STATISTICS IN MEDICINE
Statist. Med. 2008; **27**:449–461
Published online 11 June 2007 in Wiley InterScience
(www.interscience.wiley.com) DOI: 10.1002/sim.2937

Comparing odds ratios measuring improvement or deterioration across repeated exposure or treatment sessions

Pamela A. Ohman Strickland[1,2,3,4,*,†] and Shou-En Lu[1,3,4]

[1]*Department of Biostatistics, University of Medicine and Dentistry of New Jersey – School of Public Health, Piscataway, NJ, U.S.A.*
[2]*Department of Family Medicine, University of Medicine and Dentistry of New Jersey – Robert Wood Johnson Medical School, Somerset, NJ, U.S.A.*
[3]*Environmental & Occupational Health Sciences Institute, Piscataway, NJ, U.S.A.*
[4]*Cancer Institute of New Jersey, New Brunswick, NJ, U.S.A.*

SUMMARY

Unbiased estimating equations are used to estimate and compare the odds ratios representing the odds of heightened anxiety following acute exposures relative to baseline in a cross-over study examining indoor air contaminants. Although estimating equations are used in the creation of the estimates and standard errors, inference is conditional on subject. The proposed method produces estimates that are less biased and computationally simpler than pseudo-likelihood. In addition, specification of the form of the random-effects distributions is not required. Copyright © 2007 John Wiley & Sons, Ltd.

KEY WORDS: proportional odds; estimating equations; score tests; pseudo-likelihood

1. INTRODUCTION

Repeated-measures designs are widely used in biological and medical research. An individual study subject may receive one exposure (treatment) with multiple measurements taken throughout the study period or may receive different exposures (treatments) over different sessions (separated periods of time) with single or even multiple measurements taken at each session. In the latter design, usually known as the cross-over design, each individual is considered as his/her own control

*Correspondence to: Pamela A. Ohman Strickland, 683 Hoes Lane West, Room 218, P.O. Box 9, Piscataway, NJ 08854, U.S.A.
†E-mail: ohmanpa@umdnj.edu

Contract/grant sponsor: National Institute for Occupational Safety and Health (NIOSH); contract/grant number: R01 OH03691-01
Contract/grant sponsor: NIEHS Center; contract/grant number: NIEHS P30ES05022

Received 11 December 2006
Accepted 19 April 2007

and analyses implicitly control for study subjects' characteristics. The analysis of the cross-over design with pre- and post-exposure ordinal measurements within each exposure is the focus of this paper.

Consider the indoor air experiment of Fiedler *et al.* [1], conducted to examine the effects of three acute exposures on numerous health outcomes. Each individual subject was exposed in three different sessions to three exposures: volatile organic compounds (VOC), VOC plus ozone (VOCO), and masked clean air (MCA). One question of interest was whether increases (decreases) in anxiety could be attributed to exposure condition. Due to the distribution of an individual's anxiety responses, rated on a scale of 0–100, with 100 being the most intense possible, it was reasonable at each time point to categorize anxiety values into three ordinal categories. With a large number of zeros, the first category included just values of zero while two additional ordinal categories included values greater than zero. This paper develops methodology to estimate the odds ratio representing improvement or deterioration and to compare these odds ratios across exposure sessions.

In all studies with repeated measures, investigators usually choose between marginal and conditional analyses. The marginal approach typically uses the generalized estimating equations [2] and focuses on the population-averaged inference, assessing differences between exposures at each time point. Alternatively, the conditional approach uses a random-effects model, possibly with a repeated-measures covariance structure [3], to study the exposure effect at the subject-specific level.

Standard methods for estimating the conditional odds of heightened anxiety for each exposure session based on the random-effects model include maximum likelihood, restricted maximum likelihood [4, 5], penalized quasi-likelihood [6], or pseudo-likelihood [7]. These methods require the specification of the distribution for the random effects and spelling out all possible confounding factors in the regression model. On the other hand, since each individual subject received different exposures and can be considered as his/her own control, methods that compare the concordant and discordant pairs, such as McNemar's approach and conditional logistic regression [8], can be appealing.

