Should we use logistic mixed model analysis for the effect estimation in a longitudinal RCT with a dichotomous outcome variable?

Twisk, Jos W.R.; de Vente, Wieke; Apeldoorn, Adri T.; de Boer, Michiel R.

Published in:
Epidemiology Biostatistics and Public Health

DOI:
10.2427/12613

Should we use logistic mixed model analysis for the effect estimation in a longitudinal RCT with a dichotomous outcome variable?

Jos WR Twisk [1,2], Wieke de Vente [3], Adri T Apeldoorn [1,4], Michiel R de Boer [2]

(1) Department of Epidemiology and Biostatistics, VU University medical centre, Amsterdam, The Netherlands.
(2) Section Methodology and Applied Biostatistics, Department of Health Sciences, VU University, Amsterdam, The Netherlands.
(3) Department of Educational Sciences, University of Amsterdam, Amsterdam, The Netherlands.
(4) Rehabilitation Department, Medical Centre Alkmaar, Alkmaar, The Netherlands

CORRESPONDING AUTHOR: *Prof. dr JWR Twisk. - Department of Epidemiology and Biostatistics - VU university medical centre de Boelelaan 1118 - 1081 HV Amsterdam - e-mail: jwr.twisk@vumc.nl - tel: +31 (0)20 4444459 - fax: +31 (0)20 4444475*

DOI: ***10.2427/12613***
Accepted on July 6,2017

ABSTRACT

Background: Within epidemiology both mixed model analysis and GEE analysis are frequently used to analyse longitudinal RCT data. With a continuous outcome, both methods lead to more or less the same results, but with a dichotomous outcome the results are totally different. The purpose of the present study is to evaluate the performance of a logistic mixed model analysis and a logistic GEE analysis and to give an advice which of the two methods should be used.
Methods: Two real life RCT datasets with and without missing data were used to perform this evaluation. Regarding the logistic mixed model analysis also two different estimation procedures were compared to each other.
Results: The regression coefficients obtained from the two logistic mixed model analyses were different from each other, but were always higher then the regression coefficients derived from a logistic GEE analysis. Because this also holds for the standard errors, the corresponding p-values were more or less the same. It was further shown that the effect estimates derived from a logistic mixed model analysis were an overestimation of the 'real' effect estimates.
Conclusion: Although logistic mixed model analysis is widely used for the analysis of longitudinal RCT data, this article shows that logistic mixed model analysis should not be used when one is interested in the magnitude of the regression coefficients (i.e. effect estimates).

Key words: Epidemiological methods; Longitudinal studies; Logistic mixed model analysis, Logistic GEE analysis, Randomised controlled trial.

INTRODUCTION

Within epidemiology, the two most frequently used methods to analyse longitudinal data from a ramdomised controlled trial (RCT) are Generalised Estimating Equations (GEE analysis) and mixed model analysis. The latter is also known as multilevel analysis, random coefficient analysis or hierarchical linear modeling. The general idea of both

methods is that an adjustment is made for the dependency of the observations within an individual over time. In GEE analysis this adjustment is performed by modeling the within subject correlation matrix [1,2], while in mixed model analysis, this adjustment is performed by modeling the difference between the subjects (i.e. the between subject variance) [3,4]. Because the correlation within the subject is essentially the same as the difference between the subjects, the estimated regression coefficients may be expected to be the same in both methods. However, there is also another difference between the two methods. GEE analysis is known as a 'population average' approach, while mixed model analysis is known as a 'subject specific' approach [5]. This does not influence the values of the estimated regression coefficients obtained from a *linear* GEE analysis and a linear mixed mode analysis, but it does influence the values of the estimated regression coefficients obtained from a *logistic* GEE analysis and a *logistic* mixed model analysis. The difference in regression coefficients is a theoretical one, which is always in favor of a mixed model analysis, meaning that the regression coefficients obtained from a logistic mixed model analysis will always be higher (i.e. further away from zero) compared to the regression coefficients obtained from a logistic GEE analysis. This difference is based on a mathematical relationship and depends on the magnitude of the between subject variance (see equation 1) [6,7]. When there is more between subject variance, the difference between the regression coefficients will be larger.

$$\beta^{(pa)} = \left[\left(\frac{16\sqrt{3}}{15\pi}\right)^2 \sigma_b^2 + 1\right]^{-1/2} \beta^{(ss)} \qquad (1a)$$

$$\frac{16\sqrt{3}}{15\pi} = 0.588 \qquad (1b)$$

Where $\beta^{(pa)}$ is population average regression coefficient obtained from a logistic GEE analysis, σ_b^2 is between subject variance and $\beta^{(ss)}$ is subject specific regression coefficient obtained from a logistic mixed model analysis.

