Analyzing Matched 2 x 2 Tables from all Corners

Peer-reviewed author version

AERTS, Marc & MOLENBERGHS, Geert (2024) Analyzing Matched 2 x 2 Tables from all Corners. In: The American statistician,.

DOI: 10.1080/00031305.2024.2350452
Handle: http://hdl.handle.net/1942/43310

Analyzing Matched 2×2 Tables from all Corners

Marc Aerts

I-BioStat, Hasselt University, Diepenbeek, Belgium

and

Geert Molenberghs

I-BioStat, Hasselt University, Diepenbeek, Belgium

I-BioStat, KU Leuven University of Leuven, Leuven, Belgium

April 26, 2024

Abstract

Squared 2×2 tables with binary data from matched pairs are typically analysed using Cochran-Mantel-Haenszel methodology, conditional logistic regression, or random intercepts logistic regression. These are all "pair-specific" type of approaches. However, many more methods and models for clustered binary data, including marginal models and marginalisable pair-specific models, can be applied. We provide a comprehensive overview of methods and apply them all to two well-known example datasets, the prime minister's performance and the myocardial infarction datasets. The simple setting of matched binary data allows us to compare and relate different models, methods and their estimates. A technical explanation is given for why in some settings boundary estimates are obtained.

Keywords: Bahadur model, binary clustered data, generalized estimating equations, likelihood, marginal model, matched pair, random effects.

1 Introduction

In the current era of automated massive data collection and innovative data science methods, the analysis of a 2×2 square table may feel somewhat archaic. Such a square table indeed represents one of the most elementary datasets, but its underlying association structure allows a very instructive comparison of several statistical principles, methods and models, ranging from the time-honored Cochran-Mantel-Haenszel approaches to more recent techniques of marginalized multilevel models and shared random effects. It is not common practice to analyse square tables with these more elaborate models, developed for other purposes, but, through its simplicity, the setting of such basic 2×2 tables permits us to discuss and illustrate key properties and results of various approaches in an illuminating way. They allow us to bring fundamental differences to surface between analysis approaches, especially in matched settings. While for some techniques prospective and retrospective analyses lead to the same results, in particular for the association, for others there are fundamental differences, including the occurrence of boundary solutions in retrospective (outcome) matching.

Two examples from Agresti's seminal textbook on categorical data analysis (Agresti, 2002) will be used: the Prime Minister's performance data (MP data) and the Myocardial Infarction data (MI data).

Prime Minister's performance data

The data are shown in the upper part of Table 1. In a poll conducted in a random sample of 1600 voting-age British citizens, 944 indicated approval of the Prime Minister's performance in office. Six months later, of these same 1600 people, 880 indicated approval. Since relatively few people changed opinion, the diagonal of the table shows the higher counts.

Agresti illustrates the use of McNemar's test for comparing dependent proportions (showing there is a strong evidence of a drop in the approval rating), as well as the use of logit generalized linear mixed model (GLMM; normally distributed random effects in a logistic model). The off-diagonal cell counts are the informative ones, and, for a given participant, the estimated odds of approval at the second survey equals $86/150 = 0.573$ times that of the first survey. There is, as to be expected, a very strong association between the two responses, with estimated odds ratio $(794 \times 570)/(86 \times 150) = 35.084$. This GLMM is one of several models with a logistic underpinning; others will be discussed in Section 2.

Next to the original MP data, we will investigate the behaviour of the different models for what we will term the *reversed* MP dataset (as shown in the lower part of Table 1). For the sake of illustration, the outcomes of the first survey are reversed, while maintaining those of the second survey. As a consequence, the main diagonal counts in the original table become the off-diagonal ones in the reversed table, the estimated odds of approval at the second survey now equals $794/570 = 1.393$ times that at the first survey (reflecting a rise in the approval rate), and the positive association between the two responses is now clearly negative, with a sample odds ratio $1/35.084 = 0.029$. It will be interesting to investigate the performance of the models in this 'opposite' case.

Table 1: MP data: rating of performance of Prime Minister of citizens at two occasions 6 months apart. Upper table: original dataset as in (Agresti, 2002); lower table: manipulated dataset with reversed association (reversed dataset).

		Second Survey		
	First Survey	Approve	Disapprove	Total
Original dataset	Approve	794	150	944
	Disapprove	86	570	656
	Total	880	720	1600
Reversed dataset	Approve	86	570	656
	Disapprove	794	150	944
	Total	880	720	1600

Myocardial Infarction data

Table 2 shows the myocardial infarction (MI) data. A case-control study of acute myocardial infarction among Navajo Indians matched 144 MI cases according to age and gender with 144 people free of heart disease. Participants were asked whether they had ever been diagnosed as having diabetes. There are several major differences with the MP data: the number of pairs is much smaller, and it is a case-control study (rather than a longitudinal setting) in which a pair does not refer to the same subject but constitutes rather a matched (case, control) pair. For these data, we do not consider any reversed version, but we will analyse the data in both directions. With x denoting the diabetes diagnosis and y the myocardial status, the retrospective design implies $x|y$ is observed rather than $y|x$. With the symmetry property of the odds ratio in mind, it is instructive to compare the results when analysis is conducted in both directions.

Table 2: MI data: previous diagnoses of diabetes for myocardial infarction case–control pairs

	MI Cases		
MI Controls	Diabetes	No Diabetes	Total
Diabetes	9	16	25
No diabetes	37	82	119
Total	46	98	144

As discussed in (Agresti, 2002), matched pairs are typically analysed with McNemar's test (McNemar, 1947), of which the chi-squared statistic is algebraically identical to the Cochran-Mantel-Haenszel (CMH) test statistic for testing independence of binary responses of the matched pairs displayed in 'pair' specific partial tables. This connection opens the way to analyse matched pairs with a multitude of logistic regression type models. In his Chapter 10, Agresti (2002) illustrates the use of conditional logistic regression (logistic regression with conditional maximum likelihood) and of random-effects logistic regression. The Mantel-Haenszel estimate, the conditional ML, and the ML estimate for the random intercept logit model are all identical to the ratio of the off-diagonal counts. These three approaches typically suffice for analysing matched pairs in practice, but there are several other extensions of logistic regression that could be applied. Some models define the effect of interest at the pair level, others are marginal or pair-averaged, since they refer to averaging over the entire population of pairs rather than to individual pairs. Some pair-specific models easily allow the derivation of the pair-averaged effect, others do not or only approximately so. In our view, it is very instructive to apply these other models to the very simple setting of matched pairs, as it allows illustrating and explaining key features of and connections between all approaches more easily and explicitly. These insights help to understand the approaches in more complicated settings.

After a concise overview in the next section of a collection of methods and models with which matched pairs can be analysed, we will apply them to both datasets, compare them, explain differences, and discuss interrelations in Sections 3 and 4. The models will be fitted with maximum likelihood (ML) and the supplementary material provides SAS code.

2 Methods and models

2.1 The starting point

Logistic regression (LR)

In both examples, the research question of interest can be formulated in terms of logistic regression. For the MP data, $n = 1600$ citizens were surveyed twice. The counts in the 2×2 table can be represented as (x_{ij}, y_{ij}) with $i = 1, \ldots, n$, and with x the occasion indicator (and j the occasion index) defined as $x_{i1} = 0$ for the first occasion and $x_{i2} = 1$ for the second occasion, and with y_{ij} the performance approval indicator (yes=1,no=0) of participant i at occasion $j = 1, 2$. Note that $x_{ij} = x_j$ does not depend on i. Interest goes to a shift in the probability of approval from the first to the second occasion. In the logistic regression model for $y_{ij}|x_{ij} \sim \text{Bernoulli}\{P(y_{ij} = 1|x_{ij})\}$, the model for $P(y_{ij} = 1|x_{ij})$ is defined as

$$\log\left\{\frac{P(y_{ij}=1|x_{ij})}{P(y_{ij}=0|x_{ij})}\right\}=\alpha_M+\beta_M x_{ij}, \tag{1}$$

and hence this change is represented by the slope parameter β_M, or equivalently, by the odds ratio $\exp(\beta_M)$. The subscript M refers to the marginal interpretation (in contrast with a pair-specific interpretation, see further).

For the MI data, for the i-th case-control pair out of a total of $n = 144$ pairs, x_{i1} denotes the previous diabetes diagnosis (0 for negative, 1 for positive) for the control and likewise x_{i2} for the case. So j identifies the control ($j = 1$) or case ($j = 2$). Therefore, $y_{i1} \equiv 0$ and $y_{i2} \equiv 1$. In summary, x denotes the previous diabetes diagnosis, and y the case/control status. Note that now y_{ij} does not depend on i. In this situation, model (1) seems to make no sense, as y_{ij} is not a random variable. In a case-control study, it being a retrospective study, x is observed and y is fixed (suggesting a model for $x|y$), but interest goes to the potential effect of a previous diabetes diagnosis on the probability for an acute myocardial infarction, or to inference about $y|x$ as in model (1). As nicely explained in, for example, Section 5.1.4 of (Agresti (2002); see also Breslow and Day (1980)), the slope parameter β_M in model (1) is still the parameter of interest, thanks to the use of the logit link in model (1) and the identity OR($x|y$)=OR($y|x$) for the odds ratio OR.

Logistic regression with fixed pair effects (LRF)

However, model (1) nowhere reflects that observations from the same pair are most likely dependent, given that they share common pair characteristics. Following the explicit connection between McNemar's and the CMH test, model (1) can be extended with a pair-specific intercept effect α_i (pairs as strata):

$$\log\left\{\frac{P(y_{ij}=1|x_{ij})}{P(y_{ij}=0|x_{ij})}\right\}=\alpha_i+\beta_P x_{ij}. \tag{2}$$

Given the parameters α_i, the two observations on the same pair are independent, but, averaged over all pairs, the responses are correlated with a non-negative association, the intra-pair correlation, which is determined by the covariance in the pairs

$$\left\{\left(\frac{\exp(\alpha_i + \beta_P x_{i1})}{1 + \exp(\alpha_i + \beta_P x_{i1})}, \frac{\exp(\alpha_i + \beta_P x_{i2})}{1 + \exp(\alpha_i + \beta_P x_{i2})}\right)\right\}_{i=1}^{n}.$$

The larger the variation in the α_i's (relative to the value of β_P), the larger the intra-pair correlation. For the MP data, the intra-pair correlation is expected to be moderate to substantial, as the pair refers to repeated observations within the same individual. As for the MI data, the case-control pair concerns two different individuals, matched on age and gender, this intra-pair correlation is expected to be rather minor.

Model (1) is a *marginal* or *population-averaged* model. In contrast, model (2) is a *conditional* or *pair-specific* model; while it implies a marginal model, it is not directly formulated in a marginal way. As a consequence, the implied marginal model is no longer a logit type of model, as averaging

$$\frac{1}{n}\sum_{i=1}^{n} \frac{\exp(\alpha_i + \beta_P x_{ij})}{1 + \exp(\alpha_i + \beta_P x_{ij})}, \quad j = 1, 2,$$

no longer has an expit form $\exp(\cdot)/(1 + \exp(\cdot))$, due to the nonlinearity of the logit and expit functions. This hinders an easy interpretation of the marginal effect. Consequently, the slope β_M and OR $\exp(\beta_M)$ in model (1) and the slope β_P and OR $\exp(\beta_P)$ in model (2) have different interpretations, of a marginal and conditional type, respectively. In contrast, in situations where the identity link could be used, as is often the case for continuous paired data (Y_{i1}, Y_{i2}), where linear models are adopted, the pair-specific model implies a marginal linear model with the same slope, and both effects, pair-specific and pair-averaged are identical. As we will see in what follows, there are alternative formulations of the logistic regression model (2) that do allow a simple marginal interpretation as well.