For repeated ordered categorical responses, Agresti and Lang [9] proposed a generalization of the Rasch model that expresses the cumulative logit of the response distribution with subject parameters. An alternative derivation of this odds ratio with accompanying sample size calculations was presented by Ohman-Strickland and Lu [10], specifically for before–after study designs. In particular, the odds ratio estimator can be derived from an estimating equation approach for a class of random effects logistic regression for ordinal data. In this paper, we propose generalizing the estimating equation approach considered in Ohman-Strickland *et al.* and extend it to a random effects model with multilevel clustering (e.g. within-session and between-session associations). An estimating equation is formulated based on direct comparisons of concordant and discordant pair data within sessions, and a flexible working correlation structure is used for modeling between-session associations. As compared with the aforementioned random-effects models approach, one does not need to specify the distribution for the random-effects models in the proposed method, the computation is thus much easier, and the estimates are robust.

This paper is organized as follows. The proposed method is introduced in Section 2. Illustration of the proposed estimators is presented in Section 3 using the motivating data as an example. In Section 4, a simulation study demonstrates the performance of the proposed estimators. A brief discussion in Section 5 concludes this paper.

DOI: 10.1002/sim

2. METHODOLOGY

2.1. The logistic model

Consider the standard single-level polytomous logistic regression model for an ordinal variable with M categories

$$\log\lfloor \pi_{mij}(t)/\overline{\pi}_{mij}(t)\rfloor = u_i + \alpha_m - \beta_j - \theta_j t \quad \text{for } m=1,\ldots,M-1$$

representing the log-odds of a higher response for individual $i(i=1,\ldots,n)$ in exposure j $(j=1,\ldots,J)$ at time t ($t=0$ for baseline and $t=1$ for follow-up). u_i represents a random effect for individual i. In the cross-over design described in this paper, each individual (i) is subjected to each of the J exposures (e.g. active exposure or control). In this equation (dropping (t) and the i and j subscripts for simplicity), $\pi_m = \Pr(y \in A_m)$ for some set A_m, $m=1,\ldots,M-1$, and $\overline{\pi}_m = 1-\pi_m$. The sum of the first two fixed effects, $\alpha_m+\beta_j$, describes the distribution of responses at baseline for exposure j. We assume $\beta_1=0$. α_m, $m=1,\ldots,M-1$, expresses the average distribution of responses across categories at baseline for $j=1$ as $M-1$ odds, whereas β_j describes the odds of higher baseline responses for exposure j relative to exposure $j=1$. θ_j represents the log-odds ratios, comparing the odds of higher responses after the jth exposure relative to baseline for that exposure. u_i are subject-specific intercepts indicating whether, over all time points and exposures, a subject's responses tend to fall in lower or higher categories.

The value of the response before and after each exposure can fall into one of M ordinal categories. The sets A_m and their complements $\overline{A}_m$ depend on the type of ordinal logistic regression model being considered, whether proportional odds, adjacent category odds, continuous odds, or otherwise. For example, if $A_m=\{Y \leqslant m\}$ and $\overline{A}_m=\{Y>m\}$, then this represents the proportional odds model. If $A_m=\{Y=(m+1)\}$ and $\overline{A}_m=\{Y=m\}$, this represents the adjacent-categories logits model.

Alternatively, an individual's responses before and after an exposure session may be more correlated than a person's responses from separate exposure sessions. In this case, one might employ a two-level mixed model:

$$\log\lfloor \pi_{mij}(t)/\overline{\pi}_{mij}(t)\rfloor = u_i + w_{j(i)} + \alpha_m - \beta_j - \theta_j t \tag{1}$$

where the additional random component accounts for exposure session within an individuals. With this model, an individual's responses within an exposure are allowed to be more correlated than responses between exposures. Interpretations of the fixed effects remain the same.

In maximum likelihood estimation, all the nuisance parameters, including the variance components for u_i and $w_{j(i)}$, and the parameters α_m and β_j would be estimated in addition to the parameters of interest, θ_j.

2.2. The conditional odds estimating equations approach

To estimate θ, we create an estimating equation, $U(\theta)=\sum_{i=1}^{n} g_i(\theta)$ [11]. Each individual's contribution is given by $g_i(\theta)$ such that the jth element of this vector of length J is given by

$$g_{j\{i\}} = \sum_{m=1}^{M-1} [1\{Y_{ij0} \in A_m\}1\{Y_{ij1} \in \overline{A}_m\}e^{\theta_j} - 1\{Y_{ij0} \in \overline{A}_m\}1\{Y_{ij1} \in A_m\}]$$

DOI: 10.1002/sim

$$
\begin{aligned}
&= \sum_{m=1}^{M-1}[1\{Y_{ij0}\in A_m\}1\{Y_{ij1}\in \overline{A}_m\}]\times e^{\theta_j} - \sum_{m=1}^{M-1}[1\{Y_{ij0}\in \overline{A}_m\}1\{Y_{ij1}\in A_m\}] \\
&= z_{ij1}e^{\theta_j} - z_{ij2} \qquad (2)
\end{aligned}
$$