Both GEE analysis and mixed model analysis are used for the analysis of longitudinal data with a dichotomous outcome variable, but from the literature it is not clear which of the two methods should be used and which regression coefficients should be reported [7-10]. In general, it is sometimes argued that mixed model analysis should be preferred above GEE analysis because mixed model analysis is more suitable to deal with missing data [11-13].

In this paper we will illustrate the differences between the regression coefficients obtained from a longitudinal logistic GEE analysis and a longitudinal logistic mixed model analysis by using examples from two RCTs with and without missing data. The aim of the study was to evaluate the performance of both methods and to provide an advice on which of the two methods should be used and which of the results should be reported.

METHODS

Datasets

The differences between results from a logistic GEE analysis and a logistic mixed model analysis are illustrated in datasets from two RCT's. The first example dataset is derived from an RCT aimed to assess the effectiveness of a classification based treatment approach compared to usual physical therapy care in patients with subacute or chronic low back pain [14]. The outcome variable of interest was functional status, which was measured with the 10-item Oswestry Disability Index (ODI) [15], with higher scores indicating lower functional status. The maximum score on the ODI is 50 and in the present study a cut off value of 12 was used to distinguish between good (< 12) or bad (≥ 12) functional status [16]. The outcome variable was assessed at 8, 26, and 52 weeks after the start of treatment.

The second example dataset is derived from a 3-arm RCT regarding an internet-based treatment for adults with depressive symptoms [17]. Besides a waiting list (WL) group, two interventions were evaluated, i.e. an internet-based cognitive behavioral therapy (CBT) and an internet-based problem solving therapy (PST). As outcome variable self reported depression (measured with the Center for Epidemiological Studies Depression scale (CES-D)) was measured at 5, 8 and 12 weeks. The CES-D is widely used for identifying individuals with depression and a score of 16 or higher is considered to represent clinical depression.

The two datasets differ from each other in the number of groups to be compared and in the percentage of missing data (see table 1). Both studies were approved by the Medical Ethics Committee of the VU University Medical Center in Amsterdam.

Analysis

For both example datasets a logistic GEE analysis and a logistic mixed model analysis were performed. For all logistic GEE analyses, an exchangeable correlation structure was used and for all logistic mixed model analyses only a random intercept was modeled. Regarding the logistic mixed model analyses, two estimation procedures were used; a maximum likelihood procedure performed with the xtmelogit procedure in STATA [18] and a (2nd order) penalized quasi likelihood procedure performed with MLwiN [19,20].

For both datasets, the differences between the groups at the different time points were estimated simultaneously, by treating time as a categorical variable represented by dummy variables and by adding interactions between the group variable(s) and the time dummy variables to the model.

TABLE 1. Number of subjects measured at the different time-points in the two example datasets

First example dataset	**Control**	**Intervention**	
baseline	82	74	
week 8	71	68	
week 26	73	64	
week 52	71	67	
complete cases	64	62	
Second example dataset	**WL**	**CBT**	**PST**
baseline	87	88	88
week 5	71	61	52
week 8	71	51	51
week 12	63	46	42
complete cases	58	41	35

WL = waiting list, CBT = cognitive behavioral therapy, PST = problem solving therapy

TABLE 2. Regression coefficients and standard errors (between brackets) of different longitudinal logistic regression analyses regarding the low back pain intervention

		GEE	**Mixed models**	
			PQL	**ML**
week 8	intervention	-0.29 (0.34)	-0.44 (0.52)	-0.51 (0.60)
week 26	intervention	-0.04 (0.35)	-0.05 (0.54)	-0.03 (0.62)
week 52	intervention	-0.51 (0.35)	-0.76 (0.54)	-0.86 (0.62)
variance[1]			3.30	5.15

[1]*between subject variance obtained from the mixed model analyses*

To illustrate the influence of missing data on the results of the logistic GEE analysis and the logistic mixed model analysis, in both datasets, one analysis was performed on the total dataset including missing values, and one analysis was performed on a dataset with only complete cases. To evaluate the performance of the different methods, the estimated probabilities of the outcome variable were compared to the observed percentages at the different time points.

