Robustness of Generalized Linear Mixed Models for Split-Plot Designs with Binary Data

Roser Bono[1,2,*], Rafael Alarcón[3], Jaume Arnau[1], F. Javier García-Castro[4] and Maria J. Blanca[3]

1 Department of Social Psychology and Quantitative Psychology. University of Barcelona. Barcelona (Spain)
2 Institute of Neurosciences.University of Barcelona. Barcelona (Spain)
3 Department of Psychobiology and Behavioral Sciences Methods. University of Malaga. Malaga (Spain)
4 Department of Psychology, Universidad Loyola Andalucía. Seville (Spain)

Título: Robustez de los Modelos Lineales Mixtos Generalizados para Diseños Split-Plot con Datos Binarios.

Resumen: Este artículo examina la robustez del modelo lineal mixto generalizado (GLMM, por sus siglas en inglés). El GLMM estima efectos fijos y efectos aleatorios y es especialmente útil cuando la variable dependiente es binaria. También es útil cuando la variable dependiente es de medidas repetidas, ya que puede modelar la correlación. El presente estudio utilizó la simulación de Monte Carlo a fin de analizar las tasas de error de Tipo I empíricas de los GLMM en diseños split-plot. Las variables manipuladas fueron el tamaño de muestra, el tamaño de grupo, el número de medidas repetidas y la correlación entre las medidas repetidas. También se consideraron condiciones extremas, tales como muestras pequeñas, grupos no balanceados y diferente correlación en cada grupo (emparejamiento entre tamaño de grupo y correlación entre medidas repetidas). Para grupos balanceados, los resultados mostraron que el efecto grupo era robusto en todas las condiciones, mientras que para grupos no balanceados el efecto tendía a ser conservador con emparejamiento positivo y liberal con emparejamiento negativo. Con respecto a los efectos tiempo e interacción, los resultados mostraron, tanto para grupos balanceados como para no balanceados, que: (a) la prueba fue robusta con baja correlación (.2), pero conservadora para valores medios de correlación (.4 y .6), y (b) la prueba tendía a ser conservadora para emparejamiento positivo y negativo, especialmente en este último.

Palabras clave: Modelos lineales mixtos generalizados. Datos binarios. Simulación Monte Carlo. Tasa de error Tipo I.

Abstract: This paper examined the robustness of the generalized linear mixed model (GLMM). The GLMM estimates fixed and random effects, and it is especially useful when the dependent variable is binary. It is also useful when the dependent variable involves repeated measures, since it can model correlation. The present study used Monte Carlo simulation to analyze the empirical Type I error rates of GLMMs in split-plot designs. The variables manipulated were sample size, group size, number of repeated measures, and correlation between repeated measures. Extreme conditions were also considered, including small samples, unbalanced groups, and different correlation in each group (pairing between group size and correlation between repeated measures). For balanced groups, the results showed that the group effect was robust under all conditions, while for unbalanced groups the effect tended to be conservative with positive pairing and liberal with negative pairing. Regarding time and interaction effects, the results showed, for both balanced and unbalanced groups, that: (a) The test was robust with low correlation (.2), but conservative for medium values of correlation (.4 and .6), and (b) the test tended to be conservative for positive and negative pairing, especially the latter.

Keywords: Generalized linear mixed models. Binary data. Monte Carlo simulation. Type I error rate.

Introduction

Scientific researchers are faced with ever more sophisticated and innovative data analysis strategies. A good example of this is the generalized linear mixed model (GLMM), a flexible analytical strategy that has evolved from other simpler mathematical models and which has been increasingly used in recent decades. The general linear model (GLM) is the most commonly used approach in the health and social sciences (Blanca et al., 2018), where researchers often perform regression analysis, analysis of variance, and analysis of covariance. These tests require the fulfillment of a series of assumptions such as a quantitative response variable, normally distributed data, homogeneity of variance, and independence of errors, assumptions that are not always satisfied with real data. The linear mixed model (LMM) is an extension of the GLM that allows both fixed and random effects, and it is particularly useful when data are not independent (e.g., hierarchical data, repeated measures) and homogeneity of variance is not required; robust procedures for dealing with nonnormality when using this model have also been developed

(Arnau et al., 2012, 2013; Livacic-Rojas et al., 2010). Another extension of the GLM is the generalized linear model (GLIM), which is useful for non-normally distributed data (e.g., binary, multinomial, ordinal, count, etc.) and which fits data with an exponential family distribution through the use of non-linear link functions. Finally, the GLMM offers a more flexible approach to the analysis of data with a nonnormal distribution (Bolker et al., 2009; Breslow & Clayton, 1993). The GLMM is a combination of the LMM, which includes random effects, and the GLIM, which uses different link functions that transform the expected value and the linear predictor to the same scale. In summary, each of the aforementioned models has different purposes and is suitable for specific types of data.

Empirical evidence shows that in the fields of health, education, and social science it is very common for data to be non-normally distributed (Arnau et al., 2013; Arnau et al., 2014a,b; Bauer & Sterba, 2011; Blanca et al., 2013; Bono et al., 2017; Lei & Lomax, 2005; Micceri, 1989), or for there to be non-independence of observations due to nested sampling or repeated measures (Bandera & Pérez, 2018; Casals et al., 2014; Coupé, 2018; Johnson et al., 2015; Thiele & Markusen, 2012). These situations, which are very frequent in applied research, occur when repeated measurements are obtained from the same subjects (Baayen et al., 2017; Dang

et al., 2008; Noh et al., 2012; Thiele & Markusen, 2012), as in the case of longitudinal studies (Bandera & Pérez, 2018; Cho & Goodwin, 2017; Fieberg et al., 2010; Koh et al., 2019; Mowen & Culhane, 2017), or when data are hierarchically structured (Casals et al., 2014; Elosua & De Boeck, 2020; Moscatelli et al., 2012; Mowen & Culhane, 2017). The fact that these situations are so common is a primary reason for using the GLMM in applied research; additionally, it is one of the alternatives to analysis of variance for repeated measures designs when the dependent variable is categorical. Use of the GLMM for multilevel models and repeated measures are further discussed in Stroup (2013).

With non-normal distributions the GLMM uses different link functions. The most frequently used link functions are logarithmic (the linear predictor is the natural logarithm of the expected value), logit (the linear predictor is the inverse of the logistic distribution function of the expected value), and probit (the linear predictor is the inverse of the standardized normal cumulative distribution function of the expected value). The most common estimation method in GLMM is the maximum likelihood (ML) procedure, especially in the form of variants such as restricted maximum likelihood (REML), restricted subject-specific pseudo-likelihood (RSPL), and penalized quasi-likelihood (PQL). Other estimation methods are the Gauss-Hermite quadrature (GHQ), the Laplace approximation (which is a particular case of the GHQ), hierarchical likelihood (HL), and Bayesian methods based on the Markov Chain Monte Carlo (MCMC) technique. The algorithm used by the GLMM allows calculation of the best linear unbiased estimator (BLUE) of the fixed effects and the best linear unbiased predictor (BLUP) of the random effects, which shows the lowest root mean square error among all the unbiased linear predictors (Bandera & Pérez, 2018; Searle et al., 1992).

These estimation methods use very complex calculation procedures that require increasingly sophisticated statistical software, such as proc glimmix in SAS, which generalizes two other procedures: proc genmod, by including error terms that are not normally distributed, and proc mixed, by considering random effects in the model (SAS Institute, 2013). Proc glimmix is a powerful technique for modeling correlated binary outcomes (Dang et al., 2008). By default, SAS uses the RSPL approach for parameter estimation (Wolfinger & O'Connell, 1993). The RSPL approach estimates random effects models based on residual probability, and it is well suited to situations where what is of interest is the individual response rather than the population mean or marginal model.

The incorporation of newer procedures into the main statistical software packages (SAS, R, STATA, and IBM SPSS) has encouraged the use of GLMMs in a wide variety of disciplines, including biology (Bandera & Pérez, 2018; Thiele & Markussen, 2012), ecology and environmental sciences (Bolker et al., 2009; Johnson et al., 2015; Kain et al., 2015; Smith et al., 2020), medicine (Brown & Prescott, 2006; Casals et al., 2014; Cnnan et al., 1998; Platt et al., 1999;

Skrondal & Rabe-Hesketh, 2003; Witte et al., 2000), psychophysics (Moscatelli & Lacquaniti, 2011; Moscatelli et al., 2011; Moscatelli et al., 2012), psycholinguistics (Baayen et al., 2008; Coupé, 2018; Elosua & De Boeck, 2020; Quené & van den Bergh, 2008), and psychology (Aiken et al., 2015; Bono et al., 2021; Cho & Goodwin, 2017), among others.

Focusing on the field of psychology, Bono et al. (2021) observed a growing trend in the use of these analytic models, with the number of GLMM-related articles published in 2018 ($n = 88$) being more than double the figure for 2014 ($n = 39$). Bono et al. (2021) also carried out a systematic review of 80 empirical articles indexed in JCR journals during 2018, and which included a total of 118 different GLMM analyses. The results showed that the most common application of GLMMs involved the use of repeated measures designs (61.1 %), with two and three repeated measures (33.33 % and 25 % of analyses, respectively), and a sample of fewer than 500 participants (57.6 %). Regarding the characteristics of response variables, the most frequent among the studies analyzed were variables with two response categories (87.7 %). The most common distribution used was the binomial distribution (19.5 %), although in the majority of studies the distribution was not specified (52.5 %). Regarding software, SAS was the most widely used (36.4 %), with glimmix being the most common procedure (32.6 %).

More generally, the GLMM has most frequently been applied in longitudinal studies or repeated measures designs with binary response variables (Bakbergenuly & Kulinskaya, 2018; Cho & Goodwin, 2017; Gawarammana & Sooriyarachchi, 2017; Zhang et al., 2011), with count response variables (Coupé, 2018; Huang et al., 2016; Kruppa & Hothorn, 2021; Sun et al., 2019; Zhang et al., 2012), with multinomial outcomes (Jiang & Oleson, 2011) or, to a lesser extent, with continuous response variables (Lo & Andrews, 2015). It is worth noting that binary repeated measures are a common occurrence in biology, medicine, psychology, and sociology, as well as in many other practical fields (Gawarammana & Sooriyarachchi, 2017). These repeated measures are often related to each other, and it is common practice to use a first-order autoregressive covariance matrix, AR(1), to describe this serial dependence. Although the GLMM can model serial dependence between repeated binary responses, its use has not been widespread. Cho et al. (2018) show examples in the literature with binary time series and illustrate the GLMM AR(1) using empirical data.