Y_{ijt} represents the value of the response for the ith individual in the jth exposure at time t ($t=0$ at baseline and $t=1$ post-exposure). The indicator function $1\{Y_{ijt}\in A_m\}$ equals 1 if Y_{ijt} is in set A_m and 0 otherwise. (The last line serves to simplify notation in the following descriptions.) Here, as before, the parameter θ_j represents the conditional log-odds ratio representing improvement in the jth exposure. Conditional on the random effects specified in equation (1), each U_i has expectation zero. Therefore,

$$E[g_i(\theta)]=0$$

and an estimate for θ can be obtained by solving

$$U(\theta)=0$$

The estimating equations above result in an estimate of the jth log-odds ratio given by

$$\hat{\theta}_j = \log\left(\frac{\sum_{i=1}^{n} z_{ij2}}{\sum_{i=1}^{n} z_{ij1}}\right)$$

In order to demonstrate the origin of these estimating equations, observe that the odds ratio comparing time 1 with the baseline for individual i is given by

$$e^{\theta_j} = \frac{\pi_{mij}(t=1)}{\overline{\pi}_{mij}(t=1)}\times\frac{\overline{\pi}_{mij}(t=0)}{\pi_{mij}(t=0)}$$

for exposure j. This equation can be rearranged so that

$$\overline{\pi}_{mij}(t=1)\pi_{mij}(t=0)e^{\theta_j} - \pi_{mij}(t=1)\overline{\pi}_{mij}(t=0)=0$$

The observed associated quantity

$$1\{Y_{ij0}\in A_m\}1\{Y_{ij1}\in \overline{A}_m\}e^{\theta_j} - 1\{Y_{ij0}\in \overline{A}_m\}1\{Y_{ij1}\in A_m\}$$

has expectation zero for all m. Thus, the ith individual's contribution to an unbiased estimating equation for each θ_j is obtained by summing over $m=1,\ldots,M-1$. This results in (2), in which individual contributions for each of the J exposures are stacked in the vector $g_i(\theta)$. See Ohman-Strickland and Lu [10] for additional discussion. Agresti and Lang [9] derive this conditional odds ratio as a maximum likelihood estimate from the logistic regression function.

Like conditional logistic regression [8], estimation using these estimating equations will not allow us to assess the effects of matching variables, which, in this case, are the main effects of the time-independent covariates. However, by generalizing the model above, we would still be able to study interactions between time and matching variables.

DOI: 10.1002/sim

Under the framework above, inference concerning the conditional odds ratios may be made using basic estimating equation theory [11–13]. This approach will ensure appropriate correction for potential correlations between log-odds ratio estimates. In particular, as the sample size becomes large, inference may be conducted by observing that

$$\sqrt{n}(\hat{\theta}-\theta_0)\rightarrow \mathrm{N}(0, G'\Omega^{-1}G)$$

where

$$G=\lim_{n\rightarrow\infty} n^{-1}\frac{\partial}{\partial\theta}U(\theta) \quad \text{and} \quad \Omega=E[g_i(\theta)g_i(\theta)']$$

In practice, G and Ω may be consistently estimated by

$$\hat{G}_n = -\frac{1}{n}\sum_{i=1}^{n}\nabla_\theta g_i(\theta)\bigg|_{\theta=\hat{\theta}}$$

and

$$\hat{\Omega}_n = \frac{1}{n}\sum_{i=1}^{n} g_i(\theta)g_i(\theta)'\bigg|_{\theta=\hat{\theta}}$$

2.3. *Hypothesis testing*

We consider tests of two null hypotheses.

The first is

$$H_0:\ \theta_1=\cdots=\theta_J=0$$

Under this null hypothesis, none of the exposures would have an effect on the ordinal response. The alternative hypothesis is that at least one exposure does affect the response.

The second tests the null hypothesis

$$H_0:\ \theta_1=\cdots=\theta_J$$

Under this second null hypothesis, a common odds ratio is implied. We will present the optimal estimate of the common odds ratio in the context of the Score test. The alternative hypothesis is that the odds ratios differ across exposures.

