



Published in final edited form as:

Nat Rev Genet. 2013 April ; 14(4): 300. doi:10.1038/nrg2813-c2.

Response to Sul & Eskin

Alkes L. Price^{1,2,3}, **Noah A. Zaitlen**⁴, **David Reich**^{3,5}, and **Nick Patterson**³

¹Department of Epidemiology, Harvard School of Public Health, Boston, MA 02115, USA

²Department of Biostatistics, Harvard School of Public Health, Boston, MA 02115

³Program in Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, MA 02142, USA

⁴Department of Medicine, Lung Biology Center, University of California San Francisco, San Francisco, CA, 94158

⁵Department of Genetics, Harvard Medical School, Boston, MA 02115

We thank Sul & Eskin for carefully examining and confirming the limitation of standard mixed model association methods that we identified in our review, and for developing an interesting new way to address it.

In Price et al. 2010¹, 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 methods² did not fully correct for population structure, but mixed models with PC covariates³ did fully correct for population structure. We stated that “population structure is a fixed effect, and spurious associations might result if it is modeled as a random effect based on overall covariance”.

Sul & Eskin⁴ have confirmed that in this extreme simulation, standard mixed model methods do not fully correct for population structure, and that mixed models with PC covariates do fully correct for population structure. They also investigated a new approach, which is to employ a mixed model using two kinship matrices, one computed using unusually differentiated markers identified by their 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 only considered mixed models with a single random effect based on overall covariance^{2-3, 6-8}, but did not consider mixed models with multiple random effects⁴.

Another possibility, very similar to the Sul & Eskin approach, is to employ a mixed model using two kinship matrices, one computed from PC1 and one computed using the remaining PCs, based on the natural decomposition of a kinship matrix into its PCs⁹. This would also

fully correct for population structure in this extreme simulation, since Sul & Eskin showed that using a single kinship matrix computed from PC1 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 Price et al. 2010¹, we stated that standard mixed model methods are an appealing and simple approach and are sufficient to correct for stratification in many settings. Sul & Eskin indicated that the limitation we described did not arise in the Finnish and UK data sets that they analyzed. We agree that mixed models with a single random effect based on overall covariance will likely be sufficient to fully correct for population structure in most settings.

We finally note that recent work has raised additional points about mixed model methods, including inclusion vs. 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, which is however outside the scope of the current Correspondence.

Acknowledgments

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

References

1. Price AL, Zaitlen NA, Reich D, Patterson N. New approaches to population stratification in genome-wide association studies. *Nat Rev Genet.* 2010; 11:459–63. [PubMed: 20548291]
2. Kang HM, Sul JH, Service SK, Zaitlen NA, Kong SY, Freimer NB, Sabatti C, Eskin E. Variance component model to account for sample structure in genome-wide association studies. *Nat Genet.* 2010; 42:348–54. [PubMed: 20208533]
3. Zhang Z, Ersoz E, Lai CQ, Todhunter RJ, Tiwari HK, Gore MA, Bradbury PJ, Yu J, Arnett DK, Ordovas JM, Buckler ES. Mixed linear model approach adapted for genome-wide association studies. *Nat Genet.* 2010; 42:355–60. [PubMed: 20208535]
4. Sul JH, Eskin E. Mixed models can correct for population structure for genomic regions under selection. *Nat Rev Genet.* in press.
5. Yang WY, Novembre J, Eskin E, Halperin E. A model-based approach for analysis of spatial structure in genetic data. *Nat Genet.* 2012; 44:725–31. [PubMed: 22610118]
6. Segura V, Vilhjalmsón BJ, Platt A, Korte A, Seren U, Long Q, Nordborg M. An efficient multi-locus mixed-model approach for genome-wide association studies in structured populations. *Nat Genet.* 2012; 44:825–30. [PubMed: 22706313]
7. Zhou X, Stephens M. Genome-wide efficient mixed-model analysis for association studies. *Nat Genet.* 2012
8. Vilhjalmsón BJ, Nordborg M. The nature of confounding in genome-wide association studies. *Nat Rev Genet.* 2012; 14:1–2. [PubMed: 23165185]
9. Janss L, de Los Campos G, Sheehan N, Sorensen D. Inferences from genomic models in stratified populations. *Genetics.* 2012; 192:693–704. [PubMed: 22813891]
10. Sawcer S, Hellenthal G, Pirinen M, Spencer CC, Patsopoulos NA, Moutsianas L, Dilthey A, Su Z, Freeman C, Hunt SE, Edkins S, Gray E, Booth DR, Potter SC, Goris A, Band G, Oturai AB, Strange A, Saarela J, Bellenguez C, Fontaine B, Gillman M, Hemmer B, Gwilliam R, Zipp F, Jayakumar A, Martin R, Leslie S, Hawkins S, Giannoulatou E, D'Alfonso S, Blackburn H, Martinelli Boneschi F, Liddle J, Harbo HF, Perez ML, Spurkland A, Waller MJ, Mycko MP, Ricketts M, Comabella M, Hammond N, Kockum I, McCann OT, Ban M, Whittaker P, Kempainen A, Weston P, Hawkins C, Widaa S, Zajicek J, Dronov S, Robertson N, Bumpstead SJ, Barcellos