Although, as we have said, GLMMs are now increasingly used by researchers in many fields, the robustness of these models has not been studied to the same extent as is the case for the LMM (e.g., Arnau et al., 2012; Jacqmin-Gadda et al., 2007; Kowalchuk et al., 2004; Vallejo et al., 2008). Indeed, simulation studies on the robustness of GLMMs with repeated measures and binary data are still very scarce. For data of this kind, Fang and Louchin (2013) found that GLMMs hold Type I error rates better than does the GLM, while Gawarammana and Sooriyarachchi (2017) concluded that the

GLMM was satisfactory with respect to Type I error for sample sizes above 20.

A further issue to consider is that the results of simulation studies may be of little use to applied researchers in the psychological field, because the manipulated variables and their values (e.g., sample size, number of groups, number of repeated measures, correlation values, etc.) usually vary from one study to the next and are unlikely to have been chosen based on research practice; consequently, the results may not be directly applicable by researchers in this field. In addition, a standard criterion has not been used when assessing robustness, leading to different interpretations of the Type I error rate (Blanca et al., 2017, 2018). In this respect, the robustness of the GLMM has generally been interpreted in terms of how close or far the empirical alpha value obtained is from the nominal alpha value (Barker et al., 2017; Gawarammana & Sooriyarachchi, 2017; Zhang et al., 2012).

In this paper we report a simulation study conducted with the aim of providing empirical evidence on robustness (in terms of Type I error rates) of the GLMM in split-plot designs with binary response variables. To this end, we systematically manipulated a broad set of variables and range of values, based on research practice. The results of the study should help to inform the application of these models in the field of psychological or related research.

Method

A Monte Carlo simulation study was performed using the IML (interactive matrix language) module and proc glimmix of SAS 9.4 (SAS Institute, 2016). The aim was to analyze empirical Type I error rates of the GLMM in split-plot designs with binary response variables and a binomial distribution, involving two groups, two and three repeated measures, and a first-order autoregressive covariance matrix, AR(1).

Procedure

A series of macros was created that allowed generation of the data, estimation of the model, and the inference of fixed effects (adjustment by proc glimmix of SAS). These macros are available upon request from the corresponding author. Longitudinal binary data were generated with the *RandMVBinary* function (SAS/IML), developed by Wicklin (2013) and based on the algorithm of Emrich and Piedmonte (1991). This function generates multivariate binary variables with a given set of expected values and a specified correlation structure.

The following variables were manipulated:
1. Total sample size. Here we considered very small and small (N = 24, 36, 48, 60, 72, 84, and 96), moderate (N = 108, 156, 204, 252, and 300), and large (N = 348, 396, 444, and 492) sample sizes. Sample sizes less than 500 were used because they are very frequent in the field of psychology (Bono et al., 2021).

2. Coefficient of group size variation (Δn), which represents the amount of inequality in group sample size (Lix & Hinds, 2004). Following Blanca et al. (2018), different degrees of variation were considered and were grouped as null, low, medium, and high (0, .16, .33, and .50, respectively). For each N value, equal group sizes (balanced groups) and different group sizes (unbalanced groups) were considered. With balanced groups, the group sizes ranged from 12 to 246 (Tables 1-5), while with unbalanced groups the group sizes ranged from 10 to 369 (Tables 7-12).

3. Within-subject levels (K). The repeated measures were K = 2 and K = 3. According to Bono et al. (2021), these are the most commonly used in psychology (58.33 % of studies).

4. Correlation (ϕ_1) between repeated measures. The correlations with the AR(1) covariance structure were low (ϕ_1 = .2), medium (ϕ_1 = .4 and .6), and high (ϕ_1 = .8). The same correlation was established in each group. Following the simulation study of Bell and Grunwald (2011), correlations of ϕ_1 = .4 and .8 were used. We also considered low and medium correlations (ϕ_1 = .2 and ϕ_1 = .6).

5. Pairing of group size with the correlation between repeated measures (ϕ_1 = .2 and .8). A different correlation was established in each group. Pairing was null with balanced groups. With unbalanced groups, the pairing was positive when the largest group size was associated with the highest correlation and negative when the largest group size was associated with the lowest correlation.

Table 1
Empirical Type I error rates (in percentages) for the group effect in balanced groups with K = 2.

		Null pairing				
		Same correlation in both groups				Different correlation in the two groups
N	n_1, n_2	.2	.4	.6	.8	n_1 = .2; n_2 = .8
24	12, 12	4.76	3.72	3.36	3.46	2.88
36	18, 18	5.06	3.52	3.46	6.24	3.52
48	24, 24	4.86	4.14	4.44	3.72	3.40
60	30, 30	4.20	3.98	3.56	4.04	4.44
72	36, 36	5.78	3.82	4.04	5.04	4.26
84	42, 42	4.92	4.06	4.84	4.40	4.36
96	48, 48	4.36	4.64	4.26	4.94	4.22
108	54, 54	4.76	3.86	4.86	4.74	4.68
156	78, 78	4.44	4.64	4.10	4.38	3.84
204	102, 102	4.98	4.34	4.72	5.04	4.42
252	126, 126	4.94	4.42	4.98	5.02	4.50
300	150, 150	5.00	4.62	4.92	4.96	5.02
348	174, 174	5.28	4.74	4.74	4.92	4.78
396	198, 198	4.42	5.12	5.06	4.56	4.56
444	222, 222	4.82	4.74	5.00	4.90	5.20
492	246, 246	5.14	4.64	4.58	5.16	4.41

Table 2
Empirical Type I error rates (in percentages) for the time effect in balanced groups with K = 2.

			Null pairing			
		Same correlation in both groups				Different correlation in the two groups
N	n_1, n_2	.2	.4	.6	.8	$n_1 = .2; n_2 = .8$
24	12, 12	3.98	*1.82*	*0.28*	*0.04*	*1.24*
36	18, 18	3.88	*1.96*	*1.02*	*0.10*	*1.72*
48	24, 24	3.70	3.82	*2.14*	*0.04*	3.36
60	30, 30	4.62	3.48	*1.90*	*0.02*	2.84
72	36, 36	4.44	3.10	*1.78*	*0.00*	2.76
84	42, 42	3.92	3.06	*1.88*	*0.02*	3.52
96	48, 48	3.18	2.98	*1.52*	*0.06*	3.30
108	54, 54	3.06	3.32	*1.48*	*0.00*	3.12
156	78, 78	3.46	*2.36*	*0.96*	*0.00*	3.40
204	102, 102	3.10	*2.24*	*0.78*	*0.00*	2.90
252	126, 126	3.10	*2.22*	*0.50*	*0.02*	2.76
300	150, 150	3.00	*1.78*	*0.70*	*0.00*	2.98
348	174, 174	3.18	*1.86*	*0.44*	*0.00*	2.70
396	198, 198	3.02	*1.92*	*0.34*	*0.02*	3.22
444	222, 222	2.90	*1.60*	*0.60*	*0.00*	2.54
492	246, 246	2.64	*1.52*	*0.36*	*0.02*	2.60

Note. In italics = conservative.

Table 3
Empirical Type I error rates (in percentages) for the interaction effect in balanced groups with K = 2.

			Null pairing			
		Same correlation in both groups				Different correlation in the two groups
N	n_1, n_2	.2	.4	.6	.8	$n_1 = .2; n_2 = .8$
24	12, 12	3.94	1.80	*0.44*	*0.02*	*1.26*
36	18, 18	3.36	2.04	*1.10*	*0.12*	*1.60*
48	24, 24	3.80	3.84	*1.82*	*0.06*	3.58
60	30, 30	4.48	3.62	*1.86*	*0.00*	3.02
72	36, 36	3.74	3.10	*1.82*	*0.04*	2.82
84	42, 42	3.78	3.30	*1.76*	*0.02*	3.28
96	48, 48	3.48	2.86	*1.64*	*0.08*	3.42
108	54, 54	3.00	3.26	*1.52*	*0.00*	3.36
156	78, 78	2.78	*2.20*	*1.08*	*0.00*	3.32
204	102, 102	3.10	*2.22*	*0.90*	*0.04*	3.02
252	126, 126	3.36	*1.86*	*0.58*	*0.04*	2.50
300	150, 150	2.98	*1.72*	*0.50*	*0.06*	3.16
348	174, 174	3.22	*2.24*	*0.40*	*0.02*	2.74
396	198, 198	2.70	*2.10*	*0.44*	*0.02*	3.26
444	222, 222	2.84	*1.80*	*0.50*	*0.02*	2.51
492	246, 246	4.12	*1.70*	*0.54*	*0.00*	2.52

Note. In italics = conservative.

Table 4
Empirical Type I error rates (in percentages) for the group effect in balanced groups with K = 3.

			Null pairing			
		Same correlation in both groups				Different correlation in the two groups
N	n_1, n_2	.2	.4	.6	.8	$n_1 = .2; n_2 = .8$
24	12, 12	4.12	3.34	3.16	3.70	3.46
36	18, 18	4.38	3.68	4.10	3.86	4.46
48	24, 24	4.76	4.10	4.16	4.32	4.68
60	30, 30	4.48	4.32	4.44	4.68	4.32
72	36, 36	5.22	4.42	4.10	4.58	4.74
84	42, 42	4.08	4.44	4.42	4.22	5.10

			Null pairing			
		Same correlation in both groups				Different correlation in the two groups
N	n_1, n_2	.2	.4	.6	.8	$n_1 = .2; n_2 = .8$
96	48, 48	4.86	5.02	4.74	4.62	4.40
108	54, 54	5.28	5.36	4.26	4.38	4.80
156	78, 78	4.26	4.66	5.12	4.08	4.68
204	102, 102	4.96	4.60	4.20	4.62	4.66
252	126, 126	5.20	4.86	5.08	4.98	4.72
300	150, 150	5.04	4.98	5.06	4.00	4.78
348	174, 174	4.74	5.24	4.82	4.66	5.54
396	198, 198	4.48	4.66	4.46	5.14	5.08
444	222, 222	4.86	4.20	5.10	4.42	5.12
492	246, 246	5.02	4.78	4.58	4.96	5.18

Table 5
Empirical Type I error rates (in percentages) for the time effect in balanced groups with K = 3.

