LETTER

Misuse of hierarchical linear models overstates the significance of a reported association between *OXTR* and prosociality

Kogan et al. recently reported an association between an intronic variant in the oxytocin receptor gene, *OXTR*, and perceived prosociality (1). The significance of this association is described as $P < 0.001$, but the reported t statistic corresponds to $P = 3.2 \times 10^{-16}$. This level of significance is striking given the small number of genotyped individuals ($N = 23$) compared with other studies of complex trait genetics (2, 3). We show that this highly significant P value is a result of a misapplication of the statistical technique of hierarchical linear models (HLMs), and that the P value under an appropriate analysis is much larger: $P = 0.027$.

HLMs are a form of random effects model appropriate for data in which variation exists in multiple hierarchical "levels" (4). For example, academic outcomes of students from several schools depend both on the variance among schools (level 1 variance), and variance among students within each school (level 2 variance). By using an HLM to jointly estimate these two sources of variance, one could, for example, test whether socioeconomic status is correlated with a school's performance independent of the variance among its students. Crucially, HLMs assume that observations are independent, and that no correlation exists beyond the levels used in the model.

Kogan et al. (1) applied an HLM to 2,668 observations comprising the perceived prosociality of 23 genotyped individuals ("targets"), each measured by 116 observers. Level 1 of the HLM represented the variance between observers, and level 2 represented the variance across targets for each observer. While controlling for differences between observers, the model treats every observation as otherwise independent, failing to take into account the correlation that results from observing each target multiple times. It treats the data as if there were 2,668 targets, rather than 23 targets measured 116 times each.

To demonstrate that this approach can cause frequent false-positive findings, we reanalyzed raw data provided to us by Kogan et al. (1). We randomly permuted the genotypes of the 23 targets 1,000 times, and applied Kogan et al.'s HLM (1). The model falsely concluded that there was a correlation between *OXTR* genotype and prosociality at $P < 0.001$ in 55% of these null permutations, and in 10% of permutations, it reported a t statistic larger than that reported by the authors ($t = 8.22$) (1). A mixed-effects model that correctly accounts for correlation resulting from both observers and targets [using the R package lme4 (5)] yields a one-sided P value of 0.103. However, like the HLM, this model is probably more complicated than necessary, and a simple Spearman rank correlation test between the *OXTR* genotype and mean observed prosociality shows a nominally significant association ($\rho = 0.46$, two-sided $P = 0.027$).

Our reanalysis illustrates two important points. First, studies of complex trait genetics require very large sample sizes to be well powered, and a study reporting a highly significant association from such a small sample size should elicit skepticism. Second, unnecessarily complex statistical methods can obscure rather than reveal the truth. The simplest analysis is sometimes the best.

***Luke Jostins*[a], *Joseph K. Pickrell*[b], *Daniel G. MacArthur*[c], *and Jeffrey C. Barrett*[a]**

[a]*Statistical and Computational Genetics, Wellcome Trust Sanger Institute, Cambridge CB10 1HH, United Kingdom;* [b]*Department of Genetics, Harvard Medical School, Boston, MA 02115; and* [c]*Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA 02114*

1. Kogan A, et al. (2011) Thin-slicing study of the oxytocin receptor (OXTR) gene and the evaluation and expression of the prosocial disposition. *Proc Natl Acad Sci USA* 108: 19189–19192.
2. Ripke S, et al. (2011) Genome-wide association study identifies five new schizophrenia loci. *Nat Genet* 43:969–976.
3. Franke A, et al. (2010) Genome-wide meta-analysis increases to 71 the number of confirmed Crohn's disease susceptibility loci. *Nat Genet* 42:1118–1125.
4. Bryk A, Raudenbush S (1992) *Hierarchical Linear Models: Applications and Data Analysis Methods* (Sage, Thousand Oaks, CA).
5. Bates D, Maechler M (2010) *lme4: Linear Mixed-Effects Models Using S4 Classes. R Package Version 0.999375-37*. Available at http://cran.r-project.org/package=lme4. Accessed March 25, 2012.

Author contributions: L.J., D.G.M., and J.C.B. designed research; L.J. and J.K.P. performed research; L.J. and J.K.P. analyzed data; and L.J., D.G.M., and J.C.B. wrote the paper.

The authors declare no conflict of interest.

[1]To whom correspondence should be addressed. E-mail: lj4@sanger.ac.uk.