Title:

Using random effects in generalized linear mixed-effects models: assessing a common 'rule of thumb'

Dylan G.E. Gomes[1,2,*]

[1] Department of Biological Sciences, Boise State University, 1910 University Dr., Boise, ID, United States, 83725

[2] Cooperative Institute for Marine Resources Studies, Hatfield Marine Science Center, Oregon State University, 2030 SE Marine Science Dr., Newport, OR, United States, 97365

[*] Correspondence: gomesd@oregonstate.edu; dylangomes@u.boisestate.edu

Abstract

As (generalized) linear mixed-effects models (GLMMs) have become a widespread tool in ecology, the need to guide the use of such tools is increasingly important. One common 'rule of thumb' is that one needs at least five levels of a random effect. Having such few levels makes the estimation of the variance of random effects terms difficult, but it need not muddy one's ability to estimate fixed effects terms. Here, I use simulated datasets and model fitting to show that having too few random effects terms does not influence the parameter estimates or uncertainty around those estimates for fixed effects terms. Thus, it should be acceptable to use fewer levels of random effects if one is not interested in making inference about variance estimates of the random effects terms (i.e. they are 'nuisance' parameters). I also assess the potential for pseudo-replication in (generalized) linear models (LMs), when random effects are explicitly ignored and find that LMs do not show increased type-I errors compared to their mixed-effects model counterparts. These results challenge the view that it is never appropriate to model random effects terms with fewer than five levels – when inference is not being made for the random effects, it may not pose problems. Given the widespread accessibility of GLMMs, future simulation studies and further assessments of these statistical methods are necessary to understand the consequences of both violating and blindly following ‘rules of thumb.’

Keywords

Statistics, hierarchical modelling, experimental design, block-design, varying effects, quantitative, regression, ANOVA, R, programming

Introduction

While statistical analyses are becoming more complex (Low-Décarie, Chivers & Granados, 2014), advances in computing power and freely available statistical software are increasing the accessibility of such analyses. As these methods have become more complex and accessible to non-statisticians, the need to guide the use of such tools is becoming increasingly important (Bolker, 2008; Bolker et al., 2009; Zuur, Ieno & Elphick, 2010; Kéry & Royle, 2015; Kass et al., 2016; Zuur & Ieno, 2016; Harrison et al., 2018; Silk, Harrison & Hodgson, 2020). The use of generalized linear mixed-effects models (GLMM), for example, has become a widespread tool that allows one to build hierarchical models that can estimate, and thus account for, imperfect detection in biological surveys (e.g. occupancy, N-mixture, mark-recapture, etc. models) and can model correlations among data that come from groups (i.e. random effects; also known as varying effects) (Bolker, 2008; Kéry & Royle, 2015; Powell & Gale, 2015; Harrison et al., 2018; McElreath, 2020).

Generalized linear mixed-effects models are a regression type analysis that are flexible in that they can handle a variety of data generating processes such as binomial (e.g. presence / absence) and Poisson (e.g. survey counts). When the sampling distribution is Gaussian (also known as normal), this is a special case of a GLMM that is referred to as simply a linear mixed-effects model (LMM). GLMMs (and LMMs) differ from their simpler counterparts, (generalized) linear models (GLMs and LMs), in that they include random effects, in addition to the fixed effects (hence *mixed-effects*). Fixed effects (which are also often called predictors, covariates, explanatory or independent variables) are *fixed* in that the model parameters (β in equation 1 below) are fixed, or non-random, and are not drawn from a distribution. Random effects are *random* in that they are assumed to be drawn randomly from a distribution – often a Gaussian distribution – during the data-generating process. Note that one can also assign random *slopes* to variables, where the slopes of variables (not just the intercepts) are allowed to vary, and are assumed to be randomly drawn from a distribution (see Bolker, 2008; Kéry & Royle, 2015; Harrison et al., 2018).