			Null pairing			
		Same correlation in both groups				Different correlation in the two groups
N	n_1, n_2	.2	.4	.6	.8	$n_1 = .2; n_2 = .8$
24	12, 12	3.50	2.92	*2.38*	*1.52*	3.98
36	18, 18	3.61	2.90	*2.02*	3.32	3.54
48	24, 24	3.96	2.54	*1.86*	4.90	2.94
60	30, 30	3.56	2.50	*1.70*	3.54	2.64
72	36, 36	3.16	*2.36*	*1.72*	3.56	*2.10*
84	42, 42	4.00	*1.86*	*1.70*	3.14	*1.54*
96	48, 48	3.66	*2.30*	*1.16*	3.24	*1.36*
108	54, 54	3.32	*2.30*	*1.38*	3.88	*1.38*
156	78, 78	3.60	*2.21*	*1.14*	3.54	*1.04*
204	102, 102	3.38	*2.06*	*1.28*	4.42	*1.16*
252	126, 126	3.36	*2.32*	*0.84*	3.70	*1.08*
300	150, 150	3.76	*1.98*	*1.36*	3.84	*1.10*
348	174, 174	3.92	*2.36*	*1.14*	3.56	*0.94*
396	198, 198	3.38	*2.48*	*1.44*	4.00	*0.90*
444	222, 222	3.49	*2.18*	*1.32*	4.10	*1.00*
492	246, 246	3.32	*2.24*	*1.32*	3.84	*1.14*

Note. In italics = conservative.

Table 6
Empirical Type I error rates (in percentages) for the interaction effect in balanced groups with K = 3.

			Null pairing			
		Same correlation in both groups				Different correlation in the two groups
N	n_1, n_2	.2	.4	.6	.8	$n_1 = .2; n_2 = .8$
24	12, 12	4.18	2.98	*2.40*	*1.60*	4.28
36	18, 18	4.26	3.02	*2.32*	3.20	3.52
48	24, 24	3.90	*2.24*	*1.88*	5.30	2.90
60	30, 30	3.20	*2.16*	*1.58*	3.66	2.64
72	36, 36	3.42	*2.32*	*1.86*	3.62	*2.24*
84	42, 42	3.96	*1.92*	*1.32*	3.18	*1.40*
96	48, 48	3.78	*2.48*	*1.28*	3.24	*1.50*
108	54, 54	3.18	*2.04*	*1.32*	3.72	*1.24*
156	78, 78	3.30	*2.36*	*1.26*	3.58	*1.26*
204	102, 102	3.30	*2.28*	*0.88*	4.20	*1.16*
252	126, 126	3.96	*2.12*	*1.32*	3.68	*1.10*
300	150, 150	3.60	*2.32*	*0.84*	3.88	*1.06*
348	174, 174	3.34	*2.22*	*1.16*	3.52	*1.00*
396	198, 198	3.34	*2.44*	*1.26*	4.20	*1.06*
444	222, 222	3.81	*2.20*	*0.92*	3.86	*1.04*
492	246, 246	3.38	*2.38*	*1.50*	4.22	*1.10*

Note. In italics = conservative.

Table 7
Empirical Type I error rates (in percentages) for the group effect in unbalanced groups with K = 2.

N	Δn	n₁, n₂	Null pairing .2	.4	.6	.8	Positive pairing	Negative pairing
	0.16	10, 14	4.88	3.72	2.78	2.98	*2.44*	3.52
24	0.33	8, 16	3.90	*2.42*	*2.42*	*2.08*	*1.38*	3.42
	0.50	6, 18	2.88	*1.78*	*1.16*	*0.92*	*0.86*	2.52
	0.16	15, 21	4.38	3.68	3.28	3.88	3.26	3.66
36	0.33	12, 24	4.28	3.60	3.64	3.70	*2.04*	5.24
	0.50	9, 27	3.90	3.36	2.84	2.72	*1.86*	5.38
	0.16	20, 28	4.52	3.24	3.56	3.74	3.02	4.40
48	0.33	16, 32	4.74	3.62	3.80	4.18	2.84	5.12
	0.50	12, 36	5.02	4.20	4.04	4.00	*2.32*	6.52
	0.16	25, 35	4.94	4.24	4.22	4.42	3.60	4.34
60	0.33	20, 40	5.40	3.98	4.58	4.74	3.02	5.52
	0.50	15, 45	5.02	4.54	3.36	3.86	2.66	6.64
	0.16	30, 42	5.56	4.32	3.70	4.56	3.50	5.30
72	0.33	24, 48	4.72	4.36	4.64	4.74	2.82	6.20
	0.50	18, 54	5.36	4.28	4.46	4.18	2.50	6.66
	0.16	35, 49	4.50	4.24	4.22	4.56	3.60	5.52
84	0.33	28, 56	5.26	4.58	4.18	4.78	3.02	6.26
	0.50	21, 63	4.28	4.44	4.56	4.28	2.56	6.54
	0.16	40, 56	5.02	4.12	4.88	4.40	3.82	5.46
96	0.33	32, 64	5.08	4.70	4.42	4.54	3.12	6.58
	0.50	24, 72	4.80	4.46	4.76	4.68	2.54	7.00
	0.16	45, 63	5.10	4.50	4.62	4.50	3.42	5.04
108	0.33	36, 72	4.80	4.63	4.32	4.68	3.22	5.62
	0.50	27, 81	4.90	4.32	4.64	4.46	2.68	7.24
	0.16	65, 91	4.64	4.20	4.68	4.78	3.60	5.46
156	0.33	52, 104	4.82	4.68	4.78	4.18	3.26	6.64
	0.50	39, 117	5.22	4.86	4.62	4.36	2.92	7.50
	0.16	85, 119	5.20	4.74	4.56	4.70	3.38	5.50
204	0.33	68, 136	4.84	4.58	4.34	5.04	3.40	6.74
	0.50	51, 153	5.70	5.28	5.20	5.20	2.88	**7.74**
	0.16	105, 147	5.24	4.54	4.18	5.22	4.12	5.84
252	0.33	84, 168	5.06	4.42	4.94	5.14	3.50	7.20
	0.50	63, 189	5.28	5.02	5.00	4.44	2.60	**7.58**
	0.16	125, 175	5.02	4.86	4.70	4.32	4.16	6.00
300	0.33	100, 200	4.84	5.10	4.82	5.08	3.58	6.62
	0.50	75, 225	4.94	5.42	4.58	4.36	2.66	**7.68**
	0.16	145, 203	5.02	4.86	4.82	5.04	4.38	5.60
348	0.33	116, 232	5.28	4.36	4.98	4.94	3.52	6.32
	0.50	87, 261	4.90	4.40	4.88	5.06	2.54	**8.28**
	0.16	165, 231	4.84	4.90	4.86	4.96	4.06	5.46
396	0.33	132, 264	5.04	5.30	4.70	4.48	3.88	6.72
	0.50	99, 297	4.66	5.26	5.24	4.98	2.86	**7.54**
	0.16	185, 259	4.70	4.64	4.70	4.70	3.94	6.32
444	0.33	148, 296	4.92	5.08	5.12	4.70	3.64	6.58
	0.50	111, 333	4.90	5.46	4.76	4.44	2.78	**7.68**
	0.16	205, 287	4.76	4.50	5.06	5.62	4.02	5.98
492	0.33	164, 328	4.82	4.84	4.70	5.16	3.84	6.12
	0.50	123, 369	4.87	5.18	5.30	5.88	2.92	**8.58**

Note. Δn = coefficient of group size variation; null pairing: same correlation in the two groups; positive pairing: correlation .2 for n₁ and correlation .8 for n₂; negative pairing: correlation .8 for n₁ and correlation .2 for n₂; in italics = conservative; in bold = liberal.

Table 8
Empirical Type I error rates (in percentages) for the time effect in unbalanced groups with K = 2.

N	Δn	n₁, n₂	Null pairing .2	.4	.6	.8	Positive pairing	Negative pairing
	0.16	10, 14	4.66	*2.02*	*0.38*	*0.04*	*1.10*	*1.02*
24	0.33	8, 16	3.58	*1.72*	*0.76*	*0.00*	*1.24*	*0.82*
	0.50	6, 18	2.90	*1.22*	*0.20*	*0.02*	*0.58*	*0.70*
	0.16	15, 21	3.86	*2.20*	*1.32*	*0.04*	2.52	*1.58*
36	0.33	12, 24	3.46	*2.20*	*0.88*	*0.08*	2.46	*1.06*
	0.50	9, 27	3.76	*1.72*	*0.98*	*0.06*	2.12	*0.84*
	0.16	20, 28	4.18	3.50	*2.02*	*0.04*	2.94	3.08
48	0.33	16, 32	4.72	3.32	*2.38*	*0.04*	2.80	*2.22*
	0.50	12, 36	4.76	3.16	*2.06*	*0.04*	3.16	*1.90*
	0.16	25, 35	4.38	3.73	*2.02*	*0.04*	3.32	2.70
60	0.33	20, 40	4.02	3.86	*1.94*	*0.04*	2.92	2.68
	0.50	15, 45	4.00	3.16	*2.40*	*0.08*	2.70	*1.40*
	0.16	30, 42	3.92	3.36	*1.46*	*0.00*	3.32	2.30
72	0.33	24, 48	3.88	2.92	*1.38*	*0.02*	*2.04*	*1.80*
	0.50	18, 54	3.44	2.82	*1.66*	*0.02*	2.34	*1.14*
	0.16	35, 49	3.64	3.02	*1.88*	*0.02*	3.54	2.98
84	0.33	28, 56	3.50	3.68	*1.90*	*0.00*	2.80	*1.74*
	0.50	21, 63	3.18	3.70	*1.78*	*0.04*	2.14	*0.96*
	0.16	40, 56	3.78	3.08	*1.32*	*0.00*	3.50	2.46
96	0.33	32, 64	3.28	3.06	*1.32*	*0.02*	2.62	*1.62*
	0.50	24, 72	3.20	3.18	*1.18*	*0.02*	2.32	*1.02*
	0.16	45, 63	3.38	2.66	*1.34*	*0.00*	2.96	2.42
108	0.33	36, 72	2.82	2.82	*1.22*	*0.00*	2.72	*1.82*
	0.50	27, 81	3.08	2.70	*1.28*	*0.02*	2.34	*0.90*
	0.16	65, 91	2.96	*2.18*	*0.90*	*0.00*	2.70	2.50
156	0.33	52, 104	3.14	*2.48*	*0.68*	*0.00*	*1.68*	*0.92*
	0.50	39, 117	3.04	*2.32*	*0.66*	*0.00*	*1.76*	*0.36*
	0.16	85, 119	2.98	*2.22*	*0.72*	*0.04*	3.14	*2.16*
204	0.33	68, 136	3.22	*2.12*	*0.82*	*0.00*	*1.84*	*1.00*
	0.50	51, 153	3.08	*2.10*	*0.90*	*0.00*	*1.92*	*0.28*
	0.16	105, 147	3.00	*2.22*	*0.72*	*0.02*	*1.90*	*1.56*
252	0.33	84, 168	2.90	*1.84*	*0.68*	*0.02*	*1.52*	*0.80*
	0.50	63, 189	3.24	*1.84*	*0.54*	*0.02*	*1.76*	*0.16*
	0.16	125, 175	3.14	*1.82*	*0.66*	*0.02*	*1.98*	*1.76*
300	0.33	100, 200	3.76	*2.14*	*0.44*	*0.00*	*1.80*	*0.52*
	0.50	75, 225	2.80	*1.78*	*0.58*	*0.02*	*1.76*	*0.14*
	0.16	145, 203	2.92	*1.90*	*0.60*	*0.04*	2.16	*1.88*
348	0.33	116, 232	3.06	*1.80*	*0.32*	*0.02*	*1.54*	*0.46*
	0.50	87, 261	3.10	*1.62*	*0.46*	*0.02*	*1.92*	*0.14*
	0.16	165, 231	2.82	*1.62*	*0.58*	*0.00*	*1.90*	*1.54*
396	0.33	132, 264	3.64	*1.78*	*0.52*	*0.02*	*1.78*	*0.44*
	0.50	99, 297	2.74	*1.78*	*0.52*	*0.00*	*1.76*	*0.18*
	0.16	185, 259	3.26	*1.58*	*0.40*	*0.04*	*1.94*	*1.26*
444	0.33	148, 296	2.96	*1.64*	*0.38*	*0.06*	*1.56*	*0.20*
	0.50	111, 333	3.32	*1.82*	*0.56*	*0.02*	*1.88*	*0.24*
	0.16	205, 287	2.86	*1.50*	*0.44*	*0.04*	*1.96*	*1.22*
492	0.33	164, 328	3.00	*1.20*	*0.40*	*0.02*	*1.62*	*0.26*
	0.50	123, 369	3.08	*1.58*	*0.42*	*0.04*	*1.90*	*0.16*

