

# Human Demographic History Impacts Genetic Risk Prediction across Diverse Populations

Alicia R. Martin, Christopher R. Gignoux, Raymond K. Walters, Genevieve L. Wojcik, Benjamin M. Neale, Simon Gravel, Mark J. Daly, Carlos D. Bustamante, and Eimear E. Kenny\*

(The American Journal of Human Genetics 100, 635–649; April 6, 2017)

We thank Aaron Ragsdale, Dominic Nelson, and Simon Gravel for identifying a coding error in the setup of the demographic simulations shown in Figure 5 that changed the magnitude of simulation results regarding population differences. The interpretation of Figure 5 that polygenic risk scores do not generalize well across populations remains the same, although the magnitude of differences shown in Figure 5C are considerably diminished (see Ragsdale et al.<sup>1</sup> Figure 2 in this issue of AJHG for an update).

The code used for these original simulations is publicly available at the following web address: [https://github.com/armartin/ancestry\\_pipeline/blob/master/simulate\\_prs.py](https://github.com/armartin/ancestry_pipeline/blob/master/simulate_prs.py). In this code, there were three main steps in setting up the demographic component of the simulation, and the error occurred in the third step.

1. The code specified a demographic model from a previous study<sup>2</sup> that reflects the inferred history of African, East Asian, and European ancestry populations in the 1000 Genomes Project according to msprime's documentation (note that Ragsdale et al.<sup>1</sup> "Case #1" also identified a subtle model misspecification carried forward from the documentation of msprime, but it had little practical effect). This model of population history consisted of three main parameters: (1) population configurations (this specifies sample sizes, the initial population size, and growth rates), (2) migration matrices (this specifies the migration parameters among the 3 pairs of populations), and (3) demographic events (this specifies mass migration events [e.g., historical merging of these three populations], migration rate changes, and timings of these events).
2. The model setup was checked with parameters confirmed with msprime debugging tools.
3. The population configurations and migration matrices were then passed into the simulation used in the generation of Figure 5. However, these demographic event parameters (1.3 above) were erroneously *not* passed on to the simulation.

The error introduced by failing to pass the demographic event parameters resulted in three simulated populations that did not merge in the past. Instead the three modeled populations remained isolated with low levels of migration between each pair. Consequently, the simulated populations were much more genetically differentiated than is realistic in humans. Ragsdale et al.<sup>1</sup> have rerun these simulations with the demographic model specifications that we intended to use and found that the differences inferred in PRS distributions across populations in Figure 5B are vastly diminished compared to our results.

This error did not affect any other figures or empirical results in this article, and we originally interpreted Figures 4 and 5B as providing consistent empirical- and simulation-based evidence, respectively, for significant mean shifts in inferred PRS across populations relative to the true underlying distribution. The empirical findings that PRS predictions vary across populations have also been borne out in considerable empirical work based on data from diverse biobanks not available at the time of this paper's publication.<sup>3</sup>

The updated simulation analysis calls into question our original interpretation that genetic drift alone can explain nearly all of the differences. In work conducted since this paper's publication, several additional factors beyond linkage disequilibrium and allele frequency differences across populations have further highlighted how challenging the issues of PRS generalizability are. Some issues that have been explored further include residual population stratification in genome-wide association studies (GWASs), winner's curse, cohort ascertainment effects, background and negative selection, gene x gene and gene x environment interactions, and other complex factors.<sup>3–12</sup> Early evidence suggests that the relative quantitative contributions of these factors can be trait-, population-, and context-specific,<sup>10,11</sup> highlighting that calibration of PRS ought to jointly consider these complexities.

\*Correspondence: [eimear.kenny@mssm.edu](mailto:eimear.kenny@mssm.edu)

<https://doi.org/10.1016/j.ajhg.2020.08.020>

© 2017 American Society of Human Genetics.

Recent work has found variability in the magnitude of differences among PRS distributions across populations depending on the study in which GWAS summary statistics were derived.<sup>7,8</sup> We have now further considered the fact that Figure 4 in our publication showed the most pronounced PRS differences for height across populations, which were calculated with GWAS summary statistics from GIANT. The more recent work showed that GWASs of height from GIANT produced large differences in polygenic score distributions across populations but that these differences are substantially attenuated when using GWASs from the UK Biobank.<sup>7,8</sup> In-depth analyses by these studies indicate that residual population stratification in GIANT summary statistics arising from meta-analysis of smaller, heterogeneous cohorts produced larger than expected distributional differences across populations. Taken together, these results hint that the inferred PRS distribution is not necessarily expected to vary substantially across populations provided that the discovery cohort is very well controlled for population structure. However, this standard is not routinely met by GWASs, and this suggestion also needs to be examined further; specifically, a more recent study using relatively homogeneous data from the Finnish population harmonized in a manner similar to the UK Biobank still found evidence of overpredicted differences across the country.<sup>9</sup> Overall, these complexities indicate that interpreting differences in polygenic scores across populations is non-trivial.

