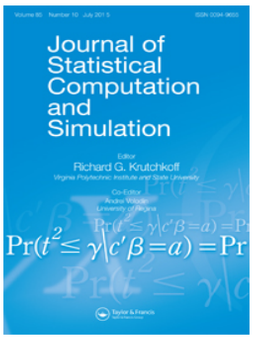

Taylor & Francis
Taylor & Francis Group

Journal of Statistical Computation and Simulation

ISSN: 0094-9655 (Print) 1563-5163 (Online) Journal homepage: www.tandfonline.com/journals/gscs20

Comparing alternating logistic regressions to other approaches to modelling correlated binary data

S.R. Bhatnagar, J. Atherton & A. Benedetti

To cite this article: S.R. Bhatnagar, J. Atherton & A. Benedetti (2015) Comparing alternating logistic regressions to other approaches to modelling correlated binary data, Journal of Statistical Computation and Simulation, 85:10, 2059-2071, DOI: 10.1080/00949655.2014.916707

To link to this article: https://doi.org/10.1080/00949655.2014.916707

Journal of Statistical Computation and Simulation, 2015
Vol. 85, No. 10, 2059–2071, http://dx.doi.org/10.1080/00949655.2014.916707

Taylor & Francis
Taylor & Francis Group

Comparing alternating logistic regressions to other approaches to modelling correlated binary data

S.R. Bhatnagar[a], J. Atherton[b] and A. Benedetti[a,c,d*]

[a]*Department of Epidemiology, Biostatistics and Occupational Health, McGill University, 1020 Ave des Pins Ouest, Montréal, Québec, Canada H3A 1A2;* [b]*Département de mathématiques, Université du Québec à Montréal, 201 Président-Kennedy Ave, Montréal, Québec, Canada H2X 3Y7;* [c]*Department of Medicine, McGill University, Montréal, Québec, Canada H3A 1A2;* [d]*Respiratory Epidemiology and Clinical Research Unit, McGill University Health Centre, 3650 St. Urbain Street, Montréal, Québec, Canada H2X 2P4*

(*Received 7 July 2013; accepted 16 April 2014*)

Alternating logistic regressions (ALRs) seem to offer some of the advantages of marginal models estimated via generalized estimating equations (GEE) and generalized linear mixed models (GLMMs). Via simulation study we compared ALRs to marginal models estimated via GEE and subject-specific models estimated via GLMMs, with a focus on estimation of the correlation structure in three-level data sets (e.g. students in classes in schools). Data set size and structure, and amount of correlation in the data sets were varied. For simple correlation structures, ALRs performed well. For three-level correlation structures, all approaches, but especially ALRs, had difficulty assigning the correlation to the correct level, though sample sizes used were small. In addition, ALRs and GEEs had trouble attaching correct inference to the mean effects, though this improved as overall sample size improved. ALRs are a valuable addition to the data analyst's toolkit, though care should be taken when modelling data with three-level structures.

Keywords: three-level data; alternating logistic regressions; generalized estimating equations; penalized quasi-likelihood; generalized linear mixed models; adaptive Gaussian Hermite quadrature; hierarchical data

AMS Subject Classification: 62J99; 62H20

1. Introduction

Increasingly, data are collected in which the standard assumption of independence between observations is not met. This could include data that consist of multiple observations on a subject over time or subjects who are clustered in some way (e.g. classes within schools, or households within neighbourhoods). For example, Katz et al. [1] assessed clustering of diarrhoea in preschool aged children at the village and household level. As computational power has grown, analytic methods have been extended to handle increasingly complex data structures.

If the association between observations on the same cluster/subject is not accounted for in the analytic strategy, inference associated with the estimated parameters may not be correct – the standard errors may be too small resulting in *p*-values that are too small and confidence intervals that are too narrow.[2]

*Corresponding author. Email: andrea.benedetti@mcgill.ca

Two broad analytic approaches are possible: (1) marginal or population-averaged models, usually estimated via generalized estimating equation (GEE), and (2) generalized linear mixed models (GLMMs), i.e. subject-specific models.[3,4]

In the marginal model, the full probability model for Y is not specified.[4] Rather, it specifies a form for the mean value of Y. The second order, or correlation structure, is considered a nuisance which must be accounted for in order to achieve correct inference.[4] Moreover, given that the choice of a correlation structure is not very important (when using the robust variance estimator); relatively few are available in commercially available software. Marginal models are very popular for binary outcomes when observations are not independent because of their computational ease, the fact that interpretation of the regression coefficients does not change depending on which levels of clustering are accounted for, and because when using the robust variance estimator, the choice of the correlation structure is not very important.[4]

GLMMs extend the linear mixed model to deal with non-continuous outcomes. In the GLMM, usually, normally distributed subject-specific random effects are incorporated into the model.[4] In this way, the second-order structure can be described. Interpretation of regression coefficients is from the subject-specific point of view. To estimate the parameters in the GLMM, the exact likelihood involves an intractable integration. Several approaches have been proposed to get around this, though the most commonly used being penalized quasi-likelihood (PQL) [5] and numerical integration via Gaussian Hermite quadrature (QUAD).[6]

With GLMMs, the association structure is also estimated, and has inference attached. On the other hand, GLMMs suffer from several shortcomings: (1) they are tricky to fit in comparison to marginal models fit via GEE; (2) estimates can be biased and it is unclear when QUAD or PQL is best; (3) only estimation via PQL can accommodate association structures more complex than compound symmetric with reasonable computing time; and (4) the interpretation of the parameters is subject-specific and changes depending on which groupings are accounted for. Moreover, the regression parameters in a GLMM are more sensitive to the random effects assumptions,[7] and even the notion of normally distributed random effects, which simplify computation, is controversial.[8]

Another option to modelling correlated binary data is alternating logistic regressions (ALRs), proposed by Carey et al.[9] The model is arrived at by alternating between estimating the mean structure and estimating the correlation structure.[9] Thus, for clusters $i = 1, 2, \ldots, m$ the binary outcome is $Y_i = (Y_{i1}, \ldots, Y_{in_i})$ with mean $E(Y_i) = \mu_i$; and the correlation structure is described in terms of the odds ratio (OR), ψ_{ijk}, such that the OR for observations in the same cluster is defined by Equation (1) [9]

$$\psi_{ijk} = \frac{\mathrm{P}(Y_{ij} = 1, Y_{ik} = 1) * \mathrm{P}(Y_{ij} = 0, Y_{ik} = 0)}{\mathrm{P}(Y_{ij} = 1, Y_{ik} = 0) * \mathrm{P}(Y_{ij} = 0, Y_{ik} = 1)}. \tag{1}$$

Defining the outcome of 1 as a success and 0 as a failure, in words, ψ_{ijk} represents the odds of success for subject j in cluster i given that subject k in cluster i was a success divided by the odds of success for subject j in cluster i given that subject k in cluster i was a failure. As with the symmetry between exposure and disease in the exposure OR and the disease OR commonly used for case–control studies and exposure based studies, respectively, the Y_{ij} and Y_{ik} are interchangeable here.