Both hypotheses can be written as special cases of the hypothesis

$$H_0:\ A\theta=0$$

where A is of rank k. For example, to test that all three odds ratios equal 1 (Hypothesis 1)

$$A=\begin{pmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & 1\end{pmatrix}$$

DOI: 10.1002/sim

with $k=3$. To test the pairwise equality of three log-odds ratios (Hypothesis 2), as is the goal in the exposure study used to motivate this methodology

$$A=\begin{pmatrix}1 & -1 & 0\\ 1 & 0 & -1\end{pmatrix}$$

with $k=2$.

2.3.1. Wald chi-square test statistics. One approach to testing hypotheses concerning θ is based on the Wald chi-square. Normal matrix theory [14, p. 353] for matrix A of rank k gives

$$A(\sqrt{n}\hat{\theta}) \sim \mathrm{N}(0, AG'\Omega^{-1}GA')$$

and, equivalently

$$A(\sqrt{n}\hat{\theta})(AG'\Omega^{-1}GA')^{-1}(\sqrt{n}\hat{\theta})'A' \sim \chi^2_k$$

2.3.2. Score test statistic. The score test is based on the statistic

$$\mathrm{U}(\theta_0)'\left[\sum_{i=1}^{n} g_i(\theta_0)g_i(\theta_0)'\right]^{-1}\mathrm{U}(\theta_0) \sim \chi^2_k$$

where θ_0 is the value of θ under the null hypothesis and k. For Hypothesis 1, θ equals the 0 vector, while for Hypothesis 2, θ must be estimated by a common odds ratio. The common odds ratio will be estimated by a linear combination of the θ_j, that is, by

$$c'\theta=\sum_{j=1}^{J} c_j\theta_j \quad \text{with} \quad \sum_{j=1}^{J} c_j=1$$

The linear estimator with the smallest variance among all linear estimators [15] uses weights given by

$$c=[e'(G_n'\Omega_n^{-1}G_n)^{-1}e]^{-1}(G_n'\Omega_n^{-1}G_n)^{-1}e$$

where e is a unit vector of length J.

3. APPLICATION

3.1. Example

In a study by Fiedler *et al.* [1], 130 healthy, non-smoking women were exposed during three 2-h exposure sessions, separated by at least a week and in random order, to MCA, VOC, and VOCO. For each condition, subjects were exposed for 135 min in a controlled environmental facility (CEF). The CEF is a 2.2 m high × 4.1 m wide × 2.7 m deep stainless-steel room, with a

DOI: 10.1002/sim

Table I. Results for assessing heightened anxiety in the exposure cross-over study, with $\hat{\theta} = (\hat{\theta}_1, \hat{\theta}_2, \hat{\theta}_3)$ estimating the log-odds of increased anxiety for CA, VOC and the VOCO exposures, respectively.

	Proposed estimating equation approach				Proc Glimmix			
			P-values (Wald and Score test for each hypothesis)				Chi-square *P*-values (Type III analysis)	
	$\hat{\theta}_j$ $(j=1,2,3)$	$\mathrm{SE}(\hat{\theta}_j)$ $(j=1,2,3)$	All OR = 1	Common OR	$\hat{\theta}_j$ $(j=1,2,3)$	$\mathrm{SE}(\hat{\theta}_j)$ $(j=1,2,3)$	All OR = 1	Common OR
20 min after exposure onset	0.42 0.55 0.48	0.26 0.24 0.26	0.099 0.1122	0.90 0.90	0.49 0.49 0.65	0.29 0.29 0.29	0.0014	0.76
50 min after exposure onset	0.90 1.33 0.87	0.34 0.39 0.37	<0.0001 0.0010	0.63 0.60	0.68 0.74 0.55	0.29 0.29 0.29	0.0013	0.88
85 min after exposure onset	0.29 1.50 0.72	0.35 0.40 0.33	0.0004 0.0005	0.049 0.029	0.19 0.86 0.61	0.30 0.29 0.29	0.0037	0.26
Post-exposure follow-up	−0.34 0.025 −0.22	0.27 0.26 0.26	0.51 0.51	0.45 0.45	−0.48 −0.029 −0.35	0.30 0.29 0.30	0.27	0.54

Note: Analyses were performed separately for each post-exposure time.

total volume of $24\,\mathrm{m}^3$, that was used to simulate space in a poorly ventilated office building. Detailed descriptions of the exposures can be found in Fiedler *et al.* [1] and Fan *et al.* [16]. Before each exposure, during exposure (20, 50, and 85 min after the start of exposure), and at follow-up (immediately following completion of exposure and in a physician's clinic located within the same building), individuals were questioned about a number of different types of symptoms, including anxiety symptoms, on a scale of 0–100. The mean of the ratings of all anxiety symptoms (feel jittery in body, feel nervous, heart palpitations, feel tense, worried) is the response for this analysis. The variability of responses was such that a majority of individuals had scores of 0 and the remaining scores were highly right skewed. Skewness of the non-zero values was more extreme than one would expect even under a Poisson model. Because of the failure of standard distributional assumptions, the values of the response were divided into three categories: zero, greater than zero to four, and greater than four. Slightly over half of the responses fell into the first category, and the remaining values were roughly divided into half between the two upper categories.