Note that model (2) has an abundant number of $n + 1$ parameters (increasing with the number of pairs) and fitting this model with fixed effects parameters α_i leads to an inconsistent ML estimate $\widehat{\beta_P} \xrightarrow{P} 2\beta_P$ (see Breslow and Day (1980) and Exercise 10.24 in Agresti (2002)). The next Section 2.2 briefly describes the two ways to consistently fit pair-specific models: i) the many nuisance parameters α_i can be eliminated by conditioning on sufficient statistics (conditional ML, confusingly with the adjective conditional as well), ii) by assuming a random intercept whose distribution represents that of the population of pair-specific α_i's, thereby avoiding the unbounded growth in number of parameters. A standard choice is the normal distribution, implying a reduction from $n+1$ to 3 parameters. The next sections describe and compare related approaches that have been developed for more general purposes, but can be applied to analyse binary matched pairs as well.

2.2 Pair-specific models

For more details on the application of conditional likelihood and random effects in this setting, we refer to Chapters 10 and 12 in Agresti (2002) and Chapter 15 in Molenberghs and Verbeke (2005).

Conditional logistic regression (CLR)

The pair-specific totals $y_{i1} + y_{i2}$ are sufficient statistics for the pair-specific intercepts α_i. Only in case this total $y_{i1} + y_{i2} = 1$ (referring to the off-diagonal cells in Table 1), the β_P parameter appears in the conditional likelihood, being a Bernoulli likelihood with odds for $(y_{i1} = 0, y_{i2} = 1)$ versus $(y_{i1} = 1, y_{i2} = 0)$ equal to $\exp(\beta_P)$, allowing so-called exact inference. As the α_i's are fully eliminated, this approach does not provide an estimate of the intra-pair correlation, nor any insights in the marginal effect of x on y.

Cochran-Mantel-Haenszel (CMH)

The non-model-based CMH test of conditional independence between y and x while controlling for z in the $2 \times 2 \times n$ table of a three-way table x, y, z (with z the pair dimension), conditions on both the totals $x_{i1} + x_{i2}$ and $y_{i1} + y_{i2}$. So, in our regression model setting, the CLR and CMH approaches essentially coincide, but CMH's original inference is based on the approximate large-sample chi-squared distribution rather than on the exact small-sample distribution. Next to testing conditional independence, Mantel and Haenszel additionally proposed an estimator for the common odds ratio.

Normal random intercept (NRI)

Here, the abundant number of α_i parameters is eliminated by considering them as an unobserved sample from a *random effects* distribution. The standard choice is the normal distribution $\alpha_i \sim N(\alpha, \sigma_u^2)$, such that, after averaging (integrating), the likelihood depends on the slope β_P and the parameters α and σ_u^2 of the normal random effect's distribution. As the unobserved α_i's are characterized by the normal random intercept distribution, this approach allows for estimates of the latent marginal intra-pair correlation ρ and the marginal effect of x on y, albeit only approximately. The variance component σ_u^2 determines the latent intra-pair correlation ρ. If $\sigma_u = 0$, then $\rho = 0$, and next the (very rough) approximation

$$\rho \approx \frac{\sigma_u^2}{\sigma_u^2 + \pi^2/3},$$

holds, based on the expression for linear models with $\pi^2/3$ the variance of the logistic distribution (of the latent continuous response variable underlying the logistic regression model). A detailed discussion of differences between the latent and manifest correlations for binary outcomes is offered by Milanzi et al. (2015). Zeger et al. (1988) proposed, for σ_u

small and with $\phi(u; \alpha, \sigma_u)$ the density of the normal distribution, the approximation

$$\int \frac{\exp(u + \beta_P x_{ij})}{1 + \exp(u + \beta_P x_{ij})} \phi(u; \alpha, \sigma_u) d\alpha \approx \frac{\exp(\beta_M x_{ij})}{1 + \exp(\beta_M x_{ij})},$$

where $\beta_M = c\beta_P$ with $c = 1/\sqrt{1 + \left(\frac{16\sqrt{3}}{15\pi}\sigma_u\right)^2}$. This allows an approximate estimate for the marginal slope and marginal OR.

Note that, for a probit model, the conditional probit model with normal random effect does imply a marginal model of probit form, with $c = 1/\sqrt{1 + \sigma_u^2}$. But a major disadvantage of the probit model is that it does not provide an odds ratio interpretation for the regression parameters.

Comparison of CLR, CMH, NRI

These pair-specific models share the same starting point and deal with the abundant number of pair-specific intercepts. It can be proven (Agresti, 2002) that the conditional ML estimate, the MH estimate and the ML NRI-estimate for the OR are all equal to the ratio of the off-diagonal counts in the 2×2 matched pairs table. Contrary to the CMH/CLR approach, the NRI approach allows the approximate estimation of the marginal slope and OR parameter and the intra-pair correlation. Also, interpretation is slightly different, as the inference from the NRI model applies to the population of pairs rather than only those sampled. The price to pay is the need to assume a parametric distribution for the α_i.

We introduce other pair-specific models, also allowing a more direct interpretation, in Sections 2.4 and 2.5, but first we discuss some genuinely marginal models.

2.3 Marginal models

Marginal models extend logistic regression by including, explicitly or implicitly, a correlation structure for the clustered data, in our case the pairs (Y_{i1}, Y_{i2}), in the estimation procedure. This extension can take place at the level of the likelihood function or at the level of the estimating equations.

Generalized estimating equations (GEE)

Generalized estimating equations (Liang and Zeger, 1986; Zeger et al., 1988) modify the estimating equations of logistic regression by introducing a so-called working correlation structure, resulting in equations that no longer derive from a (correctly specified) likelihood. Even if the working correlation structure is misspecified, it results in consistent estimates for the slope β_M (and the OR). The so-called sandwich estimator applies empirical evidence to adjust the standard errors in case the working correlation structure is inappropriate. Working correlation structures include the independence, exchangeable, unstructured, autoregressive structures. As in our case all clusters are just pairs and thus of size 2, the non-independence structures all coincide. Inference on the regression parameters is asymptotic (typically Wald type based). Formal inference on the correlation parameter is not available.

Bivariate logistic regression (BLR)

A bivariate logistic regression model extends the likelihood function by applying a bivariate probability distribution for (y_{i1}, y_{i2}). The quadrinomial distribution $P(y_{i1} = k, y_{i2} = \ell), k, \ell = 0, 1$ can be applied as such or reparameterized in terms of three parameters of interest, in our case the marginal parameters $P(y_{i1} = 1)$, $P(y_{i2} = 1)$, and an association parameter. Here, we choose the correlation $\rho = \text{cor}(y_{i1}, y_{i2})$, but the OR is another option. With the choice of correlation, the resulting model is known as the Bahadur model Bahadur (1961), with no higher order associations. All features of ML are available, including inference about the correlation parameter. Considerable detail on the Bahadur model can be found in Aerts et al. (2002).

Comparison GEE and BLR

Both approaches are marginal models. They do not consider pair-specific models and do not measure variation across the pair-specific models (otherwise than indirectly through the intra-pair correlation).

Given that, for paired data without covariates that we are considering here, there is no room to misspecify the correlation structure in GEE, unless independence would be chosen, the estimates under GEE and BLR will be identical, apart from the fact that BLR also yields a standard error for the correlation parameter. This will be exemplified in the data analysis.

2.4 Pair-specific random intercept models allowing marginal interpretation

Bridge random intercept (BRI)

The Bridge distribution for the random pair effect was developed to ensure that the conditional and marginal distributions are both of a logistic form with a simple algebraic relationship between the linear predictor functions so that the parameters in the conditional model can readily be marginally interpreted; see Wang and Louis (2003) for binary random intercept models. Replacing the normal distribution for the random intercept in the NRI model by such a bridge distribution, guarantees that the marginal logit model (1) holds but with slope β_M related to

$$\beta_M = \frac{\beta_P}{\sqrt{1 + 3\sigma_b^2/\pi^2}} = \beta_P(1 - \rho),$$

with σ_b^2 the variance of the bridge distribution and $\rho = \mathrm{corr}(y_{i1}, y_{i2}|\beta_P = 0)$ the intra-pair correlation.

Marginalized multilevel model (MMM)

The marginalized multilevel model (MMM), as introduced by Heagerty (1999), Heagerty and Zeger (2000) and extended by Griswold and Zeger (2004), specifies a separate model for the marginal and conditional means and links them using a connector function depending on covariates, marginal parameters, and the random-effect specification, allowing both a marginal and conditional interpretation of the parameters. For the "logit-probit-normal model," the conditional model can in our situation be written as, with $\Phi(\cdot)$ the standard normal cdf,

$$\Phi^{-1}(P(y_{ij} = 1|u_i)) = \delta_{ij} + u_i, \tag{3}$$

with

$$u_i \sim N(0, \sigma_m^2),$$

and

$$\delta_{ij} = (\sqrt{1 + \sigma_m^2})\Phi^{-1}(\text{expit}(\alpha + \beta_M x_{ij})).$$

The corresponding marginal model is our basic logistic regression model (1), with the slope still enjoying the log-odds ratio interpretation. Connections between the BRI and MMM are discussed in Molenberghs et al. (2013). To our knowledge, there is no (exact or approximate) analytical expression available for the intra-pair correlation ρ.

2.5 Pair-specific random probability models allowing marginal interpretation

Instead of using normal and other random effects models on the whole real line, a very natural alternative in our situation here is the use of a beta random effect on the probability scale.