Note. Δn = coefficient of group size variation; null pairing: same correlation in the two groups; positive pairing: correlation .2 for n₁ and correlation .8 for n₂; negative pairing: correlation .8 for n₁ and correlation .2 for n₂; in italics = conservative.

Table 9
Empirical Type I error rates (in percentages) for the interaction effect in unbalanced groups with K = 2.

N	Δn	n₁, n₂	Null pairing .2	.4	.6	.8	Positive pairing	Negative pairing
24	0.16	10, 14	4.42	*2.12*	*0.56*	*0.06*	*1.38*	*0.90*
	0.33	8, 16	3.68	*1.66*	*0.56*	*0.00*	*1.24*	*0.72*
	0.50	6, 18	2.90	*0.88*	*0.22*	*0.00*	*0.72*	*0.76*
36	0.16	15, 21	3.82	*1.86*	*0.12*	*0.06*	*2.08*	*1.64*
	0.33	12, 24	3.58	*1.98*	*0.80*	*0.08*	*2.22*	*1.08*
	0.50	9, 27	3.34	*1.88*	*0.82*	*0.08*	*1.82*	*0.94*
48	0.16	20, 28	4.02	3.20	*2.16*	*0.04*	2.88	2.84
	0.33	16, 32	4.70	3.28	*2.24*	*0.00*	2.82	*2.24*
	0.50	12, 36	4.02	3.36	*2.12*	*0.06*	2.90	*1.88*
60	0.16	25, 35	4.26	3.70	*2.00*	*0.04*	3.46	2.62
	0.33	20, 40	4.10	3.94	*1.98*	*0.02*	3.26	*2.48*
	0.50	15, 45	4.32	3.32	*2.32*	*0.00*	2.72	*1.56*
72	0.16	30, 42	3.86	3.34	*1.40*	*0.02*	3.22	2.74
	0.33	24, 48	4.44	2.84	*1.40*	*0.00*	*2.44*	*1.84*
	0.50	18, 54	3.54	2.98	*1.62*	*0.02*	*2.44*	*1.02*
84	0.16	35, 49	3.56	3.22	*1.94*	*0.00*	3.50	3.12
	0.33	28, 56	3.56	3.50	*1.98*	*0.00*	2.60	*1.84*
	0.50	21, 63	3.48	3.24	*1.70*	*0.02*	*2.42*	*0.90*
96	0.16	40, 56	3.50	2.90	*1.46*	*0.06*	3.52	2.60
	0.33	32, 64	3.04	3.04	*1.22*	*0.04*	*2.42*	*1.68*
	0.50	24, 72	3.68	3.08	*1.22*	*0.02*	2.56	*1.12*
108	0.16	45, 63	3.10	3.04	*1.56*	*0.06*	3.30	2.60
	0.33	36, 72	2.70	2.70	*1.22*	*0.02*	2.84	*1.84*
	0.50	27, 81	3.28	2.70	*1.16*	*0.02*	*2.22*	*0.92*
156	0.16	65, 91	2.80	2.68	*1.04*	*0.00*	2.62	*2.40*
	0.33	52, 104	2.70	*2.40*	*0.70*	*0.06*	*1.48*	*1.12*
	0.50	39, 117	2.94	*2.30*	*0.72*	*0.06*	*2.16*	*0.40*
204	0.16	85, 119	3.28	*2.24*	*0.82*	*0.00*	2.96	*2.22*
	0.33	68, 136	3.06	*2.26*	*0.84*	*0.02*	*1.84*	*1.06*
	0.50	51, 153	3.08	*1.74*	*0.88*	*0.06*	*1.98*	*0.22*
252	0.16	105, 147	3.24	*2.14*	*0.64*	*0.02*	*2.10*	*1.56*
	0.33	84, 168	2.86	*2.16*	*0.76*	*0.08*	*1.44*	*0.72*
	0.50	63, 189	3.56	*2.12*	*0.54*	*0.02*	*1.76*	*0.12*
300	0.16	125, 175	2.90	*2.00*	*0.70*	*0.00*	*2.04*	*1.58*
	0.33	100, 200	3.38	*1.72*	*0.56*	*0.02*	*1.72*	*0.50*
	0.50	75, 225	2.88	*1.74*	*0.66*	*0.02*	*1.64*	*0.20*
348	0.16	145, 203	2.98	*2.12*	*0.52*	*0.04*	*1.82*	*1.88*
	0.33	116, 232	2.86	*1.78*	*0.56*	*0.02*	*1.52*	*0.48*
	0.50	87, 261	3.68	*1.70*	*0.38*	*0.06*	*1.72*	*0.16*
396	0.16	165, 231	2.88	*1.70*	*0.54*	*0.04*	*1.76*	*1.70*
	0.33	132, 264	3.44	*1.96*	*0.60*	*0.00*	*1.38*	*0.52*
	0.50	99, 297	2.74	*1.96*	*0.38*	*0.00*	*1.40*	*0.20*
444	0.16	185, 259	3.12	*1.62*	*0.38*	*0.02*	*2.06*	*1.42*
	0.33	148, 296	2.82	*1.70*	*0.54*	*0.02*	*1.62*	*0.38*
	0.50	111, 333	3.00	*1.86*	*0.50*	*0.02*	*1.74*	*0.14*
492	0.16	205, 287	2.58	*1.32*	*0.40*	*0.00*	*1.92*	*1.38*
	0.33	164, 328	3.18	*1.78*	*0.40*	*0.00*	*1.74*	*0.32*
	0.50	123, 369	3.12	*1.74*	*0.48*	*0.06*	*1.86*	*0.12*

Note. Δn = coefficient of group size variation; null pairing: same correlation in the two groups; positive pairing: correlation .2 for n₁ and correlation .8 for n₂; negative pairing: correlation .8 for n₁ and correlation .2 for n₂; in italics = conservative.

Table 10
Empirical Type I error rates (in percentages) for the group effect in unbalanced groups with K = 3.

N	Δn	n₁, n₂	Null pairing .2	.4	.6	.8	Positive pairing	Negative pairing
24	0.16	10, 14	3.64	2.96	2.86	2.86	*2.32*	4.24
	0.33	8, 16	3.74	3.72	*2.32*	*2.40*	*1.70*	5.08
	0.50	6, 18	4.24	*2.02*	*1.24*	*1.28*	*0.52*	4.38
36	0.16	15, 21	4.38	4.34	4.26	4.02	3.14	4.42
	0.33	12, 24	4.26	3.94	4.22	4.42	*2.20*	7.02
	0.50	9, 27	4.14	4.00	3.44	3.60	*1.30*	**7.92**
48	0.16	20, 28	4.48	4.40	3.92	4.40	2.88	5.66
	0.33	16, 32	4.58	4.48	3.98	4.46	2.52	7.02
	0.50	12, 36	4.08	4.80	4.14	4.32	*1.82*	**9.00**
60	0.16	25, 35	4.88	4.44	4.44	4.44	2.88	5.72
	0.33	20, 40	5.00	4.76	4.68	4.20	2.84	**7.98**
	0.50	15, 45	4.62	4.62	3.88	4.78	*1.58*	**8.62**
72	0.16	30, 42	4.34	4.84	4.56	4.80	3.36	6.02
	0.33	24, 48	5.12	4.28	4.86	4.76	*2.34*	**7.82**
	0.50	18, 54	4.94	4.56	4.14	4.30	*1.64*	**8.64**
84	0.16	35, 49	4.36	4.60	4.32	4.76	3.26	6.34
	0.33	28, 56	4.76	5.08	4.28	4.52	2.88	**7.80**
	0.50	21, 63	4.92	4.32	4.42	4.32	*1.76*	**8.96**
96	0.16	40, 56	4.06	4.72	4.24	4.78	3.46	5.62
	0.33	32, 64	4.82	4.66	4.94	4.28	*2.48*	7.14
	0.50	24, 72	5.02	4.94	3.84	4.68	*1.80*	**9.52**
108	0.16	45, 63	4.14	4.52	4.88	4.64	3.80	5.86
	0.33	36, 72	4.70	4.70	4.62	4.96	2.70	**7.72**
	0.50	27, 81	4.74	4.76	4.70	5.38	*1.22*	**9.90**
156	0.16	65, 91	4.66	4.88	5.28	4.60	3.72	6.34
	0.33	52, 104	4.96	4.60	4.30	4.20	2.58	**8.08**
	0.50	39, 117	5.36	4.64	4.62	5.36	*1.84*	**10.84**
204	0.16	85, 119	4.94	4.86	5.22	4.90	3.78	6.50
	0.33	68, 136	5.48	5.62	5.14	4.58	2.98	**8.06**
	0.50	51, 153	4.54	4.74	4.96	5.06	*1.84*	**10.06**
252	0.16	105, 147	4.68	4.66	4.64	3.86	3.38	6.20
	0.33	84, 168	5.10	4.90	4.96	4.78	3.08	**7.78**
	0.50	63, 189	5.34	4.52	4.54	4.54	*1.98*	**9.78**
300	0.16	125, 175	4.76	5.22	4.66	4.12	4.04	6.54
	0.33	100, 200	5.22	5.06	4.72	4.84	3.02	**8.08**
	0.50	75, 225	5.82	4.76	4.74	4.40	*2.12*	**10.36**
348	0.16	145, 203	5.06	5.02	4.42	4.52	3.78	7.20
	0.33	116, 232	5.00	5.10	5.26	4.64	2.58	**8.49**
	0.50	87, 261	5.18	4.76	5.16	4.58	*1.54*	**9.76**
396	0.16	165, 231	5.18	4.58	5.06	5.12	3.16	**8.28**
	0.33	132, 264	5.08	5.18	5.06	4.30	2.84	**8.22**
	0.50	99, 297	5.24	4.82	5.26	4.64	*2.20*	**9.92**
444	0.16	185, 259	5.06	5.22	4.92	4.98	3.26	**8.18**
	0.33	148, 296	4.94	4.58	4.88	4.90	2.86	**8.98**
	0.50	111, 333	4.84	5.30	5.10	5.22	*1.66*	**9.32**
492	0.16	205, 287	5.08	5.10	5.18	4.86	3.18	**7.96**
	0.33	164, 328	5.70	4.68	4.62	4.96	2.54	**8.54**
	0.50	123, 369	4.98	5.00	4.82	5.10	*1.46*	**9.14**