In this study, investigators were primarily interested in knowing whether a difference existed in the odds ratios of increased symptoms between the three exposures. Table I presents the conditional odds ratios estimates as well as standard errors in addition to the hypothesis testing results using the proposed approach, and also the type III chi-square test based on the pseudo-likelihood approach proposed by Wolfinger and O'Connell [7], which is implemented with Proc Glimmix (an experimental procedure for SAS, version 9.0, SAS Institute, Cary, NC,

DOI: 10.1002/sim

USA, available for download at www.support.sas.com). In addition to tests of equal odds ratios between exposures, Table I also includes tests of whether all odds ratios equal one. Each test is performed at each of four follow-up time points to examine changes in anxiety from baseline.

Of particular interest is the marginal significance of the hypothesis of a common odds ratio at 85 min after exposure onset, with p-values of 0.049 and 0.029 for the Wald and Score tests, respectively. This did not approach significance using pseudo-likelihood. The odds ratios were estimated to be larger using the proposed approach as compared with the pseudo-likelihood approach.

Note that default estimation techniques were used in Proc Glimmix. In particular, estimation is conducted using restricted pseudo-likelihood [7]. In residual likelihood, the objective function is adjusted to account for the fixed effects. In addition, expansion around the current best linear unbiased predictors of the random effects was used to obtain the pseudo-likelihood. Thus, the estimates are generalized least-squares estimates.

4. SIMULATION STUDY

4.1. Study of bias and power

To examine the precision and power of this estimating approach, a series of simulation studies were conducted, mimicking the design of the exposure study detailed above with $J=3$, $M=3$, and $t=0$ or 1. To provide realistic parameter values for the simulation study, parameter estimates for model (1) applied to the data above (at the post-exposure follow-up) were obtained using the Nlmixed routine in SAS [17]. Estimates of the fixed regression parameters were as follows: $\hat{\alpha}=(\hat{\alpha}_2, \hat{\alpha}_2)=(-1.568, -0.331)$ and $\hat{\beta}=(\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3)=(0, -0.273, 0.145, 0)$ (See Section 2 for an explanation of these parameters). Assuming normal distributions for the random effects, the variance of the random intercepts was estimated to be 1.423 for the between-subject variance and 0.010 for the between-exposure variance within subjects. These estimated values were used as the basis for the simulations. In addition, for this first simulation study, the random effects distributions were assumed to be normal. Values of the log-odds ratios (θ_j, $j=1, \ldots, 3$) from the experimental data were $(0.108, 0.622, 0.441)$. However, these values were varied for the purpose of studying actual significance levels and power in the simulations.

In order to use the approach outlined in this paper, at least one discordant pair in each 'direction' is needed in order to create the estimates and standard errors. When such a pair did not exist for a particular exposure, we performed a 'zero-cell correction' by adding an additional pair to the data set in which the difference between the pre- and post-scores was $\frac{1}{2}$. The extremely small frequency within which this correction was needed is noted in the tables of results.

Table II presents the simulation results for a variety of values for $\theta=(\theta_1, \theta_2, \theta_3)$ and sample sizes, with 10 000 simulations per combination, resulting in a maximum standard error for the rejection probabilities of 0.005. For all values of θ, the estimating approach performed well. Estimates were approximately unbiased, the actual significance levels were comparable to the nominal level of 0.05 when the null hypotheses were true, and the tests showed some power to reject both null hypotheses (all odds ratios equal one and the exposures share a common odds ratio). In addition, the 'true' variance of the estimates as determined from the Monte Carlo variance was accurately estimated *via* the estimating equation approach.

DOI: 10.1002/sim

Table II. Simulation results evaluating the accuracy and precision of the proposed estimating equations approach for differing θ and sample sizes.