For ALRs, like marginal models fit via GEE, a full probability model is not specified for Y. Also similarly to marginal models fit via GEE, ALRs are computationally easy. However, in addition to inference on the mean structure, ALRs also give estimates of standard error and corresponding tests for the association among observations on the same cluster/subject, characterized by an OR.[9] Thus, in contrast to usual GEE in which the association structure is regarded as a nuisance parameter, denoted by a vector α, here it can be explored and estimated by combining the first-order

GEE for β given by Equation (2), and an offset logistic regression equation for α given by Equation (3)

$$U_1(\beta) = \sum_{i=1}^{m} \left(\frac{\partial \mu_i}{\partial \beta} \right)^{\mathrm{T}} V_i(\mu_i; \alpha)^{-1} (Y_i - \mu_i) = 0, \tag{2}$$

$$\operatorname{logit} \mathrm{P}(Y_{ij} = 1 \mid Y_{ik} = y_{ik}) = z_{ijk}^{\mathrm{T}} \alpha y_{ik} + \log \left(\frac{\mu_{ij} - \nu_{ijk}}{1 - \mu_{ij} - \mu_{ik} + \nu_{ijk}} \right). \tag{3}$$

Here, $V_i(\mu_i; \alpha)$ is a weighting square matrix of size n_i that approximates the covariance matrix of Y_i. Furthermore, z_{ijk} is a vector of pair specific covariates, α is the log OR between Y_{ij} and Y_{ik} estimated from the logistic regression model given by Equation (3) where the second term is used as an offset, $\mu_{ij} = \mathrm{P}(Y_{ij} = 1)$ and $\nu_{ijk} = \mathrm{P}(Y_{ij} = 1, Y_{ik} = 1)$. Thus, the ALR procedure iterates between (1) using Equation (2) to solve for $\hat{\beta}^{(r+1)}$ given current values $\hat{\beta}^{(r)}$ and $\hat{\alpha}^{(r)}$, and (2) using Equation (3) to solve for $\hat{\alpha}^{(r+1)}$ given $\hat{\beta}^{(r+1)}$ and $\hat{\alpha}^{(r)}$.

ALRs seem to offer some of the advantages of GLMMs (i.e. association structures that can be estimated with inference) and marginal models estimated via GEE (i.e. easy computation and interpretation). Moreover, with ALRs, the association structure can be quite complex. There is increasing interest in modelling the association parameters [10] and ALRs are increasingly used in the medical literature.[11–13] However, it seems relatively unknown how well ALRs or other approaches are able to model hierarchical data structures, especially when there are more than two levels. In this work, we aim to systematically evaluate how well ALRs perform for two- and three-level data and to compare it to marginal models estimated via GEE and GLMMs. In Sections 2 and 3 we describe simulation studies to that effect. In Section 4 we apply these methods to a real life data set looking at the association between prenatal care and community, mother, and birth-level covariates. We end with a discussion of our study in Section 5.

2. Methods

A simulation study was conducted to compare ALRs to other regression approaches to modelling correlated binary data.

2.1. *Data generation*

Two-level (e.g. subjects clustered into households) and three-level (e.g. subjects clustered into households, clustered into neighbourhoods) data sets were generated. We denote the number of clusters $i = 1, 2, \ldots, m$ and number of subjects per cluster $j = 1, 2, \ldots, n$ for the two-level data structure. For the three-level structure, we denote $i = 1, 2, \ldots, m$ clusters; $j = 1, 2, \ldots, n$ subclusters, and $k = 1, 2, \ldots, t$ subjects per subcluster.

The dichotomous independent variable, X, was generated from a Bernouilli distribution with $p = 0.5$. To generate the dichotomous outcome variable Y, first the probability of the outcome was generated from the logistic regression model given by Equation (4) for the two-level structure, and Equation (5) for the three-level structure

$$\operatorname{logit}(p) = \beta_0 + \beta_1 X + \mu_i, \tag{4}$$

$$\operatorname{logit}(p) = \beta_0 + \beta_1 X + \mu_i + \nu_{ij}, \tag{5}$$

where μ_i and ν_{ij} are random effects, corresponding to the cluster and subcluster, respectively, and generated from normal distributions with mean = 0 and variance = $\sigma^2_{\text{cluster}}$ or $\sigma^2_{\text{subcluster}}$, respectively. Then a dichotomous Y variable was generated from a Bernouilli distribution with given probability of the outcome, such that the overall prevalence of the outcome was either 0.1 or 0.5.

The total number of subjects, number of clusters and subclusters, number of subjects per cluster, variances of the random effects, and proportion of subjects with the outcome were all varied, with levels described in Table 1. For each combination of parameters, 250 and 500 data sets were generated for the two- and three-level structure, respectively.

2.2. *Data analysis*

Each data set was analysed to estimate the association between the continuous exposure X as the independent variable and the binary outcome Y as the dependent variable, using the following approaches:

(1) ALRs with one OR estimated to describe the clustering of subjects in the same cluster for the two-level data sets, and two ORs estimated to describe the clustering of subjects in the same cluster and/or subcluster for the three-level data sets.
(2) GEE using an exchangeable correlation structure (accounting for the top-level of clustering only for the three-level data sets).
(3) GLMMs estimated via PQL that included a random intercept for the two-level data sets, and two random intercepts, one for the cluster and another for the subcluster for the three-level data sets.
(4) For the two-level data sets, we also estimated GLMMs via adaptive QUAD that included a random intercept.