Correlated beta model (CBM)

When using random effects for binary data, a very natural and appealing candidate distribution to reflect the "pair-effect" is the beta distribution, also known from its pivotal role in the beta-binomial distribution. How to use and implement another distribution than the normal for the random effect has been shown by Nelson et al. (2006), using the probability integral transform (PIT)

$$v = F^{-1}(\Phi(u)), \tag{4}$$

with $u \sim N(0,1)$ and F the CDF of the random effect distribution of interest. Liu and Yu (2008) reformulates the likelihood conditional on the non-normal random effect(s) to that conditional on normal random effect(s). Consider the model, for pair $i = 1, \ldots, n$

$$y_{ij} \sim \text{Bernoulli}(\Pi_{ij}), \quad j = 1, 2,$$

with random probabilities

$$\Pi_{ij} \sim F_j^{-1}(\Phi(u_i)), u_i \sim N(0,1), \tag{5}$$

where F_1 and F_2 denote the CDF of the beta distribution with parameters α_1, β_1 and α_2, β_2 respectively, with

$$\alpha_j = \frac{1-\rho}{\rho} \frac{e^{\alpha+\beta_M x_{ij}}}{1+e^{\alpha+\beta_M x_{ij}}}, \quad j = 1, 2,$$

and

$$\beta_j = \frac{1-\rho}{\rho} \frac{1}{1+e^{\alpha+\beta_M x_{ij}}}, \quad j = 1, 2.$$

Note that, by construction, the marginal logit model holds

$$E(E(y_{ij}|\Pi_{ij})) = E(\Pi_{ij}) = \frac{\alpha_j}{\alpha_j + \beta_j} = \frac{e^{\alpha+\beta_M x_{ij}}}{1+e^{\alpha+\beta_M x_{ij}}}.$$

Straightforward calculations show that the intra-pair correlation, induced by the correlated random probabilities $\Pi_{ij} \sim F_j^{-1}(\Phi(u_i))$, is represented by

$$\rho = \mathrm{corr}(y_{i1}, y_{i2}|\beta_M = 0) = \frac{1}{1+\alpha_1+\beta_1} = \frac{1}{1+\alpha_2+\beta_2}.$$

Shared beta model (SBM)

Molenberghs et al. (2010) proposed a modeling framework for hierarchical data with both normal and conjugate random effects, following on Molenberghs et al. (2007), who had formulated the version for Poisson data. In this framework, the hierarchy in the data is captured by normal random effects in the linear predictor, with the conjugate random effect (e.g., beta for binary data, gamma for count and time-to-event data) for additional flexibility to model overdispersion. In their and subsequent applications, the conjugate random effects are observation-specific and hence independent between observations within the same cluster. Their formal framework, however, does not imply that this is the only choice. Therefore, we consider a version without normal random effects but with a pair-specific conjugate (beta) random effect. Precisely, y_{ij} is Bernoulli with random probability $\Pi_{ij} = \theta_i \kappa_j$, where

$$\kappa_j = \frac{e^{\xi_0+\xi_1 x_j}}{1+e^{\xi_0+\xi_1 x_j}}, \qquad \theta_i \sim \mathrm{Beta}(\alpha, \beta),$$

where $x_1 = 0$ and $x_2 = 1$. The pair-specific table with values $\prod_j \Pi_{ij}^{y_{ij}}(1-\Pi_{ij})^{(1-y_{ij})}$ takes the form presented in Table 3.

Table 3: The probabilities of a pair-specific table of the SBM model

	y_{i2}	
y_{i1}	1	0
1	$\theta_i^2\kappa_1\kappa_2$	$\theta_i\kappa_1(1-\theta_i\kappa_2)$
0	$\theta_i\kappa_2(1-\theta_i\kappa_1)$	$(1-\theta_i\kappa_1)(1-\theta_i\kappa_2)$

It is clear that the outcomes, given the Beta random effect, are independent. However, this is not true for the marginal probabilities, as it should. Setting $\alpha = 1/\beta$ for identifiability reasons, we have that:

$$\begin{aligned} E(\theta_i) &= \frac{1}{\beta^2+1}, \\ E(\theta_i^2) &= \frac{\beta+1}{(\beta^2+1)(\beta^2+\beta+1)}. \end{aligned}$$

Integrating over the Beta random effect, the marginal pairwise probabilities $\lambda_{i,jk}$ are as in Table 4. Probability Table 4 is valid when $\beta \geq -1$, and independence corresponds to $\beta = 0$,

Table 4: The marginal probabilities $\lambda_{i,jk}$ of the SBM model

	y_{i2}	
y_{i1}	1	0
1	$\frac{\beta+1}{(\beta^2+1)(\beta^2+\beta+1)}\kappa_1\kappa_2$	$\frac{1}{\beta^2+1}\kappa_1 - \frac{\beta+1}{(\beta^2+1)(\beta^2+\beta+1)}\kappa_1\kappa_2$
0	$\frac{1}{\beta^2+1}\kappa_2 - \frac{\beta+1}{(\beta^2+1)(\beta^2+\beta+1)}\kappa_1\kappa_2$	$1 - \frac{1}{\beta^2+1}(\kappa_1+\kappa_2) + \frac{\beta+1}{(\beta^2+1)(\beta^2+\beta+1)}\kappa_1\kappa_2$

which eases interpretation. Should β have been written as a function of α, rather than the other way around, independence would correspond to $\alpha = +\infty$, which is cumbersome. Importantly, fitting the marginal model directly allows for negative association. The log-likelihood can be expressed as:

$$\ell = \sum_{i=1}^{N} \sum_{j,k=0,1} n_{i,jk} \log(\lambda_{i,jk}).$$

If no other covariates than an indicator for case versus control are used, the log-likelihood simplifies to $\ell = \sum_{j,k=0,1} n_{jk} \log(\lambda_{jk})$. Fitting this model is straightforward.

Using the expressions in the margins of Table 4, it can be shown that

$$\log\left\{\frac{P(y_{ij}=1|x_{ij})}{P(y_{ij}=0|x_{ij})}\right\} = \operatorname{logit}\left(\frac{1}{\beta^2+1}\operatorname{expit}(\xi_0+\xi_1 x_{ij})\right), \tag{6}$$

showing that the rhs of this equation is not of the linear form $\alpha + \beta_M x_{ij}$ as in (1), unless there is no intra-pair association ($\beta = 0$, implying $\xi_1 = \beta_M$). Also, the marginal OR for the effect of x on y, can be derived as

$$\mathrm{OR}_M = \frac{\kappa_2(\beta^2 + 1 - \kappa_1)}{\kappa_1(\beta^2 + 1 - \kappa_2)}, \tag{7}$$

which reduces to $\exp(\xi_1)$ in case $\beta = 0$. The equivalent marginal slope can be obtained by $\log(\mathrm{OR}_M)$. The intra-pair correlation equals

$$\rho = \frac{\beta^3}{(\beta^2 + \beta + 1)} \sqrt{\frac{\kappa_1 \kappa_2}{(\beta^2 + 1 - \kappa_1)(\beta^2 + 1 - \kappa_2)}}.$$

3 Prime minister's performance data

3.1 Original MP data

Estimates according to the different methods are shown in Table 5. Ignoring the correlated nature of matched pairs, LR provides the correct point estimates for marginal slope (-0.163) and corresponding OR ($0.849 = (880\times 656)/(944\times 720)$). But the estimated standard error and the OR confidence intervals are too large, as LR ignores the paired nature of the data. These standard errors are corrected (from 0.072 to 0.039) in almost exactly the same way by all marginal methods, and marginalized pair-specific methods. Only the approximate marginal estimates for NRI are somewhat different (e.g., OR 0.874). This approximation works well for σ_u small, but the estimate $\hat{\sigma}_u = 5.159$ is quite large (as compared to the mean 1.242 of the random intercept). Note that the use of a probit link would imply the estimation of the marginal effect from the NRI model, as the marginal effect equals the pair-specific effect multiplied with $1/\sqrt{1 + \sigma_u^2}$ (Zeger et al., 1988).

One can observe that the loss of efficiency of GEE, as compared to the other full likelihood approaches, is seemingly negligible. Note that all estimates and standard errors of GEE-ind and GEE-exch are identical. See Supplementary Material A for explicit analytical calculations for these identical solutions, which also imply that the AIC value of logistic regression is exactly the same as the QIC of GEE-ind and GEE-exch.

The estimates for the intra-pair correlation and corresponding standard errors (if available) are close for all methods (about 0.70 and 0.02 respectively). Not unexpectedly, the intra-pair correlation is quite high in this setting. The approximate estimates for NRI are quite poor (0.89 and 0.013 respectively). Other and improved approximations for estimating the intra-pair correlation from the NRI model exist, but were not considered further in this review (Molenberghs et al., 2012).

As expected, the pair-specific slope of the LRF is incorrectly estimated as -1.113, twice the estimate -0.556 of all other pair-specific slope estimates. The estimated standard error and OR confidence interval are identical for all pair-specific models (s.e. 0.135 and OR CI (0.440;0.748)); see also Neuhaus et al. (1994) and Rice (2008). The exact CI for the CMH/CLR is a little bit wider.

The estimated variance components for the normal NRI and bridge BRI are very similar. Those of the MMM and CBM models are (not surprisingly) substantially different. The AIC value for all full likelihood models equals 3508.3. That of the CBM model is a slightly larger.

Results of the SBM model are not shown, because of the extreme high instability. No starting values and other options could be identified for obtaining the estimates.

Table 5: Original MP data: slope estimate (se), odds ratio and confidence interval, intra-pair correlation estimate, variance component estimate (if present), AIC or QIC, for the different methods. Type of method: M (marginal) and P (pair-specific), CI type: E (Exact), P (profile likelihood), and W (Wald)

method	type	slope (se)	OR	CI (type)	cor(se)	var comp	AIC/QIC*
LR	M	-0.163 (0.072)	0.849	(0.738,0.977) (P)	-	-	4372.0
LRF	P	-1.113 (0.191)	0.329	(0.225,0.477) (P)	-	-	3821.2
CMH/CLR	P	-0.556 (0.135)	0.573	(0.440,0.747) (W)	-	-	311.6
CMH/CLR	P	-0.556 (0.135)	0.573	(0.435,0.752) (E)	-	-	311.6
NRI	P	-0.556 (0.135)	0.573	(0.421,0.725) (W)		$\hat{\sigma}_u = 5.159(0.353)$	3508.3
	M$^{(\dagger)}$	-0.174 (0.042)	0.840	(0.772,0.909) (W)	0.890 (0.013)		3508.3
GEE-ind	M	-0.163 (0.039)	0.849	(0.787,0.917) (W)	-	-	4372.0*
GEE-exch	M	-0.163 (0.039)	0.849	(0.787,0.917) (W)	0.702 (-)	-	4372.0*
BLR	M	-0.163 (0.039)	0.849	(0.784,0.914) (W)	0.702 (0.018)	-	3508.3
BRI	P	-0.556 (0.135)	0.573	(0.421,0.725) (W)		$\hat{\sigma}_b = 5.907(0.396)$	3508.3
	M	-0.163 (0.039)	0.849	(0.784,0.914) (W)	0.707 (0.018)		3508.3
MMM	P	-0.507 (0.123)	0.602	(0.457,0.748) (W)		$\hat{\sigma}_m = 2.938(0.197)$	3508.3
	M	-0.163 (0.039)	0.849	(0.784,0.914) (W)	-		3508.3
CBM	M	-0.164 (0.039)	0.850	(0.783,0.914) (W)	0.706 (0.019)	$\hat{\sigma}_{\Pi_{i1}} = 0.413(0.005)$ $\hat{\sigma}_{\Pi_{i2}} = 0.418(0.005)$	3508.3

($\dagger$): $\rho \approx \sigma_{RI}^2/(\sigma_{RI}^2 + \pi^2/3)$; $\beta_M \approx \frac{\beta_P}{\sqrt{1+(\frac{16\sqrt{3}}{15\pi}\sigma_u)^2}}$ (working well only for small σ_u)

3.2 Reversed MP data

Consider the lower panel of Table 1, in which the outcomes of the first survey in the original MP data are reversed, implying a negative association between the two responses of the same participant. As compared to those of the original MP, the results of the analyses of the reversed MP data show some remarkable similarities and differences. First, some models are constrained to accommodate only a positive intra-pair correlation: NRI, BRI, MMM, CBM. These models can be extended with a particular pair-specific slope as follows:

NRI2

Referring to model (2) with the $\alpha_i = \alpha + u_i, u_i \sim N(0, \sigma_u^2)$, the random intercept is accompanied by a particular shared random slope

$$\beta_i = \beta_P + v_i, \quad v_i = -2u_i, \tag{8}$$

implying that for $x_{ij} = 0$ (first occasion) the model remains unchanged, whereas for $x_{ij} = 1$ (second occasion), we get $\alpha - u_i + \beta_P$ (as rhs in model (2)). So the same pair-specific intercept works with opposite sign on both occasions.