RESULTS

First example dataset

Table 2 shows the results of the logistic GEE analysis and the two logistic mixed model analyses performed on the first example dataset regarding the physical therapy intervention on patients with low back pain. As expected the regression coefficients obtained from the logistic GEE analysis were much lower than the ones obtained from the logistic mixed model analyses. The magnitude of the difference between the methods was more or less expected given the estimated between subject variance and the mathematical relationship shown in equation 1. Note that also the results obtained from the two logistic mixed model analyses were quite different.

To evaluate the performance of the different methods, the predicted probabilities were compared to the observed percentage of good functional status (table 3). It can be seen that most of the predicted probabilities were different from the observed percentages. However, the predicted probabilities based on the results of the logistic GEE analysis were much closer to the observed percentages compared to the predicted probabilities based on the results of the logistic mixed model analyses.

When only the complete cases were analysed (tables 4 and 5) the difference in regression coefficients between the methods was comparable to the differences observed in the analyses regarding the total dataset (i.e. including cases with missing observations). However, in the complete data the predicted probabilities obtained from the logistic GEE analysis were exactly the same as the observed percentages, while the predicted probabilities obtained from the logistic mixed model analyses were (again) too high for probabilities above 50% or too low for probabilities below 50%.

TABLE 3. Observed percentages of good functional status and predicted probabilities derived from different longitudinal logistic regression analyses regarding the low back pain intervention

		Observed	GEE	Mixed models	
				PQL	ML
week 8	usual care (n=71)	56.3	56.0	60.1	61.0
	intervention (n=68)	48.5	48.8	49.3	48.5
week 26	usual care (n=73)	39.7	38.2	32.6	29.6
	intervention (n=64)	35.9	37.3	31.4	29.0
week 52	usual care (n=71)	43.8	43.8	40.7	38.9
	intervention (n=67)	31.3	32.0	24.3	21.3

TABLE 4. Regression coefficients and standard errors (between brackets) of different longitudinal logistic regression analyses regarding the low back pain intervention from a compete case analysis

		GEE	Mixed models	
			PQL	ML
week 8	intervention	-0.44 (0.36)	-0.68 (0.57)	-0.80 (0.66)
week 26	intervention	-0.15 (0.37)	-0.19 (0.59)	-0.22 (0.67)
week 52	intervention	-0.57 (0.37)	-0.86 (0.59)	-0.97 (0.68)
variance[1]			3.63	5.62

[1]*between subject variance obtained from the mixed model analyses*

TABLE 5. Observed percentages of good functional status and predicted probabilities derived from different longitudinal logistic regression analyses regarding the low back pain intervention from a complete case analysis

		Observed	GEE	Mixed models	
				PQL	ML
week 8	usual care (n=64)	57.8	57.8	63.1	64.3
	intervention (n=62)	46.8	46.8	46.3	44.8
week 26	usual care (n=64)	39.1	39.1	33.0	30.3
	intervention (n=62)	35.5	35.5	28.9	25.9
week 52	usual care (n=64)	43.8	43.8	40.4	38.4
	intervention (n=62)	30.6	30.6	22.4	19.2

Second example dataset

Table 6 shows the results of both the logistic GEE analysis and the (two) logistic mixed model analyses performed on the second example dataset, i.e. the 3-arm RCT regarding the internet based treatment of depressive symptoms. Table 7 shows the corresponding observed percentages of depressed subjects and the predicted probabilities.

The differences between the results obtained from the different methods were comparable to the ones observed in the first example dataset, i.e. the regression coefficients obtained from the logistic mixed model analyses were much higher (i.e. further away from zero) compared to the regression coefficients obtained from the logistic GEE analysis. Again, the predicted probabilities obtained from the logistic GEE analysis were much closer to the observed percentages than the observed probabilities obtained from the logistic mixed model analyses.

The results of the analyses on a complete dataset (tables 8 and 9) also show the same picture as for the first example dataset. The predicted probabilities from the logistic GEE analysis were exactly the same as the observed percentages, while the predicted probabilities derived from the logistic mixed model analyses were (mostly) too high.