$\theta=(\theta_1,\theta_2,\theta_3)$	# pairs	Avg. $\hat{\theta}$	Avg. Var($\hat{\theta}$)	Monte Carlo Var($\hat{\theta}$)	Pr(Reject H_0: all OR = 1) Wald	Score	Pr(Reject H_0: common OR) Wald	Score
(0, 0, 0)	50	0.012	0.18	0.19	0.053	0.043	0.054	0.048
		0.002	0.20	0.21				
		0.001	0.19	0.20				
	100	0.006	0.089	0.091	0.052	0.047	0.056	0.053
		0.004	0.097	0.099				
		0.008	0.094	0.097				
	150	−0.001	0.059	0.060	0.053	0.050	0.055	0.052
		0.001	0.064	0.064				
		−0.002	0.062	0.063				
	200	0.001	0.044	0.045	0.049	0.047	0.055	0.052
		0.001	0.048	0.049				
		0.001	0.046	0.047				
(0, 0.5, 0.5)	50*	0.012	0.18	0.19	0.21	0.19	0.12	0.11
		0.52	0.23	0.24				
		0.53	0.22	0.24				
	100	0.006	0.089	0.091	0.44	0.43	0.21	0.20
		0.52	0.11	0.11				
		0.52	0.10	0.11				
	150	−0.001	0.059	0.060	0.62	0.61	0.29	0.29
		0.51	0.071	0.071				
		0.50	0.068	0.070				
	200	0.001	0.044	0.045	0.76	0.76	0.39	0.38
		0.51	0.053	0.055				
		0.51	0.051	0.052				
(0, 0.5, 1.0)	50†	0.012	0.18	0.19	0.47	0.45	0.26	0.27
		0.52	0.23	0.24				
		1.06	0.27	0.31				
	100	0.006	0.089	0.091	0.83	0.82	0.51	0.51
		0.52	0.11	0.11				
		1.04	0.13	0.14				
	150	−0.001	0.059	0.060	0.95	0.95	0.68	0.69
		0.51	0.071	0.071				
		1.01	0.083	0.087				
	200	0.001	0.044	0.045	0.99	0.99	0.81	0.82
		0.51	0.053	0.055				
		1.02	0.062	0.064				
(0.5, 0.5, 0.5)	50‡	0.54	0.20	0.22	0.33	0.30	0.051	0.052
		0.52	0.23	0.24				
		0.53	0.22	0.24				
	100	0.52	0.097	0.10	0.64	0.63	0.054	0.054
		0.52	0.11	0.11				
		0.52	0.10	0.11				

DOI: 10.1002/sim

Table II. *Continued*.

$\theta=(\theta_1,\theta_2,\theta_3)$	# pairs	Avg. $\hat{\theta}$	Avg. Var($\hat{\theta}$)	Monte Carlo Var($\hat{\theta}$)	Pr(Reject H_0: all OR = 1) Wald	Score	Pr(Reject H_0: common OR) Wald	Score
	150	0.51	0.064	0.066	0.82	0.81	0.050	0.050
		0.51	0.071	0.071				
		0.50	0.068	0.070				
	200	0.51	0.048	0.049	0.93	0.92	0.052	0.052
		0.51	0.053	0.055				
		0.51	0.051	0.052				

Note: Each simulation is based on 10 000 simulations.
*In four simulated data sets (0.04%), a correction factor for a zero count of discordant pairs was needed.
†In 15 simulated data sets (0.15%), a correction factor for a zero count of discordant pairs was needed.
‡In five simulated data sets (0.05%), a correction factor for a zero count of discordant pairs was needed.

4.2. *Comparison with pseudo-likelihood*

A second set of simulation studies compared the estimates and power of estimates produced by the estimating equation approach of this paper with results obtained from Proc Glimmix. The same parameters were used as those in the first simulation, with alterations to the random-effects distribution. Three random-effects distributions were considered in this simulation study: a normal distribution with mean zero and variance σ^2; a mixture of normals with two-thirds of subjects having a mean of -2σ, one-third having a mean of 4σ, and sharing common variance σ^2; and finally a Gamma with shape parameter equal to 1 and scale parameter $1/\sigma$. The latter Gamma was re-centered to have zero mean. Sample sizes of 50, 100, 200, and 300 were chosen in order to examine the consistency and efficiency of the estimates. Each study was based on 3600 simulations, providing a maximum standard error for the rejection probabilities of 0.008.

Results for this second simulation study are presented in Table III. In general, the proposed approach produced greater power and estimates that were biased up a little; the restricted pseudo-likelihood estimates were biased down. Particularly with larger sample sizes, the estimating equation approach produced less biased estimates. Standard errors estimated using the proposed approach, although larger, were comparable to those estimated *via* pseudo-likelihood. Non-normal random-effects distributions resulted in more biased estimates under pseudo-likelihood, particularly when the random-effects distribution was a mixture of normals.