For the two-level data, to calculate the true value of the population average β_1, we first generated 4000 subject-specific random effects from a normal distribution with mean = 0 and variance = $\sigma^2_{\text{cluster}}$. Then using the true values of β_0, β_1 and $\sigma^2_{\text{cluster}}$, and the generated values, we calculated subject-specific probabilities. Finally, we took the average probabilities and from these, calculated the true population average regression coefficient and corresponding OR. We followed a similar procedure for the three-level data but generated the random effects from a normal distribution with variance equal to the sum of $\sigma^2_{\text{cluster}}$ and $\sigma^2_{\text{subcluster}}$.

2.3. *Measures of performance*

For both the two- and three-level data simulations, we estimated type I error and power for the tests on the association ORs from ALRs. Because how the association structure is modelled can affect estimation of the regression coefficients, we compared absolute percent bias in the OR for the effect of X, and mean squared error of the regression coefficient (β_1) across all approaches. For these calculations, for ALRs and marginal models estimated via GEE, the population average regression coefficient or OR were used. We also estimated type I error and power for testing the effect of X. For ALRs and marginal models estimated via GEE, the robust variance estimate was used for these calculations. We compared the mean, standard deviation (SD), and the coefficient of variation (CV) of the association parameters across all approaches. All simulations and analyses were performed using SAS/STAT software version 9.3,[14] and default procedure values were used unless otherwise noted.

Table 1. Parameters used for data generation.

	Values
Two-level data simulation	
σ_{cluster}	0, 0.7, 1, 3
Number of clusters, number of subjects per cluster	15, 10
	30, 5
	75, 2
	15, 50
	150, 5
	375, 2
Proportion of subjects with outcome	0.1, 0.5
Three-level data simulation	
$\sigma_{\text{cluster}}, \sigma_{\text{subcluster}}$	0, 0
	0, 1
	0, 2
	1, 0
	1, 1
	1, 2
	2, 0
	2, 1
	2, 2
Number of clusters, number of subclusters per cluster, number of subjects per subcluster	10, 5, 20
	10, 10, 10
	10, 20, 5
	20, 2, 25
	20, 5, 10
	20, 10, 5
	20, 25, 2
	25, 2, 20
	25, 4, 10
	25, 10, 4
	25, 20, 2
	25, 16, 25
	25, 20, 20
	25, 16, 25
	25, 20, 20
	25, 25, 16
	50, 5, 40
	50, 10, 20
	50, 20, 10
	50, 40, 5
	100, 5, 20
	100, 10, 10
	100, 20, 5

3. Results

3.1. *Two-level data*

For data with two levels, type I error for the test of the association OR in ALRs ranged from 0.03 to 0.16 depending on data generation parameters (Table 2). In general, among data sets with the same overall sample size, type I error decreased as the size of the clusters decreased, but did not vary systematically by proportion with the outcome (results not shown). Type I error was significantly inflated when fewer large clusters made up the data set.

For data with two levels, power for the test of the association OR in ALRs decreased as the number of clusters increased in data sets of the same overall sample size. Power increased as the

Table 2. Type 1 error[a] and 95% confidence interval for testing the association OR[b,c] from ALRs for data sets with two levels and proportion with the outcome 0.1.

Sample size	Number of clusters	Type I error	95% CI[d]
150	15	0.16	(0.11, 0.20)
150	30	0.03	(0.01, 0.05)
150	75	0.04	(0.01, 0.06)
750	15	0.15	(0.10, 0.19)
750	30	0.12	(0.08, 0.16)
750	75	0.08	(0.04, 0.11)
750	150	0.03	(0.01, 0.05)
750	375	0.05	(0.02, 0.08)

[a]For each combination of data generation parameters, 250 data sets were generated.
[b]Testing if the association OR equals 1.
[c]For data with normally distributed random effects with $\sigma_{\text{cluster}} = 0$.
[d]Exact 95% confidence interval.

variance of the random effect increased, and as overall sample size increased (Figure 1). Power was greater when the proportion with the outcome was 0.5 rather than 0.1 (results not shown).

For two-level data, when the true correlation between observations in the same cluster was 0, GEE and ALRs estimated intraclass correlation coefficients (ICC) and ORs that were close to 0 or 1 on average, respectively. GLMMs estimated small covariance parameters whether PQL or QUAD was used. For all approaches, the estimated association parameters increased as the true $\sigma^2_{\text{cluster}}$ increased. GLMMs estimated via PQL under-estimated the covariance parameter; this was worse as cluster sizes decreased. GLMMs estimated via QUAD estimated covariance parameters that were close to the true values, however the CV was higher than for PQL. The CV of the estimated association parameters increased as cluster size decreased for all methods, except for the highest $\sigma^2_{\text{cluster}}$ (Table 3). In general, the CV of the estimated correlation parameter was lowest for ALR than for other methods, and lower for PQL than for QUAD, except at the highest $\sigma^2_{\text{cluster}}$. Results were similar, though coefficients of variation were somewhat higher when 10% of subjects had the outcome (results not shown).

For two-level data, the mean percent bias in $\exp(\beta_1)$, and mean squared error in β_1 was similar when estimated via ALR or GEE, as expected (results not shown).

3.2. *Three-level data*

Type I error was always inflated for the top-level OR, ranging from 0.05 to 0.24. It was more inflated as overall observations increased or number of top-level clusters increased, but did not vary by the amount of correlation in subclusters. Type I error for $\text{OR}_{\text{subcluster}}$ was similarly inflated, especially when $\sigma^2_{\text{cluster}}$ was nonzero. In that case it was near 1. Power for testing if $\log(\text{OR})_{\text{cluster}} = 0$ (Figure 2(a)) varied from 0.054 to 1, and was lower as $\sigma_{\text{subcluster}}$ increased. Power for testing $\log(\text{OR})_{\text{subcluster}}$ (Figure 2(b)) was always near 1.

For three-level data, type 1 error for β_1 was also elevated when estimated using ALRs, though this decreased as total sample size increased, and as $\sigma^2_{\text{cluster}}$ and $\sigma^2_{\text{subcluster}}$ increased. When using GEEs accounting for the top level of correlation, type I error was furthest from the nominal level (0.05) compared to all other models, across all data structures, sample sizes and variance of random effects. We also found that in general, all models performed better when the number of top-level clusters increased (results not shown).