BRI2

The same modification, but with u_i having the bridge distribution.

MMM2

Similarly to NRI2, the rhs of model (3) is changed to $\delta_{ij} + u_i + v_i x_{ij}, v_i = -2u_i$.

CBM2

Model (5) is modified into $\Pi_{ij} \sim F_j^{-1}(\Phi(u_i + v_i x_{ij})), v_i = -2u_i$.

Of course, the reversal changes the value and the interpretation of the parameter β_P; the derived marginal slope parameter β_M however remains unchanged. Results are shown in Table 6. Note that the unmodified pair-specific models NRI, BRI, MMM, and CBM all reduce to LR (since all variance components are estimated as 0.000). Switching to the modified models NRI2, BRI2, MMM2, and CBM2, we observe that all marginal (and marginalized pair-specific) models lead to essentially the same point estimates for the slope (0.565) and the OR (1.759), except for the approximate value of the NRI2 model. In this reversed case, the standard errors are corrected in the opposite direction (0.093, larger than 0.072 for LR), and the CI's for the OR are wider. Note that the estimates for the intra-pair correlation (with opposite sign) and their standard errors, the variance component estimates and the AIC values for all these models NRI2, BRI2, MMM2, and CBM2 are exactly the same as for the models NRI, BRI, MMM, and CBM applied the original MP data.

For these reversed MP data, the SMB model does not need any modification to cover a negative intra-pair correlation and now fits easily with (almost exactly) the same estimates as the NRI2, BRI2, MMM2, and CBM2 models. Note the negative variance for the θ random scale factor. The GEE-ind and GEE-exch both provide the right estimates, and without any modification GEE-exch estimates the correlation as negative (-0.702).

Finally, it can be observed that the AIC and QIC values of all (modified) models for the reversed MP data are identical to the corresponding models for the original MP data, except for conditional logistic regression (CLR). Moreover, the estimates for the variance components of the NRI2, BRI2, MMM2, and CBM2 models applied to the reversed MP data are the same as those of the corresponding NRI, BRI, MMM, and CBM models applied to the original MP data.

Table 6: Reversed MP data: estimates of effect (slope with se) and of odds ratio (with confidence interval) according to the different methods and models.

method	type	slope (se)	OR	CI (type)	cor(se)	var comp	AIC/QIC*
LR	M	0.565 (0.072)	1.759	(1.529,2.024) (P)	-	-	4372.0
LRF	P	0.663 (0.078)	1.940	(1.667,2.260) (P)	-	-	6909.9
CMH/CLR	P	0.331 (0.055)	1.393	(1.251,1.551) (W)	-	-	1856.0
CMH/CLR	P	0.331 (0.055)	1.393	(1.249,1.554) (E)	-	-	1856.0
NRI	P=M$^{(\dagger)}$	0.565 (0.071)	1.759	(1.512,2.005) (W)	0.000 (-)	$\hat{\sigma}_u = 0.000(0.043)$	4374.0
NRI2	P	1.929 (0.336)	6.879	(2.340,11.418) (W)		$\hat{\sigma}_u = 5.159(0.353)$	3508.3
NRI2	M$^{(\dagger)}$	0.604 (0.099)	1.829	(1.472,2.185) (W)	-0.890 (0.013)		3508.3
GEE-ind	M	0.565 (0.093)	1.759	(1.465,2.112) (W)	-	-	4372.0*
GEE-exch	M	0.565 (0.093)	1.759	(1.465,2.112) (W)	-0.702 (-)	-	4372.0*
BLR	M	0.565 (0.093)	1.759	(1.437,2.081) (W)	-0.702 (0.018)	-	3508.3
BRI	P=M	0.565 (0.071)	1.759	(1.512,2.005) (W)	0.000 (-)	$\hat{\sigma}_u = 0.000(0.054)$	4374.0
BRI2	P	1.924 (0.334)	6.845	(2.362,11.328) (W)		$\hat{\sigma}_b = 5.907(0.396)$	3508.3
BRI2	M	0.565 (0.093)	1.759	(1.437,2.081) (W)	-0.707 (0.018)		3508.3
MMM	P=M	0.565 (0.072)	1.759	(1.512,2.005) (W)	-	$\hat{\sigma}_m = 0(0.027)$	4374.0
MMM2	P	1.752 (0.306)	5.768	(2.310,9.225) (W)		$\hat{\sigma}_m = 2.938(0.197)$	3508.3
MMM2	M	0.565 (0.093)	1.759	(1.437,2.081) (W)	-		3508.3
CBM	M	0.565 (0.072)	1.759	(1.512,2.007) (W)	0.000 (-)	$\hat{\sigma}_{\Pi_{i1}} = 0.000$ (-)	4374.0
						$\hat{\sigma}_{\Pi_{i2}} = 0.000$ (-)	
CBM2	M	0.565 (0.094)	1.759	(1.434,2.085) (W)	-0.707 (0.018)	$\hat{\sigma}_{\Pi_{i1}} = 0.414(0.006)$	3508.3
						$\hat{\sigma}_{\Pi_{i2}} = 0.418(0.006)$	
SBM	M	0.565 (0.093)	1.759	(1.437,2.081) (W)	-0.702 (0.018)	$\hat{\sigma}^2_\theta = -0.242(0.002)$	3508.3

(†): $\rho \approx \sigma^2_{RI}/(\sigma^2_{RI} + \pi^2/3)$; $\beta_M \approx \frac{\beta_P}{\sqrt{1+(\frac{16\sqrt{3}}{15\pi}\sigma_u)^2}}$ (working well only for small σ_u)

4 Myocardial infarction data

Here we consider both directions: the design is according to $x|y$ whereas the objective of inference focuses on $y|x$. The key point is the symmetry property OR($x|y$)=OR($x|y$). The results are shown in Table 7. When using SAS `nlmixed`, the Wald CI for the slope was exponentiated to obtain the CI for the OR. The reason is the poor performance of the approximate delta-method in this particular setting. Ordinary LR, ignoring the (case,control)-pairing, leads to slope estimate 0.804 and OR estimate 2.234, irrespective of the direction. By accounting for matching a control to a case (according to age and gender), one expects to obtain a smaller standard error and more narrow confidence interval. But one expects in this situation the intra-pair correlation to be much smaller than for the MP data, and hence the gain in accuracy is expected to be rather limited. Similarly, one expects the pair-specific slope and OR to be only slightly larger than the marginal versions. And indeed, this is confirmed by the CMH/CLR results, with a slope estimate 0.838 and OR estimate 2.312. Again, identical results are obtained for both directions. All other pair-specific models (NRI, BRI) confirm these pair-specific results, but only for the direction $x|y$. Not surprisingly, as $(y_{i1}, y_{i2}) \equiv (0, 1)$, the pair-specific NRI and BRI results equal their marginal results and the LR results in the other direction $y|x$. The variance component and corresponding correlation estimates all equal 0.

Turning to the marginal estimates, the BRI slope estimate 0.804 shows a standard error estimate 0.278, only slightly smaller than the estimate 0.284 from LR. So, the matching results in an almost ignorable increase in efficiency in this case. The error of the approximation of the marginal estimates for the NRI model is relatively small. The GEE-ind (in both directions) and GEE-exch (in the $x|y$ direction) lead to the same results as the marginal estimates derived from the BRI model. The correlation estimate 0.04 from GEE-exch confirms that the correlation within a pair is close to ignorable. For GEE-exch in the direction $y|x$, one gets a slope estimate 0 and a correlation estimate -0.999 (See Supplementary Material A). The $x|y$ version of the BLR Bahadur model are identical to the GEE results. The $y|x$ version however provides different estimates for slope and OR, but the intra-pair correlation is again -1 (see Supplementary Material C). That slope/OR estimates are different is related to the restrictions of the parameter space in the Bahadur model for negative correlations (see Aerts et al. (2002)).

The results for MMM are close to those of BRI, for both directions $x|y$ and $y|x$. For the CBM and SBM marginal models, the results for $x|y$ are similar to the other marginal results, with a correlation estimate of 0.04. In the other direction $y|x$, the CBM method estimates the correlation as 0, resulting in the same slope and OR estimates as LR. In this case, SBM is identical to GEE-exch with 0 slope and correlation -1.

As one would expect, given the case-control design, all marginal models for $y|x$ estimate the marginal probability that individual j of pair i is a case as 0.5. That probability

$$P(y_{ij}=1|x_{ij}=1)P(x_{ij}=1)+P(y_{ij}=1|x_{ij}=0)P(x_{ij}=0),$$

is estimated as

$$\frac{e^{\hat{\alpha}_M+\hat{\beta}_M}}{1+e^{\hat{\alpha}_M+\hat{\beta}_M}}\times\frac{71}{288}+\frac{e^{\hat{\alpha}_M}}{1+e^{\hat{\alpha}_M}}\times\frac{217}{288}=0.5. \tag{9}$$

for different pairs of estimates $(\hat{\alpha}_M, \hat{\beta}_M)$ depending on the method. Figure 1 in Supplementary Material D shows those pairs for which equation (9) holds.

