Assortative Mating in the U.K. Through the 20th Century

Patrick Turley[1] | Ruoxi Li[2] | Miles Kimball[3]
1) ATGU, Massachusetts General Hospital 2) Center for Economic and Social Research, University of Southern California 3) Department of Economics, University of Colorado Boulder

Background

- Assortative mating (AM) is the phenomenon where mating pairs have more similar phenotypic or genotypic characteristics than would be expected if they were paired randomly.
- AM can have important consequences on health and inequality and implications.
  - AM leads to an increase in additive genetic variance for the corresponding phenotype (Crow and Kimura 1970, Crow and Felsenstein 1982). For polygenic disease traits, this would result in more people with extreme genetic values and therefore a higher prevalence of the disease (Yengo 2018).
  - AM reflects important social phenomena such as geographic mobility, racism, and classism (Mare and Maralani 2006).
- We develop a method to measure how changes in AM over time affect correlations between genotypes and therefore the variance of the polygenic score (PGS) using cross-sectional data. PGSs include traditional PGSs for various phenotypes and principal components (PCs) of the genetic data.
- AM has an important effect on the cross-sectional variance of a PGS. Indeed, essentially all cross-chromosome covariances terms contributing to the variance of a PGS are due to assortative mating at some time in the past: ½ from the parents' generation, ¼ from the grandparents' generation, etc. AM in the parents' generation can be estimated by comparing the cross-chromosome contribution to the variance of the PGS the children's generation to what it was in the parents' generation.

Data

We estimate AM use cross-sectional UK Biobank data excluding parent-child pairs and adoptees, which include 425,210 individuals born between 1938 and 1970 (Sudlow et al. 2015).

Works Cited

Crow, James Franklin, and Motoo Kimura. 1970. *An introduction to population genetics theory.* New York, Harper & Row.
Crow, James Franklin and Joseph Felsenstein. 1982. "The effect of assortative mating on the genetic composition of a population." *Social Biology* 29(1-2): 22-35.
Mare, Robert D. and Vida Maralani. 2006. "The Intergenerational Effects of Changes in Women's Educational Attainments." *American Sociological Review* 71(4): 542-64.
Sudlow et al. 2015. "UK Biobank: An Open Access Resource for Identifying the Causes of a Wide Range of Complex Diseases of Middle and Old Age." *PLoS Med* 12(3): e1001779. doi: https://doi.org/10.1371/journal.pmed.1001779.
Yengo et. al. 2018. "Imprint of Assortative Mating on the Human Genome." bioRxiv 300020. doi: https://doi.org/10.1101/300020.

Method

Key Parameter of Interest:

$R(t)$: assortative mating parameter at time t. $R(t)$ can be obtained through either spousal correlation of polygenic scores, or through our method using cross-sectional polygenic scores. Note that $R(t) = M(t) / U(t)$, where $M(t)$ is the genetic covariance of spouses conceiving a child at time t and $U(t)$ is their variance.

Key Equation:

Define cross-chromosome component of covariance as the aggregate of the covariance of all pairs of SNPs or genotypes *not* on the same chromosome. Denote $M_X(t)$ as the cross-chromosome component of covariance between the PGSs of spouses conceiving children at time t. Denote $C_X(t)$ as the cross-chromosome component of the variance of the polygenic scores due to AM. Denote G as the difference in the time of conception between a child and her parents, and denote $\mathrm{E}_G(\cdot)$ as the expectation operator over the distribution of G. Then we have:

$$M_X(t) = 2C_X(t) - E_G[C_X(t-G)] \qquad [1]$$

Derivation of Key Equation:

Denote $V(t)$ as variance of the PGSs for individuals conceived at time t, L as variance of the PGSs due to true genetic effect and linkage equilibrium, $C(t)$ as additive variance of the polygenic scores due to AM. We can decompose the variance $V(t)$ such that $V(t) = L + C(t)$. We can further decompose $C(t) = \sum_{jk} c_{jk}(t)$, where $c_{jk}(t)$ is additive variance due to AM at a given pair of genotypes *j,k*.

$c_{jk}(t)$ is related to AM in present times and in previous generations. For any pair of genotypes, j and k, they were either inherited from different parents or the same parent. If they were inherited from different parents, then $c_{jk}(t)$ measures assortative mating $m_{jk}(t)$ at genotype pair *j,k.* If they were inherited from the same parent, then $c_{jk}(t)$ measures assortative mating at genotype pair *j,k* in previous generations. For cross-chromosome genotype pair where $j \neq k$, the probability that j and k are inherited from different parents or the same parents is ½. Therefore, in mathematical notations, we have:

$$c_{jk}(t) = 1/2 m_{jk}(t) + 1/2 E_G\left[c_{jk}(t-G)\right] \qquad [2]$$

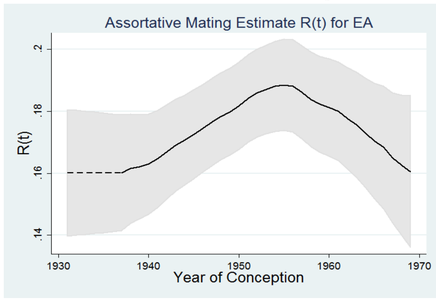
Aggregating all cross-chromosome pairs of genotypes in equation 2 and rearranging the order gives us equation 1.

Empirical Estimation of $R(t)$:

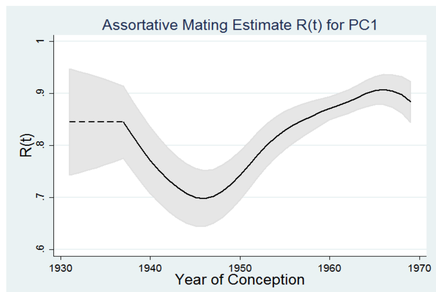
Let S_{ki} be polygenic score for an individual i on chromosome κ. We can estimate $C_X(t)$ by calculating for each individual $D_i = S_i^2 - \sum_\kappa S_{\kappa i}^2$. Then we run kernel regression using the values of D_i to produce a nonparametric estimate $\hat{C}_X(t)$. Next, we obtain $\widehat{M}_X(t)$ from equation [1]. Then, we estimate $\widehat{M}(t)$ by inflating $\widehat{M}_X(t)$ by a scaling factor λ to account for within-chromosome covariance, assuming PGS signal is evenly spread within and cross chromosomes. Finally, we use the expected population variance in the previous generation to estimate $U(t)$ and obtain AM estimate $\hat{R}(t) = \widehat{M}(t) / \widehat{U}(t)$.

Results

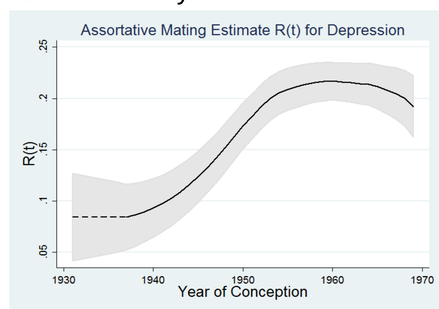
- Estimate of AM for **Educational Attainment** residualized by 20 PCs



- WWII and Estimate of AM for **PC1**



- Estimate of AM for **Depression** residualized by 20 PCs



- If AM stays at the same level past 1970, the variance of the depression PGSs will continue to increase. Corresponding prediction: an increase in the prevalence of depression.

Extensions

- Cross-trait AM.
- Addressing sample selection by mortality and response rates.
- We will validate our measure of AM using spousal pairs in other datasets.