For three-level data, for all approaches, absolute percent bias in the OR ($\exp(\hat{\beta}_1)$) and mean square error (MSE) for $\hat{\beta}_1$, decreased as sample size increased. The bias and MSE among ALRs and GEEs produced very similar results (data not shown).

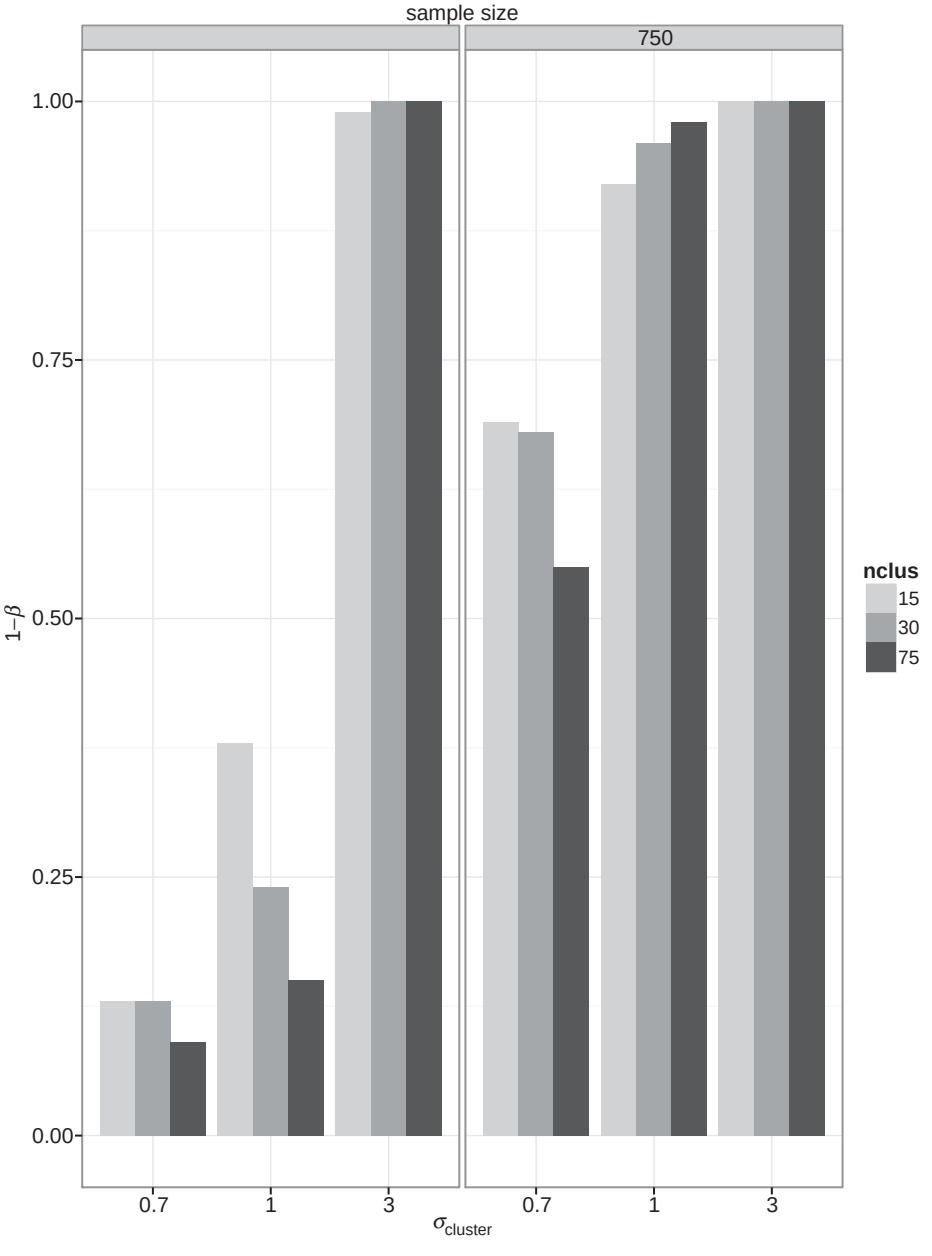

Figure 1. Power for the test of the association OR from ALRs by data set size, number of clusters and true σ_{cluster} for data sets with proportion with the outcome 0.1 for 250 iterations for two-level data.

When $\sigma^2_{\text{cluster}} = 0$ and $\sigma^2_{\text{subcluster}} = 0$, both estimated association parameters ($\hat{\sigma}^2_{\text{cluster}}$ and $\hat{\sigma}^2_{\text{subcluster}}$) were near 0 for GLMMs estimated via PQL as shown by the kernel density plots in Figure 3. $\sigma^2_{\text{cluster}}$ was under-estimated as $\sigma^2_{\text{subcluster}}$ increased (Figure 3(a)), whereas $\hat{\sigma}^2_{\text{subcluster}}$ was less influenced by $\sigma^2_{\text{cluster}}$ (Figure 3(b)).

ALRs had a harder time assigning the correct level of correlation as described by the log(ORs) between subjects in the same cluster and subcluster (Figure 4). σ_{cluster} was under-estimated as $\sigma_{\text{subcluster}}$ increased. This is most evident in the top panel of Figure 4(a) where the true σ_{cluster} is

Table 3. Comparing how different approaches estimate[a] the association structure in two-level data with $\beta_1 = 0.69$, and proportion with the outcome 0.5.

