Türkiye Klinikleri Biyoistatistik Dergisi
Turkiye Klinikleri Journal of Biostatistics Turkiye Klinikleri J Biostat. 2020;12(3):252-60

ORİJİNAL ARAŞTIRMA ORIGINAL RESEARCH

DOI: 10.5336/biostatic.2020-78912

Performances of Maximum Likelihood and Bayesian Methods Used For Estimating Three-Level Hierarchical Models: A Comparative Study

Üç Düzeyli Hiyerarşik Modelleri Tahmin Etmek İçin Kullanılan Maksimum Olabilirlik ve Bayesçi Yöntemlerin Performansı: Karşılaştırmalı Bir Çalışma

Oya KALAYCIOĞLU[a]

[a]Department of Biostatistics and Medical Informatics, Bolu Abant İzzet Baysal University Faculty of Medicine, Bolu, TÜRKİYE

ABSTRACT Objective: Hierarchical data with two or more levels are common in different fields of research including medical, educational, social and sports sciences. Maximum likelihood (ML) and Bayesian Markov Chain Monte Carlo (MCMC) estimation methods are widely used in regression analyses used for modelling these hierarchical data. However, the performances of these methods are not well studied for the estimation of three-level models. This paper aims at finding the optimal estimation technique under various combinations of number of clusters at second and third levels in three-level data sets. **Material and Methods:** A data application example is presented using a three-level dataset on football player's performance. Then, a simulation study based on the 3-level hierarchical linear model is performed for the comparison of four different maximum likelihood and Bayesian estimation approaches under various number clusters. **Results:** The data analysis and simulation study illustrate how strongly different estimation approaches affect the model parameter estimates, especially variance components. It is found that, if the main interest of the analysis is in the fixed part of the model, then any maximum likelihood or Bayesian method can be used, provided that the number of clusters at both levels are more than four. However, the main difference between these methods occurred in estimating the random terms. **Conclusion:** Results of the simulation study showed that using restricted maximum likelihood method is associated with better results for both regression coefficient and variance estimates. Obtaining valid variance estimates with Bayesian MCMC estimation requires careful consideration for defining prior distributions.

Keywords: Bayesian hierarchical modelling; mixed effects model; multilevel modelling; three-level clustering

ÖZET Amaç: Tıp, eğitim, sosyal ve spor bilimleri dahil olmak üzere farklı araştırma alanlarında iki veya daha fazla düzeyli hiyerarşik veriler yaygındır. Bu hiyerarşik verileri modellemek için kullanılan regresyon analizlerinde maksimum olabilirlik ve Bayesçi Markov Zinciri Monte Carlo (MZMC) tahmin yöntemleri yaygın olarak kullanılmaktadır. Ancak, bu yöntemlerin performansı üç düzeyli modellerin tahmini için yeterince çalışılmamıştır. Bu makale, üç düzeyli veri setlerinde ikinci ve üçüncü düzeylerindeki çeşitli küme sayısı kombinasyonları altında en iyi tahmin tekniğini bulmayı amaçlamaktadır. **Gereç ve Yöntemler:** Futbolcu performanslarına dair üç düzeyli bir veri seti kullanılarak bir veri uygulama örneği sunulmuştur. Sonrasında, dört farklı maksimum olabilirlik ve Bayesçi tahmin yaklaşımlarını farklı küme sayıları altında karşılaştırmak için, üç düzeyli hiyerarşik doğrusal modele dayanan bir simülasyon çalışması yapılmıştır. **Bulgular:** Veri analizi ve simülasyon çalışması, farklı tahmin yaklaşımlarının model parametre tahminlerini, özellikle varyans bileşenlerini ne kadar güçlü etkilediğini göstermiştir. Eğer analizin temel ilgi konusu modelin sabit kısmıysa, her iki düzeydeki küme sayısının dörtten fazla olması koşuluyla herhangi bir maksimum olabilirlik ve Bayesçi yöntemin kullanılabileceği sonucuna varılmıştır. Ancak, bu yöntemler arasındaki temel fark, rastgele terimleri tahmin etmede ortaya çıkmıştır. **Sonuç:** Simülasyon çalışmasının sonuçları, sınırlı maksimum olabilirlik yönteminin kullanılmasının hem regresyon katsayısı hem de varyans tahminleri için daha iyi sonuçlarla ilişkili olduğunu göstermiştir. Bayesçi MZMC tahmini ile geçerli varyans tahminleri elde etmek, önsel dağılımları tanımlamak için dikkatli bir değerlendirme gerektirmektedir.

