from (10) results in,

$$n_{1++} - n_{11+} - n_{+0+} \frac{\exp(\beta_{01})}{1 + \exp(\beta_{01})}. \quad (12)$$

We solve (12) for β_{01} to obtain the GEE estimate, $\tilde{\beta}_{01} = \log(n_{10+}) - \log(n_{00+})$, and substitute into (10). We then solve for β_{11} and obtain the GEE estimate,

$\tilde{\beta}_{11} = \log[(n_{00+} n_{11+}) / (n_{10+} n_{01+})]$. Thus, the solution to the estimating equations for the GEE

estimate of ψ_{YX_1} is the same as that in the ML setting ($\hat{\psi}_{YX_1} = \tilde{\beta}_{11}$). Thus, the two methods will always result in the same estimates when in the saturated model setting, and the relative efficiency of the two methods will be 1.

While we have illustrated this for the marginal log odds ratio for $Y X_1$, we note that we may permute the ordering of the subscripts and obtain the same results for the $Y X_2$ marginal log odds ratio. Similarly, if we had more than 2 predictors, then we marginalize over $K - 1$ predictors to obtain a 2×2 table similar to that given by Table 5. The results will remain the same, although summation will be taken over the $K - 1$ predictors not under consideration. This indicates that, when the model is saturated the ML estimator, $\hat{\psi}_{YX_1}$, is equal to the GEE

estimator $\tilde{\beta}_{11}$, and the relative efficiencies comparing independence GEE-based estimation method to ML estimation will always be 1.

Appendix B: Construction of C and L for 2 predictors

In the case of 2 predictors, we order the probabilities as follows with the subscripts denoting presence (1) or absence (0) of the outcome, first predictor, and second predictor, respectively. Thus, the vector of joint probabilities is ordered in lexicographical order as,

$$\boldsymbol{\pi} = \begin{pmatrix} \pi_{000} \\ \pi_{001} \\ \pi_{010} \\ \pi_{011} \\ \pi_{100} \\ \pi_{101} \\ \pi_{110} \\ \pi_{111} \end{pmatrix}.$$

The matrix, L, takes these probabilities and appropriately marginalizes over them for the ψ parameters. Since ω parameters are conditional, the lower portion of L is the identity matrix. Note that the first row of L sums over all the cells to constrain the parameter ϕ to be equal to 1. The matrix, C, then takes the appropriate contrasts of $\log(L\pi)$. This matrix is block diagonal with the upper left corresponding to β and the lower right corresponding to ω . These matrices are as follows for the 2 predictor model,

$$L = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}; C = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.5 \\ 0 & 0 & 0 & 0 & 0 & 0 & -0.5 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & -0.5 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0.5 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0.5 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & -0.5 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & -0.5 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0.5 & 1 \end{pmatrix}.$$

References

- [1]. Achenbach TM, McConaughy SH, Howell CT. Child/adolescent behavioral and emotional problems: Implications of cross-informant correlations for situational specificity. *Psychological Bulletin* 1987;101(2):213–232. [PubMed: 3562706]

- [2]. Agresti, A. Categorical Data Analysis. John Wiley and Sons; New York: 1990.
- [3]. Aitchison J, Silvey SD. Maximum-likelihood estimation of parameters subject to constraints. *Annals of Mathematical Statistics* 1957;29:813–828.
- [4]. Bergsma, WP.; Rapcsák, T. An exact penalty method for smooth equality constrained optimization with application to maximum likelihood estimation. Technical report, EURANDOM, 1 2005
- [5]. Bergsma WP, Rudas T. Marginal models for categorical data. *Annals of Statistics* 2002;30:140–159.
- [6]. Bishop, YMM.; Fienberg, SE.; Holland, PW. Discrete Multivariate Analysis: Theory and Practice. MIT Press; Cambridge, MA: 1975.
- [7]. Deming WE, Stephan FF. On a least squares adjustment of a sampled frequency table when the expected marginal totals are known. *Annals of Mathematical Statistics* 1940;11:427–444.
- [8]. Fitzmaurice GM. A caveat concerning independence estimating equations with multivariate binary data. *Biometrics* 1995;51(1):309–317. [PubMed: 7766784]
- [9]. Fitzmaurice GM, Laird NM. A likelihood-based method for analysing longitudinal binary responses. *Biometrika* 1993;80(1):141–151.
- [10]. Glonek GFV. A class of regression models for multivariate categorical responses. *Biometrika* 1996;83(1):15–28.
- [11]. Glonek GFV, McCullagh P. Multivariate logistic models. *Journal of the Royal Statistical Society —Series B* 1995;57(3):533–546.
- [12]. Grizzle JE, Starmer CF, Koch GG. Analysis of categorical data by linear models. *Biometrics* 1969;25:489–504. [PubMed: 5824401]
- [13]. Hauck WW, Donner A. Wald's test as applied to hypotheses in logit analysis. *Journal of the American Statistical Association* 1977;77:851–853.
- [14]. Horton NJ, Fitzmaurice GM. Tutorial in biostatistics: Regression analysis of multiple source data and multiple informant data from complex survey samples. *Statistics in Medicine* 2004;23:2911–2933. [PubMed: 15344194]
- [15]. Horton NJ, Laird NM, Murphy JM, Monson RR, Sobol AM, Leighton AH. Multiple informants: Mortality associated with psychiatric disorders in the Stirling County Study. *American Journal of Epidemiology* 2001;154(7):649–656. [PubMed: 11581099]
- [16]. Horton NJ, Laird NM, Zahner GEP. Use of multiple informant data as a predictor in psychiatric epidemiology. *International Journal of Methods in Psychiatric Research* 1999;8(1):6–18.
- [17]. Huber, PJ. The behaviour of maximum likelihood estimators under non-standard conditions. In: LeCam, LM.; Neyman, J., editors. *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*; University of California Press. 1967; p. 221–233.
- [18]. Lang JB, Agresti A. Simultaneously modeling joint and marginal distributions of multivariate categorical responses. *Journal of the American Statistical Association* 1994;89(426):625–632.
- [19]. Leighton, AH. My Name is Legion: The Stirling County Study of Psychiatric Disorder and Sociocultural Environment. volume I. Basic Books Inc.; New York: 1959.
- [20]. Leisenring W, Alonzo T, Pepe MS. Comparisons of predictive values of binary medical diagnostic tests for paired designs. *Biometrics* 2000;56:345–351. [PubMed: 10877288]
- [21]. Liang K-Y, Zeger SL. Longitudinal data analysis using generalized linear models. *Biometrika* 1986;73(1):13–22.
- [22]. Little, RJA.; Rubin, DB. Statistical Analysis with Missing Data. Wiley; New York: 1987.
- [23]. McCullagh, P.; Nelder, JA. Generalized Linear Models. second edition. Chapman and Hall; New York: 1989.
- [24]. Murphy JM. Continuities in community-based psychiatric epidemiology. *Archives of General Psychiatry* 1980;37:1215–1223. [PubMed: 7436683]
- [25]. Murphy JM, Neff RK, Sobol AM, Rice JX, Olivier DC. Computer diagnosis of depression and anxiety: the Stirling County Study. *Psychological Medicine* 1985;15:99–112. [PubMed: 3887448]
- [26]. Pepe MS, Whitaker RC, Seidel K. Estimating and comparing univariate associations with application to the prediction of adult obesity. *Statistics in Medicine* 1999;18:163–173. [PubMed: 10028137]
- [27]. Rudas T, Bergsma W. On applications of marginal models for categorical data. *Metron* 2004;17(1): 1–25.