$\sigma^2_{\text{cluster}}$	Sample size	n cluster	ALR[b] Mean (SD)	CV[c]	GEE[d] Mean (SD)	CV	GLMM[e] via PQL Mean (SD)	CV	GLMM[f] via QUAD Mean (SD)	CV
0	150	15	0.99(0.17)	0.17	−0.01(0.04)	−7.05	0.07(0.13)	1.77	0.06(0.12)	2.03
0	150	30	0.99(0.28)	0.28	−0.01(0.06)	−5.83	0.09(0.16)	1.67	0.10(0.19)	1.9
0	150	75	1.14(0.54)	0.48	0.01(0.11)	18.83	0.13(0.17)	1.25	0.27(0.41)	1.52
0	750	15	1.00(0.03)	0.03	0.00(0.01)	−774.6	0.02(0.03)	1.6	0.01(0.02)	1.79
0	750	75	1.00(0.08)	0.08	0.00(0.02)	−105.88	0.03(0.05)	1.5	0.03(0.05)	1.58
0	750	375	1.00(0.21)	0.21	−0.01(0.05)	−10.15	0.04(0.06)	1.49	0.08(0.14)	1.65
0.49	150	15	1.51(0.46)	0.31	0.09(0.06)	0.74	0.48(0.36)	0.77	0.49(0.43)	0.86
0.49	150	30	1.51(0.53)	0.35	0.09(0.07)	0.85	0.40(0.33)	0.83	0.51(0.50)	0.98
0.49	150	75	1.73(0.93)	0.54	0.10(0.12)	1.14	0.30(0.26)	0.85	0.74(0.82)	1.11
0.49	750	15	1.48(0.25)	0.17	0.09(0.04)	0.41	0.48(0.23)	0.47	0.46(0.22)	0.48
0.49	750	75	1.49(0.18)	0.12	0.09(0.03)	0.3	0.41(0.14)	0.34	0.48(0.17)	0.36
0.49	750	375	1.54(0.36)	0.23	0.10(0.05)	0.55	0.23(0.12)	0.55	0.53(0.34)	0.65
1	150	15	1.98(0.73)	0.37	0.15(0.08)	0.52	0.85(0.52)	0.62	0.94(0.65)	0.69
1	150	30	2.03(0.73)	0.36	0.15(0.08)	0.53	0.72(0.42)	0.58	0.99(0.67)	0.68
1	150	75	2.47(1.49)	0.6	0.18(0.13)	0.71	0.49(0.32)	0.66	1.38(1.29)	0.93
1	750	15	1.99(0.41)	0.21	0.16(0.05)	0.3	0.97(0.37)	0.38	0.95(0.37)	0.39
1	750	75	2.06(0.30)	0.15	0.17(0.04)	0.2	0.85(0.22)	0.26	1.03(0.29)	0.28
1	750	375	2.07(0.44)	0.22	0.17(0.05)	0.3	0.40(0.13)	0.32	1.04(0.42)	0.41
9	150	15	14.34(12.00)	0.84	0.51(0.11)	0.22	5.33(2.13)	0.4	10.05(8.36)	0.83
9	150	30	13.96(8.95)	0.64	0.53(0.09)	0.16	3.87(1.08)	0.28	9.88(5.01)	0.51
9	150	75	14.37(9.68)	0.67	0.54(0.10)	0.19	1.65(0.42)	0.25	9.61(5.16)	0.54
9	750	15	14.20(11.57)	0.81	0.53(0.10)	0.18	7.63(2.77)	0.36	9.19(4.28)	0.47
9	750	75	12.19(3.79)	0.31	0.54(0.05)	0.1	4.99(0.90)	0.18	9.18(2.50)	0.27
9	750	375	12.03(2.71)	0.23	0.54(0.04)	0.07	1.58(0.16)	0.1	8.84(1.72)	0.19

Notes: Results are presented for the mean (SD) and CV for association parameters estimated via ALRs, GEE, GLMMs via PQL and GLMMs via QUAD.
[a]For each combination of data generation parameters, 250 data sets were generated.
[b]Mean (SD) and CV for the OR for the association between two observations in the same cluster estimated via ALRs.
[c]CV. In some cases the CV does not equal the SD/mean as presented in the table due to rounding of the SD and mean.
[d]Mean (SD) and CV for the ICC estimated via GEE.
[e]Mean (SD) and CV for the variance of the random effect from a GLMM estimated via PQL.
[f]Mean (SD) and CV for the variance of the random effect from a GLMM estimated via adaptive QUAD.

2, however the majority of its estimated values are around 1 (for true $\sigma_{\text{subcluster}} = 2$). There is no such pattern shown for estimating $\sigma_{\text{subcluster}}$ (Figure 4(b)) as it was overestimated for large and under-estimated for small values of σ_{cluster}.

4. Example

We now consider an example using the data collected from the National Survey of Maternal and Child Health in Guatemala in 1987.[15] Rodriguez and Goldman [16] performed an analysis based on 2449 births, to 1558 mothers in 161 communities. The outcome of interest was whether or not mothers received prenatal care and covariates included community, mother, and birth-level covariates. Here we model the outcome using GEE accounting for the top level of clustering and ALRs and PQL both accounting for both levels of clustering, with the covariates found to be significant in Rodriguez and Goldman [16] (i.e. child and mother's age, ethnicity, mother and husband's education, husband's occupation, presence of a modern toilet, proportion indigenous and distance to the nearest clinic). The results are presented in Table 4. GLMM coefficients estimated via PQL were further from the null than the population average coefficients from ALR or GEE, as expected. Both GLMM and ALR attributed more correlation to babies born to the

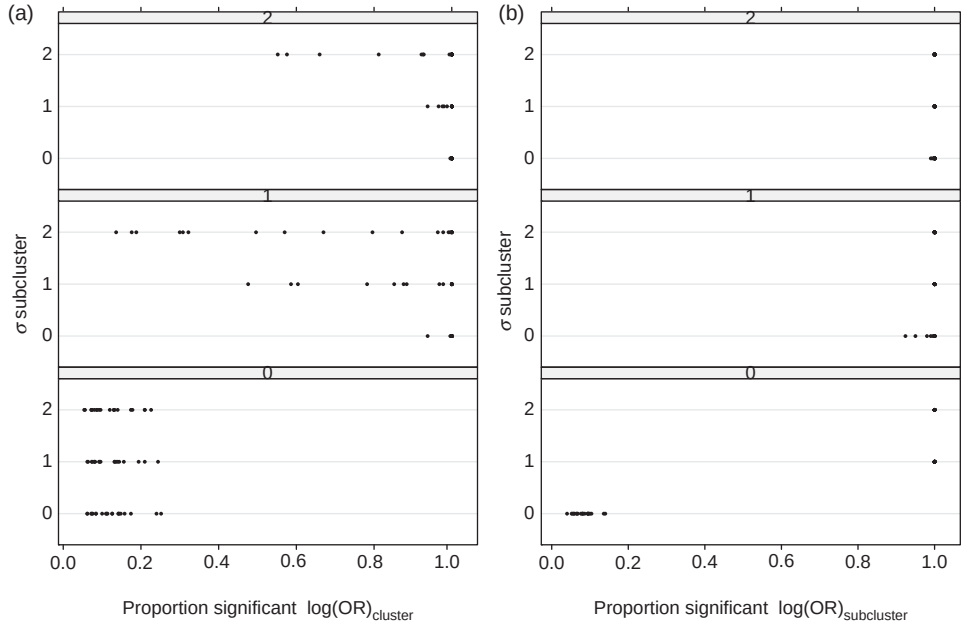

Figure 2. (a and b) Dot plot by σ_{cluster} of type I error and power for test of the association log(OR) from ALRs in three-level data, $\beta_1 = \log(2)$. (a) $\log(\text{OR})_{\text{cluster}}$ and (b) $\log(\text{OR})_{\text{subcluster}}$ are the log(OR) describing the association between subjects in the same cluster and subcluster, respectively.