Note. Δn = coefficient of group size variation; null pairing: same correlation in the two groups; positive pairing: correlation .2 for n₁ and correlation .8 for n₂; negative pairing: correlation .8 for n₁ and correlation .2 for n₂; in italics = conservative; in bold = liberal.

Table 11

Empirical Type I error rates (in percentages) for the time effect in unbalanced groups with K = 3.

N	Δn	n₁, n₂	Null pairing .2	.4	.6	.8	Positive pairing	Negative pairing
	0.16	10, 14	3.64	3.22	*2.16*	*1.32*	4.44	3.14
24	0.33	8, 16	3.22	2.64	*2.10*	*1.18*	4.32	*2.40*
	0.50	6, 18	3.52	*1.84*	*1.62*	*1.10*	4.54	*2.12*
	0.16	15, 21	3.44	2.58	*2.34*	3.52	4.66	*1.42*
36	0.33	12, 24	3.72	*2.28*	*2.32*	4.08	5.60	*1.68*
	0.50	9, 27	2.56	*2.26*	*2.12*	3.18	5.86	*1.22*
	0.16	20, 28	3.70	2.52	*2.04*	4.38	3.48	*1.76*
48	0.33	16, 32	3.72	*2.48*	*2.08*	3.78	5.54	*1.18*
	0.50	12, 36	3.36	*2.30*	*1.66*	4.54	6.52	*0.78*
	0.16	25, 35	4.20	*2.24*	*1.92*	3.00	3.26	*1.48*
60	0.33	20, 40	3.80	*2.20*	*1.72*	2.86	4.32	*1.10*
	0.50	15, 45	3.26	*2.40*	*1.74*	3.18	5.80	*0.68*
	0.16	30, 42	4.14	*2.18*	*1.52*	3.48	2.64	*1.18*
72	0.33	24, 48	3.56	*2.44*	*1.22*	3.62	4.42	*0.68*
	0.50	18, 54	3.48	*2.42*	*1.28*	3.94	4.74	*0.50*
	0.16	35, 49	3.22	*2.24*	*1.38*	3.92	2.68	*1.08*
84	0.33	28, 56	3.76	*2.40*	*1.38*	3.42	3.98	*0.44*
	0.50	21, 63	3.68	*2.20*	*1.28*	3.18	5.92	*0.38*
	0.16	40, 56	3.64	2.52	*0.96*	4.00	*2.48*	*0.84*
96	0.33	32, 64	3.62	2.82	*1.34*	4.14	3.54	*0.40*
	0.50	24, 72	3.16	*2.32*	*1.16*	3.30	4.60	*0.38*
	0.16	45, 63	3.20	*2.02*	*1.14*	4.08	*1.76*	*0.88*
108	0.33	36, 72	3.38	*2.12*	*1.40*	3.60	2.96	*0.32*
	0.50	27, 81	3.26	*2.06*	*1.00*	3.76	4.56	*0.38*
	0.16	65, 91	3.66	*2.10*	*1.28*	3.40	*1.70*	*0.80*
156	0.33	52, 104	3.52	*2.46*	*1.20*	3.68	2.80	*0.38*
	0.50	39, 117	2.82	*2.14*	*1.06*	3.70	4.18	*0.20*
	0.16	85, 119	3.62	*1.94*	*1.38*	3.92	*1.54*	*0.64*
204	0.33	68, 136	3.38	*2.32*	*1.18*	4.12	*2.40*	*0.44*
	0.50	51, 153	3.70	*2.16*	*1.02*	4.24	4.00	*0.30*
	0.16	105, 147	3.16	*2.28*	*1.06*	4.00	*1.70*	*1.00*
252	0.33	84, 168	3.06	*1.86*	*1.34*	3.56	2.56	*0.44*
	0.50	63, 189	3.32	*2.44*	*1.48*	4.06	2.94	*0.18*
	0.16	125, 175	3.64	*1.52*	*0.92*	4.24	*1.58*	*0.84*
300	0.33	100, 200	3.54	*2.20*	*0.96*	4.04	*2.30*	*0.44*
	0.50	75, 225	3.62	*2.20*	*1.20*	3.40	3.68	*0.40*
	0.16	145, 203	4.26	*2.06*	*1.44*	3.58	*1.76*	*0.76*
348	0.33	116, 232	3.44	*2.00*	*1.20*	3.42	*2.20*	*0.42*
	0.50	87, 261	3.62	*2.32*	*1.08*	3.26	2.96	*0.20*
	0.16	165, 231	3.42	2.64	*1.30*	3.32	*1.46*	*0.28*
396	0.33	132, 264	3.38	*2.06*	*1.10*	3.14	*2.48*	*0.28*
	0.50	99, 297	3.76	*2.02*	*1.38*	3.68	2.78	*0.22*
	0.16	185, 259	3.36	*1.96*	*1.24*	3.26	*1.78*	*0.36*
444	0.33	148, 296	3.64	*2.14*	*0.90*	3.24	2.58	*0.26*
	0.50	111, 333	3.54	*2.16*	*1.34*	4.68	2.98	*0.18*
	0.16	205, 287	3.54	*2.12*	*0.98*	3.22	*1.84*	*0.41*
492	0.33	164, 328	3.56	*2.20*	*1.08*	3.14	2.76	*0.36*
	0.50	123, 369	4.10	*2.06*	*1.16*	4.10	2.82	*0.20*

Note. Δn = coefficient of group size variation; null pairing: same correlation in the two groups; positive pairing: correlation .2 for n₁ and correlation .8 for n₂; negative pairing: correlation .8 for n₁ and correlation .2 for n₂; in italics = conservative.

Table 12

Empirical Type I error rates (in percentages) for the interaction effect in unbalanced groups with K = 3.

N	Δn	n₁, n₂	Null pairing .2	.4	.6	.8	Positive pairing	Negative pairing
	0.16	10, 14	3.50	3.22	*2.26*	*1.34*	4.42	3.30
24	0.33	8, 16	3.24	*2.46*	*2.02*	*1.30*	4.28	*2.34*
	0.50	6, 18	3.70	*1.66*	*1.62*	*1.04*	4.82	*2.06*
	0.16	15, 21	3.86	3.02	2.60	3.48	4.84	*1.32*
36	0.33	12, 24	3.66	2.50	*2.46*	4.18	6.02	*1.70*
	0.50	9, 27	2.84	*2.18*	*2.22*	3.14	6.12	*1.26*
	0.16	20, 28	3.96	*2.24*	*2.30*	4.50	3.84	*1.92*
48	0.33	16, 32	3.36	2.50	*2.12*	3.80	5.70	*1.48*
	0.50	12, 36	3.42	*1.96*	*1.62*	4.42	6.78	*0.70*
	0.16	25, 35	4.02	*2.28*	*1.82*	3.24	2.98	*1.56*
60	0.33	20, 40	3.94	*2.24*	*1.84*	2.93	3.88	*1.00*
	0.50	15, 45	3.46	*2.30*	*1.68*	3.36	6.06	*0.60*
	0.16	30, 42	3.78	*1.98*	*1.56*	3.60	2.80	*1.20*
72	0.33	24, 48	3.32	*1.94*	*1.32*	3.86	4.96	*0.76*
	0.50	18, 54	3.12	*2.16*	*1.50*	3.90	5.88	*0.66*
	0.16	35, 49	3.66	*2.16*	*1.56*	3.84	2.50	*1.16*
84	0.33	28, 56	3.72	*2.20*	*1.16*	3.10	4.04	*0.42*
	0.50	21, 63	3.44	*1.70*	*1.34*	3.22	5.42	*0.54*
	0.16	40, 56	3.58	*2.46*	*1.08*	3.70	*2.44*	*1.02*
96	0.33	32, 64	3.76	*2.16*	*1.46*	3.84	3.34	*0.62*
	0.50	24, 72	3.10	*2.10*	*1.18*	3.30	4.84	*0.36*
	0.16	45, 63	3.64	*2.10*	*1.40*	3.82	*2.20*	*0.94*
108	0.33	36, 72	3.24	*2.14*	*1.30*	3.94	3.16	*0.50*
	0.50	27, 81	3.76	*2.42*	*1.30*	3.78	4.62	*0.46*
	0.16	65, 91	3.10	*2.13*	*1.16*	3.26	*2.02*	*0.64*
156	0.33	52, 104	3.20	*2.26*	*1.32*	3.56	2.68	*0.40*
	0.50	39, 117	3.16	*1.90*	*1.22*	3.82	4.24	*0.24*
	0.16	85, 119	3.28	*2.26*	*1.38*	4.16	*1.44*	*0.78*
204	0.33	68, 136	3.12	*2.12*	*1.10*	3.76	*2.14*	*0.28*
	0.50	51, 153	3.32	*2.30*	*0.90*	4.04	3.30	*0.30*
	0.16	105, 147	3.28	*2.44*	*1.44*	4.10	*1.44*	*0.64*
252	0.33	84, 168	3.32	*1.92*	*1.28*	3.64	2.66	*0.56*
	0.50	63, 189	3.08	*2.26*	*1.10*	4.04	3.28	*0.14*
	0.16	125, 175	3.68	*2.00*	*1.36*	4.54	*1.66*	*0.60*
300	0.33	100, 200	3.56	*2.20*	*1.28*	4.06	*2.18*	*0.42*
	0.50	75, 225	3.92	*2.20*	*1.14*	3.26	3.36	*0.18*
	0.16	145, 203	3.52	*2.32*	*1.12*	3.50	*1.34*	*0.74*
348	0.33	116, 232	3.52	*2.26*	*1.00*	3.46	*2.26*	*0.38*
	0.50	87, 261	3.56	*2.20*	*1.02*	3.50	3.32	*0.24*
	0.16	165, 231	3.38	*2.40*	*1.12*	3.34	*1.34*	*0.40*
396	0.33	132, 264	3.30	*2.06*	*1.08*	3.06	*2.28*	*0.54*
	0.50	99, 297	3.72	*1.88*	*1.20*	4.00	2.58	*0.10*
	0.16	185, 259	3.20	*2.20*	*1.24*	3.22	*1.96*	*0.44*
444	0.33	148, 296	3.72	*2.20*	*0.92*	3.50	*2.14*	*0.38*
	0.50	111, 333	3.68	*2.02*	*1.00*	4.12	*2.04*	*0.16*
	0.16	205, 287	3.30	*2.08*	*1.34*	3.18	*1.92*	*0.50*
492	0.33	164, 328	3.06	*2.16*	*1.08*	3.00	*2.04*	*0.36*
	0.50	123, 369	4.46	*1.98*	*1.30*	4.08	*2.21*	*0.10*

