# Genetic Selectivity in Migration: Polygenic Insights into Heterogeneous Outcomes of Domestic Migration in the US
## Extended Abstract

Yuxin Liu
Sociology Department, Princeton University
yl9858@princeton.edu

### *Introduction*

Studies on migration outcomes often face significant identification challenges, particularly selection bias and unobserved confounders such as education, family socioeconomic status (SES), and health. Genetic data offers a potential solution to some of these issues, as the innate nature of genetic traits helps mitigate the empirical absence of pre-migration data. The broader purpose of this research is to understand (a) the effect of genetic endowment on various types of movement; and (b) the role of genetic selectivity on migration outcomes.

Using AddHealth data, this study first examines the direction and magnitude of genetic selectivity of U.S. internal migrants in the U.S. by employing polygenic scores (PGS) for migration distance, educational attainment, health, and risk-taking behavioral traits. Beyond the aggregated level picture, I specifically attend to the different genetic profiles and selectivity among various types of migrants, especially long-distance and short-distance migrants. Does the migration distance PGS monotonically predict actual migration patterns? Are there opposite types of selectivity for long- and short-distance migrants? Based on these descriptive findings, the study then proposes an improved identification strategy for assessing the causal impact of the observed genetic selection criteria on mobility behaviors.

### *Theoretical Framework*

Two competing models exist for the understanding of migration selectivity and outcomes. The first approach suggests that migration requires determination and resources and therefore is positively selected and leads to advantageous results (Chiswick, 1999; Feliciano, 2005; Abramitzky & Boustan, 2017; Aksoy & Poutvaara, 2021). The second approach, by contrast, argues that migration produces broken social ties and interrupted routines, thus bringing about negative consequences such as increased delinquency. The second approach also suggests that migration attracts risk-seeking people and those with disadvantaged socioeconomic status, especially among short-distance, frequent, and compelled movers (Garboden et al., 2017; Kogan et al., 2011; Sharkey & Sampson, 2010).

The two theories, in fact, are not mutually exclusive; they point in conflicting directions only because they are built on different sets of assumptions. The first positive model assumes that migrants are voluntarily choosing to relocate to achieve specific economic goals, and once they succeed their improved situation will yield better outcomes. This applies more effectively to long-distance migrants, especially those involved in border crossings, i.e., international migrants. On the other hand, the second approach attends to the fact that many migrants are less self-driven and more dependent on environmental constraints. Short-distance, more frequent migration is usually included in that interpretation. This study attempts to use empirical causal analysis to evaluate these two seemingly conflicting theories by disentangling the genetic endowment of migrants and the migration outcomes.

### *Methods*

The PGI for educational attainment will be accessed directly from the corresponding data, as it is often calculated by the data providers. The PGI for migration distance will be calculated using the GWAS results from Furuya et al. (2023). For crime-prone behavioral traits, PGI will include those for Antisocial personality disorder (Rautiainen et al., 2016), Aggressive behavior (Pappa et al., 2016; Barnes et al., 2019), and Risk-prone behavior (Ksinan et al., 2022). Additional PGI, such as those related to health-related traits, will be included where available, depending on the specific datasets used.

This study will proceed in three main steps. First, I will replicate the method utilized by Furuya et al. (2023b) and synthesize the GWAS summary data from several previous studies to analyze the genetic correlation between migration distance, educational attainment, health, and crime-prone behavioral traits. This analysis will provide a descriptive overview of the relationships between these traits and serve as a validity check by comparing results with existing literature.

To address portability concerns, I will assess the robustness of the PGI for migration distance and other key phenotypes across the datasets to address the portability concerns, especially among the population that has greater ancestry distance from the training sample, in most cases the UK Biobank (Wang et al., 2024). In particular, I will evaluate whether individuals with higher PGI for migration distance are indeed more likely to migrate longer distances within each dataset. This step serves as an additional validation of the PGI and its association with the corresponding traits in the study population.

Second, I will compare the PGI for educational attainment, socioeconomic status, crime-prone traits, and health measures across different migration groups. These comparisons will include migrants versus non-migrants, as well as long-distance, less frequent migrants versus short-distance, more frequent migrants. Depending on the data structure, more nuanced categorizations, such as high SES and low SES migrants, might also be compared. This comprehensive analysis will provide a detailed and nuanced understanding of the genetic selectivity of migrants across diverse contexts.

Lastly, I will evaluate whether the genetic selectivity found in previous steps is indeed causally correlated with migration behavior by adopting a within-family design. By controlling for parents' genotype, the randomly inherited genetic variants can serve as a natural experiment to address the identification issues of confounding and reverse causation (Davies et al., 2018). Mendelian imputation will be utilized to calculate the parents' genotype when one of the parents' data is missing (Young et al., 2022).