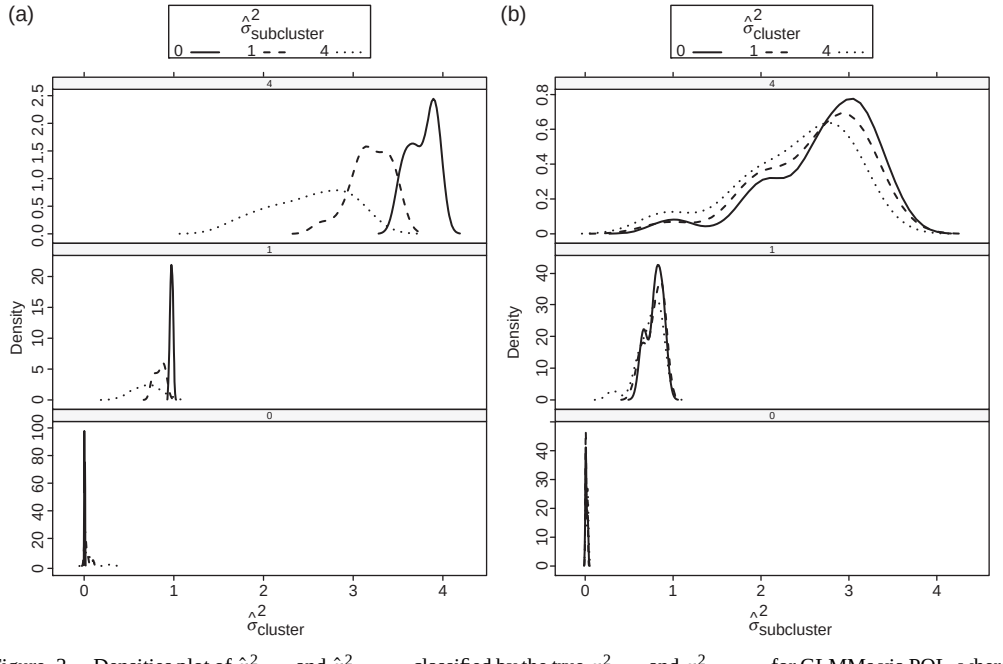

Figure 3. Densities plot of $\hat{\sigma}^2_{\text{cluster}}$ and $\hat{\sigma}^2_{\text{subcluster}}$ classified by the true $\sigma^2_{\text{cluster}}$ and $\sigma^2_{\text{subcluster}}$, for GLMMs via PQL, where $\hat{\sigma}^2_{\text{cluster}}$ is the estimated variance of $\sigma^2_{\text{cluster}}$ and $\hat{\sigma}^2_{\text{subcluster}}$ is the estimated variance of $\sigma^2_{\text{subcluster}}$. (a) Density plot of PQL estimated variance by $\sigma^2_{\text{cluster}}$ and (b) density plot of PQL estimated variance by $\sigma^2_{\text{subcluster}}$.

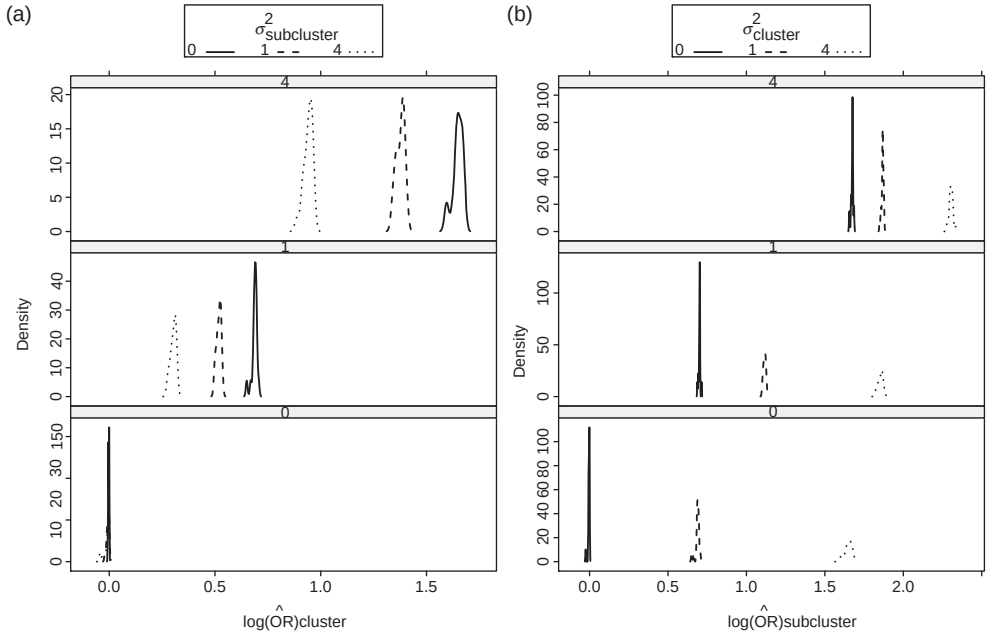

Figure 4. Densities plot of $\log(\widehat{\text{OR}})_{\text{cluster}}$ and $\log(\widehat{\text{OR}})_{\text{subcluster}}$ for ALR classified by the true $\sigma^2_{\text{cluster}}$ and $\sigma^2_{\text{subcluster}}$, where $\log(\widehat{\text{OR}})_{\text{cluster}}$ describes the correlation between subjects in the same cluster and $\log(\widehat{\text{OR}})_{\text{subcluster}}$ describes the correlation between subjects in the same subcluster. (a) Density plot of ALR $\log(\widehat{\text{OR}})_{\text{cluster}}$ by $\sigma^2_{\text{cluster}}$ and (b) density plot of ALR $\log(\widehat{\text{OR}})_{\text{subcluster}}$ by $\sigma^2_{\text{subcluster}}$.

same mother than to babies born in the same community, as expected. Results from the simulation suggests that the PQL estimated $\hat{\sigma}_{\text{community}}$ and $\hat{\sigma}_{\text{mother}}$ are likely to be under-estimated. The log(ORs) estimated via ALR were also likely to be under-estimated since the association between two observations in the same family was much stronger than the community level association.