Table 7: MI data: estimates of effect (slope with se) and of odds ratio (with confidence interval; profile likelihood PI, exact E, Wald W, or exponentiated Wald e^W) according to the different methods and models.

method	order	type	slope (se)	OR	CI (type)	cor(se)	var comp	(A/Q*)IC ($x\|y, y\|x$)
LR	$x\|y$ & $y\|x$	M	0.804 (0.284)	2.234	(1.292,3.938) (P)	-	-	394.9, 317.3
LRF	$x\|y$ & $y\|x$	P	1.677 (0.423)	5.348	(2.380,12.579) (P)	-	-	419.8, 672.1
CMH/CLR	$x\|y$ & $y\|x$	P	0.838 (0.299)	2.312	(1.286,4.157) (W)	-	-	66.9, 193.1
CMH/CLR	$x\|y$ & $y\|x$	P	0.838 (0.299)	2.313	(1.255,4.453) (E)	-	-	
NRI	$x\|y$	P	0.838 (0.299)	2.312	(1.280,4.178) (e^W)		$\hat{\sigma}_u = 0.490(0.557)$	319.1
	$x\|y$	M$^{(\dagger)}$	0.784 (0.273)	2.190	(1.276,3.760) (e^W)	0.068 (0.144)		319.1
	$y\|x$	P=M$^{(\dagger)}$	0.804 (0.284)	2.234	(1.276,3.913) (e^W)	0.000 (-)	$\hat{\sigma}_u = 0.000(0.121)$	396.9
GEE-ind	$x\|y$ & $y\|x$	M	0.804 (0.278)	2.234	(1.296,3.852) (W)	-	-	317.3*, 392.9*
GEE-exch	$x\|y$	M	0.804 (0.278)	2.234	(1.296,3.852) (W)	0.04	-	317.3*
GEE-exch	$y\|x$	M	0.000 (0.000)	1.000	-	-0.999	-	399.3*
BLR	$x\|y$	M	0.804 (0.278)	2.234	(1.290,3.870) (e^W)	0.040 (0.085)	-	319.1
BLR	$y\|x$	M	0.528 (0.321)	1.695	(0.899,3.197) (e^W)	-1 (0)	-	202.9
BRI	$x\|y$	P	0.838 (0.299)	2.312	(1.280,4.178) (e^W)		$\hat{\sigma}_b = 0.536(0.612)$	319.1
BRI	$x\|y$	M	0.804 (0.278)	2.234	(1.290,3.870) (e^W)	0.041 (0.088)		319.1
BRI	$y\|x$	P=M	0.804 (0.284)	2.234	(1.276,3.913) (e^W)	0.000 (-)	$\hat{\sigma}_b = 0.000(0.148)$	396.9
MMM	$x\|y$	P	0.836 (0.297)	2.307	(1.282,4.153) (e^W)		$\hat{\sigma}_u = 0.286(0.325)$	319.1
MMM	$x\|y$	M	0.804 (0.278)	2.234	(1.290,3.870) (e^W)	-		319.1
MMM	$y\|x$	P=M	0.804 (0.284)	2.234	(1.276,3.913) (e^W)	-		396.9
CBM	$x\|y$	M	0.804 (0.278)	2.234	(1.290,3.871) (e^W)	0.040 (0.086)	$\hat{\sigma}_{\Pi_{i1}} = 0.076(0.082)$ $\hat{\sigma}_{\Pi_{i2}} = 0.093(0.100)$	319.1
CBM	$y\|x$	M	0.804 (0.283)	2.234	(1.276,3.912) (e^W)	0.000 (-)	$\hat{\sigma}_{\Pi_{i1}} = 0.000(-)$ $\hat{\sigma}_{\Pi_{i2}} = 0.000(-)$	396.9
SBM	$x\|y$	M	0.804 (0.278)	2.234	(1.290,3.870) (e^W)	0.040 (0.085)		319.1
SBM	$y\|x$	M	0 (-)	1	(-,-) (W)	-1 (0.000)		205.6

($\dagger$): $\rho \approx \sigma^2_{RI}/(\sigma^2_{RI} + \pi^2/3)$; $\beta_M \approx \frac{\beta_P}{\sqrt{1+(\frac{16\sqrt{3}}{15\pi}\sigma_u)^2}}$ (working well only for small σ_u)

5 Further discussion

Several methods and models for analysing binary data from matched pairs have been reviewed, some of which are standard approaches for matched 2×2 tables that have been in use for a long time, while others have been developed more recently for more general hierarchical data structures. The application of the latter approaches to this basic setting allows us to i) illustrate that they all can be applied to such basic data, ii) compare the resulting estimates from different types of models (pair specific, marginal, marginalisable), iii) discuss methodological connections and insights between all approaches. In particular, there are settings where always a boundary solution is obtained. Theoretical arguments are given to explain this.

Note that almost all methods produce results that are identical as soon as association is accommodated in some form. This is reassuring in the sense that it offers a good amount of robustness of model choice in the matched pair setting considered here. The user may then select the simplest and/or most readily available tool, such as, for example GEE, with implementations in R (`geepack`) and SAS (the GENMOD and GEE procedures) among others. Of course, when data would be incomplete, then the likelihood-based methods have the advantage of being valid under missingness at random, provided regularity conditions hold, whereas the semi-parametric methods are only valid under missingness completely at random. Covariates can easily be included in the regression models, whether conditional, marginal, or random-effects based, while an ad-hoc method such as Cochran-Mantel-Haenszel does not allow for general covariate settings. Of note, the equality of results across classes of methods do not automatically carry over to these more general settings.

An obvious question is to what extent our findings extend to more general settings, such as multinomial and ordinal data, inclusion of explanatory variables, longitudinal data with more than 2 occasions, case-control studies with multiple controls per case, etc. Partial results and insights are available, scattered over literature, but an equivalent exhaustive study including all methods and models discussed here, is considered as an interesting but challenging avenue of further research.

In this paper, we have considered a number of likelihood-based and semi-parametric methods, but not Bayesian alternatives, even though categorical data models and contingency tables have received attention in the literature (Albert and Chib, 1993; Polson et al., 2013; Ishwaran and James, 2002). Arguably, the Bayesian framework deserves a separate, proper treatment, where the choice of priors (non-informative, informative, and regularizing to accommodate constraints) is examined. Informative priors might be based, for example, on historic studies, whether also of a matched pairs nature or not. Also, MCMC based estimation methods will not automatically lead to identical estimates whereas frequentist methods discussed here do so, always or for certain designs and/or data configurations.

Supplementary Materials

Supplementary materials include: A) a derivation of the point and variance estimators for GEE, prospective pairs, B) a derivation of the point and variance estimators for GEE, retrospective pairs, C) a derivation of the point estimators for BLR, retrospective pairs, D) a figure of $\hat{\alpha}_M$ and $\hat{\beta}_M$ for which equation (9) holds, E) SAS code for the reversed MP data.

Acknowledgments

We would like to thank the editor and referees for very helpful feedback and suggestions. Any remaining errors are the responsibility of the authors.

Disclosure Statement

This article did not receive specific funding, and the authors have no competing interests to declare.

References

Aerts, M., G. Molenberghs, L. M. Ryan, and H. Geys (2002). *Topics in modelling of clustered data*. CRC Press.

Agresti, A. (2002). In *Categorical Data Analysis, 2nd Edition*. Wiley.

Albert, J. and S. Chib (1993). Bayesian analysis of binary and polychotomous response data. *Journal of the American Statistical Association 88*, 669–679.

Bahadur, R. R. (1961). A representation of the joint distribution of responses to n dichotomous items. *Studies in item analysis and prediction 6*, 158–168.

Breslow, N. and N. E. Day (1980). In *Statistical Methods in Cancer Research, Vol. I, The Analysis of Case-Control Studies.* Lyon: IARC.

Ishwaran, H. and L. F. James (2002). Approximate dirichlet process computing in finite normal mixtures. *Journal of Computational and Graphical Statistics 11*, 508–532.

Liang, K.-Y. and S. Zeger (1986). Longitudinal data analysis using generalized linear models. *Biometrika 73*, 13–22.

Liu, L. and Z. Yu (2008). A likelihood reformulation method in non-normal random effects models. *Statistics in medicine 27*(16), 3105–3124.

McNemar, Q. (1947). Note on the sampling error of the difference between correlated proportions or percentages. *Psychometrika 12*, 153–157.

Milanzi, E., G. Molenberghs, A. Alonso, G. Verbeke, and P. D. Boeck (2015). Reliability measures in item response theory: Manifest versus latent correlation functions. *British Journal of Mathematical and Statistical Psychology 68*, 43–64.

Molenberghs, G., M. G. Kenward, G. Verbeke, S. Iddi, and A. Efendi (2013). On the connections between bridge distributions, marginalized multilevel models, and generalized linear mixed models.

Molenberghs, G. and G. Verbeke (2005). In *Models for Discrete Longitudinal Data*. New York: Springer.

Molenberghs, G., G. Verbeke, and C. G. Demétrio (2007). An extended random-effects approach to modeling repeated, overdispersed count data. *Lifetime data analysis 13*, 513–531.

Molenberghs, G., G. Verbeke, C. G. Demétrio, and A. Vieira (2010). A family of generalized linear models for repeated measures with normal and conjugate random effects. *Statistical Science 35*, 325–347.

Molenberghs, G., G. Verbeke, S. Iddi, and C. G. Demétrio (2012). A combined beta and normal random-effects model for repeated, overdispersed binary and binomial data. *Journal of Multivariate Analysis 111*, 94–109.

Nelson, K. P., S. R. Lipsitz, G. M. Fitzmaurice, J. Ibrahim, M. Parzen, and R. Strawderman (2006). Use of the probability integral transformation to fit nonlinear mixed-effects models with nonnormal random effects. *Journal of Computational and Graphical Statistics 15*(1), 39–57.

Neuhaus, J., J. Kalbfleisch, and W. Hauck (1994). Conditions for consistent estimation in mixed-effects models for binary matched-pairs data. *Can. J. Stat. 22*, 139–148.

Polson, N. G., J. G. Scott, and J. Windle (2013). Bayesian inference for logistic models using pólya–gamma latent variables. *1339-1349 108*.

Rice, K. (2008). Equivalence between conditional and random-effects likelihoods for pair-matched case-control studies. *J. Amer. Stat. Ass. 103*, 385–396.

Wang, Z. and T. Louis (2003). Matching conditional and marginal shapes in binary random intercept models using a bridge distribution function. *Biometrika 90*, 765–775.

Zeger, S., K.-Y. Liang, and P. Albert (1988). Models for longitudinal data: A generalized estimating equation approach. *Biometrics 44*, 1049–1060.

Supplementary Materials

A Point and variance estimators for GEE, prospective pairs

In this supplementary material, we are concerned with the case where all pairs have $x_{i1} = 0$ and $x_{i2} = 1$. We will show that the GEE estimator is independent of the working correlation chosen and that the same is true for the robust variance estimator. The naive variance estimator of the intercept is still independent of the working correlation structure, but not that of the key regression parameter β_M.

We use the following notation:

$$P(Y_{ij} = 1|x_{ij}) = \frac{e^{\alpha+\beta_M x_{ij}}}{1 + e^{\alpha+\beta_M x_{ij}}}$$

where $x_{i1} = 0$ and $x_{i2} = 1$. Let the number of pairs with $Y_{i1} = r$ and $Y_{i2} = s$ be m_{rs}, $(r, s = 0, 1)$, with

$$\sum_{r=0}^{1}\sum_{s=0}^{1} m_{rs} = n.$$

The two marginal probabilities are:

$$\mu_{i1} = \mu_1 = \frac{e^{\alpha}}{1 + e^{\alpha}}, \tag{10}$$
$$\mu_{i2} = \mu_2 = \frac{e^{\alpha+\beta_M}}{1 + e^{\alpha+\beta_M}}. \tag{11}$$

The corresponding variances are $v_j = \mu_j(1 - \mu_j)$, $(j = 1, 2)$.

The generalized estimating equations take the form:

$$U = \sum_{i=1}^{n} D_i' V_i^{-1}(\boldsymbol{Y}_i - \boldsymbol{\mu}_i), \tag{12}$$

with $\boldsymbol{Y}_i = (Y_{i1}, Y_{i2})'$ and, in this case, $\boldsymbol{\mu}_i = \boldsymbol{\mu} = (\mu_1, \mu_2)'$. Further,

$$
\begin{aligned}
D_i = D &= \frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\beta}} = \begin{pmatrix} v_1 & 0 \\ v_2 & v_2 \end{pmatrix}, && (13) \\
V_i = V &= \begin{pmatrix} \sqrt{v_1} & 0 \\ 0 & \sqrt{v_2} \end{pmatrix} \cdot \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \cdot \begin{pmatrix} \sqrt{v_1} & 0 \\ 0 & \sqrt{v_2} \end{pmatrix} \\
&= \begin{pmatrix} v_1 & \rho\sqrt{v_1 v_2} \\ \rho\sqrt{v_1 v_2} & v_2 \end{pmatrix}, && (14)
\end{aligned}
$$

with ρ the assumed working correlation.