Anahtar kelimeler: Bayesçi hiyerarşik modelleme; karma etkili modeller; çok düzeyli modelleme; üç-düzeyli kümeleme

Correspondence: Oya KALAYCIOĞLU
Department of Biostatistics and Medical Informatics, Bolu Abant İzzet Baysal University Faculty of Medicine, Bolu, TÜRKİYE
E-mail: oyakalaycioglu@ibu.edu.tr

Peer review under responsibility of Turkiye Klinikleri Journal of Biostatistics.

Received: 16 Sep 2020 ***Received in revised form:*** 22 Oct 2020 ***Accepted:*** 04 Nov 2020 ***Available online:*** 21 Dec 2020

In many fields of research, it is common to see hierarchical data sets where observations are nested within higher units and cannot be assumed to be independent. An hierarchical structure of data can occur when the individual observations are nested within different groups, or when repeated measurements are taken from the same individuals as in longitudinal studies, or both.[1] The correlation between observations into account, a common approach is to use hierarchical (i.e. mixed-effects, multilevel) models.[2] When data consists of three-level hierarchy, clustering must be accounted by the analysis model in order to ensure that resulting estimates are reliable. There are various likelihood and Bayesian Markov Chain Monte Carlo (MCMC) estimation techniques to fit three-level models, however the literature on estimation techniques to be used in the three-level context are limited.[3]

These estimation techniques were well compared for two-level data.[4-6] However, to the best of our knowledge, there is no to limited work which compares likelihood and Bayesian methods in the three-level context. The purpose of this study is to compare and discuss the performance of likelihood based and Bayesian MCMC estimation methods for the estimation of three-level linear hierarchical model, for different combinations of number of clusters at level-2 and level-3. Thus, it is aimed to contribute to the methodological literature in this area.

The study is motivated with a football players performance data set, which has a three-level hierarchical structure (i.e. football players (level-1) are nested within teams (level-2), teams are nested within leagues (level-3)). The data analysis demonstrates how strongly different estimation approaches affect the relevant results. In addition, the performance of likelihood and Bayesian estimation approaches were examined for modelling three-level hierarchical dataset with a simulation study. The layout of the paper is as follows. In Section 2 the fundamentals of three level hierarchical model and estimation techniques are given. The data which motivated this work and are used as a basis for comparing the likelihood and Bayesian methods are introduced in Section 3. In Section 4, the assessment of the techniques is illustrated via simulation study. Section 5 provides a discussion and recommendations.

MATERIAL AND METHODS

THREE-LEVEL HIERARCHICAL LINEAR MODEL

A three-level hierarchical model was used to analyse data and for the simulation study, where a random term is added to the model for each cluster implying a random intercepts model given below:[7]

$$y_{ijk} = \beta_0 + \sum_{q=1}^{p} \beta_q x_{(q)ijk} + v_k + u_{jk} + e_{ijk} \qquad (1)$$

$$[v_k] \sim N(0, \Omega_v),\ [u_{jk}] \sim N(0, \Omega_u),\ [e_{ijk}] \sim N(0, \Omega_e)$$

In the model above, y_{ijk} corresponds to the dependent variable of i^{th} level-1 unit, which is nested within the j^{th} level-2 cluster, which is further nested in k^{th} level-3 cluster. The $i \times q$ matrix of $x_{(q)ijk}$ represents the value for an independent variable q (q=*1,…,p*) for the corresponding y_{ijk}. The regression coefficients (β_0, ..., β_p) and three variance estimates (σ_v^2, σ_u^2 and σ_e^2) are referred as "fixed" and "random" effect parameters, respectively. The assumptions of the model are normality, independence of error terms at cluster and observation level, linearity and homoscedasticity. The expected correlation between the two randomly chosen units within the same cluster is assessed with intra-class correlations (ICC) in hierarchical models. The ICC for 2^{nd} and 3^{rd} levels are calculated using the formulae below.[8]

$$\rho_{level-2} = \frac{\sigma_u^2}{\sigma_v^2 + \sigma_u^2 + \sigma_e^2}$$

$$\rho_{level-3} = \frac{\sigma_v^2}{\sigma_v^2 + \sigma_u^2 + \sigma_e^2}$$

LIKELIHOOD ESTIMATION METHODS IN HIERARCHICAL MODELLING

The two main likelihood-based estimation techniques are Maximum Likelihood (ML) and Restricted (Residual) Maximum Likelihood (REML) estimation. In ML method, both regression coefficients (fixed effects) and the variance components (random effects) are included in the likelihood function, whereas the likelihood function in REML contains only the variance components, and the regression coefficients are estimated in a second estimation step.[8] ML estimation yields simultaneous estimation of fixed and random components by maximizing the likelihood function of the data.[9] On the other hand, in REML the estimation of fixed and random components is separated.[4]

