The International Journal of Biostatistics

Volume 7, Issue 1

2011

Article 27

Relative Risk Estimation in Cluster Randomized Trials: A Comparison of Generalized Estimating Equation Methods

Lisa N. Yelland, *University of Adelaide*

Amy B. Salter, *University of Adelaide*

Philip Ryan, *University of Adelaide*

Recommended Citation:

Yelland, Lisa N.; Salter, Amy B.; and Ryan, Philip (2011) "Relative Risk Estimation in Cluster Randomized Trials: A Comparison of Generalized Estimating Equation Methods," *The International Journal of Biostatistics*: Vol. 7: Iss. 1, Article 27.

DOI: 10.2202/1557-4679.1323

Relative Risk Estimation in Cluster Randomized Trials: A Comparison of Generalized Estimating Equation Methods

Lisa N. Yelland, Amy B. Salter, and Philip Ryan

Abstract

Relative risks have become a popular measure of treatment effect for binary outcomes in randomized controlled trials (RCTs). Relative risks can be estimated directly using log binomial regression but the model may fail to converge. Alternative methods are available for estimating relative risks but these have generally only been evaluated for independent data. As some of these methods are now being applied in cluster RCTs, investigation of their performance in this context is needed. We compare log binomial regression and three alternative methods (expanded logistic regression, log Poisson regression and log normal regression) for estimating relative risks in cluster RCTs. Clustering is taken into account using generalized estimating equations (GEEs) with an independence or exchangeable working correlation structure. The results of our large simulation study show that the log binomial GEE generally performs well for clustered data but suffers from convergence problems, as expected. Both the log Poisson GEE and log normal GEE have advantages in certain settings in terms of type I error, bias and coverage. The expanded logistic GEE can perform poorly and is sensitive to the chosen working correlation structure. Conclusions about the effectiveness of treatment often differ depending on the method used, highlighting the need to pre-specify an analysis approach. We recommend pre-specifying that either the log Poisson GEE or log normal GEE will be used in the event that the log binomial GEE fails to converge.

KEYWORDS: relative risk, log binomial regression, generalized estimating equation, cluster randomized trial, simulation

Author Notes: The authors would like to thank members of the PoCT Trial Management Committee which included Justin Beilby, Janice Gill, Briony Glastonbury, Roger Killeen, Pamela McKittrick, Caroline Laurence, Mark Shephard, Andrew St John, David Thomas, Phil Tideman, Rosy Tirimacco and Paul Worley. This work was supported by a Commonwealth funded Australian Postgraduate Award.

1 Introduction

The odds ratio has traditionally been used to describe the effect of treatment on binary outcomes in randomized controlled trials (RCTs). The usefulness of the odds ratio is now a matter of debate (Lee, 1994; Sinclair and Bracken, 1994; Walter, 2000; Schechtman, 2002; Cummings, 2009), with some authors expressing a clear preference for the relative risk (Lee, 1994; Sinclair and Bracken, 1994; Cummings, 2009). As a result, relative risks have been reported in many recent RCTs (e.g. Boardman et al., 2004; Green et al., 2008; Ahmad et al., 2009).

Relative risks can be difficult to estimate, particularly when covariate adjustment is required. Log binomial regression has been recommended for estimating relative risks (Skov et al., 1998) but this method may fail to converge (Zou, 2004). Many alternative methods have been suggested (Wacholder, 1986; Flanders and Rhodes, 1987; Schouten et al., 1993; Lee, 1994; Zhang and Yu, 1998; Barros and Hirakata, 2003; Deddens et al., 2003; McNutt et al., 2003; Zou, 2004; Carter et al., 2005; Lumley et al., 2006; Localio et al., 2007; Yu and Wang, 2008; Penman and Johnson, 2009) and numerous comparisons have been made between these methods for independent data (Skov et al., 1998; Barros and Hirakata, 2003; Deddens et al., 2003; McNutt et al., 2003; Zou, 2004; Carter et al., 2005; Blizzard and Hosmer, 2006; Petersen and Deddens, 2008; Yu and Wang, 2008; Kleinman and Norton, 2009; Penman and Johnson, 2009; Yelland et al., 2011). In contrast, comparisons in the context of clustered data are limited (Santos et al., 2008). Given that many of these methods for estimating relative risks are now being applied in cluster RCTs (e.g. Mason et al., 2007; Zeng et al., 2008), investigation of their performance in this context is needed.

A key feature of cluster RCTs is that observations taken on subjects from the same cluster are likely to be dependent. Failure to account for this dependence in the analysis can lead to biased standard errors of parameter estimates and possibly false conclusions about the effectiveness of treatment (Cannon et al., 2001). A popular way to account for dependence in the analysis is to use generalized estimating equations (GEEs) (Liang and Zeger, 1986). This population-averaged approach estimates the marginal or average covariate effect on the outcome. The alternative is to use a cluster-specific approach, which estimates the covariate effect conditional on the value of a random cluster effect (Neuhaus et al., 1991). Use of cluster-specific models to estimate the effect of covariates which remain constant within clusters has been questioned (Pendergast et al., 1991; Neuhaus, 1992). The population-averaged GEE approach may be better suited to estimating the effect of such covariates (Neuhaus, 1992; Austin, 2007), including treatment in cluster RCTs.

In this article, we compare the performance of log binomial regression and several alternative methods for estimating relative risks in the context of cluster RCTs, using GEEs to account for dependence. In Section 2, the methods for estimating relative risks are described and details of the simulation study performed to compare them are provided. Simulation results are presented in Section 3 and an example dataset is described and analyzed in Section 4. We conclude with a discussion and recommendations in Section 5.

2 Methods

2.1 Setting and notation

Consider a two-group parallel cluster RCT comparing a new treatment to a standard or control treatment. Let there be M independent clusters recruited and randomly allocated to the ‘treatment’ or ‘control’ group. Let Y_{ij} be a binary outcome for subject j ($j = 1, \dots, N_i$) in cluster i ($i = 1, \dots, M$), where $Y_{ij} = 1$ if the outcome is a ‘success’, however this is defined, and $Y_{ij} = 0$ otherwise. The probability of success for subject j in cluster i (π_{ij}) is assumed to depend on a vector of known covariates $\mathbf{X}_{ij} = (X_0, X_{1i}, X_{2ij}, \dots, X_{Kij})$, where $X_0 = 1$, X_{1i} is a binary cluster level treatment group indicator, coded as $X_{1i} = 1$ for treatment and $X_{1i} = 0$ for control, and $X_{2ij}, \dots, X_{Kij}$ are $K - 1$ possible cluster and/or individual level baseline covariates; here we restrict our attention to the case where $K = 2$. The aim of the analysis is to estimate the relative risk of success comparing treatment to control, while potentially adjusting for pre-specified categorical and/or continuous baseline covariates.

To achieve this aim, four methods are considered as described in Section 2.2: the log binomial GEE, expanded logistic GEE, log Poisson GEE and log normal GEE, each with an independence or an exchangeable working correlation structure. All methods are implemented using SAS v9.2 (Cary, NC, USA).

2.2 Statistical methods

The log binomial regression model, which has been recommended for estimating relative risks for independent data (Skov et al., 1998), is a generalized linear model combining a log link with a binomial distribution. The log binomial GEE is an extension of this model to the case of clustered data. The GEE model may be written as

$$\log(\pi_{ij}) = \mathbf{X}_{ij}\boldsymbol{\beta}, \quad (1)$$

where $\boldsymbol{\beta}$ is a column vector of model parameters and the data are assumed to follow a binomial distribution conditional on the covariate values. The relative risk comparing treatment to control is estimated by $\exp(\hat{\beta}_1)$, where $\hat{\beta}_1$ is the estimated coefficient of the treatment indicator X_{1i} . This method can be implemented using SAS PROC GENMOD. A unique cluster identifier is specified as the *subject* in the *repeated* statement to account for dependence at the cluster level using GEEs.