DISCUSSION

In this paper we compared the performance of a logistic GEE analysis with the performance of logistic mixed model analysis applied on two longitudinal RCT datasets. Based on the results (i.e. the comparison

TABLE 6. Regression coefficients and standard errors (between brackets) obtained from different logistic longitudinal data analyses performed on the 3-arm RCT regarding the internet based treatment of depressive symptoms

		GEE	Mixed models	
			PQL	ML
week 5	CBT	-0.48 (0.46)	-0.74 (0.73)	-0.87 (0.76)
	PST	-1.35 (0.44)	-2.07 (0.72)	-2.37 (0.77)
week 8	CBT	-1.02 (0.40)	-1.60 (0.66)	-1.87 (0.73)
	PST	-0.80 (0.40)	-1.22 (0.68)	-1.43 (0.73)
week 12	CBT	-1.20 (0.42)	-1.90 (0.72)	-2.19 (0.78)
	PST	-1.08 (0.44)	-1.67 (0.74)	-1.89 (0.78)
variance[1]			4.38	5.84

CBT = cognitive behavioral therapy, PST = problem solving therapy
[1]between subject variance obtained from the mixed model analyses

TABLE 7. Observed percentages of depressed subjects and predicted probabilities obtained from different logistic longitudinal data analyses performed on the 3-arm RCT regarding the internet based treatment of depressive symptoms

		Observed	GEE	Mixed models	
				PQL	ML
week 5	WL (n=71)	84.5	85.2	93.8	95.9
	CBT (n=61)	77.0	78.1	87.9	90.8
	PST (n=52)	57.7	59.9	65.7	68.7
week 8	WL (n=71)	76.1	76.8	86.4	90.1
	CBT (n=51)	54.9	54.2	56.3	58.4
	PST (n=51)	58.8	59.8	65.3	68.4
week 12	WL (n=63)	82.5	80.4	90.1	92.9
	CBT (n=46)	56.5	55.2	57.5	59.6
	PST (n=42)	54.8	58.3	63.2	66.4

WL = waiting list, CBT = cognitive behavioral therapy, PST = problem solving therapy

between observed and predicted probabilities), we can conclude that the regression coefficients obtained from a logistic mixed model analysis are too high and should therefore not be used as effect measure.

There are several papers in which a logistic GEE analysis (a population average approach) is compared to a logistic mixed model analysis (a subject specific approach). Most of these comparisons were made on cross-sectional data with clustering of data on for instance neighborhood level, school level, etc. Although the directions of the differences were comparable to the ones observed in the present study, the magnitude of the differences was, in general, much lower [21-23]. This is due to the fact that the between cluster differences in these cross-sectional studies are much lower than the between cluster (i.e. subject) differences within a longitudinal study. It was already mentioned that the magnitude of the differences between the results of the two methods depend on the magnitude of the between cluster/subject variance (see equation 1). Surprisingly, in none of the papers comparing logistic GEE analysis with logistic mixed model analysis, a recommendation is provided which of the two methods should be used. It is sometimes argued that preferring one method above the other depends on the question to be answered [8,24]. In general, if one is interested in the regression coefficient, i.e. the effect estimation, a population average approach should be used and when one is interested in estimating the heterogeneity between subjects in a longitudinal study or between clusters in a cross-sectional study, a subject-specific approach should be used. In longitudinal RCTs, one is not interested in the heterogeneity between subjects, but one is interested in the effect estimation, taking into account the dependency of the observations within the subjects and treat it as a nuisance. For this purpose, logistic GEE analysis provides a valid estimate of the coefficient, while logistic mixed model analysis does not.

One of the arguments against the use of a logistic GEE analysis is that the results of a logistic GEE analysis are biased when there are missing data, especially when the missing data are not completely at random, i.e. not MCAR [11-13]. In most longitudinal RCTs, there is missing data and in most longitudinal RCTs, the missing data are not MCAR, so it is a common believe that a logistic GEE analysis should not be used in those situations. Although this argument is theoretically true, it should be realised that

TABLE 8. Regression coefficients and standard errors (between brackets) obtained from different logistic longitudinal data analyses performed on the 3-arm RCT regarding the internet based treatment of depressive symptoms from a complete case analysis

		GEE	Mixed models	
			PQL	ML
week 5	CBT	-0.89 (0.58)	-1.26 (0.87)	-1.41 (0.89)
	PST	-2.10 (0.55)	-3.12 (0.86)	-3.46 (0.93)
week 8	CBT	-1.10 (0.45)	-1.66 (0.72)	-1.89 (0.78)
	PST	-1.18 (0.46)	-1.79 (0.75)	-2.03 (0.82)
week 12	CBT	-1.40 (0.46)	-2.14 (0.73)	-2.40 (0.81)
	PST	-1.28 (0.48)	-1.92 (0.77)	-2.17 (0.83)
variance[1]			3.79	4.94

CBT = cognitive behavioral therapy, PST = problem solving therapy
[1]between subject variance obtained from the mixed model analyses