The advantages of random effects are multifold; they allow one to combine information (as in a meta-analysis), deal with spatiotemporal autocorrelation, use partial pooling to borrow strength from other populations or groups, account for grouping or blocked designs, and estimate population-level parameters, among others (Kéry & Royle, 2015). If we are interested in the variability of a population (of individuals, groups, sites, or populations), it is difficult to estimate this variation with too few levels of individuals, groups, sites, or populations (i.e. random effects terms).

> "When the number of groups is small (less than five, say), there is typically not enough information to accurately estimate group-level variation" (Gelman & Hill, 2006).

> "...if interest lies in measuring the variation among random effects, a certain number is required...To obtain an adequate estimate of the among-population heterogeneity – that is, the variance parameter – at least 5 - 10 populations might be required" (Kéry & Royle, 2015).

> "With <5 levels, the mixed model may not be able to estimate the among-population variance accurately." (Harrison et al., 2018).

This 'rule of thumb' that random effects terms should have at least five levels (i.e. groups) is backed by limited empirical evidence (Harrison, 2015), but it is intuitive that too few draws from distribution will hinder one's ability to estimate the variance of that distribution. Indeed, in each of the above segments of quoted text, the authors suggest that at least 5 levels are needed for *estimation of group-level, or among-population, variance*. However, this rule is often adhered to out of context, where authors or reviewers suggest that one cannot use random effects terms if they do not contain at least five levels.

Simulations by Harrison (2015) demonstrate that random effects variance can be biased more strongly when the levels of random effects terms are low, yet in this work it appears that slope (beta) estimates for fixed effects terms are generally not more biased with only three random effects levels compared to five. There are many cases (and some would argue that in *most cases*, see below) in which the variance of random effects is not directly of interest to the research question at hand.

> "...in the vast majority of examples of random-effects (or mixed) models in ecology, the random effects do *not* have a clear ecological interpretation. Rather, they are merely abstract constructs invoked to explain the fact that some measurements are more similar to each other than others are – i .e., to model correlations in the observed data" (Kéry & Royle, 2015).

Thus, it is unclear whether or not it is appropriate to use random effects when there are fewer than five grouping levels in situations where one does not directly care about 'nuisance' among-population variance, but instead is interested in estimates and the variance (i.e. uncertainty) of predictor variables (i.e. fixed effects). The current state of practice in ecology is to drop the random effects terms such that we are now using generalized linear models where we are not grouping observations (we drop the **M**ixed-effects from the GL**M**M to become GLM). I question whether we are choosing to accept pseudoreplication of repeat-measures (Hurlbert, 1984; Kéry

& Royle, 2015), rather than inaccurate estimates of among-population variance. In cases where one does not care about among-population variance, this tradeoff may be non-existent, but little research exists to support this. Here, I perform simulations to assess whether *fixed effects* estimates are more biased when the accompanying *random effects* consist of fewer than five levels; I also ask whether using an alternative model without random effects (GLMs) leads to higher type I errors (demonstrating a 'significant' effect when in fact one does not exist).

Methodology

All simulation of datasets and model fitting was done in R v4.0.4 (R Core Team, 2017), all visualizations were completed using the aid of R package `ggplot2` (Wickham, 2011), and all code is available from Zenodo at http://doi.org/10.5281/zenodo.4679101.

Data generation

I used a modified version of code from Harrison (2015), to explore the importance of varying two parameters in a linear mixed-effect model (LMM): the number of observations in a dataset (30, 60, or 120), and the number of levels of the random intercept term (3, 5, or 20). This was done by generating a response variable y_i from the following equation:

$$y_i = \alpha_{j(i)} + \beta_1 X_{1_i} + \beta_2 X_{2_i} + \varepsilon_i \quad [1]$$

$$\alpha_j \sim Normal(\mu, \sigma)$$

[2]

Where $\alpha_{j(i)}$ is the intercept for site (or population) *j* to which observation (or individual) *i* belongs. Thus, each observation shared a site-level intercept, which were drawn from a normal distribution with mean (μ) = 0 and standard deviation (σ) = 0.5. β_1 and β_2 are the slope parameters for two generic predictor variables (X_{1_i} and X_{2_i} respectively), which were both drawn from a normal distribution with μ = 0 and σ = 0.5, which mimics standardized variables that are centered by their mean and scaled by two standard deviations (Gelman, 2008). The error term ε_i is unique to each observation *i* that is drawn from a normal distribution with μ_ε= 0 and σ_ε = 0.25 (same as equation 2 above).