The expanded logistic GEE is an extension of the method described by Schouten et al. (1993) for estimating relative risks based on independent data. The authors recognized that by manipulating the data, logistic regression can be used to estimate relative risks directly. An expanded dataset is created, where subjects achieving a successful outcome are duplicated and the outcome is changed to a failure for these duplicates. If there are X successes out of N subjects in the original dataset, the probability of success in this dataset is $\pi = X/N$, and the probability of success in the expanded dataset is $\pi^* = X/(X + N) = \pi/(1 + \pi)$.

This implies that $\pi = \pi^*/(1 - \pi^*)$, i.e. the probability of success in the original dataset equals the odds of success in the expanded dataset. As some subjects are included twice in the expanded dataset, once where the outcome is a success and once where the outcome is a failure, use of this method results in a negative dependence in the data at the subject level. This can be accounted for using GEEs in SAS PROC GENMOD by specifying a unique subject identifier as the *subject* in the *repeated* statement. To extend this method to the case where subjects are nested within clusters, we use GEEs to account for dependence at the cluster level, as for the log binomial GEE. Dependence at the subject level does not need to be specified separately, as subjects are perfectly nested within clusters (Miglioretti and Heagerty, 2007). Since the probability of success in the original dataset equals the odds of success in the expanded dataset, fitting the logistic GEE model

$$\log\left[\pi_{ij}/(1 - \pi_{ij})\right] = \mathbf{X}_{ij}\boldsymbol{\beta} \quad (2)$$

to the expanded dataset results in consistent estimates of the parameters in the log binomial GEE model (1), apart from possibly the intercept (Schouten et al., 1993). Thus, the relative risk comparing treatment to control is estimated by $\exp(\hat{\beta}_1)$,

where $\hat{\beta}_1$ is now the estimated coefficient of the treatment indicator X_{1i} from the logistic GEE model (2) applied to the expanded dataset.

The log Poisson GEE is an extension of the modified Poisson regression approach suggested by Zou (2004) to overcome the convergence problems of the log binomial regression model. For independent data, Zou (2004) suggested fitting a generalized linear model with a log link and replacing the binomial distribution with a Poisson distribution. To avoid overestimation of standard errors, robust variance estimation was suggested. This can be achieved in SAS PROC GENMOD by specifying a unique subject identifier as the *subject* in the *repeated* statement (Spiegelman and Hertzmark, 2005), even though there is only one observation per subject. We consider a straightforward extension of this method for clustered data, using GEEs to account for dependence by instead specifying a unique cluster identifier as the *subject* in the *repeated* statement, and refer to this method as the log Poisson GEE. Since this model has the same form as model (1), the relative risk comparing treatment to control is again estimated by $\exp(\hat{\beta}_1)$.

The log normal GEE also has the same form as model (1). It is identical to the log Poisson GEE, except that the Poisson distribution is replaced with a normal distribution. This approach was proposed by Lumley et al. (2006) in the context of independent data, with robust variance estimation suggested to avoid standard errors being over- or underestimated.

Each of the above methods uses GEEs to account for clustering. GEEs require specification of a working correlation structure to model the dependence in the data. Although the choice of structure will not impact on the consistency of the parameter estimates when the mean model is correctly specified (Liang and Zeger, 1986), choosing a structure closer to the truth can lead to gains in efficiency (Fitzmaurice, 1995; Mancl and Leroux, 1996). Here we consider both independence and exchangeable (equal correlation) working correlation structures. This allows investigation of the potential efficiency gains resulting from applying an exchangeable structure in a cluster RCT setting where exchangeability may be a reasonable assumption.

GEEs are known to underestimate the variance of the parameter estimates when the number of clusters is small (e.g. less than 40 (Lu et al., 2007)), resulting in increased type I error rates (Mancl and DeRouen, 2001) and reduced confidence interval coverage (Lu et al., 2007). Methods have been proposed to correct for the bias in small samples, including that of Mancl and DeRouen (2001). This method provides a bias-corrected covariance estimate, used to perform Wald tests, and has been recommended for cluster RCTs (Lu et al., 2007). Here we consider both moderate and small numbers of clusters (50 and 20 respectively), with bias corrections made in the latter case using the method of

Mancl and DeRouen (2001), as implemented in the SAS macro diag103.sas (Hammill and Preisser, 2006).

2.3 Simulation study

The relative performance of the different methods for estimating relative risks was studied by simulation. For each simulation scenario, 1000 datasets were generated with 50 clusters of (average) size 10, or 20 clusters of (average) size 25, to maintain an (average) total sample size of 500. Cluster sizes were either fixed or generated according to a Poisson distribution. Subjects were randomly assigned an outcome with probability

$$\pi_{ij} = \exp(\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2ij} + u_i), \quad (3)$$

where X_{1i} was a binary cluster level indicator for treatment, X_{2ij} was a binary or continuous cluster or individual level covariate, and u_i was a random cluster effect, used to induce dependence between outcomes of subjects from the same cluster. Clusters were randomized to the treatment or control group using blocked randomization. If π_{ij} exceeded 1 for given values of the treatment indicator, covariate and random cluster effect, new values of the covariate and/or random cluster effect were generated until $\pi_{ij} \leq 1$ (note that new values rarely had to be generated in practice). Although (3) is a cluster-specific model, population-averaged GEEs can be used to estimate the treatment relative risk $\exp(\beta_1)$, since the cluster-specific and population-averaged treatment parameters coincide in this case (Ritz and Spiegelman, 2004).

The expected baseline risk was 0.1 for each simulation scenario, while the treatment and covariate relative risks were 1, 1.25 or 2. Random cluster effects were generated from a normal distribution with mean 0 and variance 0.01, 0.1 or 0.2. This corresponds to an intraclass correlation coefficient (ICC) between 0.001 and 0.055, depending on the treatment and covariate effects (Yelland et al., 2011). ICCs in this range are typical for outcomes from cluster RCTs in practice (Smeeth and Ng, 2002; Elley et al., 2005; Parker et al., 2005). Cluster level covariates were generated from either a Bernoulli distribution with probability 0.5, or a normal distribution with mean 0.5 and variance 0.25. This choice allows direct comparison between results for the binary and continuous covariate, since the mean and variance are equal in this case. Binary individual level covariates were generated from a beta-binomial distribution with an average cluster-specific prevalence of 0.5 and an ICC of 0.05 or 0.2 (Austin, 2007). Continuous individual

level covariates were generated using the model $X_{2ij} = \mu + a_i + e_{ij}$, with overall mean $\mu = 0.5$, and normally distributed random cluster effects (a_i) and error terms (e_{ij}), each with mean 0. Variances were chosen such that the total variance was 0.25 and the ICC was 0.05 or 0.2.

Each simulated dataset was analyzed using the log binomial GEE, expanded logistic GEE, log Poisson GEE and log normal GEE. Analyses were performed with an independence and an exchangeable working correlation structure, and with and without adjustment for the baseline covariate. When the number of clusters was 20, analyses were also performed both with and without the small sample bias correction of Mancl and DeRouen (2001).