Both methods which are used as estimation methods in hierarchical linear mixed models have their own merits. McCulloch and Neuhaus (2014) argue that ML method has a demerit of estimating biased estimates for random effect parameters and REML estimates have less bias.[10] They point out the growing preference for REML due to the following merits. First, REML produces unbiased estimates for random parameters when the number of observations in clusters are equal (i.e., balanced data set). Second, since fixed parameters (β) are not involved while forming the likelihood, there is no loss of degrees of freedoms occurs due to the estimation of fixed effects.[11] Finally, REML estimates are less sensitive to outliers in the data than ML.[12] Boedeker (2017) agrees that in theory REML leads to better estimates, especially when the number of clusters is small.[4]

Likelihood estimates are often obtained through iterative procedures such as Iterative Generalized Least Squares (IGLS) which is designed specifically for hierarchical models although they can be adapted to other models.[13] The iteration usually starts from 'reasonable' fixed parameter estimates, typically those obtained from ordinary least squares fit where the variance parameters are assumed to be 0. The next step follows with computational procedure that tries to improve starting values to produce better estimates.[8] When two consecutive estimates for each parameter are sufficiently close to each other, the program concludes that *convergence* has been achieved and that no more iteration is needed. The IGLS algorithm which is modified to produce REML estimates is named as Restricted Iterative Generalized Least Squares (RIGLS). Since in REML the likelihood function is restricted to include only the variance components, the number of parameters to be optimized is reduced in RIGLS and this can improve the convergence properties.[14]

BAYESIAN MCMC ESTIMATION METHODS IN HIERARCHICAL MODELLING

Bayesian estimation is commonly being used for the estimation of hierarchical models, as it is well suited for connecting the information within and across higher levels of hierarchy.[15] Bayesian approach incorporates parameter uncertainty through the prior distribution defined for a parameter estimate, and a random sample of representative parameter values from the posterior distribution is obtained at the estimation stage.

One of the main tasks while fitting Bayesian models is to define the priors for model parameters. It is a common approach to use noninformative priors when there is no prior knowledge on the distributions of parameters. When defining noninformative priors for variance parameters, typical choices are inverse-gamma prior where the hyperparameters of the distribution are very small (i.e., $\alpha = 0.001$, $\beta = 0.001$) and uniform distribution on variance scale.[16,17]

After defining the prior distributions, posterior distributions of both fixed and random parameters need to be specified, where $\beta_0, \dots, \beta_q$ and $\sigma_e{}^2, \sigma_u{}^2, \sigma_v{}^2$ represent the fixed and random parameters, respectively. This is done by combining the additional information in the form of a prior probability distribution and data.

$$p\left(\beta_{0,} \dots, \beta_q, \ \sigma_e{}^2, \sigma_u{}^2, \sigma_v{}^2 \middle| Y, \boldsymbol{X}\right) \propto p\left(\beta_{0,} \dots, \beta_q, \ \sigma_e{}^2, \sigma_u{}^2, \sigma_v{}^2\right). p\left(Y \middle| \boldsymbol{X}, \beta_{0,} \dots, \beta_q, \sigma_e{}^2, \sigma_u{}^2, \sigma_v{}^2\right),$$

where *Y* represents the response variable and ***X*** is the matrix of *k* explanatory variables.

DATA EXAMPLE

This study is motivated by the modelling football players' performance data set which was obtained from the whoscored.com database (https://www.whoscored.com). The data set contains information from 15 major Leagues of different countries consisting of teams and 5723 players in total. The variables that were used in the study were the players' rating scores, and various performance variables including number of goals, shots, tackles, assists, passes per 90 minutes and players' position. The primary objective of the analysis was to examine to identify whether the number of goals, shots, tackles, assists, passes per 90 minutes and players' positions affect their ratings. The structure of the dataset emerged the need of using 3-level hierarchical regression model, as the players are nested within teams, and teams are nested within leagues (Figure 1). The number of teams at each league varied from 16 to 28 and 25 players nested within each team on average.

The three-level hierarchical model is given below:

$$rating_{ijk} = \beta_0 + \beta_1 goal_{ijk} + \beta_2 assist_{ijk} + \beta_3 shot_{ijk} + \beta_4 pass_{ijk} + \beta_5 tackle_{ijk} + \beta_6 midfielder_{ijk} + \beta_6 forward_{ijk} + v_k + u_{jk} + e_{ijk} \qquad (1)$$

$$[v_k] \sim N(0, \Omega_v), \qquad [u_{jk}] \sim N(0, \Omega_u), \qquad [e_{ijk}] \sim N(0, \Omega_e)$$

where i^{th} football player is nested in team *j* which is also nested in league *k*.

