REPLY TO RISK AND ZHU: Mixed-effects modeling as a principled approach to heritability analysis with repeat measurements

Tian Ge\textsuperscript{a,b,c,1}, Avram J. Holmes\textsuperscript{a,d,e}, Randy L. Buckner\textsuperscript{a,e,f,g}, Jordan W. Smoller\textsuperscript{b,c}, and Mert R. Sabuncu\textsuperscript{a,h,i}

In ref. 1 we demonstrated that lumping together stable effects (e.g., unique environment) and transient effects (e.g., measurement error) in heritability analysis can be problematic when comparing traits with different levels of heritability. In particular, the conventional approach that averages repeat measurements (denoted as $h^2_{\text{avg}}$) underestimates the heritability of a trait in the presence of biological transients. We proposed a linear mixed-effects model that leverages repeat measurements to explicitly account for intra- and intersubject variation of a trait and produces unbiased heritability estimates (denoted as $h^2_{\text{rep}}$).

Risk and Zhu (2) point out that under our modeling assumptions the bias of $h^2_{\text{avg}}$ depends both on the magnitude of intrasubject variation and the number of repeat measurements. Specifically, the bias decreases as the measurement error reduces and the number of repeat measurements increases. Therefore, Risk and Zhu (2) suggest that averaging repeat measurements remains an acceptable approach when the intrasubject variability is low and/or a larger number of repeat measurements are available.

We agree with Risk and Zhu (2) that averaging more repeat measurements cancels out more intrasubject variation and reduces the bias of $h^2_{\text{avg}}$. The resting-state functional MRI example in ref. 1 represents a prototype case where a substantial amount of measurement noise exists in the data and only two scanning sessions were collected for each participant. In other scenarios where the measurement has higher test-retest reliability and/or more repeat measurements are available the difference between $h^2_{\text{rep}}$ and $h^2_{\text{avg}}$ may be substantially smaller.

However, we also note that (i) averaging repeat measurements only reduces but does not eliminate the bias; (ii) averaging repeat measurements cannot dissociate intra- and intersubject variation; and (iii) in many applications the measurement noise of a trait can be substantial and the number of repeat measurements may be limited due to cost, study design, and other factors. Therefore, we believe that the proposed statistical model is a principled approach to explicitly model and correct for the effect of intrasubject fluctuations on heritability estimates and can serve as a general approach to study the intra- and intersubject variation of a trait using repeat measurements across a range of situations.


\textsuperscript{a}Athinoula A. Martinos Center for Biomedical Imaging, Massachusetts General Hospital, Harvard Medical School, Charlestown, MA 02129; \textsuperscript{b}Psychiatric and Neurodevelopmental Genetics Unit, Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA 02114; \textsuperscript{c}Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, MA 02138; \textsuperscript{d}Department of Psychology, Yale University, New Haven, CT 06520; \textsuperscript{e}Department of Psychiatry, Massachusetts General Hospital, Cambridge, MA 02138; \textsuperscript{f}Department of Psychology, Harvard University, Cambridge, MA 02138; \textsuperscript{g}School of Electrical and Computer Engineering, Cornell University, Ithaca, NY 14853; and \textsuperscript{h}School of Electrical and Computer Engineering, Cornell University, Ithaca, NY 14853

The authors declare no conflict of interest.
Published under the PNAS license.

To whom correspondence should be addressed. Email: tge1@mgh.harvard.edu.