The methods were compared based on the following properties, determined for each simulation scenario: convergence rate, calculated as the percentage of simulations where the iterative model fitting algorithm converged to a solution; type I error rate, calculated as the percentage of Wald tests which resulted in a rejection of the null hypothesis of no treatment effect at the 5% level when the null hypothesis was true; power, calculated as the percentage of Wald tests which resulted in a rejection of the null hypothesis of no treatment effect at the 5% level when the null hypothesis was false; median percent relative bias in the estimated relative risk, where relative bias was calculated as the estimated relative risk minus the true relative risk, divided by the true relative risk; coverage rate, calculated as the percentage of 95% confidence intervals containing the true relative risk; maximum prevalence of invalid predicted probabilities (i.e. $\hat{\pi}_{ij} > 1$); and median relative efficiency of the independence compared to the exchangeable working correlation structure, where relative efficiency was calculated as the variance of the estimated treatment coefficient obtained assuming exchangeable correlation, divided by the variance obtained assuming independence. Where a method failed to converge for a particular simulated dataset, results from that dataset were excluded for that method when making comparisons.

Given the large number of simulation scenarios considered (1296 in total), it is not possible to present the full results of the simulation study in this article (full results are available from the first author on request). Instead, we summarize the simulation results across scenarios by calculating the percentage of scenarios where convergence problems occurred, and where the type I error and coverage rate differed significantly from the nominal level, as well as the median and interquartile range (IQR) for the median percent relative bias. For 1000 simulated datasets, type I error rates less than 3.6% or greater than 6.4% differ significantly ($p < 0.05$) from the nominal level of 5% based on a normal approximation test for a proportion. Similarly, coverage rates less than 93.6% or greater than 96.4% differ significantly from the nominal level of 95%. Results are reported by the number of clusters (50 or 20) and the cluster size (fixed or varying). For 20 clusters, only

the results of the bias corrected analyses are presented. To assess whether differences between methods were important, the median (IQR) difference between the best and worst methods was determined for both power and the median percent relative bias.

Results are reported overall and for the subset of simulated datasets where the log binomial GEE failed to converge. This subset is of particular interest, as alternative methods are often only considered when the log binomial GEE does not converge. Since convergence sometimes differed depending on whether an independence or exchangeable working correlation structure was used, results for methods with an independence structure in this subset are based on simulated datasets where the log binomial GEE failed to converge when independence was assumed. Similarly, results for methods with an exchangeable structure in this subset are based on simulated datasets where the log binomial GEE failed to converge when exchangeability was assumed. To avoid conclusions being drawn based on small samples, results are only presented for simulation scenarios where the log binomial GEE failed to converge for a minimum of 50 simulated datasets.

3 Simulation Results

3.1 Convergence

No convergence problems occurred for the expanded logistic GEE, log Poisson GEE or log normal GEE when an independence working correlation structure was used (Table 1). Other methods all suffered from convergence problems in some scenarios. For a fixed cluster size, the log binomial GEE with independence or exchangeable working correlation failed to converge most often, while the expanded logistic GEE was the most commonly affected method for varying cluster size. Convergence problems occurred more often when the variance of the random effects was 0.01, the covariate relative risk was 2, the cluster size varied or adjusted analyses were performed, particularly when the covariate was continuous. Convergence sometimes failed for unadjusted analyses but only when the cluster size varied. For the log binomial GEE, convergence rates fell as low as 89% when the number of clusters was 50, and as low as 8% when the number of clusters was 20. In contrast, convergence rates remained above 95% for the expanded logistic GEE, and above 99% for the log Poisson GEE and log normal GEE, independent of the number of clusters.

Table 1: Number (%)^a of simulation scenarios where convergence problems occurred.

M	N_i	Cluster Size	Method	Working Correlation	
				Independence	Exchangeable
50	10	Fixed	Log Binomial	67 (20.7)	67 (20.7)
			Expanded Logistic	0 (0.0)	2 (0.6)
			Log Poisson	0 (0.0)	0 (0.0)
			Log Normal	0 (0.0)	0 (0.0)
50	10	Varying	Log Binomial	66 (20.4)	86 (26.5)
			Expanded Logistic	0 (0.0)	259 (79.9)
			Log Poisson	0 (0.0)	35 (10.8)
			Log Normal	0 (0.0)	29 (9.0)
20	25	Fixed	Log Binomial	94 (29.0)	94 (29.0)
			Expanded Logistic	0 (0.0)	4 (1.2)
			Log Poisson	0 (0.0)	0 (0.0)
			Log Normal	0 (0.0)	1 (0.3)
20	25	Varying	Log Binomial	104 (32.1)	160 (49.4)
			Expanded Logistic	0 (0.0)	175 (54.0)
			Log Poisson	0 (0.0)	88 (27.2)
			Log Normal	0 (0.0)	92 (28.4)

M = number of clusters, N_i = (average) cluster size.

^aOut of the 324 simulation scenarios considered for each of the four settings presented (50 or 20 clusters and fixed or varying cluster size).

3.2 Type I error

All methods produced some type I error rates which differed significantly from the nominal level, with the log normal GEE generally having the most problems (Table 2). Type I error problems were especially common for the expanded logistic GEE with exchangeable correlation when there were 50 clusters of varying size. Type I error rates were often too large when the covariate relative risk was 2, the variance of the random effects was 0.1 or 0.2, the number of clusters was 50, or an individual level covariate was present and the cluster size varied. Type I error rates ranged from 4.5% to 9.1% when the number of clusters was 50, and 2.1% to 8.4% when the number of clusters was 20. When small sample bias corrections were not made, type I errors were significantly above the nominal level for nearly every method in every scenario involving 20 clusters, ranging from 6.1% to 14.7%.

Table 2: Number (%)^a of simulation scenarios where the type I error rate differed significantly from the nominal level.

M	N_i	Cluster Size	Method	Working Correlation	
				Independence	Exchangeable
50	10	Fixed	Log Binomial	16 (14.8)	16 (14.8)
			Expanded Logistic	11 (10.2)	32 (29.6)
			Log Poisson	12 (11.1)	12 (11.1)
			Log Normal	29 (26.9)	31 (28.7)
50	10	Varying	Log Binomial	47 (43.5)	55 (50.9)
			Expanded Logistic	46 (42.6)	76 (70.4)
			Log Poisson	47 (43.5)	55 (50.9)
			Log Normal	60 (55.6)	64 (59.3)
20	25	Fixed	Log Binomial	8 (7.4)	8 (7.4)
			Expanded Logistic	10 (9.3)	12 (11.1)
			Log Poisson	7 (6.5)	9 (8.3)
			Log Normal	12 (11.1)	16 (14.8)
20	25	Varying	Log Binomial	18 (16.7)	13 (12.0)
			Expanded Logistic	18 (16.7)	19 (17.6)
			Log Poisson	20 (18.5)	18 (16.7)
			Log Normal	30 (27.8)	27 (25.0)

M = number of clusters, N_i = (average) cluster size.

^aOut of the 108 simulation scenarios considered where treatment had no effect for each of the four settings presented (50 or 20 clusters and fixed or varying cluster size).

3.3 Coverage

Coverage rates were significantly below the nominal level for between 5.6% and 72.8% of simulation scenarios, depending on the method and the sample size (Table 3). Problems with coverage were more common for the log normal GEE in general, and for the expanded logistic GEE with exchangeable correlation when there were 50 clusters of varying size. Coverage was often a problem when the covariate relative risk was 2, the variance of the random effects was 0.2, the number of clusters was 50, or an individual level covariate was present and the cluster size varied. The minimum coverage rate was 89.4% and 85.8%, and the maximum was 96.0% and 97.9%, for scenarios with 50 and 20 clusters respectively. In the absence of small sample bias corrections in scenarios with 20 clusters, coverage rates were almost always significantly lower than the nominal level and ranged between 85.3% and 93.9%.

Table 3: Number (%^a) of simulation scenarios where the coverage rate differed significantly from the nominal level.