5. Discussion

In this paper, we compared ALRs to several other approaches to modelling correlated binary data, with particular emphasis on estimating the correlation structure of the data. The correlation structure might give some insight into disease transmission, aetiology, risk factors and can inform the targeting of interventions.[1] Methods for continuous data do not extend naturally to binary data. For example, when the outcome is continuous, the association among observations on the same cluster/subject is well-characterized via the ICC.[4] However, when the outcome is binary, the range of ICC values depends on the probabilities of positive responses.[17] Modelling the association between subjects in the same cluster using the OR, as is done in ALRs, may be a more natural characterization.[4,9] Additionally, hypothesis testing of the variance parameters in GLMMs (and indeed linear mixed models) is problematic because variance parameters are usually constrained to be positive. ALRs also avoid this problem.

We found that ALRs behaved similarly to marginal models estimated via GEE for mean effects. We also found that type I error and power for tests of the association parameters in ALRs were reasonable for one-level structures, but type I error was inflated and power was lower for more complex data structures.

In general, results from the simulation study on three-level data suggested that it was difficult to ascribe clustering to the correct level, particularly for ALRs. Subcluster log(ORs) increased

Table 4. Parameter estimates of prenatal care among women for GEE top level, ALR and PQL methods.

	GEE TOP	ALR	PQL
Fixed effects			
Individual			
Child aged 3–4 years	−0.14 (−0.29, 0.01)	−0.18 (−0.3, −0.06)	−0.21 (−0.44, 0.02)
Mother aged greater 25 years	0.18 (−0.05, 0.41)	0.18 (0.01, 0.36)	0.22 (−0.04, 0.47)
Family			
Indigenous, no Spanish	−1.23 (−1.89, −0.57)	−1.18 (−1.81, −0.56)	−1.36 (−2.05, −0.67)
Indigenous Spanish	−0.57 (−0.95, −0.18)	−0.63 (−0.99, −0.27)	−0.70 (−1.2, −0.2)
Mother's education primary	0.42 (0.16, 0.68)	0.42 (0.16, 0.68)	0.49 (0.19, 0.78)
Mother's education secondary or better	0.96 (0.37, 1.54)	0.97 (0.38, 1.55)	1.16 (0.4, 1.91)
Husband's education secondary or better	0.57 (0.06, 1.07)	0.70 (0.22, 1.17)	0.78 (0.21, 1.34)
Husband agricultural employee	−0.21 (−0.52, 0.09)	−0.32 (−0.61, −0.03)	−0.32 (−0.64, 0)
Modern toilet in household	0.39 (0.01, 0.78)	0.48 (0.1, 0.86)	0.58 (0.15, 1.02)
Community			
Proportion indigenous, 1981	−1.05 (−1.67, −0.43)	−0.88 (−1.46, −0.3)	−1.17 (−1.91, −0.43)
Distance to nearest clinic	−0.01 (−0.02, −0.01)	−0.01 (−0.02, 0.00)	−0.01 (−0.02, −0.01)
Association structure			
Family	–	$\alpha_2 = 4.61\ (4.07, 5.15)$[a]	$\sigma_1^2 = 1.61\ (0.19)$[b]
Community	ICC = 0.17[c]	$\alpha_1 = 0.59\ (0.38, 0.80)$[d]	$\sigma_2^2 = 0.80\ (0.18)$[b]

[a]Log(OR) and 95% CI for the association between two observations in the same family estimated via ALRs.
[b]Variance of the random effect and standard error from a GLMM estimated via PQL.
[c]ICC estimated via GEE accounting for the top level of clustering, i.e. families within communities.
[d]Log(OR) and 95% CI for the association between two observations in the same community estimated via ALRs.

as cluster level correlation increased, whereas cluster level log(ORs) decreased as subcluster correlation increased. GLMMs estimated via PQL performed somewhat better though variance components were under-estimated.

While there is a large literature on the performance of GLMMs (estimated via PQL or numerical integration), this work is usually directed at the mean response parameters or focuses on simple correlation structures. For example, Moineddin et al. [18] found that with two levels, the variance components were extremely overestimated with small groups, and were slightly under-estimated with moderate group size for GLMMs (estimated using QUAD). On the other hand, for two-level GLMMs, PQL under-estimated the variance components when the denominator was small,[19, 20] which we also observed. Other research has focused on bias in the regression parameters, depending on cluster size, number of clusters, and heterogeneity across clusters.[8,9,21] In other work, a three-level GLMM was developed and results of a limited simulation study suggested that the association parameters were reasonably well estimated.[22] However, in at least one application using a three-level GLMM, the estimated association parameters were quite different depending on whether estimation was by QUAD or PQL.[23]

One of the challenges inherent in comparing these approaches is that the models all have different assumptions and the resulting parameters often have different interpretations. Despite this, we believe that if the correlation structure is of scientific interest, it is helpful to know the performance of the various approaches in terms of estimating it. Of note, while in theory it is possible to estimate a nested random effects model via adaptive QUAD, it is often difficult, at least on run-of-the-mill computers, to complete unless the design matrix is suitably small. In our investigations, GLMMs estimated via adaptive QUAD for two levels of clustering did not converge. This may depend on the software package used or features of the generated data.

We compared the performance of the approaches via simulation study. The models and their associated estimation techniques are complicated, and thus the theoretical derivation of the properties of the estimated model parameters is challenging. Indeed, simulation studies are necessary here and offer an important tool.[24] However, we have made many simplifications. We only included

one dichotomous independent variable, and only considered one- and two-level exchangeable correlation structures. For the three-level data structure simulations, we were limited to small sample sizes. Still, we believe that understanding performance in the simplest case can still provide insight to more complicated scenarios. Another weakness of our work is that data were generated from a random effects model. Thus, it is difficult to assess whether ALRs estimated the association structure correctly when σ_{cluster} or $\sigma_{\text{subcluster}}$ were greater than 0. Despite this, our investigations have allowed us to assess performance.