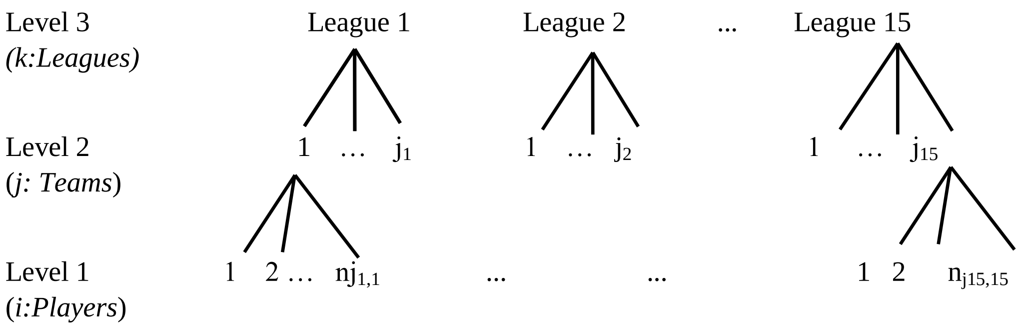

FIGURE 1: Nested structure of football players ratings' data.

The model above was fitted with ML, REML and Bayesian estimation methods using MLwiN version 3.03 which is a specialized software for fitting hierarchical models.[13] As there is no prior knowledge on the distributions of fixed and random effects as in this study, for Bayesian model fitting noninformative priors were used. For regression coefficients diffuse uniform priors were used ($p(\beta) \propto 1$) and for variance components inverse-gamma and uniform priors were defined as shown:

i. Diffuse inverse-gamma priors for variance components (equivalent to assume gamma priors for *precision* parameter $\lambda = \frac{1}{\sigma^2}$)

$$p\left({}^{1}/_{\sigma_v^2}\right) \propto \Gamma(\varepsilon,\varepsilon)\,,\ \ p\left({}^{1}/_{\sigma_u^2}\right) \propto \Gamma(\varepsilon,\varepsilon)\,,\ \ p\left({}^{1}/_{\sigma_e^2}\right) \propto \Gamma(\varepsilon,\varepsilon)$$

where ε is very small such as 0.001.

ii. Diffuse uniform priors for variance components.

$$p(\sigma_v^2) \propto U(0,\tfrac{1}{\varepsilon})\,,\ p(\sigma_u^2) \propto U\left(0,\tfrac{1}{\varepsilon}\right)\,,\ p(\sigma_e^2) \propto U(0,\tfrac{1}{\varepsilon})$$

where ε is a small positive real number.

To investigate the effect of different choice of diffuse variance priors on parameter estimates, Bayesian models were fitted with the priors given in (i) and (ii), in turn. For all the parameters, it was concluded that running 3 independent chains for 10000 times were enough to obtain sufficiently small Monte Carlo Standard Error (MCSE). The first 2000 iterations of each chain were discarded from the analysis as burn-in period.

TABLE 1: Estimates and their standard errors (given in parentheses) for three-level hierarchical model with random intercepts fitted by four different approaches.

Parameter estimates	Estimation methods			
	Likelihood based		*Bayesian MCMC*	
	ML	REML	InvGamma[a]	Uniform[b]
Intercept (β_0)	6.204 (0.017)	6.204 (0.017)	6.203 (0.018)	6.202 (0.018)
Goal (β_1)	0.874 (0.028)	0.874 (0.028)	0.875 (0.028)	0.874 (0.028)
Shot (β_2)	0.051 (0.006)	0.051 (0.006)	0.051 (0.006)	0.050 (0.006)
Assist (β_3)	0.567 (0.026)	0.567 (0.026)	0.567 (0.026)	0.567 (0.026)
Pass (β_4)	0.008 (0.001)	0.008 (0.001)	0.008 (0.001)	0.009 (0.001)
Tackle (β_5)	0.088 (0.004)	0.088 (0.004)	0.089 (0.004)	0.088 (0.004)
Position				
Defender	*Reference*			
Midfielder (β_6)	-0.238 (0.008)	-0.238 (0.008)	-0.238 (0.008)	-0.237 (0.008)
Forward (β_7)	-0.272 (0.013)	-0.272 (0.013)	-0.273 (0.013)	-0.273 (0.013)
Level-3 variance (σ_v^2)	0.001 (0.001)	0.001 (0.001)	0.002 (0.001)	0.003 (0.002)
Level-2 variance (σ_u^2)	0.002 (0.001)	0.002 (0.001)	0.003 (0.001)	0.004 (0.001)
Level-1 variance (σ_e^2)	0.053 (0.001)	0.053 (0.001)	0.053 (0.001)	0.053 (0.001)
ICC (level-3)	0.018	0.018	0.034	0.050
ICC (level-2)	0.036	0.036	0.054	0.070
-2xlog-likelihood	-338.753	-338.733	-	-
Deviance (MCMC)	-	-	-383.875	-383.928