For all simulated datasets, parameter values were fixed at β_1 = 2 and β_2 = 0, meaning X_{2_i} does not have a linear relationship with, or is only randomly related to, the response variable y_i. This allows for an assessment of type-I error rate, since any significant *p* values for this β_2 slope parameter are erroneous.

Model fitting simulations

For each of the nine combinations of scenarios (30, 60, or 120 observations by 3, 5, or 20 random intercept levels), I simulated 10,000 datasets. Each dataset was fit with a linear mixed-effect model (LMM) and a linear model (LM). All model fitting was done with R functions \`lmer\` (LMM) or \`lm\` (LM) in the package \`lme4\` or in \`base\` R, respectively (Bates et al., 2007; R Core Team, 2017).

```
#LMM:
m1 <- lmer(y ~ x1 + x2 + (1|Site))
```

R Code

Where x_1 and x_2 are fixed effects (see equation 1), and `(1|Site)` is the syntax for specifying a random intercept ($\alpha_{j(i)}$ in equation 1). In ecology, we often fit independent sites as unique levels of a random effect, so I use site here for demonstration purposes. But site can be replaced with individual, group, population, etc.

Often the recommendation, if one has fewer than 5 levels of random effects terms ($j \leq 5$ in $\alpha_{j(i)}$), is to fit the random effects as fixed effects (LMM becomes LM), specified in R as:

```
#LM:
m2 <- lm(y ~ x1 + x2 + Site)
```

R Code

and mathematically defined as:

$$y_i = \beta_1 X_{1_i} + \beta_2 X_{2_i} + \beta_3 Site_{1(i)} + \beta_4 Site_{2(i)} + \cdots \beta_{n+2} Site_{n(i)} + \varepsilon_i$$

[3]

Now a β term is estimated for each site (or population) level independently. Site parameters no longer come from a normal distribution (as in equation 2), but instead are considered fixed, hence *fixed effects*.

Thus, both a LMM and a LM were fit to each simulated dataset (n = 10,000) of each of the nine combinations (30, 60, or 120 observations by 3, 5, or 20 random intercept levels) of data-generation. This allowed for a comparison of the type-I error rates of LMMs and LMs, the latter of which ignores the blocked structure of data (i.e. site-level grouping).

Type-I error calculation

Type-I error rate was calculated as the proportion of 10,000 models that a 'significant' p value of ≤ 0.05 was obtained for the β_2 parameter estimate in which the true value of that parameter was set to be 0. I sampled (with replacement) 10,000 p value 'observations' from each group of 10,000 models to produce a new proportion of type-I error; this process was repeated 1,000 times, and the bootstrapped 95% confidence intervals were calculated as the 0.025 and 0.975 quantiles of those 1,000 replications (see code).

Results

Estimating model parameters and uncertainty

Linear mixed models were able to resurrect simulated fixed effect relationships with no noticeable patterns in bias, regardless of number of levels of random effects or sample size. That is, both mean model parameter estimates (β_1 and β_2) were centered on their true values (Table 1; Figure 1). The uncertainty around these estimates generally decreased as sample size increased. For example, doubling the sample size from 30 observations to 60 observations lead to a decrease by 36.6% and 35.5% in parameter estimate uncertainty (for β_1 and β_2 respectively; Table 1; Figure 1). Another doubling to 120 observations lead to a further decrease in uncertainty by 33.4% and 32.9%, respectively. The number of levels of random effects appears to be relatively non-important in resurrecting model parameter estimates within these simulation scenarios (Table 1; Figure 1); instead there were small, likely negligible, increases in uncertainty around fixed effect parameter estimates as the number of levels of random effects increased.