M	N_i	Cluster Size	Method	Working Correlation	
				Independence	Exchangeable
50	10	Fixed	Log Binomial	81 (25.0)	82 (25.3)
			Expanded Logistic	71 (21.9)	117 (36.1)
			Log Poisson	73 (22.5)	73 (22.5)
			Log Normal	104 (32.1)	108 (33.3)
50	10	Varying	Log Binomial	148 (45.7)	167 (51.5)
			Expanded Logistic	141 (43.5)	236 (72.8)
			Log Poisson	145 (44.8)	156 (48.1)
			Log Normal	180 (55.6)	198 (61.1)
20	25	Fixed	Log Binomial	22 (6.8)	22 (6.8)
			Expanded Logistic	21 (6.5)	24 (7.4)
			Log Poisson	18 (5.6)	20 (6.2)
			Log Normal	34 (10.5)	44 (13.6)
20	25	Varying	Log Binomial	57 (17.6)	55 (17.0)
			Expanded Logistic	62 (19.1)	61 (18.8)
			Log Poisson	65 (20.1)	61 (18.8)
			Log Normal	89 (27.5)	87 (26.9)

M = number of clusters, N_i = (average) cluster size.

^aOut of the 324 simulation scenarios considered for each of the four settings presented (50 or 20 clusters and fixed or varying cluster size).

3.4 Power

When unadjusted analyses were performed, power was the same for the log binomial GEE, log Poisson GEE and log normal GEE with an independence working correlation structure. These methods also had the same power when an exchangeable structure was used, provided the cluster size was fixed. For these settings, the expanded logistic GEE tended to have slightly higher power compared to other methods when the number of clusters was 50, but slightly lower power when only 20 clusters were considered. Power differed between methods otherwise and no method consistently produced the highest or lowest power (data not shown). Differences in power between methods were generally small; the median (IQR) difference in power between the best and worst method was 0.78% (0.40-1.15) and 0.75% (0.40-1.38) when the number of clusters was 50 and 20 respectively.

3.5 Bias

For unadjusted analyses and fixed cluster size, the median percent relative bias was the same for all methods except the expanded logistic GEE with exchangeable correlation, which tended to produce biases that were slightly larger in magnitude compared to other methods. For unadjusted analyses and varying cluster size, the median percent relative bias was the same for all methods assuming independence and similar for methods assuming exchangeable correlation. Where the median percent relative bias was the same across methods, the treatment effect estimates were also the same for each of the individual simulated datasets. Bias otherwise differed, with no method consistently having the smallest or largest median percent relative bias. The median (IQR) difference in the median percent relative bias between the best and worst method was 0.52% (0.28-0.81) for 50 clusters and 0.57% (0.30-1.04) for 20 clusters, indicating that differences in bias between methods were generally small. All methods tended to slightly overestimate the relative risk, except when there were 20 clusters of fixed size (Table 4). The different methods produced similar results overall in terms of the median and IQR for the median percent relative bias.

3.6 Invalid predicted probabilities

No invalid predicted probabilities occurred when unadjusted analyses were performed, when adjustment was made for a binary covariate, or for the log binomial GEE with either working correlation structure. For other methods, adjusting for a continuous covariate often led to some predicted probabilities exceeding 1, particularly for scenarios where the covariate relative risk was 2. Invalid predicted probabilities occurred for between 26.5% and 30.2% of scenarios where adjustment was made for a continuous covariate, depending on the method and the chosen correlation structure. The expanded logistic GEE with an exchangeable working correlation structure often had the largest maximum prevalence of invalid predicted probabilities for a given scenario (data not shown). The prevalence of invalid predicted probabilities was generally quite small, however, with a maximum of only 6.5% across all methods and scenarios.

Table 4: Median (IQR) for the median percent relative bias in the estimated relative risk.

M	N_i	Cluster Size	Method	Working Correlation	
				Independence	Exchangeable
50	10	Fixed	Log Binomial	0.6 (0.0, 1.5)	0.6 (0.0, 1.5)
			Expanded Logistic	0.6 (0.0, 1.5)	0.8 (0.1, 1.5)
			Log Poisson	0.6 (0.0, 1.5)	0.6 (0.0, 1.5)
			Log Normal	0.8 (0.0, 1.7)	0.7 (0.0, 1.7)
50	10	Varying	Log Binomial	0.5 (-0.2, 0.9)	0.4 (-0.2, 1.0)
			Expanded Logistic	0.5 (-0.1, 0.9)	0.4 (-0.1, 1.0)
			Log Poisson	0.5 (-0.1, 0.8)	0.4 (-0.2, 1.0)
			Log Normal	0.5 (-0.1, 1.0)	0.4 (-0.2, 1.0)
20	25	Fixed	Log Binomial	0.0 (-0.5, 0.4)	0.0 (-0.5, 0.3)
			Expanded Logistic	0.0 (-0.3, 0.5)	0.0 (-0.4, 0.4)
			Log Poisson	0.0 (-0.3, 0.4)	0.0 (-0.2, 0.4)
			Log Normal	0.0 (0.0, 0.8)	0.0 (0.0, 0.8)
20	25	Varying	Log Binomial	0.6 (-0.1, 1.3)	0.7 (-0.1, 1.3)
			Expanded Logistic	0.8 (0.2, 1.3)	0.7 (0.2, 1.3)
			Log Poisson	0.8 (0.1, 1.3)	0.8 (0.2, 1.3)
			Log Normal	0.9 (0.2, 1.6)	1.0 (0.3, 1.6)

M = number of clusters, N_i = (average) cluster size.

3.7 Relative efficiency

The relative efficiency of the independence compared to the exchangeable working correlation structure was always 1 for the log binomial GEE, log Poisson GEE and log normal GEE when the cluster size was fixed and either an unadjusted analysis was performed or adjustment was made for a cluster level covariate. Relative efficiencies varied across simulated datasets otherwise but the median remained very close to 1 for all methods in all scenarios (data not shown).

3.8 Differing conclusions

The conclusion about the effectiveness of treatment differed depending on the method used for up to 16.8% of simulated datasets for a given scenario. Inconsistencies were more common when the number of clusters was 20, the cluster size varied or adjusted analyses were performed. In some cases, the different conclusions resulted from p-values that were marginally smaller or marginally greater than 0.05; however, large differences in p-values were also often observed. The log binomial GEE, expanded logistic GEE and log Poisson

GEE with an independence working correlation structure generally led to the same conclusions about the effectiveness of treatment (data not shown).

3.9 Unadjusted versus adjusted analyses

The impact of adjusting for a covariate varied between methods. More convergence problems occurred for adjusted analyses, especially for the log binomial GEE. Convergence problems occurred for this method in 331 (51.1%) and 356 (54.9%) of the 648 scenarios in total where adjustment was made for a covariate, compared to none and 51 (7.9%) of the 648 scenarios where no adjustment was made, using an independence and an exchangeable working correlation structure respectively. The log normal GEE was most affected by the type of analysis otherwise, with adjusted analyses producing substantially more problems with both type I error and coverage rates compared to unadjusted analyses. When an independence working correlation structure was specified, type I error rates differed significantly from the nominal level for this method in 87 (40.3%) of the 216 adjusted scenarios where treatment had no effect, compared to 44 (20.4%) of the 216 unadjusted scenarios. For an exchangeable correlation structure, type I error problems were seen in 96 (44.4%) adjusted scenarios compared to 42 (19.4%) unadjusted scenarios. Significant differences in coverage rates from the nominal level increased from 153 (23.6%) and 158 (24.4%) scenarios when an unadjusted analysis was performed, to 254 (39.2%) and 279 (43.1%) scenarios when an adjusted analysis was performed, for independence and exchangeable correlation respectively. Changes in type I error and coverage due to adjustment for other methods were generally small. Covariate adjustment had little impact on bias for all methods. Adjusted analyses tended to be more powerful than unadjusted analyses for scenarios with 50 clusters, but less powerful for 20 clusters. The median (IQR) change in power due to adjustment across all methods was 0.3% (-0.1% to 1.0%) and -0.3% (-1.1% to 0.3%) for scenarios with 50 and 20 clusters respectively.