$\varepsilon = 0.001$, [a]Inverse-gamma $(\varepsilon, \varepsilon)$ and diffuse uniform priors were defined for variance components and regression coefficients, respectively.
[b]Uniform(0, ε^{-1}) and diffuse uniform priors were defined for variance components and regression coefficients, respectively. ML: Maximum Likelihood. REML: Restricted Maximum Likelihood. MCMC: Markov Chain Monte Carlo. ICC: Intra-class correlation

The point estimates and standard deviations obtained from both likelihood and Bayesian approaches are given in the Table 1. With the likelihood-based methods the point estimates were predicted at the end of iterative procedures IGLS and RIGLS. Besides, in the Bayesian estimation procedure, the estimates of the regression coefficients and variance components are the posterior means computed from the posterior distribution of each parameter.

The ICC at the 2nd and 3rd levels were estimated to be low, however the likelihood ratio tests favoured the larger model with three levels over the fixed effects or 2-level models. In terms of coefficient estimates results are similar among different approaches, but there are some differences between methods in terms of variance component estimates and ICC, especially for level-3 variance. With ML method, the level-3 variance was estimated to be 0.001, however it was estimated as 0.003 with relatively large standard error with Bayesian estimation using uniform priors for variance components. This result may be due to two reasons. First, the variance component parameter depends on the number of clusters and the number of clusters at the 3rd level (i.e. players) and 2nd level (i.e. teams) were not enough to yield accurate estimates for random parameters. Second, the true values of higher-level variance parameters were small and they were overestimated when uniform priors were defined for random effect parameters which were close to zero. The effect of diffuse N(0,1) priors on fixed effect parameters is small, since they were based upon total sample size (N_{total}=*5723*).

SIMULATION STUDY

A simulation study based on the 3-level hierarchical linear model given in Model (1) was conducted to compare the performance of different estimation methods through a Monte Carlo simulation study by using 1000 simulations. For simplicity the model consisted of a normally distributed continuous variable as an independent variable. A continuous outcome was generated where the true value was defined as β_1=0.20. 16 different data sets were generated for the combinations of four different cluster sizes by varying the number clusters at the 2nd and 3rd levels and keeping the number of units at each level-2 as 20. The ICC at the 2nd and 3rd levels were defined as 0.10 and 0.30, respectively.

Each data set was fitted using the estimation methods used for the data analysis in Section 2. For comparing the accuracy of parameters, relative bias and coverage rates were calculated. Relative bias was calculated with $\left(\frac{\hat{\beta}-\beta_{true}}{\beta_{true}}\right) \times 100$ and coverage of nominal 95% confidence intervals (CI) (i.e. credible intervals (CrI) for Bayesian estimation) were calculated as the proportion of times that the 95% CIs include true estimate for i=1,…,1000.

RESULTS

Table 2 shows the results of the simulations for the parameter estimates for the fixed regression coefficient obtained with four different estimation techniques. When the number of clusters at both levels are more than 12, all methods provided negligible bias. When the number of clusters at both levels were small, i.e. 4, the methods showed a negligible upward bias. Under the same condition, the coverage rates were poor with all methods, but we note that Bayesian estimation resulted in slightly better coverage rates.

TABLE 2: Simulation results for the fixed effect estimate of the continuous variable.