LF, Ravindrarajah R, Abraham R, Alfredsson L, Ardlie K, Aubin C, Baker A, Baker K, Baranzini SE, Bergamaschi L, Bergamaschi R, Bernstein A, Berthele A, Boggild M, Bradfield JP, Brassat D, Broadley SA, Buck D, Butzkueven H, Capra R, Carroll WM, Cavalla P, Celius EG, Cepok S, Chiavacci R, Clerget-Darpoux F, Clysters K, Comi G, Cossburn M, Cournu-Rebeix I, Cox MB, Cozen W, Cree BA, Cross AH, Cusi D, Daly MJ, Davis E, de Bakker PI, Debouverie M, D'Hooghe MB, Dixon K, Dobosi R, Dubois B, Ellinghaus D, Elovaara I, Esposito F, Fontenille C, Foote S, Franke A, Galimberti D, Ghezzi A, Glessner J, Gomez R, Gout O, Graham C, Grant SF, Guerini FR, Hakonarson H, Hall P, Hamsten A, Hartung HP, Heard RN, Heath S, Hobart J, Hoshi M, Infante-Duarte C, Ingram G, Ingram W, Islam T, Jagodic M, Kabesch M, Kermodé AG, Kilpatrick TJ, Kim C, Klopp N, Koivisto K, Larsson M, Lathrop M, Lechner-Scott JS, Leone MA, Leppa V, Liljedahl U, Bomfim IL, Lincoln RR, Link J, Liu J, Lorentzen AR, Lupoli S, Macchiardi F, Mack T, Marriott M, Martinelli V, Mason D, McCauley JL, Mentch F, Mero IL, Mihalova T, Montalban X, Mottershead J, Myhr KM, Naldi P, Ollier W, Page A, Palotie A, Pelletier J, Piccio L, Pickersgill T, Piehl F, Pobywajlo S, Quach HL, Ramsay PP, Reunanen M, Reynolds R, Rioux JD, Rodegher M, Roesner S, Rubio JP, Ruckert IM, Salvetti M, Salvi E, Santaniello A, Schaefer CA, Schreiber S, Schulze C, Scott RJ, Sellebjerg F, Selmaj KW, Sexton D, Shen L, Simms-Acuna B, Skidmore S, Sleiman PM, Smestad C, Sorensen PS, Sondergaard HB, Stankovich J, Strange RC, Sulonen AM, Sundqvist E, Syvanen AC, Taddeo F, Taylor B, Blackwell JM, Tienari P, Bramon E, Tourbah A, Brown MA, Tronczynska E, Casas JP, Tubridy N, Corvin A, Vickery J, Jankowski J, Villoslada P, Markus HS, Wang K, Mathew CG, Wason J, Palmer CN, Wichmann HE, Plomin R, Willoughby E, Rautanen A, Winkelmann J, Wittig M, Trembath RC, Yaouanq J, Viswanathan AC, Zhang H, Wood NW, Zuvich R, Deloukas P, Langford C, Duncanson A, Oksenberg JR, Pericak-Vance MA, Haines JL, Olsson T, Hillert J, Ivinson AJ, De Jager PL, Peltonen L, Stewart GJ, Hafler DA, Hauser SL, McVean G, Donnelly P, Compston A. Genetic risk and a primary role for cell-mediated immune mechanisms in multiple sclerosis. *Nature*. 2011; 476:214–9. [PubMed: 21833088]

11. Lippert C, Listgarten J, Liu Y, Kadie CM, Davidson RI, Heckerman D. FaST linear mixed models for genome-wide association studies. *Nat Methods*. 2011; 8:833–5. [PubMed: 21892150]
12. Listgarten J, Lippert C, Kadie CM, Davidson RI, Eskin E, Heckerman D. Improved linear mixed models for genome-wide association studies. *Nat Methods*. 2012; 9:525–6. [PubMed: 22669648]
13. Mefford J, Witte JS. The Covariate's Dilemma. *PLoS Genet*. 2012; 8:e1003096. [PubMed: 23162385]