### ***Hypothesis***

1. Long-distance migrants are genetically positively selected in terms of health, socioeconomic status, and crime-prone traits, while short-distance, more frequent movers have negative genetic selectivity in terms of health, socioeconomic status, and crime-prone traits.
2. Those with a higher genetic propensity to move long distances, regardless of their actual migration actions, also have a higher genetic propensity for better socioeconomic status and health.
3. The positive and negative genetic endowments found in previous sections are causally related to long- and short-distance mobility behaviors.

### ***Significance***

The findings of this paper can bridge the gap between two competing theories of migration outcomes. By demonstrating that long-distance migration is positively selected while short-distance migration is more likely to attract individuals with risk-prone or disadvantaged profiles, the study will highlight the coexistence of these theoretical frameworks under different migration contexts. This nuanced understanding will help refine sociological theories on migration and mobility.

Through the application of within-family designs and Mendelian imputation, the study aims to establish causal relationships between genetic selectivity and migration outcomes. This approach will disentangle the effects of genetic endowment from environmental and parental influences, providing robust evidence for how genetic traits influence migration behaviors.

The genetic insights from this research have the potential to inform policies aimed at addressing the socioeconomic and health disparities associated with migration. By identifying the genetic and environmental factors that contribute to different migration outcomes, policymakers can better target interventions to support vulnerable migrant populations, particularly those engaged in short-distance or frequent moves.

## Bibliography

Abramitzky, R., & Boustan, L. (2017). Immigration in American Economic History. *Journal of Economic Literature*, *55*(4), 1311–1345. https://doi.org/10.1257/jel.20151189

Aksoy, C. G., & Poutvaara, P. (2021). Refugees' and irregular migrants' self-selection into Europe. *Journal of Development Economics*, *152*, 102681. https://doi.org/10.1016/j.jdeveco.2021.102681

Chiswick, B. (1999). Are Immigrants Favorably Self-Selected? *American Economic Review*, *89*(2), 181–185. https://doi.org/10.1257/aer.89.2.181

Davies, N. M., Holmes, M. V., & Davey Smith, G. (2018). Reading Mendelian randomisation studies: A guide, glossary, and checklist for clinicians. *BMJ (Clinical Research Ed.)*, *362*, k601. https://doi.org/10.1136/bmj.k601

Feliciano, C. (2005a). Does Selective Migration Matter? Explaining Ethnic Disparities in Educational Attainment among Immigrants' Children. *The International Migration Review*, *39*(4), 841–871.

Furuya, S., Liu, J., Sun, Z., Lu, Q., & Fletcher, J. M. (2023a). The Big (Genetic) Sort? A Research Note on Migration Patterns and Their Genetic Imprint in the United Kingdom. *Demography*, *60*(6), 1649–1664. https://doi.org/10.1215/00703370-11054960

Furuya, S., Liu, J., Sun, Z., Lu, Q., & Fletcher, J. M. (2023b). Understanding Internal Migration: A Research Note Providing an Assessment of Migration Selection With Genetic Data. *Demography*, *60*(6), 1631–1648. https://doi.org/10.1215/00703370-11053145

Garboden, P. M., Leventhal, T., & Newman, S. (2017). Estimating the Effects of Residential Mobility: A Methodological Note. *Journal of Social Service Research*, *43*(2), 246–261. https://doi.org/10.1080/01488376.2017.1282392

Kogan, I., Kalter, F., Liebau, E., & Cohen, Y. (2011). Individual Resources and Structural Constraints in Immigrants' Labour Market Integration. In M. Wingens, M. Windzio, H. de Valk, & C. Aybek (Eds.), *A Life-Course Perspective on Migration and Integration* (pp. 75–100). Springer Netherlands. https://doi.org/10.1007/978-94-007-1545-5_4

Kihara, T. (2022). Socioeconomic selectivity of Japanese migration to the continental United States during the Age of Mass Migration. *Journal of Ethnic and Migration Studies*, *48*(11), 2577–2600. https://doi.org/10.1080/1369183X.2021.1963219

Sharkey, P., & Sampson, R. J. (2010). Destination Effects: Residential Mobility and Trajectories of Adolescent Violence in a Stratified Metropolis. *Criminology*, *48*(3), 639–681. https://doi.org/10.1111/j.1745-9125.2010.00198.x

Wang, J. Y., Lin, N., Zietz, M., Mares, J., Narasimhan, V. M., Rathouz, P. J., & Harpak, A. (2024). Three Open Questions in Polygenic Score Portability. *bioRxiv*, 2024.08.20.608703. https://doi.org/10.1101/2024.08.20.608703

Young, A. I., Nehzati, S. M., Benonisdottir, S., Okbay, A., Jayashankar, H., Lee, C., Cesarini, D., Benjamin, D. J., Turley, P., & Kong, A. (2022). Mendelian imputation of parental genotypes improves estimates of direct genetic effects. *Nature Genetics*, *54*(6), 897–905. https://doi.org/10.1038/s41588-022-01085-0