

Response to Sul and Eskin

Alkes L. Price, Noah A. Zaitlen, David Reich and Nick Patterson

We thank Sul and Eskin (Mixed models can correct for population structure for genomic regions under selection. *Nature Reviews Genetics* 26 Feb 2013 (doi:10.1038/nrg2813-c1))¹ for carefully examining and confirming the limitation of standard mixed model association methods that we identified in our Progress article (New approaches to population stratification in genome-wide association studies. *Nature Reviews Genetics* 11, 459–463 (2010))² and for developing an interesting new way to address it.

In our article², we investigated the limits of mixed model methods by considering an extreme simulation in which most markers had low population differentiation ($F_{ST} = 0.01$), but a small fraction of markers were unusually differentiated (allele frequency difference = 0.6). We found that standard mixed model methods³ did not fully correct for population structure, but mixed models with principal component covariates⁴ did fully correct for population structure. We stated that “population structure is a fixed effect, and spurious associations might result if it is modelled as a random effect based on overall covariance”.

Sul and Eskin¹ have confirmed that, in this extreme simulation, standard mixed model methods do not fully correct for population structure and that mixed models with principal component covariates do fully correct for population structure. They also investigated a new approach, which is to use a mixed model using two kinship matrices: one computed using unusually differentiated markers identified by their spatial ancestry analysis (SPA) method⁵, and one computed using the remaining markers. They reported that this approach also fully corrects for population structure in this simulation. Thus, population stratification (a fixed effect in this simulation) can be addressed using random effects in a way that we had not previously considered: our review considered only mixed models with a single random effect based on overall

covariance^{3,4,6–8} but did not consider mixed models with multiple random effects¹.

Another possibility, very similar to the Sul and Eskin¹ approach, is to use a mixed model that uses two kinship matrices — one computed from principal component 1, and one computed using the remaining principal components; this approach is based on the natural decomposition of a kinship matrix into its principal components⁹. This would also fully correct for population structure in this extreme simulation, as Sul and Eskin¹ showed that using a single kinship matrix computed from principal component 1 fully corrects for population structure.

A broader question is whether the limitation of standard mixed model methods that arises in this extreme simulation is a major concern in empirical studies. In our article², we stated that standard mixed model methods are an appealing and simple approach and are sufficient to correct for stratification in many settings. Sul and Eskin¹ indicated that the limitation we described did not arise in the Finnish and UK data sets that they analysed. We agree that mixed models with a single random effect based on overall covariance will probably be sufficient to correct for population structure fully in most settings.

Finally, we note that recent work has raised additional points about mixed model methods, including inclusion versus exclusion of the candidate marker in the kinship matrix, use of only a small subset of markers in computing the kinship matrix and effects of case–control ascertainment^{10–13}. We believe that these are important points that merit further investigation, but this is outside the scope of the current Correspondence.

Alkes L. Price is at the Department of Epidemiology, Harvard School of Public Health, Boston, Massachusetts 02115, USA; the Department of Biostatistics, Harvard School of Public Health, Boston, Massachusetts 02115, USA; and the Program in Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, Massachusetts 02142, USA.

Noah A. Zaitlen is at the Department of Medicine, Lung Biology Center, University of California San Francisco, San Francisco, California 94158, USA.

David Reich is at the Program in Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, Massachusetts 02142, USA; and the Department of Genetics, Harvard Medical School, Boston, Massachusetts 02115, USA.

Nick Patterson is at the Program in Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, Massachusetts 02142, USA.

Correspondence to A.L.P.
e-mail: aprice@hsph.harvard.edu

doi:10.1038/nrg2813-c2
Published online 26 February 2013

1. Sul, J. H. & Eskin, E. Mixed models can correct for population structure for genomic regions under selection. *Nature Rev. Genet.* 26 Feb 2013 (doi:10.1038/nrg2813-c1).
2. Price, A. L., Zaitlen, N. A., Reich, D. & Patterson, N. New approaches to population stratification in genome-wide association studies. *Nature Rev. Genet.* 11, 459–463 (2010).
3. Kang, H. M. *et al.* Variance component model to account for sample structure in genome-wide association studies. *Nature Genet.* 42, 348–354 (2010).
4. Zhang, Z. *et al.* Mixed linear model approach adapted for genome-wide association studies. *Nature Genet.* 42, 355–360 (2010).
5. Yang, W. Y., Novembre, J., Eskin, E. & Halperin, E. A model-based approach for analysis of spatial structure in genetic data. *Nature Genet.* 44, 725–731 (2012).
6. Segura, V. *et al.* An efficient multi-locus mixed-model approach for genome-wide association studies in structured populations. *Nature Genet.* 44, 821–830 (2012).
7. Zhou, X. & Stephens, M. Genome-wide efficient mixed-model analysis for association studies. *Nature Genet.* 44, 821–824 (2012).
8. Vilhjalmsdottir, B. J. & Nordborg, M. The nature of confounding in genome-wide association studies. *Nature Rev. Genet.* 14, 1–2 (2013).
9. Janss, L., de Los Campos, G., Sheehan, N. & Sorensen, D. Inferences from genomic models in stratified populations. *Genetics* 192, 693–704 (2012).
10. Sawcer, S. *et al.* Genetic risk and a primary role for cell-mediated immune mechanisms in multiple sclerosis. *Nature* 476, 214–219 (2011).
11. Lippert, C. *et al.* FaST linear mixed models for genome-wide association studies. *Nature Methods* 8, 833–835 (2011).
12. Listgarten, J. *et al.* Improved linear mixed models for genome-wide association studies. *Nature Methods* 9, 525–526 (2012).
13. Mefford, J. & Witte, J. S. The covariate’s dilemma. *PLoS Genet.* 8, e1003096 (2012).

Acknowledgements

We are grateful to E. Eskin, P. Visscher, J. Yang and M. Goddard for helpful discussions.

Competing interests statement

The authors declare no competing financial interests.

FURTHER INFORMATION

Alkes L. Price’s homepage: <http://www.hsph.harvard.edu/alkes-price>

ALL LINKS ARE ACTIVE IN THE ONLINE PDF