3.10 Subset where log binomial GEE failed to converge

The log binomial GEE failed to converge for a minimum of 50 simulated datasets in 116 (9.0%) scenarios assuming independence and 121 (9.3%) scenarios assuming exchangeability, all of which involved adjustment for a continuous covariate. For these scenarios, the median (IQR) number of simulated datasets where the log binomial GEE failed to converge was 105 (78-627). The performance of the expanded logistic GEE, log Poisson GEE and log normal GEE in this subset is shown in Table 5. All methods performed similarly in terms of convergence, with no problems observed when independence was assumed but problems for the majority of scenarios (between 62.0% and 65.3%) when

exchangeability was assumed. The log normal GEE had the fewest problems with type I error and coverage, and tended to have smaller median percent relative bias compared to the expanded logistic GEE and log Poisson GEE. The log normal GEE also typically had the lowest power (data not shown). The median (IQR) difference between the best and worst method was 2.6% (1.6-4.3) for the median percent relative bias and 4.5% (2.6-6.8) for power, indicating that differences between methods could be important in this subset.

Table 5: Number (%) of simulation scenarios where convergence problems occurred, and where the type I error and coverage rates differed significantly from the nominal level, and median (IQR) for the median percent relative bias in the estimated relative risk for the subset of simulation scenarios where the log binomial GEE failed to converge.

Statistical Property	Method	Working Correlation	
		Independence	Exchangeable
Convergence Problems: N (%)	Expanded Logistic	0 (0.0)	79 (65.3)
	Log Poisson	0 (0.0)	75 (62.0)
	Log Normal	0 (0.0)	75 (62.0)
Type I Error Problems: N (%)	Expanded Logistic	13 (38.2)	15 (40.5)
	Log Poisson	11 (32.4)	16 (43.2)
	Log Normal	4 (11.8)	7 (18.9)
Coverage Problems: N (%)	Expanded Logistic	32 (27.6)	38 (31.4)
	Log Poisson	30 (25.9)	35 (28.9)
	Log Normal	18 (15.5)	20 (16.5)
Median Percent Relative Bias: Median (IQR)	Expanded Logistic	4.1 (0.7-9.9)	3.7 (0.7, 9.5)
	Log Poisson	3.9 (0.6-9.4)	3.5 (0.6, 8.9)
	Log Normal	2.8 (0.2-7.9)	2.1 (-0.1, 6.8)

4 Example

As an example, we consider data from the Point of Care Testing (PoCT) in General Practice Trial (Laurence et al., 2008), a cluster RCT of PoCT versus pathology laboratory testing in an Australian general practice setting. Practices were randomized to the intervention or control group using blocked randomization within strata defined by geographic location. Intervention practices were provided with PoCT devices for measuring pathology results, while practices in the control group continued to provide standard care which involved patient blood samples being analyzed at a pathology laboratory. Patients were recruited through the practices and were eligible to participate if they were aged 18 years and over and had diabetes and/or hyperlipidaemia and/or were taking

anticoagulant medication. The primary outcome for the trial was whether the final test result was within or outside a pre-specified target range (Bubner et al., 2009), based on clinical guidelines, for each of seven tests included in the trial. We consider one of these tests, glycated haemoglobin (HbA1c), for illustration. Outcome data were available for 1592 diabetic patients (912 intervention, 680 control) from 53 practices (30 intervention, 23 control) with an average cluster size of approximately 30 patients. The ICC for this outcome was 0.04, indicating a weak positive dependence between outcomes of patients from the same practice.

The data were analyzed using each of the methods described in Section 2. Both unadjusted and adjusted analyses were performed, with adjustment made for each of the following variables separately: geographic location (urban or non-urban), a binary cluster level covariate with 33% urban; gender, a binary patient level covariate with 54% male; and age, a continuous patient level covariate which was approximately normally distributed with mean 65 years and standard deviation 11 years. There was some evidence to indicate that the risk of having a final HbA1c result within the target range was greater for females compared to males, and very strong evidence that the risk increased with age. Geographic location did not appear to have an effect on the outcome.

In the intervention group, 64.5% of patients had their final HbA1c result within the target range compared to 50.4% of control patients. The estimated relative risk for whether the final HbA1c result was within target range varied depending on the method of analysis and choice of covariate for adjustment, ranging from 1.21 to 1.34 (Table 6). When an independence working correlation structure was assumed, all methods produced an identical estimate of the treatment effect for the unadjusted analysis, as expected from the simulation study. Results were also similar between these methods for adjusted analyses, no matter which variable was adjusted for. When an exchangeable structure was used, the log binomial GEE, log Poisson GEE and log normal GEE produced similar treatment effect estimates which were smaller than the corresponding estimates under independence, while the expanded logistic GEE produced the largest estimates. This pattern occurred quite often in the simulation study. The conclusion was the same independent of the method used to estimate the relative risk and the covariate (if any) that was controlled for in the analysis: intervention patients were significantly more likely to be in target range compared to control patients.

To provide an example with fewer clusters, 20 practices from the PoCT trial were randomly selected and the outcome was re-analyzed for this subset. The results, corrected for the small sample bias, are shown in Table 7. The pattern seen in the relative risk estimates is similar to that seen when all 53 practices were included in the analysis, with estimates ranging from 1.26 to 1.36. However, the log binomial GEE with either an independence or exchangeable correlation

structure now fails to converge when adjustment is made for age. This method often suffered from convergence problems in the simulation study, especially when adjustment was made for a continuous covariate. As expected, the 95% confidence intervals are wider compared to analysis of the full dataset. The lower limit of the 95% confidence intervals ranges from 0.89 to 1.01 and the effect of treatment is no longer significant at the 5% level, unless adjustment is made for age and analysis is performed using either the expanded logistic GEE with an exchangeable correlation structure, or the log normal GEE with an independence structure. Use of different methods often led to different conclusions being drawn about the effect of treatment in the simulation study.

Table 6: Relative risk (95% confidence interval) from analysis of whether the final HbA1c test result was within target range for diabetic patients in the PoCT Trial based on all 53 practices.

Covariate Adjustment	Method	Working Correlation	
		Independence	Exchangeable
None	Log Binomial	1.28 (1.10, 1.48)	1.22 (1.06, 1.41)
	Expanded Logistic	1.28 (1.10, 1.48)	1.33 (1.14, 1.54)
	Log Poisson	1.28 (1.10, 1.48)	1.22 (1.06, 1.42)
	Log Normal	1.28 (1.10, 1.48)	1.22 (1.06, 1.41)
Geographic Location	Log Binomial	1.27 (1.09, 1.48)	1.21 (1.04, 1.41)
	Expanded Logistic	1.27 (1.09, 1.48)	1.32 (1.13, 1.53)
	Log Poisson	1.27 (1.09, 1.48)	1.21 (1.04, 1.41)
	Log Normal	1.27 (1.08, 1.48)	1.21 (1.04, 1.41)
Gender	Log Binomial	1.28 (1.11, 1.48)	1.24 (1.07, 1.43)
	Expanded Logistic	1.28 (1.10, 1.48)	1.33 (1.14, 1.54)
	Log Poisson	1.28 (1.11, 1.48)	1.23 (1.06, 1.42)
	Log Normal	1.28 (1.11, 1.48)	1.23 (1.07, 1.43)
Age	Log Binomial	1.29 (1.11, 1.49)	1.24 (1.07, 1.43)
	Expanded Logistic	1.29 (1.11, 1.49)	1.34 (1.15, 1.56)
	Log Poisson	1.29 (1.11, 1.49)	1.23 (1.07, 1.42)
	Log Normal	1.29 (1.11, 1.49)	1.24 (1.07, 1.42)

5 Discussion

In this article, we compared methods for estimating relative risks in the context of cluster RCTs, using GEEs with an independence or exchangeable working correlation structure to account for dependence. The results of our simulation study confirm that convergence problems can occur for the log binomial model, as shown previously in the context of independent data (Deddens et al., 2003;

Blizzard and Hosmer, 2006; Yu and Wang, 2008; Yelland et al., 2011). This highlights the need to consider alternative methods for estimating relative risks in cluster RCTs, such as the expanded logistic GEE, log Poisson GEE and log normal GEE. These methods were also shown to suffer from convergence problems when applied to clustered data but only when an exchangeable working correlation structure was assumed. In general, our simulation study shows that there is variability between methods for estimating relative risks in terms of type I error, coverage, power and bias. Observed differences between methods were often small, however, and no single method performed best overall. This suggests that a number of methods may be reasonable for estimating relative risks from clustered data in practice.