All LMM estimates of the distribution mean (μ) were unbiased, regardless of number of levels of random effects or sample size (Table 1; Figure 2A). The random effects variance (σ), however, was not centered at the true value, and it was more biased with fewer levels of random effects, whereas sample size did not affect this bias (Table 1; Figure 2B). That is, with only three levels of random effects the magnitude of the bias was 12.2% of the true value. Increasing to five levels of random effects nearly halved this bias to 6.4%, and increasing to 10 levels halved the bias again to 3.2% of the true value. Averaged across numbers of random effects terms, estimates were biased by about 7% regardless of sample size (7.1%, 7.4%, and 7.2% for N = 30, 60, and 120 respectively).

The uncertainty around random effects estimates (μ and σ) showed the reverse pattern as the fixed effects. That is uncertainty generally decreased with an increased number of random effects levels, whereas sample size did little to alleviate this uncertainty (Table 1; Figure 2). Increasing the number of random effects levels from 3 to 5, and then from 5 to 10, decreased the uncertainty for μ by 22.4% and 29.1%, respectively, and for σ by 26.6% and 29.8% respectively.

Type-I errors

For all simulated datasets, both LMM and LM produced type-I error rates around the typical α = 0.05, with 95% confidence intervals overlapping this value. Neither sample size, nor the number of random effects levels seemed to influence the type-I error rate. Furthermore, dropping the random effects structure (using a LM instead of a LMM) did not increase the probability type-I errors (Figure 3).

Discussion

The work presented here demonstrates that i) fixed effects estimates are not more biased when the levels of an accompanying random effect have fewer than five ($n \leq 5$) levels, but population-level variance estimates are and ii) type-I error rates are not increased by using LM instead of LMM, contrary to previous expectations.

These results suggest that fixed effects parameter estimation is not strongly influenced by, nor biased by, the number of levels of random effects terms, but uncertainty in those estimates is much more strongly influenced by sample size. While this pattern may appear to contradict the decreased uncertainty around beta estimates in Figure 2 of Harrison (2015), this instead is due to differences in the way that sample size was handled between that work and the current work. Harrison (2015) coded each random effect level to be associated with a fixed number of observations (N=20), such that each additional random effect level yielded an increased sample size. Here, sample size (i.e. number of observations) has been separated from the number of random effects terms.

Despite this difference, the estimation of random effects terms (μ and σ) in the simulations presented here suggest consistent patterns with Harrison (2015) and support previous ‘rules of thumb’ and simulations suggesting that fewer than five levels of random effects terms can make estimation of population-level variance difficult (Gelman & Hill, 2006; Harrison, 2015; Kéry &

Royle, 2015; Harrison et al., 2018). Thus, the combination of these results suggest that using fewer than five levels of random effects is acceptable when one is only interested in estimating fixed effects parameters; in other words, when inference about the variance of random effects terms (e.g. sites, individuals, populations) is not of direct interest, but instead are used to group data, as in a block design of a study. In these cases, however, caution should be taken in reporting the variance estimates for such population-level parameters – as this information can later be taken out of context of the question at hand.

Interestingly, type-I errors were not more likely in any situation. This possibly suggests that mis-specified linear models that are theoretically missing a random effect are relatively robust to this omission – at least in some simple cases such as the scenarios presented here. While this perhaps alleviates some concern over inflated type-I errors due to pseudoreplication while ignoring the grouped nature of repeat-measures studies and non-independent data, this should not be taken as evidence to purposefully omit random effects when such a structure is appropriate. Instead, it warrants future investigation and further simulation studies with more thorough scenarios and more complex data structures.

Often researchers or reviewers cite this 'rule of thumb' as to why one should not use a mixed-effects model, leaving others to fight their case as to why they ignored such a rule. This is likely exacerbated by the fact that authors or peer-reviewers can easily point out that this 'rule of thumb' exists (Gelman & Hill, 2006; Harrison, 2015; Kéry & Royle, 2015; Harrison et al., 2018), but may find it more difficult or time-consuming to make a nuanced argument against following such a pervasive rule. Hopefully the results presented here will challenge that view, and allow the fitting of random effects when inference is not being made for the random effects. It is critical to note that these results are far from comprehensive. Given the widespread accessibility of GLMMs, future simulation studies and further assessments of these statistical methods are necessary to understand the consequences of both violating and blindly following 'rules of thumb.'