We have omitted two alternative approaches that might also be of interest. Second-order GEEs (GEE2s) extend the theory of GEEs to estimate the second-moment parameters, which allows inference about the regression parameters and the association parameters.[24–26] The association structure can be characterized via ORs,[26] as in ALRs. For this reason, and because interpretation is from the population average perspective, this method is closely related method to ALRs, but is computationally more tricky.[2] Moreover, GEE2s assume that the higher order moments are equal to 0, while ALRs make no assumptions about higher order moments.[2] Moreover, GEE2s are not available in most software packages. We also did not compare GLMMs estimated using Bayesian approaches, which might be particularly useful when the number of clusters is small or the model is complicated.[27,28]

Certainly, when choosing from among possible methods to model correlated binary data, there are several considerations including: whether the method is implemented in commercial software; the interpretation of the parameters; whether the second-order structure is of interest scientifically; if the method can incorporate the correlation structure of interest, and finally, the statistical properties of the model for both the mean structure and the second-order structure. Our results suggest that ALRs do offer a middle road between GLMMs and marginal models estimated via GEE, but also that caution is needed, no matter the approach used, when modelling three-level data structures.

References

[1] Katz J, Carey VJ, Zeger SL, Sommer A. Estimation of design effects and diarrhea clustering within households and villages. Am J Epidemiol. 138;1993:994–1006.
[2] Molenberghs G, Verbeke G. Models for discrete longitudinal data. Springer series in statistics. New York: Springer; 2005.
[3] Aerts M, Molenberghs G, Ryan LM, Geys H. Topics in modelling of clustered data. Boca Raton: CRC Press; 2002.
[4] Diggle P, Heagerty P, Liang K-Y, Zeger SL. Analysis of longitudinal data. 2nd ed. Oxford statistical science series, 25. Oxford: Oxford University Press; 2002.
[5] Breslow N, Clayton D. Approximate inference in generalized linear mixed models. J Am Stat Assoc. 88;1993:9–25.
[6] Pinheiro J, Bates D. Approximations to the log-likelihood function in the nonlinear mixed-effects model. J Comput Graph Stat. 4;1995:12–35.
[7] Heagerty PJ, Zeger SL. Marginalized multilevel models and likelihood inference. Stat Sci. 15;2000:1–19.
[8] Carlin J, Wolfe R. A case study on the choice, interpretation and checking of multilevel models for longitudinal binary outcomes. Biostatistics. 2;2001:397–416.
[9] Carey V, Zeger SL, Diggle P. Modeling multivariate binary data with alternating logistic regressions. Biometrika. 80;1993:517–526.
[10] Yi GY, He W, Liang H. Analysis of correlated binary data under partially linear single-index logistic models. J Multivariate Anal. 100;2009:278–290.
[11] Zhan Q, Cruickshanks KJ, Klein BE, Klein R, Huang GH, Pankow JS, Gangnon RE, Tweed TS. Generational differences in the prevalence of hearing impairment in older adults. Am J Epidemiol. 171;2010:260–266.
[12] Riddle DL, Kong X, Jiranek WA. Factors associated with rapid progression to knee arthroplasty: complete analysis of three-year data from the osteoarthritis initiative. Joint Bone Spine. 79;2012:298–303.
[13] Ambalavanan N, Walsh M, Bobashev G, Das A, Levine B, Carlo WA, Higgins RD. Intercentre differences in bronchopulmonary dysplasia or death among very low birth weight infants. Pediatrics. 127;2011:e106–e116.
[14] SAS/STAT. SAS Institute Inc. 9.3 for windows. Cary, NC: SAS/STAT; 2011.
[15] Pebley AR, Goldman N, Rodriguez G. Prenatal and delivery care and childhood immunization in guatemala: do family and community matter? Demography. 33;1996:231–247.
[16] Rodriguez G, Goldman N. Improved estimation procedures for multilevel models with binary response: a case-study. J R Stat Soc Ser A. 164;2001:339–355.

[17] Prentice RL. Correlated binary regression with covariates specific to each binary observation. Biometrics. 44;1988:1033–1048.
[18] Moineddin R, Matheson FI, Glazier RH. A simulation study of sample size for multilevel logistic regression models. BMC Med Res Methodol. 7;2007:34.
[19] Bates D, Sarkar D. Lme4: linear mixed effects models using s4 classes. R help; 2007.
[20] Breslow N, Lin X. Bias correction in generalized linear mixed models with a single component of dispersion. Biometrika. 82;1995;91.
[21] Breslow N. Whither PQL? Proceedings of the second seattle symposium in biostatistics. Lecture notes in statistics, 179, New York: Springer; 2003. p. 1–22.
[22] Wang K, Lee AH, Hamilton G, Yau KK. Multilevel logistic regression modelling with correlated random effects: application to the smoking cessation for youth study. Stat Med. 25;2006:3864–3876.
[23] Olsen MK, DeLong ER, Oddone EZ, Bosworth HB. Strategies for analyzing multilevel cluster-randomized studies with binary outcomes collected at varying intervals of time. Stat Med. 27;2008:6055–6071.
[24] Fitzmaurice GM, Lipsitz SR. A model for binary time series data with serial odds ratio patterns. Appl Stat. 44;1995:51–61.
[25] Prentice RL, Zhao LP. Estimating equations for parameters in means and covariances of multivariate discrete and continuous responses. Biometrics. 47;1991:825–839.
[26] Qaqish B, Liang K. Marginal models for correlated binary responses with multiple classes and multiple levels of nesting. Biometrics. 48;1992:939–950.
[27] Gelman A, Hill J. Data analysis using regression and multilevel/hierarchical models. Cambridge: Cambridge University Press; 2006.
[28] Localio AR, Berlin JA, Have TR. Longitudinal and repeated cross-sectional cluster-randomization designs using mixed effects regression for binary outcomes: bias and coverage of frequentist and Bayesian methods. Stat Med. 25;2006:2720–2736.