We observe a reversal in the relative sizes of the standard errors from the two approaches from what was observed in the data analysis example. We had hypothesized that the difference in standard errors between the two approaches in the data analysis example was potentially due to the misspecification of the random-effects distribution in the pseudo-likelihood approach. Unfortunately, the simulations studies do not shed any additional light on this. An appropriate random-effects distribution for the data analysis example may have been sufficiently irregular so as not to be mimicked by any easily specified distribution.

In separate simulations not reported in this paper, Glimmix was used to obtain maximum likelihood estimates using pseudo-likelihood estimation. These were notably poor with respect to bias and power for the cases considered here, particularly with the non-normal random-effects distributions.

DOI: 10.1002/sim

Table III. Simulation results comparing the proposed estimating equation approach with GLIMMIX when $\theta = (0, 0.5, 1.0)$.

		Estimating equation				GLIMMIX			
				Score power (SE)				Power (SE)	
Random-effects distribution	N	Bias($\hat{\theta}_j$) ($j = 1, 2, 3$)	Avg. SE($\hat{\theta}_j$) ($j = 1, 2, 3$)	All OR = 1	Common OR	Bias($\hat{\theta}_j$)	Avg. SE($\hat{\theta}_j$) ($j = 1, 2, 3$)	All OR = 1	Common OR
$\theta = (0, 0.5, 1.0)$									
Normal		0.012	0.43			0.011	0.40		
	50*	0.026	0.47	0.45	0.27	−0.029	0.43	0.45	0.23
		0.054	0.52			−0.063	0.45		
		0.011	0.30			0.008	0.28		
	100	0.022	0.33	0.83	0.51	−0.026	0.30	0.82	0.48
		0.044	0.36			0.054	0.31		
		0.001	0.21			0.001	0.20		
	200	0.008	0.23	0.99	0.82	−0.039	0.21	0.99	0.81
		0.019	0.25			−0.074	0.22		
		0.002	0.17			−0.002	0.16		
	300	0.000	0.19	0.99	0.94	−0.043	0.17	0.99	0.94
		0.010	0.20			−0.079	0.18		
Normal mixture		0.008	0.63			0.008	0.53		
	50†	0.067	0.67	0.23	0.18	−0.070	0.56	0.18	0.10
		0.10	0.71			−0.16	0.57		
		0.011	0.43			0.009	0.37		
	100‡	0.015	0.45	0.53	0.31	−0.092	0.39	0.45	0.22
		0.075	0.49			−0.15	0.39		
		−0.003	0.30			−0.001	0.26		
	200	0.017	0.32	0.86	0.55	−0.088	0.27	0.81	0.45
		0.020	0.33			−0.18	0.28		
		0.005	0.24			0.004	0.21		
	300	0.012	0.25	0.98	0.72	−0.088	0.22	0.96	0.65
		0.016	0.27			−0.18	0.23		
Gamma		0.014	0.43			0.015	0.41		
	50§	0.016	0.49	0.41	0.26	−0.046	0.45	0.40	0.22
		0.071	0.54			−0.065	0.46		
		−0.003	0.30			−0.003	0.28		
	100	0.023	0.34	0.79	0.50	−0.033	0.31	0.77	0.46
		0.033	0.37			−0.083	0.32		
		−0.005	0.21			−0.005	0.20		
	200	0.004	0.24	0.98	0.80	−0.047	0.22	0.98	0.78
		0.008	0.26			−0.098	0.23		

Table III. *Continued.*

Random-effects distribution	N	Estimating equation: Bias($\hat{\theta}_j$) ($j=1,2,3$)	Avg. SE($\hat{\theta}_j$) ($j=1,2,3$)	Score power (SE): All OR = 1	Common OR	GLIMMIX: Bias($\hat{\theta}_j$)	Avg. SE($\hat{\theta}_j$) ($j=1,2,3$)	Power (SE): All OR = 1	Common OR
		−0.001	0.17			0.000	0.16		
	300	−0.003	0.19	0.99	0.94	−0.053	0.18	0.99	0.93
		0.010	0.21			−0.095	0.18		