Given the constant nature of matrices (13)–(14), (12) simplifies to:

$$
\begin{aligned}
U &= D'V^{-1}\left[m_{11}\begin{pmatrix} 1-\mu_1 \\ 1-\mu_2 \end{pmatrix} + m_{10}\begin{pmatrix} 1-\mu_1 \\ -\mu_2 \end{pmatrix} + m_{01}\begin{pmatrix} -\mu_1 \\ 1-\mu_2 \end{pmatrix} + m_{00}\begin{pmatrix} -\mu_1 \\ -\mu_2 \end{pmatrix}\right] \\
&= D'V^{-1}\begin{pmatrix} m_{11} + m_{10} - n\mu_1 \\ m_{11} + m_{01} - n\mu_2 \end{pmatrix},
\end{aligned}
$$

from which it follows that

$$
\widehat{\mu}_1 = \frac{m_{11} + m_{10}}{n}, \qquad \widehat{\mu}_2 = \frac{m_{11} + m_{01}}{n},
$$

and, using (10) and (11):

$$
\begin{aligned}
\widehat{\alpha} &= \log\left(\frac{m_{11} + m_{10}}{m_{00} + m_{01}}\right), && (15) \\
\widehat{\beta}_M &= \log\left(\frac{m_{11} + m_{01}}{m_{00} + m_{10}}\right) - \log\left(\frac{m_{11} + m_{10}}{m_{00} + m_{01}}\right). && (16)
\end{aligned}
$$

Clearly, the solution to the estimating equations does not depend on the matrices D and V. In particular, the estimators do not depend on the working correlation.

Turning to variance estimation, define:

$$
\begin{aligned}
I_0 &= \sum_{i=1}^{n} D_i' V_i^{-1} D_i = nD'V^{-1}D, && (17)\\
I_1 &= \sum_{i=1}^{n} D_i' V_i^{-1} \mathrm{var}(\boldsymbol{Y}_i) V_i^{-1} D_i = nD'V^{-1}WV^{-1}D, && (18)
\end{aligned}
$$

where

$$
W = \frac{1}{n} \sum_{i=1}^{n} \mathrm{var}(\boldsymbol{Y}_i).
$$

The naive (a.k.a. model based) variance estimator is $V_{n,\beta}(\rho) = I_0^{-1}$ and the robust (a.k.a. empirically corrected) estimator is $V_{r,\beta}(\rho) = I_0^{-1} I_1 I_0^{-1}$. It follows from (17), (13), and (14), that

$$
\begin{aligned}
I_0 &= \frac{n}{1-\rho^2} \begin{pmatrix} v_1 - 2\rho\sqrt{v_1 v_2} + v_2 & -\rho\sqrt{v_1 v_2} + v_2 \\ -\rho\sqrt{v_1 v_2} + v_2 & v_2 \end{pmatrix}, \\
V_{n,\beta}(\rho) &= \frac{1}{n v_1 v_2} \begin{pmatrix} v_2 & \rho\sqrt{v_1 v_2} - v_2 \\ \rho\sqrt{v_1 v_2} - v_2 & v_1 - 2\rho\sqrt{v_1 v_2} + v_2 \end{pmatrix}. && (19)
\end{aligned}
$$

Clearly, the naive variance of $\widehat{\alpha}$ does not depend on the working correlation, but the variance of $\widehat{\beta}_M$ does and obviously also the covariance (and correlation) between them.

Before computing the robust variance, it is helpful to first consider the moment-based estimator of the correlation, which is also what would be used should an exchangeable working structure be adopted (as well as other structures with non-zero correlation, given that they coincide for paired data, such as auto-regressive, and unstructured).

$$
\begin{aligned}
\rho_e &= \frac{1}{n} \sum_{i=1}^{n} \frac{(y_{i1} - \mu_1)(y_{i2} - \mu_2)}{\sqrt{v_1 v_2}} \\
&= \frac{1}{n\sqrt{v_1 v_2}} \left[m_{11}(1-\mu_1)(1-\mu_2) - m_{10}(1-\mu_1)\mu_2 - m_{01}\mu_1(1-\mu_2) + m_{00}\mu_1\mu_2 \right] && (20)
\end{aligned}
$$

To calculate W, consider

$$\sum_{i=1}^{n} \mathrm{var}(\boldsymbol{Y}_i) = m_{11} \begin{pmatrix} 1-\mu_1 \\ 1-\mu_2 \end{pmatrix} (1-\mu_1, 1-\mu_2) + m_{10} \begin{pmatrix} 1-\mu_1 \\ -\mu_2 \end{pmatrix} (1-\mu_1, -\mu_2)$$
$$+ m_{01} \begin{pmatrix} -\mu_1 \\ 1-\mu_2 \end{pmatrix} (-\mu_1, 1-\mu_2) + m_{00} \begin{pmatrix} -\mu_1 \\ -\mu_2 \end{pmatrix} (-\mu_1, -\mu_2)\,.$$

Some simple algebra produces:

$$W = \begin{pmatrix} v_1 & \rho_e\sqrt{v_1 v_2} \\ \rho_e\sqrt{v_1 v_2} & v_2 \end{pmatrix},$$

which equals (14) for the specific case when the exchangeable correlation (20) is used.

With exchangeable correlation, $W = V$, hence $I_1 = I_0$ and robust and naive variances are equal.

Furthermore, it is easy to show that all robust variance estimators are equal, regardless of the working correlation. It follows from the fact that D and V are constant across observations and, moreover, that D is a square, invertible matrix:

$$I_0^{-1} I_1 I_0^{-1} = \frac{1}{n} D^{-1} V (D')^{-1} D' V^{-1} W V^{-1} D D^{-1} V (D')^{-1} = \left[n D' W^{-1} D \right]^{-1}.$$

So, in conclusion, the point estimators are independent of the working correlation chosen, with the same holding for the robust variance estimator. The naive variance estimator for $\widehat{\alpha}$ is also independent of the correlation, but this is not true for the naive variance of $\widehat{\beta}_M$ – the key parameter – which is correct only when the moment-based estimator for the correlation is used.

B Point and variance estimators for GEE, retrospective pairs

Consider now the 'reverse' situation of the previous supplementary material, where for every pair $y_{i1} = 0$ and $y_{i2} = 1$, and where hence the data are characterized by n_{rs}, the pairs with $x_{i1} = r$ and $x_{i2} = s$, for $r, s = 0, 1$.

Considering GEE in this case, leads to the system of equations:

$$
\begin{aligned}
(1-\rho)n_{11}(1-2\mu_2) - [n_{10}\mu_2 - n_{01}(1-\mu_2)] &= \sqrt{\frac{v_2}{v_1}}\rho[n_{10}(1-\mu_1) - n_{01}\mu_1], && (21)\\
(1-\rho)n_{00}(1-2\mu_1) - [n_{01}\mu_1 - n_{10}(1-\mu_1)] &= \sqrt{\frac{v_1}{v_2}}\rho[n_{01}(1-\mu_2) - n_{10}\mu_2], && (22)
\end{aligned}
$$

where $\mu_1 = P(Y_{ij} = 1|x_{ij} = 0)$, $\mu_2 = P(Y_{ij} = 1|x_{ij} = 1)$, and $v_r = \mu_r(1-\mu_r)$, $(r = 1, 2)$.

Starting from $\rho = 0$, (21)–(22) immediately leads to:

$$
\begin{aligned}
\widehat{\mu_1} &= \frac{n_{00} + n_{10}}{2n_{00} + n_{10} + n_{01}},\\
\widehat{\mu_2} &= \frac{n_{11} + n_{01}}{2n_{11} + n_{01} + n_{10}},
\end{aligned}
$$

which, unsurprisingly, coincides with the logistic regression solution.

For general ρ, it follows that a solution is found by setting $\rho = -1$, $\mu_1 = \mu_2 = 1/2$, a degenerate solution. This is not unexpected, because in every pair one outcome is 0 and the other is 1, so maximal heterogeneity. This implies that, while GEE fitting in the setting of Supplementary Material A is straightforward and an interior solution is guaranteed unless one of the counts is equal to zero, in the current, matched-pairs setting, GEE with arbitrary (CS) correlation is infeasible.

Incidentally, the same can be same when employing the bivariate normal likelihood (even though the outcomes are binary). CS in this case takes the form $V = \sigma^2 I + dJ$, with I and J the two-dimensional identity and one matrix, respectively. Because the determinant

$|V| = \sigma^2(\sigma^2 + 2d)$ and the inverse takes the form

$$V^{-1} = \frac{1}{\sigma^2}\left(I - \frac{d}{\sigma^2 + 2d}J\right),$$

the kernel of the log-likelihood takes the form:

$$\begin{aligned} \ell &= -\frac{n}{2}\log[\sigma^2(\sigma^2 + 2d)] \\ &\quad -\sum_{r=0}^{1}\sum_{s=0}^{1}\frac{n_{rs}}{2}(-\mu_{r+1}; 1 - \mu_{s+1})\frac{1}{\sigma^2(\sigma^2 + 2d)}\begin{pmatrix} d + \sigma^2 & -d \\ -d & d + \sigma^2 \end{pmatrix}\begin{pmatrix} -\mu_{r+1} \\ 1 - \mu_{s+1} \end{pmatrix}. \end{aligned}$$

Now choosing $\mu_1 = \mu_2 = 1/2$, σ^2 arbitrary, and $d = -\frac{\sigma^2}{2} + \varepsilon$, leads to

$$\ell = -\frac{n}{2}\log(2\sigma^2) - \frac{n}{2}\log\varepsilon - \frac{n}{4\sigma^2},$$

from which it follows that

$$\ell \overset{\varepsilon \to +\infty}{\longrightarrow} +\infty,$$

which is coherent with the GEE result. This explains why both GEE and mixed-model software on data of this type either fails to converge or produces results very close to the above.