		Estimation methods							
		Likelihood based				*Bayesian MCMC*			
Number of clusters		ML		REML		InvGamma[a]		Uniform[b]	
Level-3	Level-2	*Rel.bias*	*Cov.*	*Rel.bias*	*Cov.*	*Rel.bias*	*Cov.*	*Rel.bias*	*Cov.*
4	4	3.75	88.7	3.75	87.3	3.30	92.6	3.50	91.2
4	8	0.68	97.1	0.49	97.3	1.04	94.1	1.06	97.2
4	12	-0.23	96.1	-0.25	96.4	2.40	94.8	2.37	95.7
4	15	-0.01	96.0	-0.02	96.2	2.12	95.3	2.00	95.9
8	4	-0.06	94.1	-0.06	94.0	0.82	95.1	0.73	95.1
8	8	-0.16	96.6	-0.16	95.1	1.04	96.2	0.99	95.4
8	12	0.38	96.1	0.38	96.0	1.28	94.3	1.29	95.6
8	15	0.43	96.2	0.62	96.2	2.16	94.3	2.00	95.6
12	4	-0.31	93.4	-0.31	93.2	0.27	93.4	0.20	92.6
12	8	0.72	95.8	0.72	96.4	1.65	93.4	1.56	93.1
12	12	0.09	94.4	0.09	94.4	0.41	93.6	0.39	94.7
12	15	0.09	95.6	0.09	95.1	1.07	94.5	1.50	95.1
15	4	0.72	94.3	0.72	92.8	0.23	93.2	0.20	93.5
15	8	0.72	94.8	0.72	95.8	1.07	93.2	1.00	94.2
15	12	0.72	95.8	0.62	95.6	0.41	94.8	0.39	95.8
15	15	0.47	95.5	0.47	95.7	0.47	95.0	0.61	95.8

$\varepsilon = 0.001$, [a]Inverse-gamma $(\varepsilon, \varepsilon)$ and diffuse uniform priors were defined for variance components and regression coefficients, respectively.
[b]Uniform(0, ε^{-1}) and diffuse uniform priors were defined for variance components and regression coefficients, respectively. Rel.bias: Relative bias, Cov: 95% Nominal Coverage. ML: Maximum Likelihood. REML: Restricted Maximum Likelihood. MCMC: Markov Chain Monte Carlo.

The relative bias of the variance components was assessed across the different sizes of clusters at both levels (Table 3, Table 4, Table 5). In general, the choice of the estimation technique and priors did not show any significant effect on the variance component estimates at level-1 and level-2, as all methods provided unbiased estimates under different combinations of number of clusters. The only exception was the scenario where both levels consist of 4 clusters as ML estimation resulted in a considerable bias.

For the estimation of the 3rd level variances, $\sigma_v{}^2$, REML estimates were superior to the other methods on all aspects. However, Bayesian estimation with inverse-gamma priors benefitted from higher number of level-3 clusters and resulted in unbiased level-3 variance estimates if the number of level-3 clusters is more than 8. ML estimation is unbiased for $\sigma_v{}^2$, if the there are at least 12 level-3 clusters and 8 level-2 clusters, provided that there are 20 level-1 units. Using uniform priors for variance parameters resulted in biased estimates for $\sigma_v{}^2$ for all combinations of number of clusters. When the cluster size at 3rd level increased, the relative bias obtained with uniform priors considerably decreased.

TABLE 3: Relative bias for the level-1 variance component (σ_e^2).

Number of clusters		Estimation methods			
		Likelihood based		*Bayesian MCMC*	
Level-3	Level-2	ML	REML	InvGamma[a]	Uniform[b]
4	4	-1.35	0.02	1.04	1.11
4	8	0.61	0.15	0.91	0.97
4	12	-0.71	-0.07	0.11	0.13
4	15	-0.51	0.04	0.10	0.13
8	4	0.28	0.01	0.43	0.49
8	8	-0.60	-0.09	0.18	0.25
8	12	-0.35	-0.12	0.27	0.29
8	15	-0.18	0.08	0.10	0.11
12	4	-0.12	0.14	0.22	0.23
12	8	-0.11	0.14	-0.08	-0.02
12	12	0.15	0.22	-0.02	0.10
12	15	0.10	0.21	0.04	0.05
15	4	-0.07	0.11	0.20	0.26
15	8	-0.09	0.15	-0.07	-0.18
15	12	0.10	0.14	-0.01	0.09
15	15	0.08	0.10	0.04	0.04

$\varepsilon = 0.001$, [a]Inverse-gamma $(\varepsilon, \varepsilon)$ and diffuse uniform priors were defined for variance components and regression coefficients, respectively.
[b]Uniform(0, ε^{-1}) and diffuse uniform priors were defined for variance components and regression coefficients, respectively. ML: Maximum Likelihood. REML: Restricted Maximum Likelihood. MCMC: Markov Chain Monte Carlo.

TABLE 4: Relative bias for the level-2 variance component (σ_u^2).