Table 7: Relative risk (95% confidence interval) from analysis of whether the final HbA1c test result was within target range for diabetic patients in the PoCT Trial based on a random sample of 20 practices.

Covariate Adjustment	Method	Working Correlation	
		Independence	Exchangeable
None	Log Binomial	1.31 (0.98, 1.77)	1.27 (0.94, 1.73)
	Expanded Logistic	1.31 (0.98, 1.77)	1.34 (0.99, 1.81)
	Log Poisson	1.31 (0.98, 1.77)	1.27 (0.93, 1.72)
	Log Normal	1.31 (0.98, 1.77)	1.27 (0.94, 1.73)
Geographic Location	Log Binomial	1.30 (0.93, 1.83)	1.27 (0.89, 1.80)
	Expanded Logistic	1.30 (0.93, 1.84)	1.32 (0.95, 1.84)
	Log Poisson	1.30 (0.93, 1.83)	1.26 (0.89, 1.79)
	Log Normal	1.30 (0.93, 1.83)	1.26 (0.89, 1.79)
Gender	Log Binomial	1.33 (0.99, 1.79)	1.31 (0.97, 1.78)
	Expanded Logistic	1.31 (0.97, 1.77)	1.33 (0.99, 1.80)
	Log Poisson	1.31 (0.97, 1.77)	1.28 (0.94, 1.74)
	Log Normal	1.33 (0.99, 1.78)	1.30 (0.96, 1.77)
Age	Log Binomial	Did not converge	Did not converge
	Expanded Logistic	1.33 (0.99, 1.79)	1.36 (1.01, 1.84)
	Log Poisson	1.33 (0.99, 1.79)	1.30 (0.96, 1.74)
	Log Normal	1.34 (1.01, 1.78)	1.32 (0.99, 1.75)

Only one previous study compared methods for estimating relative risks in the context of clustered data (Santos et al., 2008). The authors conducted a simulation study to compare log Poisson regression and several other methods based on fitting a logistic regression model and using different methods of standardization (Flanders and Rhodes, 1987) to calculate the relative risk. Clustering was accounted for in the analysis using random effects models, rather than GEEs. The log Poisson approach was inferior to the logistic regression

approach in terms of coverage but this may be due to the chosen simulation model. The data were simulated under a logistic model which assumes a constant odds ratio, rather than a constant relative risk. The authors did not specify how the 'true' relative risk was calculated from this model in order to determine coverage. A range of options are available, such as calculating a conditional relative risk that applies to subjects with a specified covariate pattern (Flanders and Rhodes, 1987). Choosing a different method for calculating the true relative risk may have produced different results, as demonstrated previously for independent observations (Yelland et al., 2011). We simulated data under a log binomial model which assumes a constant relative risk and found that the log Poisson approach performed well in this setting.

The expanded logistic GEE performed somewhat differently compared to other methods when an exchangeable working correlation structure was specified. This method regularly failed to converge and had problems with both type I error and coverage for the majority of the simulation scenarios considered with 50 clusters of varying size. Further, while other methods produced identical treatment effect estimates in scenarios where the cluster size was fixed and unadjusted analyses were performed, the expanded logistic GEE with exchangeable correlation produced estimates that differed. In the example dataset, the estimated relative risk was also somewhat larger for this method compared to estimates based on other methods. The expanded logistic GEE requires the data to be modified prior to analysis such that some subjects have multiple outcomes that are perfectly negatively correlated. Although outcomes were positively correlated within clusters for the simulation study, the negative correlation within subjects often resulted in negative estimates of the working correlation parameter for this method. This may explain the observed differences between the expanded logistic GEE with exchangeable correlation and other methods; using GEEs to analyze negatively correlated data can lead to convergence problems or produce unreliable results and an independence working correlation structure may be preferable in this case (Hanley et al., 2000).

When the number of clusters is small, GEEs underestimate the variance of parameter estimates and bias corrections are required to maintain appropriate type I error and coverage rates (Mancl and DeRouen, 2001; Lu et al., 2007). This was confirmed by the current study, where failure to apply a small sample bias correction in scenarios involving 20 clusters almost always resulted in type I error rates that were too high and coverage rates that were too low compared to the nominal level. Small sample corrections are currently recommended when the number of clusters is less than 40 (Lu et al., 2007). Interestingly, our simulation study showed that problems with type I error and coverage were common when the number of clusters was 50 and the cluster size varied. Further investigation of these scenarios revealed that many of these problems could be avoided by

applying bias correction techniques intended for small samples. This suggests that bias correction methods may also have a role to play when the number of clusters is moderate and the cluster size varies. A previous simulation study showed that when no bias correction was applied, type I error rates were more inflated for varying cluster sizes compared to fixed cluster sizes, although the number of clusters did not exceed 40 (Mancl and DeRouen, 2001). Further research is needed to determine the range of circumstances under which bias correction techniques are beneficial for GEEs.

Adjustment for baseline covariates is often of interest in analyses of cluster RCTs. We found that covariate adjustment resulted in more convergence problems for the log binomial GEE, suggesting that alternative methods for estimating relative risks are more likely to be required if adjusted analyses are planned. Of the different alternative methods considered, the log normal GEE was most sensitive to the choice between unadjusted and adjusted analyses, with poorer performance in terms of type I error and coverage observed in the latter setting.

Recommendation of a single method for estimating relative risks in cluster RCTs is difficult, since no single method showed superior performance in the simulation study across all scenarios or based on all statistical properties. In general, we do not recommend using the expanded logistic GEE. This method was proposed for independent data at a time when software for estimating parameters from models with non-canonical links, such as the log link for binary data, was not readily available. The method is more difficult to explain and implement compared to other methods and we found no advantages in terms of performance in the simulation study. In fact, the expanded logistic GEE often performed relatively poorly compared to other methods when an exchangeable correlation structure was specified.

Excluding the expanded logistic GEE, we found that the performance of other methods did not differ greatly from each other in the simulation study overall, suggesting that all may be acceptable for use in practice. The log binomial GEE generally performed well but this method may not be an option in practice due to convergence problems, leaving the log Poisson GEE and log normal GEE for consideration. Overall, the log Poisson GEE performed slightly better than the log normal GEE in terms of type I error, coverage and bias. However, the opposite was true in the subset where the log binomial GEE failed to converge. While the latter results may be of greater interest, since alternative methods are often only used when the log binomial GEE does not converge, these findings should be interpreted with some caution as they are based on a relatively small subset (approximately 9% of all simulation scenarios considered with a median of only 105 simulated datasets per scenario). We therefore suggest that both the log Poisson GEE and log normal GEE are worthy of consideration for

analyzing cluster RCTs in practice, either as the chosen method for estimating relative risks or as an alternative method to be used in the event that the log binomial GEE fails to converge. If both unadjusted and adjusted analyses are planned, the log Poisson GEE may be preferable, since the performance of the log normal GEE is more sensitive to the choice between unadjusted and adjusted analyses.