Acknowledgements

I would like to thank the National Science Foundation (NSF) for funding (GRFP 2018268606), and to Boise State University for general support.

References

Bates D, Sarkar D, Bates MD, Matrix L. 2007. The lme4 package. *R package version* 2:74.

Bolker BM. 2008. *Ecological models and data in R*. Princeton University Press.

Bolker BM, Brooks ME, Clark CJ, Geange SW, Poulsen JR, Stevens MHH, White J-SS. 2009. Generalized linear mixed models: a practical guide for ecology and evolution. *Trends in ecology & evolution* 24:127–135.

Gelman A. 2008. Scaling regression inputs by dividing by two standard deviations. *Statistics in medicine* 27:2865–2873.

Gelman A, Hill J. 2006. *Data analysis using regression and multilevel/hierarchical models*. Cambridge university press.

Harrison XA. 2015. A comparison of observation-level random effect and Beta-Binomial models for modelling overdispersion in Binomial data in ecology & evolution. *PeerJ* 3:e1114.

Harrison XA, Donaldson L, Correa-Cano ME, Evans J, Fisher DN, Goodwin CE, Robinson BS, Hodgson DJ, Inger R. 2018. A brief introduction to mixed effects modelling and multi-model inference in ecology. *PeerJ* 6:e4794.

Hurlbert SH. 1984. Pseudoreplication and the design of ecological field experiments. *Ecological monographs* 54:187–211.

Kass RE, Caffo BS, Davidian M, Meng X-L, Yu B, Reid N. 2016. *Ten simple rules for effective statistical practice*. Public Library of Science.

Kéry M, Royle JA. 2015. *Applied Hierarchical Modeling in Ecology: Analysis of distribution, abundance and species richness in R and BUGS: Volume 1: Prelude and Static Models*. Academic Press.

Low-Décarie E, Chivers C, Granados M. 2014. Rising complexity and falling explanatory power in ecology. *Frontiers in Ecology and the Environment* 12:412–418.

McElreath R. 2020. *Statistical rethinking: A Bayesian course with examples in R and Stan*. CRC press.

Powell LA, Gale GA. 2015. *Estimation of Parameters for Animal Populations*. Caught Napping Publications, Lincoln, NE.

R Core Team. 2017. *R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing*.

Silk MJ, Harrison XA, Hodgson DJ. 2020. Perils and pitfalls of mixed-effects regression models in biology. *PeerJ* 8:e9522.

Wickham H. 2011. ggplot2. *Wiley Interdisciplinary Reviews: Computational Statistics* 3:180–185.

Zuur AF, Ieno EN. 2016. A protocol for conducting and presenting results of regression-type analyses. *Methods in Ecology and Evolution* 7:636–645.

Zuur AF, Ieno EN, Elphick CS. 2010. A protocol for data exploration to avoid common statistical problems. *Methods in ecology and evolution* 1:3–14.