The simulations in this study provided an early guide for how we think about PRS differences across populations, and we regret that this error produced an oversimplified explanation to large inferred differences across populations. Since this study's publication, a community-maintained repository of demographic models has been developed in *stdpopsim* with quality-control procedures in place to potentially prevent such modeling implementation mistakes in the future.<sup>13</sup> The central message that human history impacts PRS remains unchanged, as do implications that large studies encompassing more diverse ancestries are needed to more equitably increase PRS accuracies. We thank our colleagues for identifying this error and helping correct the record. Although this error was unfortunate, its identification and correction are testament to the power of open science and public sharing of code and data.

## References

1. Ragsdale, A.P., Nelson, D., Gravel, S., and Kelleher, J. (2020). Lessons learned from bugs in models of human history. *Am. J. Hum. Genet.* 107.
2. Gravel, S., Henn, B.M., Gutenkunst, R.N., Indap, A.R., Marth, G.T., Clark, A.G., Yu, F., Gibbs, R.A., Bustamante, C.D.; and 1000 Genomes Project (2011). Demographic history and rare allele sharing among human populations. *Proc. Natl. Acad. Sci. USA* 108, 11983–11988.
3. Martin, A.R., Kanai, M., Kamatani, Y., Okada, Y., Neale, B.M., and Daly, M.J. (2019). Clinical use of current polygenic risk scores may exacerbate health disparities. *Nat. Genet.* 51, 584–591.
4. Novembre, J., and Barton, N.H. (2018). Tread Lightly Interpreting Polygenic Tests of Selection. *Genetics* 208, 1351–1355.
5. Rosenberg, N.A., Edge, M.D., Pritchard, J.K., and Feldman, M.W. (2018). Interpreting polygenic scores, polygenic adaptation, and human phenotypic differences. *Evol. Med. Public Health* 2019, 26–34.
6. Durvasula, A., and Lohmueller, K.E. (2019). Negative selection on complex traits limits genetic risk prediction accuracy between populations. *bioRxiv*.
7. Sohail, M., Maier, R.M., Ganna, A., Bloemendal, A., Martin, A.R., Turchin, M.C., Chiang, C.W., Hirschhorn, J., Daly, M.J., Patterson, N., et al. (2019). Polygenic adaptation on height is overestimated due to uncorrected stratification in genome-wide association studies. *eLife* 8, e39702.
8. Berg, J.J., Harpak, A., Sinnott-Armstrong, N., Joergensen, A.M., Mostafavi, H., Field, Y., Boyle, E.A., Zhang, X., Racimo, F., Pritchard, J.K., and Coop, G. (2019). Reduced signal for polygenic adaptation of height in UK Biobank. *eLife* 8, e39725.
9. Kerminen, S., Martin, A.R., Koskela, J., Ruotsalainen, S.E., Havulinna, A.S., Surakka, I., Palotie, A., Perola, M., Salomaa, V., Daly, M.J., et al. (2019). Geographic Variation and Bias in the Polygenic Scores of Complex Diseases and Traits in Finland. *Am. J. Hum. Genet.* 104, 1169–1181.
10. Wang, Y., Guo, J., Ni, G., Yang, J., Visscher, P.M., and Yengo, L. (2020). Theoretical and empirical quantification of the accuracy of polygenic scores in ancestry divergent populations. *bioRxiv*. <https://doi.org/10.1101/2020.01.14.905927>.
11. Bitarello, B.D., and Mathieson, I. (2020). Polygenic scores for height in admixed populations. *bioRxiv*. <https://doi.org/10.1101/2020.04.08.030361>.
12. Mostafavi, H., Harpak, A., Agarwal, I., Conley, D., Pritchard, J.K., and Przeworski, M. (2020). Variable prediction accuracy of polygenic scores within an ancestry group. *eLife* 9, e48376.
13. Adrion, J.R., Cole, C.B., Dukler, N., Galloway, J.G., Gladstein, A.L., Gower, G., Kyriazis, C.C., Ragsdale, A.P., Tsambos, G., Baumdicker, F., et al. (2020). A community-maintained standard library of population genetic models. *bioRxiv*. <https://doi.org/10.1101/2019.12.20.885129>.