C Point estimators for BLR, retrospective pairs

Consider the same setting as in Supplementary Material B, with now the Bahadur model (BLR) applied. Adopt the same notation. The log-likelihood takes the form:

$$\begin{aligned}
\ell &= n_{11}\left[\log(1-\mu_2)+\log\mu_2+\log(1-\rho)\right] \\
&\quad +n_{10}\left[\log(1-\mu_2)+\log\mu_1+\log\left(1-\rho\sqrt{\frac{1-\mu_1}{\mu_1}}\sqrt{\frac{\mu_2}{1-\mu_2}}\right)\right] \\
&\quad +n_{01}\left[\log(1-\mu_1)+\log\mu_2+\log\left(1-\rho\sqrt{\frac{\mu_1}{1-\mu_1}}\sqrt{\frac{1-\mu_2}{\mu_2}}\right)\right] \\
&\quad +n_{00}\left[\log(1-\mu_1)+\log\mu_1+\log(1-\rho)\right]. \qquad (23)
\end{aligned}$$

To ensure that the correlation satisfies its range restriction, write

$$\rho = \frac{e^z-1}{e^z+1},$$

which of course implies that

$$z = \log\left(\frac{1+\rho}{1-\rho}\right).$$

Further, adopt the notation

$$\theta_k^2 = \frac{\mu_k}{1-\mu_k},$$

for $k = 1, 2$, implying

$$\mu_k = \frac{\theta_k^2}{1+\theta_k^2}, \qquad 1-\mu_k = \frac{1}{1+\theta_k^2}.$$

Using the new notation, log-likelihood (23) transforms to

$$\begin{aligned}
\ell &= n_{11}\left[-2\log(1+\theta_2^2)+\log\theta_2^2+\log 2-\log(e^z+1)\right] \\
&\quad +n_{10}\left[-\log(1+\theta_2^2)-\log(1+\theta_1^2)+\log\theta_1^2+\log\left(1-\frac{e^z-1}{e^z+1}\cdot\frac{\theta_2}{\theta_1}\right)\right] \\
&\quad +n_{01}\left[-\log(1+\theta_1^2)-\log(1+\theta_2^2)+\log\theta_2^2+\log\left(1-\frac{e^z-1}{e^z+1}\cdot\frac{\theta_1}{\theta_2}\right)\right] \\
&\quad +n_{00}\left[-2\log(1+\theta_1^2)+\log\theta_1^2+\log 2-\log(e^z+1)\right]. \qquad (24)
\end{aligned}$$

Calculating the score equation for z produces:

$$\frac{\partial\ell}{\partial z}=-(n_{11}+n_{00})\frac{e^z}{e^z+1}-n_{10}\frac{\frac{2e^z}{(e^z+1)^2}\cdot\frac{\theta_2}{\theta_1}}{1-\frac{e^z-1}{e^z+1}\cdot\frac{\theta_2}{\theta_1}}-n_{01}\frac{\frac{2e^z}{(e^z+1)^2}\cdot\frac{\theta_1}{\theta_2}}{1-\frac{e^z-1}{e^z+1}\cdot\frac{\theta_1}{\theta_2}}.$$

If $z\to-\infty$, then $e^z\to 0$, and $\frac{\partial\ell}{\partial z}\to 0$. The three functions in z, present in the expression for this derivative are monotonically increasing in z, for all values of θ_1 and θ_2. Given the minus signs, the denominators decrease, which implies that the derivative monotonically decreases with z, hence $z=-\infty$, corresponding to $\rho=-1$, is the only solution.

Calculating the other two derivatives, and plugging in the solution for the correlation parameter, yields:

$$\begin{aligned}
\frac{\partial\ell}{\partial\theta_1} &= n_{00}\left(\frac{-4\theta_1}{1+\theta_1^2}+\frac{2\theta_1}{\theta_1^2}\right)+n_{10}\left(\frac{2\theta_1}{\theta_1^2}-\frac{2\theta_1}{1+\theta_1^2}-\frac{\frac{\theta_2}{\theta_1^2}}{1+\frac{\theta_2}{\theta_1}}\right) \\
&\quad +n_{01}\left(-\frac{2\theta_1}{1+\theta_1^2}+\frac{\frac{1}{\theta_2}}{1+\frac{\theta_1}{\theta_2}}\right), \qquad (25) \\
\frac{\partial\ell}{\partial\theta_2} &= n_{11}\left(\frac{-4\theta_2}{1+\theta_2^2}+\frac{2\theta_2}{\theta_2^2}\right)+n_{01}\left(\frac{2\theta_2}{\theta_2^2}-\frac{2\theta_2}{1+\theta_2^2}-\frac{\frac{\theta_1}{\theta_2^2}}{1+\frac{\theta_1}{\theta_2}}\right) \\
&\quad +n_{10}\left(-\frac{2\theta_2}{1+\theta_2^2}+\frac{\frac{1}{\theta_1}}{1+\frac{\theta_2}{\theta_1}}\right). \qquad (26)
\end{aligned}$$

Consider the potential solution $\mu_1=\mu_2=0.5$, corresponding to $\theta_1=\theta_2=1$. This is a solution only if $n_{01}=n_{10}$, but not otherwise, unlike in the GEE case (Supplementary

Material B). Hence, the probabilities are to be found by solving (25)–(26). Given that most but not all terms drop for $\theta_1 = \theta_2 = 1$, the solution will typically not deviate a lot from it. Indeed, for these values, (25) equals $0.5(n_{10} - n_{01}$, with (26) its opposite.

D $\hat{\alpha}_M$ and $\hat{\beta}_M$ for which equation (9) holds

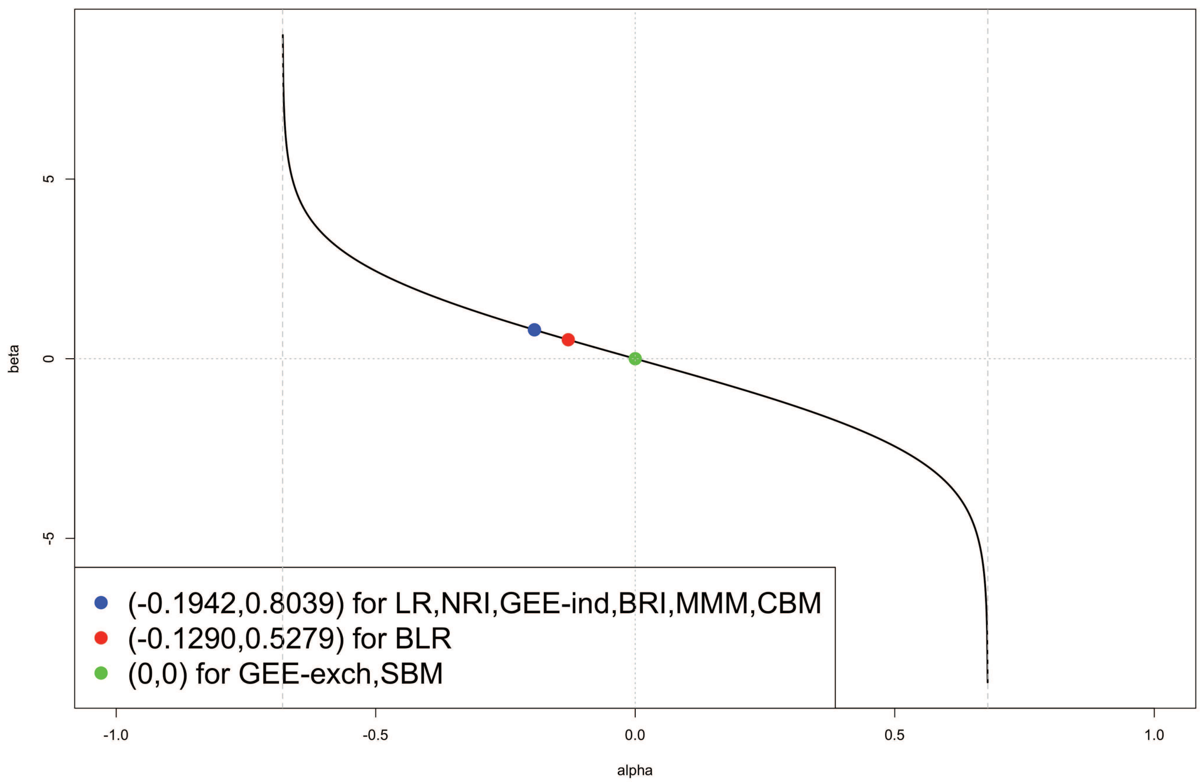

Figure 1: Formula (9): slope β_M as a function of the intercept α_M, with vertical asymptotes at $\text{logit}((0.5 - 71/288)/(217/288)) = -0.679$ and $\text{logit}(0.5/(217/288)) = 0.679$.