Whichever method is chosen to estimate relative risks in practice, it is important that the method is pre-specified. Given that different methods often resulted in different conclusions being drawn about the effectiveness of treatment in the simulation study, pre-specification avoids the possibility of a method being selected based on the favorability of the results.

When using GEEs to account for clustering, a choice must be made about which working correlation structure to use. An independence working correlation structure has been recommended for RCTs with clustered data (Dahmen and Ziegler, 2004). Our results suggest that this approach is reasonable for cluster RCTs, since the estimated treatment effect obtained using an independence working correlation structure was highly efficient compared to the more realistic exchangeable structure, consistent with theoretical results (Mienl and Leroux, 1996). Further, assuming independence had the advantage of producing fewer problems with convergence, type I error and coverage compared to assuming exchangeability.

Type I error and power were calculated for the simulation study based on a Wald test of the null hypothesis of no treatment effect. The Wald test was chosen as this could be performed when small sample bias corrections were applied and led to conclusions that were consistent with the 95% confidence intervals for the treatment effect. Use of the score test may be preferable in practice, however. We investigated results based on the score test for scenarios where the number of clusters was 50 and hence small sample bias corrections were not applied. The score test produced fewer type I error problems overall compared to the Wald test but did not alter our recommendations.

We simulated data under a cluster-specific model, using normally distributed random effects to induce correlation, but analyzed the data using population-averaged GEEs. This approach is valid for the log link since the parameters of the cluster-specific and population-averaged models coincide, apart from the intercept, but may not be suitable for other links (Ritz and Spiegelman, 2004). For instance, the parameters of the two models differ for the logit link (Neuhaus et al., 1991) and failure to recognize this may produce misleading results (e.g. Colak and Ozdamar, 2007). We also simulated data such that the effect of the covariate on the outcome was the same both between and within clusters. In practice, this may not be the case and possible differences in between

and within cluster covariate effects should be investigated (Neuhaus and Kalbfleisch, 1998).

A limitation of this study is that we only considered extensions of some of the available methods for estimating relative risks for independent data to the case of clustered data. These methods were chosen as they were easily extended to account for clustering and were simple to implement using available software, making them more likely to be used in practice. The chosen methods have also been shown to perform well in the context of independent data (Yelland et al., 2011). Methods that perform poorly with independent data would also be expected to perform poorly with clustered data when the dependence is small, as is typically the case in cluster RCTs (Smeeth and Ng, 2002; Elley et al., 2005; Parker et al., 2005), and were therefore not considered here. A further limitation is that we simulated data under a log binomial model only, which assumes a constant relative risk. Simulating data under an alternative model (e.g. a logistic model) is complicated by the fact that the relative risk is no longer constant. In order to determine the ‘true’ relative risk for comparison with simulation results, standardization (Flanders and Rhodes, 1987) is required and results may depend on the method of standardization. Log Poisson and log normal regression have been shown to perform relatively well under different simulation models in an independent data setting when the true relative risk is determined using marginal standardization (Yelland et al., 2011).

In conclusion, the relative risk is now a popular measure of effect for binary outcomes in RCTs, including cluster RCTs. Relative risks can be estimated using log binomial regression with GEEs to account for clustering but this model may fail to converge. When this occurs, the log Poisson GEE and log normal GEE are both reasonable alternatives for estimating relative risks. If the log binomial GEE is pre-specified as the method of analysis for a binary outcome in a cluster RCT, we recommend pre-specifying that either the log Poisson GEE or log normal GEE will be used in the event that the log binomial GEE fails to converge.

References

- Ahmad, F., Hogg-Johnson, S., Stewart, D. E., Skinner, H. A., Glazier, R. H. and Levinson, W. (2009). Computer-assisted screening for intimate partner violence and control: a randomized trial. *Annals of Internal Medicine* 151(2): 93-102.
- Austin, P. C. (2007). A comparison of the statistical power of different methods for the analysis of cluster randomization trials with binary outcomes. *Statistics in Medicine* 26(19): 3550-3565.

- Barros, A. J. and Hirakata, V. N. (2003). Alternatives for logistic regression in cross-sectional studies: an empirical comparison of models that directly estimate the prevalence ratio. *BMC Medical Research Methodology* 3: 21.
- Blizzard, L. and Hosmer, D. W. (2006). Parameter estimation and goodness-of-fit in log binomial regression. *Biometrical Journal* 48(1): 5-22.
- Boardman, L. A., Steinhoff, M. M., Shackelton, R., Weitzen, S. and Crowthers, L. (2004). A randomized trial of the Fischer cone biopsy excisor and loop electrosurgical excision procedure. *Obstetrics and Gynecology* 104(4): 745-750.
- Bubner, T. K., Laurence, C., Gialamas, A., Yelland, L. N., Ryan, P., Willson, K. J., Tideman, P., Worley, P. and Beilby, J. J. (2009). Effectiveness of point-of-care testing for therapeutic control of chronic conditions: results from the PoCT in General Practice Trial. *Medical Journal of Australia* 190(11): 624-626.
- Cannon, M. J., Warner, L., Taddei, J. A. and Kleinbaum, D. G. (2001). What can go wrong when you assume that correlated data are independent: an illustration from the evaluation of a childhood health intervention in Brazil. *Statistics in Medicine* 20(9-10): 1461-1467.
- Carter, R. E., Lipsitz, S. R. and Tilley, B. C. (2005). Quasi-likelihood estimation for relative risk regression models. *Biostatistics* 6(1): 39-44.
- Colak, E. and Ozdamar, K. (2007). Comparison of generalized estimating equations' performance in clustered binary observations. *Trakya Universitesi Tip Fakultesi Dergisi* 24(3): 222-227.
- Cummings, P. (2009). The relative merits of risk ratios and odds ratios. *Archives of Pediatrics & Adolescent Medicine* 163(5): 438-445.
- Dahmen, G. and Ziegler, A. (2004). Generalized estimating equations in controlled clinical trials: hypotheses testing. *Biometrical Journal* 46(2): 214-232.
- Deddens, J., Petersen, M. R. and Lei, X. (2003). Estimation of prevalence ratios when PROC GENMOD does not converge. *Proceedings of the 28th Annual SAS Users Group International Conference*, Paper 270-28. SAS Institute Inc., Cary, NC.

- Elley, C. R., Kerse, N., Chondros, P. and Robinson, E. (2005). Intraclass correlation coefficients from three cluster randomised controlled trials in primary and residential health care. *Australian and New Zealand Journal of Public Health* 29(5): 461-467.
- Fitzmaurice, G. M. (1995). A caveat concerning independence estimating equations with multivariate binary data. *Biometrics* 51(1): 309-317.
- Flanders, W. D. and Rhodes, P. H. (1987). Large sample confidence intervals for regression standardized risks, risk ratios, and risk differences. *Journal of Chronic Diseases* 40(7): 697-704.
- Green, B. B., Cook, A. J., Ralston, J. D., Fishman, P. A., Catz, S. L., Carlson, J., Carrell, D., Tyll, L., Larson, E. B. and Thompson, R. S. (2008). Effectiveness of home blood pressure monitoring, web communication, and pharmacist care on hypertension control: a randomized controlled trial. *Journal of the American Medical Association* 299(24): 2857-2867.
- Hammill, B. G. and Preisser, J. S. (2006). A SAS/IML software program for GEE and regression diagnostics. *Computational Statistics & Data Analysis* 51(2): 1197-1212.
- Hanley, J. A., Negassa, A. and Edwardes, M. D. D. (2000). GEE analysis of negatively correlated binary responses: a caution. *Statistics in Medicine* 19(5): 715-722.
- Kleinman, L. C. and Norton, E. C. (2009). What's the risk? A simple approach for estimating adjusted risk measures from nonlinear models including logistic regression. *Health Services Research* 44(1): 288-302.
- Laurence, C., Gialamas, A., Yelland, L., Bubner, T., Ryan, P., Willson, K., Glastonbury, B., Gill, J., Shephard, M. and Beilby, J. (2008). A pragmatic cluster randomised controlled trial to evaluate the safety, clinical effectiveness, cost effectiveness and satisfaction with point of care testing in a general practice setting - Rationale, design and baseline characteristics. *Trials* 9: 50.
- Lee, J. (1994). Odds ratio or relative risk for cross-sectional data. *International Journal of Epidemiology* 23(1): 201-203.
- Liang, K. Y. and Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika* 73(1): 13-22.