- [28]. Whitaker RC, Wright JA, Pepe MS, Seidel KD, Dietz WH. Predicting obesity in young adulthood from childhood and parental obesity. The New England Journal of Medicine 1997;337(13):869–873. [PubMed: 9302300]

Figure 1. Asymptotic relative efficiency as a function of $\omega_{X_1X_2}$ and $\omega_{YX_1X_2}$ when $\psi_{YX} = 0.4$ Table 1: Asymptotic Relative Efficiencies for 2 Predictor Case

Table 1
Asymptotic Relative Efficiencies for 2 Predictor Case

Ψ_{YX}	Minimum ARE	$1 \leq OR_{X_1X_2} \leq 2^{\dagger}$	$1 \leq OR_{X_1X_2} \leq 5^{\dagger}$
0	0.915	0.925	0.922
0.1	0.913	0.922	0.920
0.5	0.906	0.913	0.911
1.0	0.900	0.905	0.904
2.0	0.894	0.896	0.895
3.0	0.891	0.893	0.892
4.0	0.890	0.892	0.891

$^{\dagger} OR_{X_1X_2}$ represents the marginal predictor-predictor OR.

Table 2
Asymptotic Relative Efficiencies for 3 Predictor Case

ψ_{YX}	Minimum ARE	$1 \leq OR_{X_a X_b} \leq 2^{\ddagger}$	$1 \leq OR_{X_a X_b} \leq 5^{\ddagger}$
0	0.800	0.819	0.815
0.1	0.801	0.816	0.811
0.5	0.791	0.805	0.800
1.0	0.782	0.795	0.789
2.0	0.773	0.791	0.776
3.0	0.769	0.787	0.772
4.0	0.765	0.784	0.769

$\ddagger OR_{X_a X_b}$ represents any marginal predictor-predictor OR.

Table 3

Number Deaths in Stirling County Study (Total number of subjects per group in parentheses)

GP		0	1
DPAX	0	162(N =748)	31 (N = 90)
	1	27(N = 81)	12(N = 34)

Table 4

Saturated and Reduced Model Results

Parameter	Coefficient	$\hat{\beta}$	$\widehat{se}(\hat{\beta})$	$\hat{\beta}$	$\widehat{se}(\hat{\beta})$
ψ_Y	β_1	-1.13	0.0755	-1.13	0.0755
$\psi_{X_1 Y_1}$	β_2	-1.99	0.0994	-1.99	0.0994
$\psi_{X_1 Y_2}$	β_3	-1.90	0.0963	-1.90	0.0963
$\psi_{Y X_1}$	β_4	0.539	0.213	0.564*	0.159*
$\psi_{Y X_2}$	β_5	0.586	0.206		
$\omega_{X_1 X_2}$	β_6	1.40	0.246	1.40	0.246
$\omega_{Y X_1 X_2}$	β_7	-0.555	0.491	-0.555	0.491

* In the reduced model $\beta_4 = \beta_5$.

Table 5

Marginal Y-X Table of Counts

	Y = 0	Y = 1	
$X_{I=0}$	n_{00+}	n_{10+}	n_{+0+}
$X_{I=1}$	n_{01+}	n_{11+}	n_{+1+}
	n_{0++}	n_{1++}	N