## Harrison et al (2017) Type I Error Rate Simulation Methodology

For every simulation we generated a dataset of 500 rows, featuring a selection of continuous and/or categorical fixed effects, one random intercept grouping, and a Gaussian response variable.

Within these datasets, each of the desired number of continuous explanatory variables was produced by randomly drawing 500 times from a randomly selected normal distribution. The potential means of these distributions were 0, 10, 30 and 50 and the standard deviations 1, 5, 20, 30, 50 and 75. Similarly each categorical explanatory variable was created by assigning each datapoint a random category, with the number of categories for each variable varying between two and five.

The response variable was always randomly drawn from a normal distribution with a mean of zero and a standard deviation of fifty, and the random effect was always a six level factor in continuous blocks of random length.

Each full model consisted of zero to fifteen continuous variables and zero to five categorical variables, fitted using a Gaussian error structure.  1000 simulations were run for each combination of these numbers of categorical and continuous variables (n = 256 different scenarios) .

 All possible models using the terms in the full model were ranked via their Akaike information criterion (AIC) using the ‘dredge’ function in *MuMIn* (Barton 2016), after which model averaging was carried out on models within <6 Δ AIC of the top model (Richards 2008).

The p values from the model averaging results were then used to determine the number of Type I errors produced. We considered a type I error if the 95% confidence intervals from the model averaged parameter estimates of a coefficient did not cross zero.

Code for downstream processing of these results, including plotting, is provided as a R Markdown document on Figshare (see Data Accessibility).