- Localio, A. R., Margolis, D. J. and Berlin, J. A. (2007). Relative risks and confidence intervals were easily computed indirectly from multivariable logistic regression. *Journal of Clinical Epidemiology* 60(9): 874-882.
- Lu, B., Preisser, J. S., Qaqish, B. F., Suchindran, C., Bangdiwala, S. and Wolfson, M. (2007). A comparison of two bias-corrected covariance estimators for generalized estimating equations. *Biometrics* 63(3): 935-941.
- Lumley, T., Kronmal, R. and Ma, S. (2006). Relative risk regression in medical research: models, contrasts, estimators, and algorithms. *UW Biostatistics Working Paper Series: Working Paper* 293.
- Mancl, L. A. and DeRouen, T. A. (2001). A covariance estimator for GEE with improved small-sample properties. *Biometrics* 57(1): 126-134.
- Mancl, L. A. and Leroux, B. G. (1996). Efficiency of regression estimates for clustered data. *Biometrics* 52(2): 500-511.
- Mason, S., Knowles, E., Colwell, B., Dixon, S., Wardrope, J., Gorringer, R., Snooks, H., Perrin, J. and Nicholl, J. (2007). Effectiveness of paramedic practitioners in attending 999 calls from elderly people in the community: cluster randomised controlled trial. *British Medical Journal* 335: 919.
- McNutt, L. A., Wu, C. T., Xue, X. N. and Hafner, J. P. (2003). Estimating the relative risk in cohort studies and clinical trials of common outcomes. *American Journal of Epidemiology* 157(10): 940-943.
- Miglioretti, D. L. and Heagerty, P. J. (2007). Marginal modeling of nonnested multilevel data using standard software. *American Journal of Epidemiology* 165(4): 453-463.
- Neuhaus, J. M. (1992). Statistical methods for longitudinal and clustered designs with binary responses. *Statistical Methods in Medical Research* 1(3): 249-73.
- Neuhaus, J. M. and Kalbfleisch, J. D. (1998). Between- and within-cluster covariate effects in the analysis of clustered data. *Biometrics* 54(2): 638-645.
- Neuhaus, J. M., Kalbfleisch, J. D. and Hauck, W. W. (1991). A comparison of cluster-specific and population-averaged approaches for analyzing correlated binary data. *International Statistical Review* 59(1): 25-35.

- Parker, D. R., Evangelou, E. and Eaton, C. B. (2005). Intraclass correlation coefficients for cluster randomized trials in primary care: the cholesterol education and research trial (CEART). *Contemporary Clinical Trials* 26(2): 260-267.
- Pendergast, J. F., Gange, S. J., Newton, M. A., Lindstrom, M. J., Palta, M. and Fisher, M. R. (1991). A survey of methods for analyzing clustered binary response data. *International Statistical Review* 64(1): 89-118.
- Penman, A. D. and Johnson, W. D. (2009). Complementary log-log regression for the estimation of covariate-adjusted prevalence ratios in the analysis of data from cross-sectional studies. *Biometrical Journal* 51(3): 433-442.
- Petersen, M. R. and Deddens, J. A. (2008). A comparison of two methods for estimating prevalence ratios. *BMC Medical Research Methodology* 8: 9.
- Ritz, J. and Spiegelman, D. (2004). Equivalence of conditional and marginal regression models for clustered and longitudinal data. *Statistical Methods in Medical Research* 13(4): 309-323.
- Santos, C. A. S. T., Fiaccone, R. L., Oliveira, N. F., Cunha, S., Barreto, M. L., do Carmo, M. B. B., Moncayo, A. L., Rodrigues, L. C., Cooper, P. J. and Amorim, L. D. (2008). Estimating adjusted prevalence ratio in clustered cross-sectional epidemiological data. *BMC Medical Research Methodology* 8: 80.
- Schechtman, E. (2002). Odds ratio, relative risk, absolute risk reduction, and the number needed to treat - Which of these should we use? *Value in Health* 5(5): 431-436.
- Schouten, E. G., Dekker, J. M., Kok, F. J., Leccesie, S., Vanhouwelingen, H. C., Pool, J. and Vandenbroucke, J. P. (1993). Risk ratio and rate ratio estimation in case-cohort designs: hypertension and cardiovascular mortality. *Statistics in Medicine* 12(18): 1733-1745.
- Sinclair, J. C. and Bracken, M. B. (1994). Clinically useful measures of effect in binary analyses of randomized trials. *Journal of Clinical Epidemiology* 47(8): 881-889.
- Skov, T., Deddens, J., Petersen, M. R. and Endahl, L. (1998). Prevalence proportion ratios: estimation and hypothesis testing. *International Journal of Epidemiology* 27(1): 91-95.

- Smeeth, L. and Ng, E. S. W. (2002). Intraclass correlation coefficients for cluster randomized trials in primary care: data from the MRC Trial of the Assessment and Management of Older People in the Community. *Controlled Clinical Trials* 23(4): 409-421.
- Spiegelman, D. and Hertzmark, E. (2005). Easy SAS calculations for risk or prevalence ratios and differences. *American Journal of Epidemiology* 162(3): 199-200.
- Wacholder, S. (1986). Binomial regression in GLIM: estimating risk ratios and risk differences. *American Journal of Epidemiology* 123(1): 174-184.
- Walter, S. D. (2000). Choice of effect measure for epidemiological data. *Journal of Clinical Epidemiology* 53(9): 931-939.
- Yelland, L. N., Salter, A. B. and Ryan, P. (2011). Relative risk estimation in randomized controlled trials: a comparison of methods for independent observations. *The International Journal of Biostatistics* 7(1): 5.
- Yelland, L. N., Salter, A. B., Ryan, P. and Laurence, C. O. (2011). Adjusted intraclass correlation coefficients for binary data: methods and estimates from a cluster-randomized trial in primary care. *Clinical Trials* 8: 48-58.
- Yu, B. B. and Wang, Z. Q. (2008). Estimating relative risks for common outcome using PROC NLP. *Computer Methods and Programs in Biomedicine* 90(2): 179-186.
- Zeng, L. X., Cheng, Y., Dang, S. N., Yan, H., Dibley, M. J., Chang, S. Y. and Kong, L. Z. (2008). Impact of micronutrient supplementation during pregnancy on birth weight, duration of gestation, and perinatal mortality in rural western China: double blind cluster randomised controlled trial. *British Medical Journal* 337: a2001.
- Zhang, J. and Yu, K. F. (1998). What's the relative risk? A method of correcting the odds ratio in cohort studies of common outcomes. *Journal of the American Medical Association* 280(19): 1690-1691.
- Zou, G. Y. (2004). A modified Poisson regression approach to prospective studies with binary data. *American Journal of Epidemiology* 159(7): 702-706.