E SAS code for reversed MP data

The data analyses for this paper was generated using SAS software, Version 9.4 of the SAS System for Windows. Copyright ©2020 SAS Institute Inc. SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc., Cary, NC, USA.

```
data spm;
infile "spmrev.dat";
  input survey approval pair;
run;

/* LR: logistic regression ignoring the matching*/
proc logistic data=spm descending;
  model approval=survey/link=logit clodds=both expb;
run;

/* LRF: logistic regression with fixed effects for pairs*/
proc logistic data=spm descending;
  class pair;
  model approval=survey pair/link=logit clodds=both expb;
run;

/* CMH/CLR: approach*/
proc freq data=spm;
    tables pair*survey*approval / cmh alpha=0.05;
    exact COMOR;
run;
proc logistic data=spm;
   strata pair;
   model approval(event='1')=survey;
   exact survey / estimate=both;
```

```
run;

/* NRI: random intercept logistic regression with normal distribution*/
proc nlmixed data=spm qpoints=100;
  parms alpha=0.36 beta=-0.16 sig=0.5;
  pi=exp(alpha+u+beta*survey)/(1+exp(alpha+u+beta*survey));
  pi0=exp(alpha)/(1+exp(alpha));
  pi1=exp(alpha+beta)/(1+exp(alpha+beta));
  pii0=(1+exp(alpha))**(-2);
  pii1=(1+exp(alpha+beta))**(-2);
  model approval~binary(pi);
  random u~normal(0,sig*sig) subject=pair;
  estimate 'OR' exp(beta);
  estimate 'rho' sig*sig/(sig*sig+constant("pi")**2/3);
  estimate 'rho0' sig*sig*(pi0**2)*pii0/(sig*sig*(pi0**2)*pii0+pi0*(1-pi0));
  estimate 'rho1' sig*sig*(pi1**2)*pii1/(sig*sig*(pi1**2)*pii1+pi1*(1-pi1));
  estimate 'marginal slope' beta/sqrt(1+((16*sqrt(3)/(15*constant("pi")))**2)*(sig**2));
  estimate 'marginal OR' exp(beta/sqrt(1+((16*sqrt(3)/(15*constant("pi")))**2)*(sig**2)));
run;

/* NRI2: random intercept and shared random slope logistic regression with normal distribution*/
proc nlmixed data=spm qpoints=100;
  pi=exp(alpha+u+(beta-2*u)*survey)/(1+exp(alpha+u+(beta-2*u)*survey));
  pi0=exp(alpha)/(1+exp(alpha));
  pi1=exp(alpha+beta)/(1+exp(alpha+beta));
  pii0=(1+exp(alpha))**(-2);
  pii1=(1+exp(alpha+beta))**(-2);
  model approval~binary(pi);
  random u~normal(0,sig*sig) subject=pair;
  estimate 'OR' exp(beta); /*slightly different CI than exponentiating the interval for beta*/
  estimate 'rho' sig*sig/(sig*sig+constant("pi")**2/3);
  estimate 'rho0' sig*sig*(pi0**2)*pii0/(sig*sig*(pi0**2)*pii0+pi0*(1-pi0));
```

```
  estimate 'rho1' sig*sig*(pi1**2)*pii1/(sig*sig*(pi1**2)*pii1+pi1*(1-pi1));
  estimate 'marginal slope' beta/sqrt(1+((16*sqrt(3)/(15*constant("pi")))**2)*(sig**2));
  estimate 'marginal OR' exp(beta/sqrt(1+((16*sqrt(3)/(15*constant("pi")))**2)*(sig**2)));
run;

/* GEE-ind: generalized estimating equations for the repeated measures*/
proc genmod data=spm descending;
  class pair;
  model approval =  survey / dist=bin link=logit;
  repeated subject = pair/type=ind covb ecovb;
  estimate 'beta' survey 1 -1 / exp;
run;

/* GEE-exch: generalized estimating equations for the repeated measures*/
proc genmod data=spm descending;
  class pair;
  model approval =  survey / dist=bin link=logit;
  repeated subject = pair/type=exch covb ecovb;
  estimate 'beta' survey 1 -1 / exp;
run;

data spmw;
  infile "spmwrev.dat";
  input app1 app2;
run;

/* BLR: bivariate Bahadur model with correlation*/
proc nlmixed data=spmw qpoints=200;
  parms alpha=0.36 beta=-0.16 rho=0.01;
  p1s = 1/(1+exp(-alpha));
  ps1 = 1/(1+exp(-alpha-beta));
  p11 = p1s*ps1+rho*sqrt(p1s*(1-p1s)*ps1*(1-ps1));
```

```
  p10 = p1s - p11;
  p01 = ps1 - p11;
  p00 = 1 - p1s - ps1 + p11;
  ll = app1*app2*log(p11) + app1*(1-app2)*log(p10) + (1-app1)*app2*log(p01)
       + (1-app1)*(1-app2)*log(p00);
  model ll ~ general(ll);
  estimate 'OR' exp(beta);
run;

/* BRI: random intercept logistic regression with bridge distribution */
proc nlmixed data=spm qpoints=100;
  parms alpha=0.36 beta=-0.16 s1=0.5;
  pi=constant("pi");
  uni = probnorm(b/s1);
  phi = 1.0/sqrt(1 + 3/pi/pi*s1*s1);
  Bl = 1/phi*log(sin(pi*uni*phi)/sin(phi*pi*(1-uni)));
  tmp = alpha+beta*survey;
  expeta = exp(Bl + tmp);
  p = expeta/(1 + expeta);
  model approval~binary(p);
  random b~normal(0,s1*s1) subject=pair;
  estimate 'OR' exp(beta);
  estimate 'proportionality constant' phi;
  estimate 'intra-pair correlation' 1-phi;
  estimate 'marginal beta' beta*phi;
  estimate 'marginal OR' exp(phi*beta);
  estimate 'sqrt variance component' pi*SQRT((phi**(-2)-1)/3);
run;

/* BRI2: random intercept and shared random slope logistic regression with bridge distribution */
proc nlmixed data=spm qpoints=100;
  parms alpha=0.36 beta=-0.16 s1=0.5;
```

```
  pi=constant("pi");
  uni = probnorm(b/s1);
  phi = 1.0/sqrt(1 + 3/pi/pi*s1*s1);
  Bl = 1/phi*log(sin(pi*uni*phi)/sin(phi*pi*(1-uni)));
  tmp = alpha+(beta-2*Bl)*survey;
  expeta = exp(Bl + tmp);
  p = expeta/(1 + expeta);
  model approval~binary(p);
  random b~normal(0,s1*s1) subject=pair;
  estimate 'OR' exp(beta);
  estimate 'proportionality constant' phi;
  estimate 'intra-pair correlation' 1-phi;
  estimate 'marginal beta' beta*phi;
  estimate 'marginal OR' exp(phi*beta);
  estimate 'sqrt variance component' pi*SQRT((phi**(-2)-1)/3);
run;


/* MMM: Logistic-probit-normal */
PROC NLMIXED data=spm qpoints=200; /* tech=quanew method=hardy;*/
   parms alpha=0.36 beta=-0.163  tau=3;
   eta_m=alpha+beta*survey;
   pi_m = 1/(1+exp(-eta_m));
   delta = sqrt(1+(tau*tau)) * probit(pi_m);
   eta_c = delta + b;
   pi_c = probnorm(eta_c);
   MODEL approval ~ binary(pi_c);
   RANDOM b ~ NORMAL(0,tau*tau) SUBJECT=pair;
   estimate 'OR' exp(beta);
   estimate 'pair-specific slope' SQRT(1+tau*tau)*beta;
   estimate 'pair-specific OR' exp(SQRT(1+tau*tau)*beta);
run;
```

```
/* MMM2: modified logistic-probit-normal */
PROC NLMIXED data=spm qpoints=200; /* tech=quanew method=hardy;*/
   parms alpha=0.36 beta=-0.163  tau=3;
   eta_m=alpha+beta*survey;
   pi_m = 1/(1+exp(-eta_m));
   delta = sqrt(1+(tau*tau)) * probit(pi_m);
   eta_c = delta + b -2*b*survey;
   pi_c = probnorm(eta_c);
   MODEL approval ~ binary(pi_c);
   RANDOM b ~ NORMAL(0,tau*tau) SUBJECT=pair;
   estimate 'OR' exp(beta);
   estimate 'pair-specific slope' SQRT(1+tau*tau)*beta;
   estimate 'pair-specific OR' exp(SQRT(1+tau*tau)*beta);
run;

/* CBM: logistic regression with random pi with beta distribution unchanged
        but with rho transformed to assure postivity*/
proc nlmixed data=spm method=GAUSS NOAD fd qpoints=50;
   parms alpha=-0.364 beta=0.565 rhoeta=-10;
   rho=exp(rhoeta)/(1+exp(rhoeta));
   pi = exp(alpha+beta*survey)/(1 + exp(alpha+beta*survey)) ;
   alpha_re = pi*(1-rho)/rho;
   beta_re = (1- pi)*(1-rho)/rho;
   alpha_1 = exp(alpha)/(1 + exp(alpha))*(1-rho)/rho;
   beta_1 = 1/(1 + exp(alpha))*(1-rho)/rho;
   alpha_2 = exp(alpha+beta)/(1 + exp(alpha+beta))*(1-rho)/rho;
   beta_2 = 1/(1 + exp(alpha+beta))*(1-rho)/rho;
   prob = CDF('NORMAL',b) ;
   p = quantile('BETA',prob,alpha_re,beta_re);
   MODEL approval ~ binary(p);
   random b~normal(0,1) subject=pair;
```

```
   estimate 'OR' exp(beta);
   estimate 'sqrt variance component beta 1' SQRT(alpha_1*beta_1/(((alpha_1+beta_1)**2)
                                              *(alpha_1+beta_1+1)));
   estimate 'sqrt variance component beta 2' SQRT(alpha_2*beta_2/(((alpha_2+beta_2)**2)
                                              *(alpha_2+beta_2+1)));
   estimate 'rho' 1/(alpha_1+beta_1+1);
   estimate 'rho2' 1/(alpha_2+beta_2+1);
   estimate 'alpha1' alpha_1;
   estimate 'alpha2' alpha_2;
   estimate 'beta1' beta_1;
   estimate 'beta2' beta_2;
run;

/* CBM2: modified logistic regression with random pi with beta distribution unchanged
         but with rho transformed to assure postivity*/
proc nlmixed data=spm method=GAUSS NOAD fd qpoints=50;
   parms alpha=-0.364 beta=0.565 rho=0.7;
   pi = exp(alpha+beta*survey)/(1 + exp(alpha+beta*survey)) ;
   alpha_re = pi*(1-rho)/rho;
   beta_re = (1- pi)*(1-rho)/rho;
   alpha_1 = exp(alpha)/(1 + exp(alpha))*(1-rho)/rho;
   beta_1 = 1/(1 + exp(alpha))*(1-rho)/rho;
   alpha_2 = exp(alpha+beta)/(1 + exp(alpha+beta))*(1-rho)/rho;
   beta_2 = 1/(1 + exp(alpha+beta))*(1-rho)/rho;
   prob = CDF('NORMAL',b-2*b*survey) ;
   p = quantile('BETA',prob,alpha_re,beta_re);
   MODEL approval ~ binary(p);
   random b~normal(0,1) subject=pair;
   estimate 'OR' exp(beta);
   estimate 'sqrt variance component beta 1' SQRT(alpha_1*beta_1/(((alpha_1+beta_1)**2)
                                              *(alpha_1+beta_1+1)));
   estimate 'sqrt variance component beta 2' SQRT(alpha_2*beta_2/(((alpha_2+beta_2)**2)
```

```
                                              *(alpha_2+beta_2+1)));
   estimate 'rho' 1/(alpha_1+beta_1+1);
   estimate 'rho2' 1/(alpha_2+beta_2+1);
   estimate 'alpha1' alpha_1;
   estimate 'alpha2' alpha_2;
   estimate 'beta1' beta_1;
   estimate 'beta2' beta_2;
run;


data approval;
input appr1 appr2 aantal aantal2;
cards;
1 1 794  86
1 0 150 570
0 1  86 794
0 0 570 150
;
run;

/* SBM model */
proc nlmixed data=approval;
   parms xi0=0.1 xi1=-0.1 beta=0.1;
   mmb=1/(1+beta**2);
   vvb=(1+beta)/(beta**2+beta+1)/(beta**2+1);
   k1=exp(xi0+xi1)/(1+exp(xi0+xi1));
   k2=exp(xi0)/(1+exp(xi0));
   alpha=1/beta;
   cell11=vvb*k1*k2;
   cell10=mmb*k1-vvb*k1*k2;
   cell01=mmb*k2-vvb*k1*k2;
   cell00=1-mmb*(k1+k2)+vvb*k1*k2;
```

```
   if appr1=1 and appr2=1 then ll=aantal2*log(cell11);
   if appr1=1 and appr2=0 then ll=aantal2*log(cell10);
   if appr1=0 and appr2=1 then ll=aantal2*log(cell01);
   if appr1=0 and appr2=0 then ll=aantal2*log(cell00);
   model aantal ~ general(ll);
   estimate 'alpha' 1/beta df=1599;
   estimate 'difference in proportions' cell10-cell01 df=1599;
   estimate 'odds ratio' cell11*cell00/cell10/cell01 df=1599;
   estimate 'inverse odds ratio' cell10*cell01/cell11/cell00 df=1599;
   estimate 'k1' k1 df=1599;
   estimate 'k2' k2 df=1599;
   estimate 'cell11' cell11 df=1599;
   estimate 'cell10' cell10 df=1599;
   estimate 'cell01' cell01 df=1599;
   estimate 'cell00' cell00 df=1599;
   estimate 'OR2' k2*(1-mmb*k1)/(k1*(1-mmb*k2)) df=1599;
   estimate 'betaM' log(k2*(1-mmb*k1)/(k1*(1-mmb*k2))) df=1599;
   estimate 'rho' beta**3/(1+beta**2)/(beta**2+beta+1)*sqrt(k1*k2/((1-k1/(beta**2+1))
                *(1-k2/(beta**2+1))))  df=1599;
   estimate 'vartheta1' vvb-mmb**2 df=1599;
   estimate 'vartheta2' alpha*beta/((1+alpha+beta)*(alpha+beta)**2) df=1599;
run;
```