Number of clusters		Estimation methods			
		Likelihood based		*Bayesian MCMC*	
Level-3	Level-2	ML	REML	InvGamma[a]	Uniform[b]
4	4	-13.13	-2.74	4.36	2.57
4	8	-2.94	-2.92	-0.50	1.95
4	12	-1.73	-1.71	0.01	0.64
4	15	-1.16	-1.45	0.12	0.18
8	4	2.61	2.20	1.37	1.56
8	8	-1.44	-1.42	0.43	0.53
8	12	-1.14	-1.14	0.54	0.37
8	15	-0.55	-0.53	0.18	0.15
12	4	-2.15	-2.14	0.87	0.72
12	8	-0.74	-0.74	-0.53	-0.40
12	12	0.15	0.17	0.30	0.39
12	15	-0.11	0.12	0.25	0.66
15	4	-0.60	-0.52	0.15	0.16
15	8	-0.25	0.28	0.25	0.25
15	12	0.10	0.14	-0.01	0.09
15	15	0.08	0.10	0.04	0.04

$\varepsilon = 0.001$, [a]Inverse-gamma $(\varepsilon, \varepsilon)$ and diffuse uniform priors were defined for variance components and regression coefficients, respectively.
[b]Uniform(0, ε^{-1}) and diffuse uniform priors were defined for variance components and regression coefficients, respectively. ML: Maximum Likelihood. REML: Restricted Maximum Likelihood. MCMC: Markov Chain Monte Carlo.

TABLE 5: Relative bias for the level-3 variance component (σ_v^2).

Number of clusters		Estimation methods			
		Likelihood based		*Bayesian MCMC*	
Level-3	Level-2	ML	REML	InvGamma[a]	Uniform[b]
4	4	-22.84	-4.27	38.43	206.17
4	8	-11.86	-1.26	13.44	56.29
4	12	-12.23	-3.09	10.73	36.23
4	15	-16.93	-2.71	15.31	25.31
8	4	-13.56	0.88	8.51	63.34
8	8	-14.58	-4.52	-9.04	34.03
8	12	-5.15	1.28	-1.83	16.64
8	15	-6.91	3.31	-2.47	15.00
12	4	-14.11	-5.54	3.34	23.04
12	8	-7.26	3.22	-2.76	21.26
12	12	-6.43	-1.96	0.71	12.35
12	15	-4.31	1.31	1.31	12.14
15	4	-13.09	-3.31	2.86	29.13
15	8	-8.73	-1.93	2.47	32.58
15	12	-4.91	1.56	0.87	15.48
15	15	-3.31	1.10	1.10	11.37

$\varepsilon = 0.001$, [a]Inverse-gamma $(\varepsilon, \varepsilon)$ and diffuse uniform priors were defined for variance components and regression coefficients, respectively.
[b]Uniform(0, ε^{-1}) and diffuse uniform priors were defined for variance components and regression coefficients, respectively. ML: Maximum Likelihood. REML: Restricted Maximum Likelihood. MCMC: Markov Chain Monte Carlo.

DISCUSSION

When the estimation techniques were compared in terms of bias and coverage through a simulation study, all methods provided unbiased regression coefficient estimates. The only scenario where the regression coefficients were estimated with low coverage was when the number of clusters were specified as 4 at both levels. In previous studies, at least 15 clusters were recommended for obtaining accurate estimates for continuous outcomes in two-level datasets.[18] However, the findings presented in our study showed that the suggestions regarding the number of clusters for two-level data sets are not directly adaptable to three-level context as the nesting structure is more complicated with three-level data structures.

In general, variance estimates of a three-level model obtained with REML were less biased compared to ML and Bayesian methods, regardless of number of clusters. Given that REML performed approximately the same with other methods for the regression coefficients, the findings of this study agree with the growing preference of REML. However, it is important to note that with REML only differences in random part (variance components) can be compared with likelihood tests and when the interest is assessing the models which differ only in the fixed part it is not recommended to use.[8]

In terms of Bayesian methods, it was concluded that the choice of priors for variance parameters has a high impact on variance estimates. When uniform priors were used both for fixed and random parameters, the estimates for the 3^{rd} level variance parameters were positively biased. With 2-level models it is known that a uniform prior tends to give a biased variance estimates at level-2, when the number of level 2 units is small.[13] This study showed that, with small number of clusters at both levels in a three-level model, uniform priors resulted in upward bias in level-3 variances, however level-2 variances were estimated precisely.

There exist several limitations of this study. First, it was aimed at assessing the effect of the estimation techniques for different number of clusters at Level-2 and Level-3. Therefore, simulated the datasets were generated with varying the number of clusters at both levels, keeping the level-1 units and ICCs fixed at all scenarios. In addition, using regression coefficients at higher levels, interaction terms or random slopes were avoided, as extending the model to include additional parameters may bring the issue of convergence with Bayesian estimation.