Note: Each simulation is based on 3600 simulated data sets with 50, 100, 200 or 300 subjects each. The three random-effects distributions are a normal, $N(0, \sigma^2)$; a mixture of normals, $\frac{2}{3} \times N(-2\sigma, \sigma^2) + \frac{1}{3} \times N(4\sigma, \sigma^2)$; and a Gamma re-centered to a zero mean, $G(1, 1/\sigma) - \sigma$.
*Four out of 3600 simulated data sets (0.11%) required a correction factor for a zero count of discordant pairs.
†113 out of 3600 simulated data sets (3.13%) required a correction factor for a zero count of discordant pairs.
‡Five out of 3600 simulated data sets (0.14%) required a correction factor for a zero count of discordant pairs.
§10 out of 3600 simulated data sets (0.28%) required a correction factor for a zero count of discordant pairs.

5. DISCUSSION

The estimating equation approach to conditional information from different sessions in a cross-over trial is promising. While simple, it provides accurate and precise estimates, regardless of the underlying random-effects distributions. The estimates are asymptotically unbiased, providing an alternative to the pseudo-likelihood estimators of Wolfinger and O'Connell [7]. In addition, the estimates have similar standard errors relative to the pseudo-likelihood estimators.

In the estimation procedure, we proposed a working independence structure for estimation of the variance–covariance matrix. It would be possible to employ an alternative working structure. The estimator would still be a consistent estimator even if the structure of the working correlation structure were misspecified [2].

Extending the proposed approach to handle within-session covariates would be straightforward and similar to handling within-session covariates in conditional logistic regression.

ACKNOWLEDGEMENTS

The authors wish to acknowledge Nancy Fiedler, PhD, for use of her data obtained *via* a project funded by a grant from the National Institute for Occupational Safety and Health (NIOSH) R01 OH03691-01 and the NIEHS Center Grant, NIEHS P30ES05022.

REFERENCES

1. Fiedler N, Laumbach R, Kelly-McNeil K, Lioy P, Fan Z-H, Zhang J, Ottenweller J, Ohman-Strickland P, Kipen H. Health effects of a mixture of indoor air volatile organics, their ozone oxidation products and stress. *Environmental Health Perspectives* 2005; **11**:1542–1548.
2. Liang KY, Zeger SL. Longitudinal data analysis using generalized linear models. *Biometrika* 1986; **73**:13–22.
3. Laird N, Ware J. Random-effects models for longitudinal data. *Biometrics* 1982; **38**:963–974.
4. Drum ML, McCullagh P. REML estimation with exact covariance in the logistic mixed model. *Biometrics* 1993; **49**:677–689.

5. Liao JG, Lipsitz SR. A type of restricted maximum likelihood estimator for variance components in generalized linear mixed models. *Biometrika* 2002; **89**:401–409.
6. Breslow NE, Clayton DG. Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association* 1993; **88**:9–25.
7. Wolfinger R, O'Connell M. Generalized linear mixed models: a pseudo-likelihood approach. *Journal of Statistical Computation and Simulation* 1993; **4**:233–243.
8. Breslow N. Covariance adjustment of relative-risk estimates in matched studies. *Biometrics* 1982; **38**:661–683.
9. Agresti A, Lang JB. A proportional odds model with subject-specific effects for repeated ordered categorical responses. *Biometrika* 1993; **80**:527–534.
10. Ohman-Strickland PA, Lu S-E. Estimates, power and sample size calculations for two-sample ordinal outcomes under before–after study designs. *Statistics in Medicine* 2003; **22**:1807–1818.
11. Godambe VP. An optimum property of regular maximum likelihood estimation. *Annals of Mathematical Statistics* 1960; **31**:1208–1211.
12. Huber PJ. The behavior of maximum likelihood estimates under non-standard conditions. *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, vol. 1, Berkeley, CA, U.S.A., 1967; 221–233.
13. Barndorff-Nielsen OE, Cox DR. *Inference and Asymptotics*. Chapman & Hall: London, 1994.
14. Searle SR. *Matrix Algebra Useful for Statistics*. Wiley: New York, 1982.
15. Wei LJ, Johnson WE. Combining dependent tests with incomplete repeated measurements. *Biometrika* 1985; **72**:359–364.
16. Fan Z, Lioy P, Weschler C, Fiedler N, Kipen H, Zhang J. Ozone-initiated reactions with mixtures of volatile organic compounds under simulated indoor conditions. *Environmental Science and Technology* 2003; **37**:1811–1821.
17. SAS Institute. *SAS/STAT User's Guide*, *Version 8.* SAS Publishing: Cary, NC, 2002.

DOI: 10.1002/sim