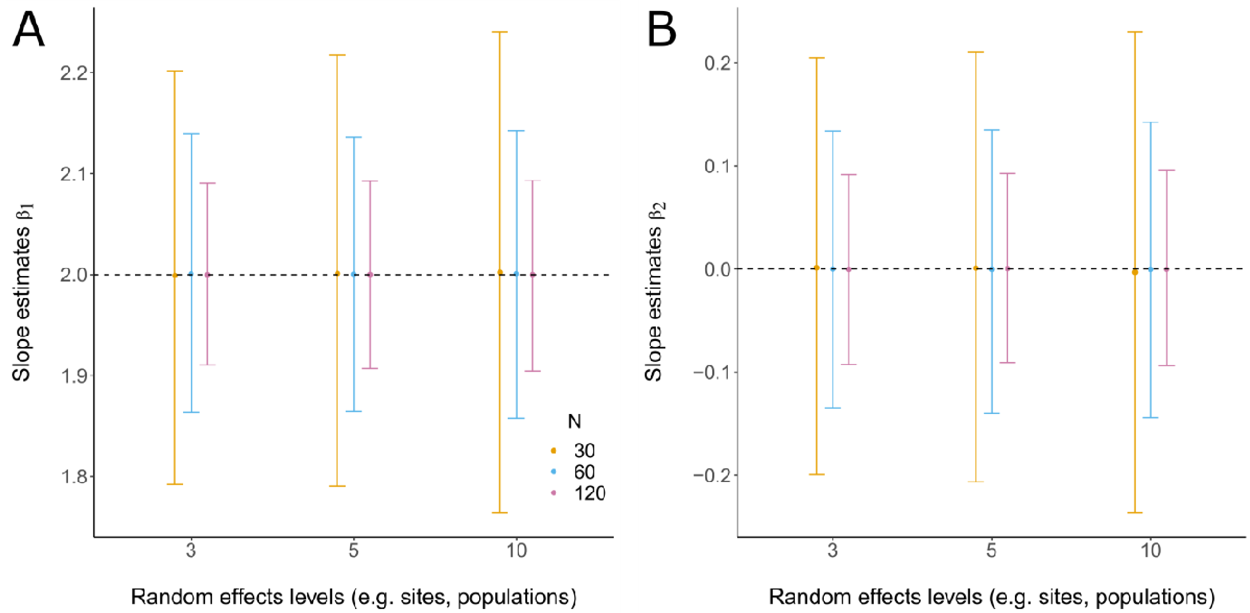

Figure 1: Fixed effects model estimates for simulated data. Each point is the mean estimate for 10,000 models (and datasets), whereas error bars are 95% confidence intervals. N = the number of observations (i.e. number of rows) in each dataset. Dashed lines indicate the true value. In all scenarios the bias in parameter estimates are negligible. As the sample size increases, our certainty around the parameter estimates (β) increases, but the number of random effects has a relatively minor effect on estimating β. When sample sizes (N) are low, parameter uncertainty increases with increasing levels of random effects (assuming a consistent N).