Note. Δn = coefficient of group size variation; null pairing: same correlation in the two groups; positive pairing: correlation .2 for n₁ and correlation .8 for n₂; negative pairing: correlation .8 for n₁ and correlation .2 for n₂; in italics = conservative.

Five thousand replications of each combination of the above variables were performed at a significance level of .05.

Once the data had been generated, the GLMM was applied using proc glimmix. The classification variables were subject, group, and time. The fixed effects of the model were group, time, and interaction (group x time). The RSPL approach was used because its focus is the individual response. A binomial distribution was fitted with the logit link function, which is widely used for binary responses (Malik et al., 2020). For the adjustment of degrees of freedom, we used the Kenward-Roger approximation (Kenward & Roger, 2009), specifically the improved (KR2) method in SAS. This approach uses a less biased precision estimator and is recommended when analyzing repeated measures data (Stroup et al., 2018). This correction further reduces the precision estimator bias for the fixed and random effects under nonlinear covariance structures. Finally, the random effects were intercept and the slope over time. For the time variable, the AR(1) covariance structure was fitted. This correlation structure is close to an autoregressive structure, which is usually expected in repeated measures data (Gawarammana & Sooriyarachchi, 2017; Dang et al., 2008; Liu et al., 2012).

Data analysis

Robustness was evaluated by means of Bradley's (1978) criterion, according to which a test is considered robust when the empirical Type I error rate is between .025 and .075 for a nominal alpha level of .05. When the empirical Type I error rate is above the upper limit, the test is considered liberal, and when it is below the lower limit it is considered conservative. The effects of the group factor, time factor, and interaction factor (group x time) were analyzed for each combination of manipulated variables.

Results

Balanced groups

In balanced groups, with $K = 2$ and $K = 3$, the group effect was robust in all conditions (Tables 1 and 4), including when the groups had different correlation values between repeated measures (null pairing between correlation and group size).

For $K = 2$, time and interaction effects were robust with correlation of .2, but the results became increasingly conservative as the correlation increased (Tables 2 and 3). They were also generally robust with null pairing.

For $K = 3$, time and interaction effects were robust with correlation of .2 and .8, while with correlation of .4 and .6 the results were conservative in the majority of conditions (Tables 5 and 6). They were also conservative with null pairing between correlation and group size in the majority of conditions.

Unbalanced groups

In unbalanced groups, for $K = 2$ and null pairing the group effect was robust in almost all conditions (Table 7), except with a small sample size ($N = 24$), where it can become conservative. For positive pairing between group size and correlation the test tended to be conservative with small sample sizes, whereas for negative pairing it tended to be liberal when inequality of group sample size was high in moderate and large sample sizes. For $K = 3$, it was conservative with positive pairing and liberal for negative pairing (Table 10).

For $K = 2$, time and interaction effects were robust with correlation of .2. However, the effects were increasingly conservative as the correlation increased (Tables 8 and 9), especially with correlation of .6 and .8. They were also conservative with both positive and negative pairing between group size and correlation.

For $K = 3$, time and interaction effects were robust with correlation .2 and .8, although with correlation of .8 and very small samples they were conservative. Results were generally conservative with correlation of .4 and .6 (Tables 11 and 12). They also tended to be conservative for positive pairing between group size and correlation as the sample size increased (from $N = 96$), and for negative pairing in all conditions.

Discussion

The aim of this study was to examine the empirical Type I error rates associated with the use of GLMMs under the most common conditions found in research in psychology (sample sizes less than 500, two groups, two and three repeated measures, and binary data). The RSPL approach in proc glimmix was used because its focus is on individual responses. In view of the fact that longitudinal data are normally correlated, the AR(1) covariance structure was fitted, and because the simulated samples were smaller than 500 we used the improved Kenward-Roger approach.

The results of this study provide researchers with useful information about the performance of these models, especially under extreme conditions (e.g., small samples, unbalanced groups, different correlation between repeated measures in each group). The most important conclusions in split-plot designs are:

1. For the group effect and balanced groups, the GLMM was always robust, including when the correlation between repeated measures of each group was different.
2. For the group effect and unbalanced groups, the GLMM can be (a) conservative when pairing between group size and the correlation between repeated measures is positive, and (b) liberal when this pairing is negative. With negative pairing the test became systematically more liberal as the inequality of group sample size increased.
3. For the time and interaction effects, the empirical Type I error rates of the GLMM depend on the correlation

between repeated measures, the pairing between group size and correlation, the inequality of group sample sizes, and the number of repeated measures. Specifically:

a. The GLMM was robust with low correlation between repeated measures (.2) for all conditions studied.

b. When correlations between repeated measures were equal in the groups, the GLMM was generally conservative for medium values of correlations (.4 and .6). However, for high values (.8) it was only conservative with two repeated measures.

c. When correlations between repeated measures were different in the groups, the GLMM was generally conservative for both positive and negative pairing in unbalanced designs. It was also conservative for three repeated measures and null pairing in balanced designs.

d. As the inequality of group sample sizes increased, the GLMM was systematically more conservative for negative pairing.

For time and interaction effects, the GLMM analyzed in this study is generally robust with low correlation between repeated measures. However, with medium correlation it tends to be conservative, and researchers would therefore need to exercise caution before accepting the null hypothesis based on their results. We also found that the test is robust with $K = 3$ when the correlation between measures is .8. It is also important to note that for the group effect the test may become very liberal in the case of negative pairing between group size and correlation, and hence any significant differences between groups would need to be interpreted with care.

Dang et al. (2008) found that robustness of the GLMM in terms of power estimates depends on the within-subject correlation of the binary outcomes. Specifically, if the random effects remain the same but the within-subject correlations increase, the estimated power decreases. Although other simulation studies have examined the robustness of the GLMM with binary responses (e.g., Fang & Louchin, 2013; Gawarammana & Sooriyarachchi, 2017), their results cannot be compared with our findings here as the variables manipulated and study objectives were different.

Gawarammana and Sooriyarachchi (2017) showed that with three repeated measures, a binary response variable, the logit link function, and quadrature estimation, proc glimmix yielded satisfactory results with respect to Type I error for sample sizes above 20. Here we used proc glimmix in SAS, although other options with binary data are proc nlmixed, which is useful for nonlinear mixed models (Zhang et al., 2011), and proc genmod, which is useful for generalized estimating equations (Gawarammana & Sooriyarachchi, 2017; Landerman et al., 2011). Zhang et al. (2012) analyzed the use of proc nlmixed with binary data, obtaining robust results under correct model assumptions and with a relatively small number of random effects. Gawarammana and Sooriyarachchi (2017) concluded that proc genmod only works well

when the sample size is very large (250 or over). In the review by Bono et al. (2021), none of the empirical studies they analyzed reported using proc nlmixed or proc genmod, and hence they did not compare the two procedures. It should be noted nonetheless that when applied to the modeling of binary responses, different procedures within a package may give quite different results (Zhang et al., 2011).

The robustness of the GLMM also differs depending on the parameter estimation method used (Zhang et al., 2012). Here we only used the RSPL, as this is well suited to the objectives of research in psychology. It is important to note, however, that empirical studies in the field of psychology that are based on the GLMM rarely report the estimation method used (Bono et al., 2021). Furthermore, few studies focused on the GLMM have examined methods of correcting the degrees of freedom for small samples (Bell & Grunwald, 2011). Li and Redden (2015) compared the Kenward-Roger procedure with other methods of adjusting the degrees of freedom when the GLMM is used to analyze binary outcomes in small sample cluster-randomized trials. They found that the Kenward-Roger method can provide tests with very conservative Type I error rates.

Many of the results derived from simulation studies may be of little use to applied researchers in the psychological field, because the manipulated variables and their values (e.g., sample size, number of groups, number of repeated measures, correlation, etc.) show considerable heterogeneity and are rarely chosen based on research practice. Hence our aim in this simulation study was to provide empirical evidence about the robustness (in terms of Type I error) of the GLMM in split-plot designs with binary response variables. To achieve this aim, we systematically manipulated a broad set of variables and range of values, based on research practice in psychology and related fields.

Our focus here has been on the analysis of repeated binary data. However, the GLMM can be considered for other types of outcomes, such as repeated ordinal responses (Lin, 2010; Lin & Chen, 2016) or repeated count measures (Bell & Grunwald, 2011; Huang et al., 2016; Zhang et al., 2012). Other important issues related to the performance of GLMMs that are not discussed in this paper are the approach chosen to deal with missing data or attrition (Amatya & Bhaumik, 2018; Miller et al., 2020; Noh et al., 2012), overdispersion (Johnson et al., 2015), misspecification of the shape of a random effects distribution (Litière et al, 2007; McCulloch & Neuhaus, 2011), link function misspecification (Yu & Huang, 2019), and the impact of different random effects covariance matrices for longitudinal data (Hoque & Torabi, 2018). Although much more research is still required in this area, the results obtained in the present study provide useful information about the performance of GLMMs in psychological research settings. This is important as applied researchers can now use statistical software and specific packages (e.g. SAS, R, and SPSS) to conduct GLMM analyses.