TABLE 9. Observed percentages of depressed subjects and predicted probabilities obtained from different logistic longitudinal data analyses performed on the 3-arm RCT regarding the internet based treatment of depressive symptoms from a complete case analysis

		Observed	GEE	Mixed models	
				PQL	ML
week 5	WL (n=58)	89.7	89.7	96.1	97.3
	CBT (n=41)	78.0	78.0	87.5	89.8
	PST (n=35)	51.4	51.4	52.0	53.1
week 8	WL (n=58)	77.6	77.6	86.7	89.6
	CBT (n=41)	53.7	53.7	55.2	56.6
	PST (n=35)	51.4	51.4	52.0	53.1
week 12	WL (n=58)	81.0	81.0	89.1	92.4
	CBT (n=41)	51.2	51.2	49.0	52.3
	PST (n=35)	54.3	54.3	54.6	54.2

WL = waiting list, CBT = cognitive behavioral therapy, PST = problem solving therapy

the percentage of missing data must be very high to have a detrimental influence on the validity of the results of a GEE-analysis [5] and that a logistic mixed model analysis is only valid in situations when missing data is missing at random (MAR) and when the model is correctly specified (i.e. with a random intercept and with all necessary random slopes) [5]. In the analysis performed on the example datasets it is not clear what the impact of the missings is on the estimation of the effect of the intervention. However, looking at the predicted probabilities from both the logistic GEE analysis and the logistic mixed model analyses, the influence of missing data is not very big. In all analyses the comparison between the predicted probabilities and the observed frequencies was in favor of the logistic GEE analysis. This is despite the fact that the missing data in both datasets was not completely at random [14,17] and that the percentage of missing data in the second example dataset was relatively high. There might be theoretical situations with larger amounts of MAR data in which logistic mixed model analysis might outperform logistic GEE analysis. However, longitudinal RCTs usually have less than 25% missing data.

It is sometimes argued that logistic GEE analysis and logistic mixed model analysis can be used interchangeable, because both the regression coefficients and the standard errors are higher in a more or less systematical manner when they are derived from a logistic mixed model analysis compared to a logistic GEE analysis. Consequently, the p-values and the answer to the question whether there is a significant difference between the intervention(s) and the control group is similar between the two statistical methods. When one is only interested in hypothesis testing, this is a valid argument, but nowadays, especially in epidemiology the major interest is in the estimation of the magnitude of the effect of the intervention(s) (i.e. regression coefficients and confidence intervals) rather than in hypothesis testing. And because the effect estimates are highly different between the two methods, one should make a careful choice between the two methods irrespective of the level of significance.

The comparisons in this paper also show that the results obtained from a logistic mixed model analysis vary considerably depending on the estimation procedure used.

There was a remarkable difference in the results obtained from a penalised quasi likelihood approach compared to the results obtained from a maximum likelihood approach. From the literature there is some evidence that the penalised quasi likelihood approach is slightly better than the maximum likelihood approach [25], which is more or less confirmed by our results. Nevertheless, both methods are frequently used. The difference observed between the two estimation procedures is a further indication that the results of a logistic mixed model analysis should be interpreted with great caution.

The present study deals with longitudinal data. As been mentioned before, mixed model analysis is also used in cross-sectional studies where individual data is clustered within for instance neighborhoods or schools. In those situations the same problems occur, although the differences between the results obtained from a logistic GEE analysis and a logistic mixed model analysis are less pronounced, due to the lower between cluster variance. When a longitudinal multicenter trial is performed, besides the clustering of the repeated measurements within the subjects, there is also clustering on the center level. When the number of centers is relatively large, a logistic GEE analysis can not be used anymore because within a (logistic) GEE analysis it is not possible to take into account clustering on more than one level. When the number of centers is relatively small, the center could be added as a covariate to the model. Mixed model analysis is capable of dealing with clustering on more than one level, so when also the clustering on the center level must be taken into account, a (logistic) mixed model analysis should be used with the same 'problems' as has been shown in the present paper. The simplest solution to this 'problem' is to ignore the clustering on the center level and to use a logistic GEE analysis. The effect of this ignoring approach depends, of course, on the magnitude of the between center variance. An alternative solution is to use a logistic mixed model analysis taking into account both the clustering on the subject level and on the center level and to transform the obtained subject specific regression coefficients into population average regression coefficients by using equation 1. However, in the latter the estimated regression coefficients will still highly depend on the estimation procedure used.