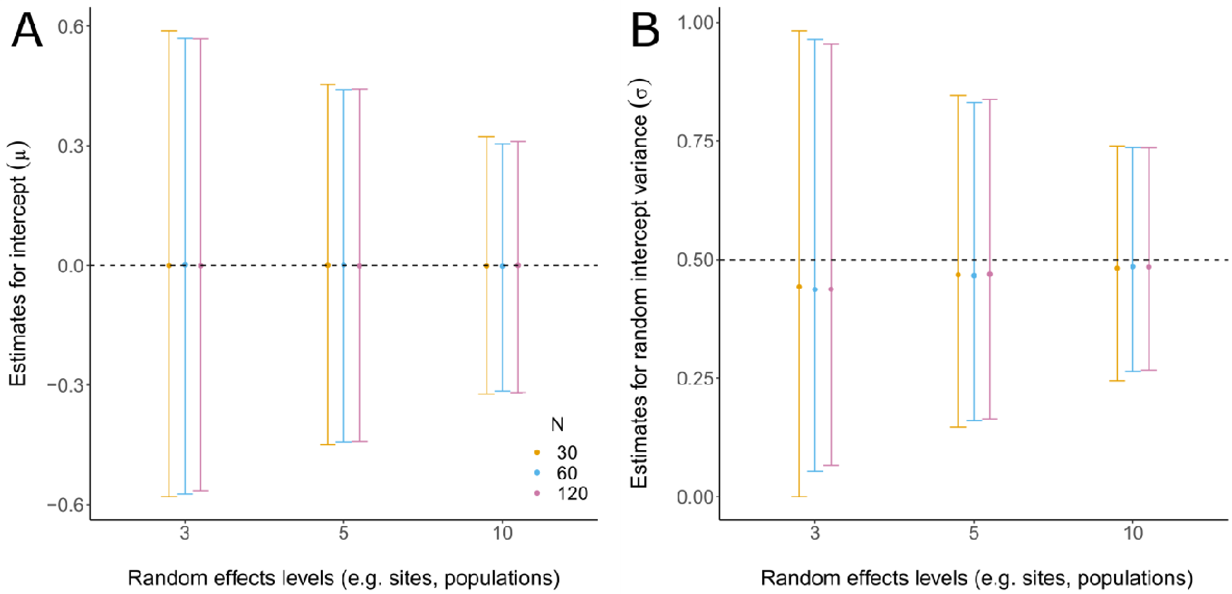

Figure 2: Random effects model estimates for simulated data. Each point is the mean estimate for 10,000 models (and datasets), whereas error bars are 95% confidence intervals. N = the number of observations (i.e. number of rows) in each dataset. Dashed lines indicate the true value. A) As the number of random effects levels increases, the uncertainty around the mean (μ) decreases. Sample size has a relatively minor effect on estimating μ. B) As the number of random effects levels increases, the bias and uncertainty around the random effects variance (σ) decreases. Sample size has a small, but relatively minor effect on estimating σ. The bias in σ starts to approach the starting (simulated) σ = 0.5 as the number of random effects reaches 10.