It should be noted that although we simulated a broad set of variables and range of values, the conclusions to be drawn are limited by the range of evaluated conditions. For example, we studied only one model for binary outcomes, and we did not evaluate model performance with alternative link functions or estimation methods. These limitations are potentially fruitful directions for future research. Finally, the robustness of the GLMM can also be studied in terms of power (Chen et al., 2016; Stroup, 2013). Despite the increasing use of GLMMs, there are few articles reporting a power analysis for these models (e.g., Dang et al., 2008; Jiang & Oleson, 2011; Johnson et al., 2015; Kain et al., 2015). A further study analyzing the robustness of the GLMM in terms of power and calculating the optimal sample, as proposed for the LMM (Vallejo et al., 2019), would therefore provide a useful complement to the present results.

References

Aiken, L. S., Mistler, S. A., Coxe, S., & West, S. G. (2015). Analyzing count variables in individuals and groups: Single level and multilevel models. *Group Process & Intergroup Relations, 18*(3), 290–314. https://doi.org/10.1177/1368430214556702

Amatya, A., & Bhaumik, D. K. (2018). Sample size determination for multilevel hierarchical designs using generalized linear mixed models. *Biometrics, 74*(2), 673–684. https://doi.org/10.1111/biom.12764

Arnau, J., Bono, R., Blanca, M. J., & Bendayan, R. (2012). Using the linear mixed model to analyze non-normal data distributions in longitudinal designs. *Behavior Research Methods, 44*(4), 1224–1238. https://doi.org/10.3758/s13428-012-0196-y

Arnau, J., Bendayan, R., Blanca, M. J., & Bono, R. (2013). The effect of skewness and kurtosis on the robustness of linear mixed models. *Behavior Research Methods, 45*(3), 873–879. https://doi.org/10.3758/s13428-012-0306-x

Arnau, J., Bendayan, R., Blanca, M. J., & Bono, R. (2014a). The effect of skewness and kurtosis on the Kenward-Roger approximation when group distributions differ. *Psicothema, 26*(2), 279–285. https://doi.org/10.7334/psicothema2013.174

Arnau, J., Bendayan, R., Blanca, M. J., & Bono, R. (2014b). Should we rely on the Kenward–Roger approximation when using linear mixed models if the groups have different distributions? *British Journal of Mathematical and Statistical Psychology, 67*, 408–429. https://doi.org/10.1111/bmsp.12026

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59*(4), 390–412. https://doi.org/10.1016/j.jml.2007.12.005

Baayen, H., Vasishth, S., Kliegl, R., & Bates, D. (2017). The cave of shadows: Addressing the human factor with generalized additive mixed models. *Journal of Memory and Language, 94*, 206–234. https://dx.doi.org/10.1016/j.jml.2016.11.006

Bakbergenuly, I., & Kulinskaya, E. (2018). Meta-analysis of binary outcomes via generalized linear mixed models: A simulation study. *BMC Medical Research Methodology, 18*(70), 1–18. https://doi.org/10.1186/s12874-018-0531-9

Bandera, E., & Pérez, L. (2018). Los modelos lineales generalizados mixtos. Su aplicación en el mejoramiento de plantas [Generalized linear mixed models: Their application in plant breeding]. *Cultivos Tropicales, 39*(1), 127–133.

Barker, D., D'Este, C., Campbell, M. J., & McElduff, P. (2017). Minimum number of clusters and comparison of analysis methods for cross sectional stepped wedge cluster randomised trials with binary outcomes: A simulation study. *Trials, 18*(119), 1–11. https://doi.org/10.1186/s13063-017-1862-2

Bauer, D. J., & Sterba, S. K. (2011). Fitting multilevel models with ordinal outcomes: Performance of alternative specifications and methods of estimation. *Psychological Methods, 16*(4), 373–390. https://doi.org/10.1037/a0025813

Bell, M. L., & Grunwald, G. K. (2011). Small sample estimation properties of longitudinal count models. *Journal of Statistical Computation and Simulation, 81*(9), 1067–1079. https://doi.org/10.1080/00949651003674144

Blanca, M. J., Alarcón, R., Arnau, J., Bono, R., & Bendayan, R. (2017). Non-normal data: Is ANOVA still a valid option? *Psicothema, 29*(4), 552–557. https://doi.org/10.7334/psicothema2016.383

Blanca, M. J., Alarcón, R., Arnau, J., Bono, R., & Bendayan, R. (2018). Effect of variance ratio on ANOVA robustness: Might 1.5 be the limit? *Behavior Research Methods, 50*, 937-962. https://doi.org/10.3758/s13428-017-0918-2

Blanca, M. J., Arnau, J., López-Montiel, D., Bono, R., & Bendayan, R. (2013). Skewness and kurtosis in real data samples. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences, 9*(2), 78–84. https://doi.org/10.1027/1614-2241/a000057

Bolker, B. M., Brooks, M. E., Clark, C. J., Geange, S. W., Poulsen, J. R., Stevens, M. H. H., & White, J. S. (2009). Generalized linear mixed models: A practical guide for ecology and evolution. *Trends in Ecology and Evolution, 24*(2), 127–135. https://doi.org/10.1016/j.tree.2008.10.008

Bono, R., Alarcón, R., & Blanca, M. J. (2021). Report quality of generalized linear mixed models in psychology: A systematic review. *Frontiers in Psychology, 12,* Article 666182. https://doi.org/10.3389/fpsyg.2021.666182

Bono, R., Blanca, M. J., Arnau, J., & Gómez-Benito, J. (2017). Non-normal distributions commonly used in health, education, and social sciences: A systematic review. *Frontiers in Psychology, 8*, Article1602. https://doi.org/10.3389/fpsyg.2017.01602

Bradley, J. V. (1978). Robustness? *British Journal of Mathematical and Statistical Psychology, 31*(2), 144–152. https://doi.org/10.1111/j.2044-8317.1978.tb00581.x

Breslow, N. E., & Clayton, D. G. (1993). Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association, 88*(421), 9–25. https://doi.org/10.2307/2290687

Brown, H., & Prescott, R. (2006). *Applied mixed models in medicine.* (2nd ed.). John Wiley & Sons.

Casals, M., Girabent-Farrés, M., & Carrasco J. L. (2014). Methodological quality and reporting of generalized linear mixed models in clinical medicine (2000-2012): A systematic review. *PLoS One, 9*, Article e112653. https://doi.org/10.1371/journal.pone.0112653

Chen, T., Lu, N., Arora, J., Katz, I., Bossarte, R., He, H., Xia, Y., Zhang, H., & Tu, X.M. (2016). Power analysis for cluster randomized trials with binary outcomes modeled by generalized linear mixed-effects models. *Journal of Applied Statistics, 43*(6), 1104–1118. https://doi.org/10.1080/02664763.2015.1092109

Cho, S. J., Brown-Schmidt, S., & Lee, W. Y. (2018). Autoregressive generalized linear mixed effect models with crossed random effects: An application to intensive binary time series eye-tracking data. *Psychometrika, 83*(3), 751–771. https://doi.org/10.1007/s11336-018-9604-2

Cho, S., & Goodwin, A. P. (2017). Modeling learning in doubly multilevel binary longitudinal data using generalized linear mixed models: An application to measuring and explaining word learning. *Psychometrika, 82*(3), 846–870. https://doi.org/10.1007/s11336-016-9496-y

Roser Bono et al.

Cnnan, A., Laird, N. M., & Slasor, P. (1998). Tutorial in biostatistics: Using the general linear mixed model to analyse unbalanced repeated measures and longitudinal data. *Statistics in Medicine, 16*(20), 2349–2380. https://doi.org/10.1002/(sici)1097-0258(19971030)16:20<2349::aid-sim667>3.0.co;2-e

Coupé, C. (2018). Modeling linguistic variables with regression models: Addressing non-gaussian distributions, non-independent observations, and non-linear predictors with random effects and generalized additive models for location, scale, and shape. *Frontiers in Psychology, 9, Article* 513. http://doi.org/10.3389/fpsyg.2018.00513

Dang, Q., Mazumdar, S., & Houck, P. R. (2008). Sample size and power calculations based on generalized linear mixed models with correlated binary outcomes. *Computer Methods and Programs in Biomedicine, 91*(2), 122–127. https://doi.org/10.1016/j.cmpb.2008.03.001

Elosua, P., & De Boeck, P. (2020). Educational assessment issues in linguistically diverse contexts: A case study using a generalised linear mixed model. *Language, Culture and Curriculum, 33*(3), 305–318. https://doi.org/10.1080/07908318.2019.1662432

Emrich, L. J., & Piedmonte, M. R. (1991). A method for generating high-dimensional multivariate binary variables. *American Statistician, 45*(4), 302–304. https://doi.org/10.2307/2684460

Fang, L., & Louchin, T. M. (2013). Analyzing binomial data in split-plot design: classical approach or modern techniques? *Communications in Statistics –Simulation and Computation, 42*(4), 727–740. https://doi.org/10.1080/03610918.2011.650264

Fieberg, J., Matthiopoulos, J., Hebblewhite, M., Boyce, M. S., & Frair, J. L. (2010). Correlation and studies of habitat selection: Problem, red herring or opportunity? *Philosophical Transactions of the Royal Society B, 365*, 2233–2244. https://doi.org/10.1098/rstb.2010.0079

Gawarammana, M. B. M. B. K., & Sooriyarachchi, M. R. (2017). Comparison of methods for analyzing binary repeated measures data: A simulation-based study. *Communications in Statistics – Simulation and Computation, 46*(3), 2103–2120. https://doi.org/10.1080/03610918.2015.1035445

Hoque, E., & Torabi, M. (2018). Modeling the random effects covariance matrix for longitudinal data with covariates measurement error. *Statistics in Medicine, 37*(28), 4167–4184. https://doi.org/10.1002/sim.7908

Huang, L., Tang, L, Zhang, B., Zhang, Z., & Zhang, H. (2016). Comparison of different computational implementations on fitting generalized linear mixed-effects models for repeated count measures. *Journal of Statistical Computation and Simulation, 86*(12), 2392–2404. https://doi.org/10.1080/00949655.2015.1111376

Jacqmin-Gadda, H., Sibillot, S., Proust, C., Molina, J. M., & Thiébaut, R. (2007). Robustness of the linear mixed model to misspecified error distribution. *Computational Statistics and Data Analysis, 51*(10), 5142–5154. https://doi.org/10.1016/j.csda.2006.05.021

Jiang, D., & Oleson, J. J. (2011). Simulation study of power and sample size for repeated measures with multinomial outcomes: An application to sound direction identification experiments (SDIE). *Statistics in Medicine, 30*(19), 2451–2466. https://doi.org/10.1002/sim.4302

Johnson, P. C. D., Barry, S. J. E., Ferguson, H. M., & Müller, P. (2015). Power analysis for generalized linear mixed models in ecology and evolution. *Methods in Ecology and Evolution 6*, 133–42. https://doi.org/10.1111/2041-210X.12306