CONCLUSIONS

This paper shows that logistic GEE analysis outperforms logistic mixed model analysis for longitudinal RCT data regarding the estimated regression coefficients (i.e the effect estimates). It is also shown that the regression coefficients obtained from a longitudinal logistic mixed model analysis are an overestimation of the actual regression coefficients. It is therefore advised to use a longitudinal logistic GEE analysis for the effect estimation in longitudinal RCTs.

Acknowledgements

We thank Pim Cuijpers for providing the data of one of the example datasets.

References

1. Liang K-Y, Zeger SL. Longitudinal data analysis using generalised linear models. Biometrica 1986;73:45–51.
2. Zeger SL, Liang K-Y. Longitudinal data analysis for discrete and continuous outcomes. Biometrics 1986;42:121–30.
3. Laird NM, Ware JH. Random effects models for longitudinal data. Biometrics 1982;38:963–74.
4. Goldstein H. Multilevel statistical models. London: Edward Arnold;1985.
5. Twisk JWR. Applied longitudinal data analysis for epidemiology. A practical guide (2nd edition). Cambridge UK, Cambridge University Press; 2013.
6. Have TR ten, Ratcliffe SJ, Reboussin BA, Miller ME. Deviations from the population-averaged cluster-specific relationship for clustered binary data. Statistical Methods in Medical Research 2004;13:3-16.
7. Heo M, Leon AC. Comparison of statistical methods for analysis of clustered binary outcomes. Stat Med 2005;24:911-923.
8. Hu FB, Goldberg J, Hedeker D, Flay BR, Pentz M. Comparison of population-averaged and subject-specific approaches for analyzing repeated binary outcomes. Am J Epidemiol 1998;147:694-703.
9. Hubbard AE, Ahern J, Fleischer NL, et al. To GEE or not to GEE. Comparing population average and mixed models for estimating the associations between neighborhood risk factors and health. Epidemiology 2010;21:467-74.
10. Masaoud E, Stryhn H. A simulation study to assess statistical methods for binary repeated measures data. Preventive Veterinary Medicine 2010;93:81-97.
11. Little RJA. Modelling the drop-out mechanism repeated measures studies. Journal of the American Statistical Association 1995;90:1112–21.
12. Albert PS. Longitudinal data analysis (repeated measures) in clinical trials. Statistics in Medicine 1999;18:1707–32.
13. Omar RZ, Wright EM, Turner RM, Thompson SG. Analysing repeated measurements data: a practical comparison of methods. Statistics in Medicine, 1999;18:1587–603.
14. Apeldoorn AT, Ostelo RW, Helvoirt H van, et al. A randomized controlled trial on the effectiveness of a classification-based system for subacute and chronic low back pain. Spine 2012;37:1347-56.
15. Fairbank JC, Pynsent PB: The Oswestry Disability Index. Spine 2000;25:2940-52.
16. Tonosu J, Takeshita K, Hara N, Matsudaira K, Kato S, Masuda K, Chikuda H. The normative score and the cut-off value of the Oswestry Disabiltiy Index (ODI). Eur J Spine 2012;21:1596-602.
17. Warmerdam L, Straten A van, Twisk J, Riper H, Cuijpers P. Internet-based treatment for adults with depressive symptoms: randomized controlled trial. J Med Internet Res 2008;10:e44.
18. STATA. Stata reference manual, release 7. College Station, Texas: Stata Press; 2001.
19. Goldstein H. Rasbash J, Plewis I, Draper D, Browne W, Yang M, Woodhouse G, Healy M. A user's guide to MLwiN. London:

Institute of Education; 1998.
20. Rasbash J, Browne W, Goldstein H, et al. A user's guide to MLwiN, 2nd edn. London: Institute of Education; 1999.
21. Bellamy SL, Gibberd R, Hancock L, et al. Analysis of dichotomous outcome data for community intervention studies. Stat Methods Med Res 2000;9:135.
22. Kim H-Y, Preisser JS, Rozier RG, Valiyaparambi JV. Multilevel analysis of group randomized trials with binary data. Community Dent Oral Epidemiol 2006;34:241-251.
23. Ma J, Thabane L, Kaczorowski J, et al. Comparison of Bayesian and classical methods in the analysis of cluster randomized controlled trials with a binary outcome: The Community Hypertension Assessment Trial (CHAT). BMC Medical Research Methodology 2009;9:37.
24. Subramanian SV. The relevance of multilevel statistical methods for identifying causal neighbourhood effects. Social Science & Medicine 2004;58:1961-1967.
25. Twisk JWR. Applied multilevel analysis. A practical guide. Cambridge UK, Cambridge University Press; 2006.

*