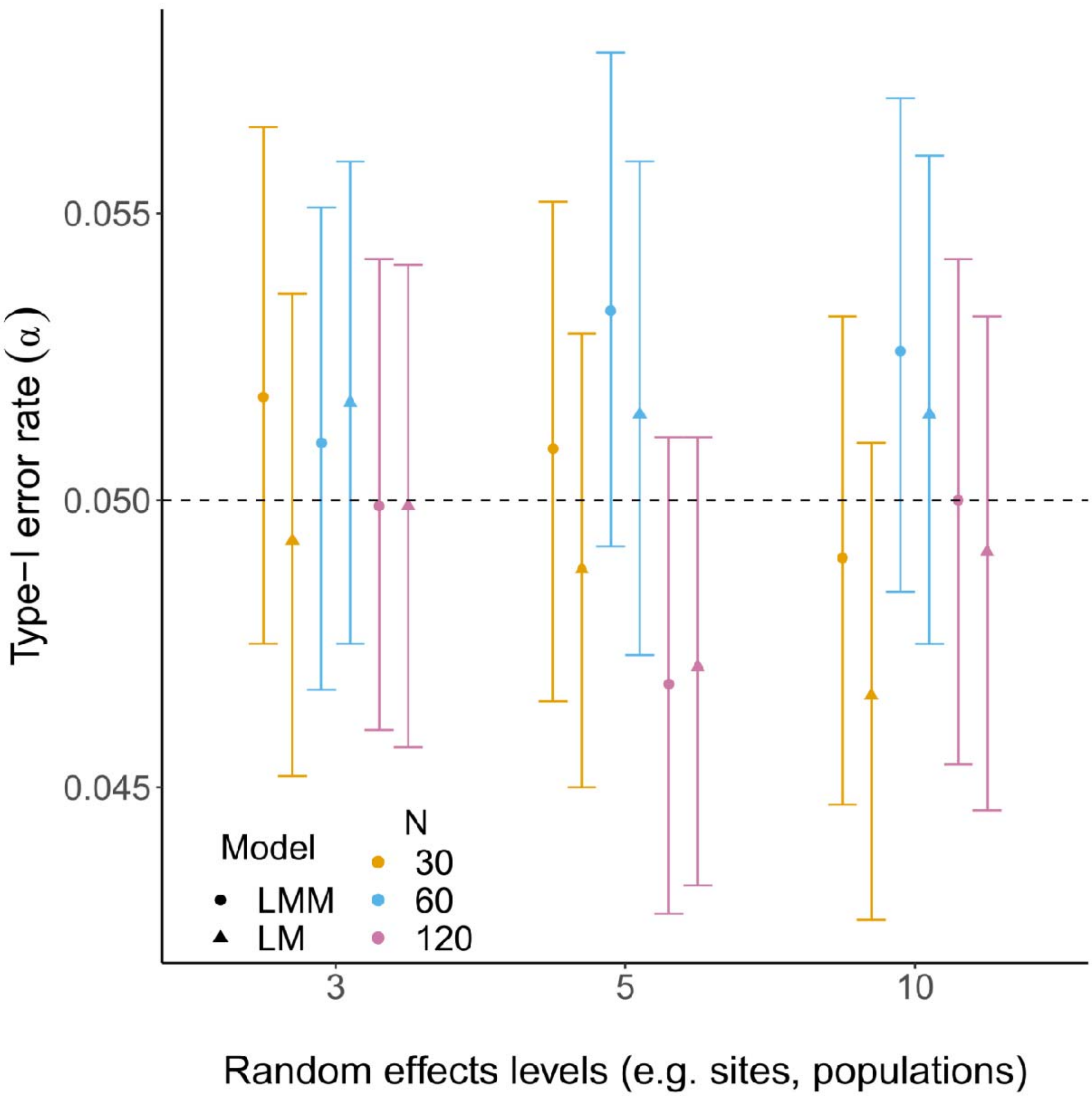

Figure 3: Type-I error for various linear models (LM) and linear mixed-effects models (LMM). Type-I error rate was calculated as the proportion of models (n = 10,000) in which a ‘significant’ p value of ≤ 0.05 was obtained for a parameter estimate in which the true value of that parameter was set to be 0 (Figure 1B); each point represents this proportion. To generate error bars as 95% confidence intervals, I used bootstrapping to replicate this process 1,000 times (see methods). N = the number of observations (i.e. number of rows) in each dataset. Symbols indicate model type (LM vs LMM). Dashed lines indicate the true alpha value (0.05).

Table 1: Model estimates from 10,000 simulated datasets. The number of levels of random effects (RE) was varied (3, 5, or 10), as was the number of observations in the dataset (N = 30, 60, or 120). The true (T) values for the data generation process (equation 1) are indicated in the second header row underneath the estimated parameter labels (fixed effects: β_1, β_2; random effects: μ, σ). The mean of 10,000 model estimates (β_1, β_2, μ, σ) are indicated for the respective models below the true values. Lower and upper bounds on 95% confidence intervals for each parameter is calculated as the 0.025 and 0.975 quantiles, respectively, of 1,000 bootstrapped replications (see methods).

RE		β_1	β_1 95% CI		β_2	β_2 95% CI		μ	μ 95% CI		σ	σ 95% CI	
levels	N	T = 2	Lower	Upper	T = 0	Lower	Upper	T = 0	Lower	Upper	T = 0.5	Lower	Upper
3	30	1.999	1.792	2.202	0.001	-0.199	0.205	0.000	-0.580	0.589	0.443	0.000	0.982
5	30	2.001	1.790	2.217	0.001	-0.207	0.211	0.001	-0.450	0.454	0.468	0.147	0.846
10	30	2.002	1.764	2.241	-0.003	-0.237	0.230	-0.001	-0.322	0.324	0.482	0.245	0.739
3	60	2.001	1.864	2.139	0.000	-0.135	0.134	0.002	-0.573	0.569	0.437	0.053	0.964
5	60	2.000	1.865	2.136	0.000	-0.140	0.135	0.001	-0.443	0.441	0.466	0.161	0.831
10	60	2.001	1.858	2.143	0.000	-0.144	0.142	-0.002	-0.314	0.305	0.485	0.265	0.736
3	120	2.000	1.910	2.090	0.000	-0.093	0.091	-0.001	-0.565	0.568	0.438	0.067	0.954
5	120	2.000	1.907	2.093	0.000	-0.091	0.093	-0.001	-0.442	0.443	0.469	0.164	0.838
10	120	2.000	1.905	2.093	0.000	-0.094	0.096	0.000	-0.318	0.311	0.485	0.267	0.735