CONCLUSION

In this paper, the performances of maximum likelihood and Bayesian model-based approaches that had the potential to accommodate the individual dependency in the nested structure of 3-level hierarchical dataset were investigated. In the motivating example which is based on a three-level data, in terms of estimating regression coefficients all methods provided similar results. However, the main difference between likelihood and Bayesian methods was observed in the highest-level variance in 3-level hierarchical linear model. The simulation study based on a hypothetical three-level data set with a normally distributed variable showed that, all methods provide unbiased estimates with acceptable nominal coverage rates for the regression coefficient. The only exception was the scenario where the number of clusters at level-2 and level-3 are as small as four. If the random part is the main interest of the analysis, then the use of REML may be preferred. Using Bayesian MCMC estimation requires careful consideration of prior distributions for obtaining accurate variance estimates.

Although the methods under varying number of clusters at level-2 and level-3 with several simulation scenarios were assessed in this study, it should be avoided to overgeneralize these results which are based on a single hypothetical data set.

Source of Finance

During this study, no financial or spiritual support was received neither from any pharmaceutical company that has a direct connection with the research subject, nor from a company that provides or produces medical instruments and materials which may negatively affect the evaluation process of this study.

Conflict of Interest

No conflicts of interest between the authors and / or family members of the scientific and medical committee members or members of the potential conflicts of interest, counseling, expertise, working conditions, share holding and similar situations in any firm.

Authorship Contributions

This study is entirely author's own work and no other author contribution.

REFERENCES

1. Olvera Astivia OL, Gadermann A, Guhn M. The relationship between statistical power and predictor distribution in multilevel logistic regression: a simulation-based approach. BMC Med Res Methodol. 2019;19(1):97.[Crossref] [PubMed] [PMC]
2. Gelman A. Multilevel (hierarchical) modeling: what it can and cannot do. Technometrics. 2019;48(3):432-5.[Crossref]
3. McNeish D, Wentzel KR. Accommodating small sample sizes in three-level models when the third level is incidental. Multivariate Behav Res. 2017;52(2):200-15.[Crossref] [PubMed]
4. Boedeker P. Hierarchical linear modeling with maximum likelihood, restricted maximum likelihood, and fully Bayesian estimation. Practical Assessment, Research and Evaluation. 2017;22.[Crossref]
5. Browne WJ, Draper D. A comparison of Bayesian and likelihood-based methods for fitting multilevel models. Bayesian Analysis. 2006;1(3):473-514.[Crossref]
6. Gumedze FN, Dunne TT. Parameter estimation and inference in the linear mixed model. Linear Algebra and its Applications. 2011;435(8):1920-44.[Crossref]
7. Goldstein H. Multilevel Statistical Models. London: Institute of Education; 1999.
8. Hox JJ, Moerbeek M, Van de Schoot R. Multilevel Analysis: Techniques and Applications. Routledge; 2017.[Crossref] [PMC]
9. Corbeil RR, Searle SR. Restricted maximum likelihood (REML) estimation of variance components in the mixed model. Technometrics. 1976;18(1):31-8.[Crossref]
10. McCulloch CE, Neuhaus JM. Generalized linear mixed models. Wiley StatsRef: Statistics Reference Online, 2014.[Crossref]
11. Jiang J, Rao J, Gu Z, Nguyen T. Fence methods for mixed model selection. The Annals of Statistics. 2008;36(4):1669-92.[Crossref]
12. Verbyla AP. Modelling variance heterogeneity: residual maximum likelihood and diagnostics. Journal of the Royal Statistical Society: Series B (Methodological). 1993;55(2):493-508.[Crossref]
13. Browne WJ. MCMC Estimation in MLwiN v3.03. Centre for Multilevel Modelling, University of Bristol; 2019.
14. Jones A, Huddleston E. SAS/STAT 9.2 User's Guide. 2nd ed. Cary, NC: SAS Institute; 2009.[Link]
15. Townsend J, Busemeyer J, Kruschke J, Vanpaemel W. Bayesian Estimation in Hierarchical Models. In: Busemeyer JR, Wang Z, Townsend JT, Eidels A, eds. The Oxford Handbook of Computational and Mathematical Psychology. New York, NY: Oxford University Press; 2015.
16. Daniels MJ. A prior for the variance in hierarchical models. The Canadian Journal of Statistics. 1999;27:567-78.[Crossref]
17. Gelman A, Shalizi CR. Philosophy and the practice of Bayesian statistics. Br J Math Stat Psychol. 2013;66(1):8-38.[Crossref] [PubMed] [PMC]
18. McNeish D, Stapleton L. The effect of small sample size on two-level model estimates: a review and illustration. Educational Psychology Review. 2016;28(2):295-314.[Crossref]