Kain, M. P., Bolker, B. M., & McCoy, M. W. (2015). A practical guide and power analysis for GLMMs: Detecting among treatment variation in random effects. *PeerJ, 3,* Article e1226. https://doi.org/10.7717/peerj.1226

Kenward, M. G., & Roger, J. H. (2009). An improved approximation to the precision of fixed effects from restricted maximum likelihood. *Computational Statistics and Data Analysis, 53*(7), 2583–2595. https://doi.org/10.1016/j.csda.2008.12.013

Koh, H., Li, Y., Zhan, X., Chen, J., & Zhao, N. (2019). A distance-based kernel association test based on the generalized linear mixed model for correlated microbiome studies. *Frontiers in Genetics, 10,* Article 458. https://doi.org/10.3389/fgene.2019.00458

Kowalchuk, R. K., Keselman, H. J., Algina, J., & Wolfinger, R. D. (2004). The analysis of repeated measurements with mixed-model adjusted *F* tests. *Educational and Psychological Measurement*, *64*(2), 224–242. https://doi.org/10.1177/0013164403260196

Kruppa, J., & Hothorn, L. (2021). A comparison study on modeling of clustered and overdispersed count data for multiple comparisons. *Journal of Applied Statistics*, *48*(16), 3220–3232. https://doi.org/10.1080/02664763.2020.1788518

Landerman, L. R., Mustillo, S. A., & Land, K. C. (2011). Modeling repeated measures of dichotomous data: Testing whether the within-person trajectory of change varies across levels of between-person factors. *Social Science Research*, *40*(5), 1456–1464. https://doi.org/10.1016/j.ssresearch.2011.05.006

Lei, M., & Lomax, R. G. (2005). The effect of varying degrees on nonnormality in structural equation modeling. *Structural Equation Modeling, 12*(1), 1–27. https://doi.org/10.1207/s15328007sem1201_1

Li, P., & Redden, D. T. (2015). Comparing denominator degrees of freedom approximations for the generalized linear mixed model in analyzing binary outcome in small sample cluster-randomized trials. *BMC Medical Research Methodology, 15*(38), 1–12. https://doi.org/10.1186/s12874-015-006-x

Lin, K. C. (2010). Goodness-of-fit tests for modeling longitudinal ordinal data. *Computational Statistics and Data Analysis, 54*(7), 1872–1880. https://doi.org/10.1016/j.csda.2010.02.013

Lin, K. C., & Chen, Y. J. (2016). Goodness-of-fit- tests of generalized linear mixed models for repeated ordinal responses. *Journal of Applied Statistics, 43*(11), 2053–2064. https://doi.org/10.1080/02664763.2015.1126568

Litière, S., Alonso, A., & Molenberghs, G. (2007). Type I and Type II error under random-effects misspecification in generalized liner mixed models. *Biometrics, 63*(4), 1038–1044. https://doi.org/10.1111/j.1541-0420.2007.00782.x

Liu, S., Rovine, M. J., & Molenaar, P. C. (2012). Selecting a linear mixed model for longitudinal data: Repeated measures analysis of variance, covariance pattern model, and growth curve approaches. *Psychological Methods, 17*(1), 15–30. https://doi.org/10.1037/a0026971

Livacic-Rojas, P., Vallejo, G., & Fernández, P. (2010). Analysis of Type I error rates of univariate and multivariate procedures in repeated measures designs. *Communications in Statistics — Simulation and Computation, 39*(3), 624–640. https://doi.org/10.1080/03610910903548952

Lix, L. M., & Hinds, A. M. (2004). Multivariate contrasts for repeated measures designs under assumptions violations. *Journal of Modern Applied Statistical Methods, 3*(2), 333–344. https://doi.org/10.22237/jmasm/1099267620

Lo, S., & Andrews, S. (2015). To transform or not transform: Using generalized linear mixed models to analyses reaction time data. *Frontiers in Psychology, 6,* Article 1171. https://doi.org./10.3389/fpsyg.2015.01171

Malik, W. A., Marco-Llorca, C., Berendzen, K, & Piepho, H. P. (2020). Choice of link and variance function for generalized linear mixed models: A case study with binomial response in proteomics. *Communications in Statistics – Theory and Methods, 49*(17), 4313–4332. https://doi.org/10.1080/03610926.2019.1599021

McCulloch, C. E., & Neuhaus, J. M. (2011). Misspecifying the shape of a random effects distribution: Why getting it wrong may not matter. *Statistical Science, 26*(3), 388–402. https://doi.org/10.1214/11-STS361

Micceri, T. (1989). The unicorn, the normal curve, and other improbable creatures. *Psychological Bulletin, 105*(1), 156–166. https://doi.org/10.1037/0033-2909.105.1.156

Miller, M. L., Roe, D. J., Hu, C., & Bell, M. L. (2020). Power difference in a χ2 test vs generalized linear mixed model in the presence of missing data: A simulation study. *BMC Medical Research Methodology, 20*(50), 1–12. https://doi-org.sire.ub.edu/10.1186/s12874-020-00936-w

Moscatelli, A., & Lacquaniti, F. (2011). The weight of time: Gravitational force enhances discrimination of visual motion duration. *Journal of Vision, 11*(4), 1–17. https://doi.org/10.1167/11.4.5

Moscatelli, A., Mezzetti, M., & Lacquaniti, F. (2012). Modeling psychophysical data at the population-level: The generalized linear mixed model. *Journal of Vision 12*(26), 1–17. https://doi.org/10.1167/12.11.26

Moscatelli, A., Polito, L., & Lacquaniti, F. (2011). Time perception of action photographs is more precise than that of still photographs. *Experimental Brain Research, 210*(1), 25–32. https://doi.org./10.1007/s00221-011-2598-y

Mowen, T. J., & Culhane, S. E. (2017). Modeling recidivism within the study of offender reentry: Hierarchical generalized linear models and lagged

dependent variable models. *Criminal Justice and Behavior, 44*(1), 85–102. https://doi.org/10.1177/0093854816678647

Noh, M., Wu, L., & Lee, Y. (2012). Hierarchical likelihood methods for nonlinear and generalized linear mixed models with missing data and measurement errors in covariates. *Journal of Multivariate Analysis, 109*, 42–51. http://doi.org/10.1016/j.jmva.2012.02.011

Platt, R. W., Leroux, B. G., & Breslow, N. (1999). Generalized linear mixed models for meta-analysis. *Statistics in Medicine, 18*(6), 643–654. https://doi.org/10.1002/(SICI)1097-0258(19990330)18:6<643::AID-SIM76>3.0.CO;2-M

Quené, H., & van den Bergh, H. (2008). Example of mixed-effects modeling with crossed random effects and with binomial data. *Journal of Memory and Language, 59*(4), 413–425. https://doi.org/10.1016/j.jml.2008.02.002

SAS Institute Inc. (2013). *The GLIMMIX procedure. In SAS/STAT® 13.1 User's Guide.* SAS Institute Inc.

SAS Institute Inc. (2016). *SAS/STAT® 14.2 User's Guide.* SAS Institute Inc.

Searle, M. P., Waters, D. J., Rex, D. C., & Wilson, R. N. (1992). Pressure, temperature and time constraints on Himalayan metamorphism from eastern Kashmir and western Zanskar. *Journal of the Geological Society, 149*(5), 753–773. https://doi.org./10.1144/gsjgs.149.5.0753

Skrondal, A., & Rabe-Hesketh, S. (2003). Some applications of generalized linear latent and mixed models in epidemiology: Repeated measures, measurement error and multilevel modeling. *Norwegian Journal of Epidemiology, 13*(2), 265–278.

Smith, L. M., Stroup, W. W., & Marx, D. B. (2020). Poisson cokriging as a generalized linear mixed model. *Spatial Statistics, 35*, Article 100399. https://doi.org/10.1016/j.spasta.2019.100399

Stroup, W. W. (2013). *Generalized linear mixed models. Modern concepts, methods and applications.* Taylor and Francis.

Stroup, W. W., Milliken, G. A., Claassen, E. A., & Wolfinger, R. D. (2018). *SAS for mixed models: Introduction and basic applications.* SAS Institute Inc.

Sun, S., Zhu, J., Mozaffari, S., Ober, C., Chen, M., & Zhou, X. (2019). Heritability estimation and differential analysis of count data with generalized linear mixed models in genomic sequencing studies. *Bioinformatics, 35*(3), 487-496. https://doi.org/10.1093/bioinformatics/bty644

Thiele, J., & Markusen, B. (2012). Potential of GLMM in modelling invasive spread. *CAB Reviews, 7*(16), 1–10. https://doi.org/10.1079/PAVSNNR20127016

Vallejo, G., Ato, M., Fernández, M. P., & Livacic-Rojas, P. E. (2019). Sample size estimation for heterogeneous growth curve models with attrition. *Behavior Research Methods, 51*(3), 1216–1243. https://doi.org/10.3758/s13428-018-1059-y

Vallejo, G., Ato, M., & Valdés, T. (2008). Consequences of misspecifying the error covariance structure in linear mixed models for longitudinal data. *Methodology: European Journal of Research for the Behavioral and Social Sciences, 4*(1), 10–21. https://doi.org/10.1027/1614-2241.4.1.10

Wicklin, R. (2013). *Simulating data with SAS.* SAS Institute Inc.

Witte, J. S., Greenland, S., Kim, L., & Arab, L. (2000). Multilevel modeling in epidemiology with GLIMMIX. *Epidemiology, 11*(6), 684–688. https://doi.org/10.1097/00001648-200011000-00012

Wolfinger, R., & O'Connell, M. (1993). Generalized linear models: A pseudo-likelihood approach. *Journal of Statistical Computation and Simulation, 48*(3-4), 233–243. https://doi.org/10.1080/00949659308811554

Yu, S., & Huang, X. (2019). Link misspecification in generalized linear mixed models with a random intercept for binary responses. *Test, 28*(3), 827–843. https://doi.org/10.1007/s11749-018-0602-6

Zhang, H., Lu, N., Feng, C., Thurston, S. W., Xia, Y., Zhu, L., & Tu, X. M. (2011). On fitting generalized linear mixed-effects models for binary responses using different statistical packages. *Statistics in Medicine, 30*(20), 2562–2572. https://doi.org/10.1002/sim.4265

Zhang, H., Yu, Q., Feng, C., Gunzler, D., Wu, P., & Tu, X. M. (2012). A new look at the difference between the GEE and the GLMM when modeling longitudinal count responses. *Journal of Applied Statistics, 39*(9), 2067–2079. https://doi.org/10.1080/02664763.2012.700452