Wave IV Education Genetic Risk Score

Introduction

This guide describes the education genetic risk score (GRS_EDU) for Add Health twin and full sibling respondents who provided saliva samples at Wave IV. This variable is the weighted sum of risk alleles identified in the Rietveld et al. (2013) genome-wide association study for the respondents of the Add Health Sibling Pairs sample. An overview of the predictive performance of the variable (amongst the black and white respondents) can be found in Domingue et al. (2015).

Construction

GRS_EDU

After quality controls (additional details can be found in McQueen et al., 2014), the genetic database included 1,886 individuals with valid data on 940,862 single nucleotide polymorphisms.

Genotyping was conducted with the Illumina HumanOmni1-Quad v1 platform using DNA extracted (via Oragene saliva collection) from individuals at Wave IV. Complete details of the QC process that resulted in the data used here are available in McQueen et al. (2014). We removed 18,665 SNPs from the original panel of genetic data (based on a missingness threshold of 5% computed in a sample of individuals who contained information on at least 90% of all available SNPs) to arrive at a genetic database consisting of 940,862 SNPs. In the original QC analysis, 74 genetic samples (some of which may have been duplicates included as part of the QA process) were dropped due to missingness concerns. SNPs in the Add Health Sibling Pairs genetic database were matched to SNPs with reported results in the educational attainment GWAS (Rietveld et al., 2013). Over 2/3 of the SNPs in the Add Health genetic database were included in the GWAS results (642,627 SNPs). For each of these SNPs, a loading was calculated as the number of education associated alleles multiplied by the effect-size estimated in the original GWAS. SNPs with relatively large p-values will have small effects (and thus be down weighted in creating the composite), so we do not impose a p-value threshold. Sensitivity analyses describing an alternative approach in which SNPs were "clumped" is described in the sensitivity analysis of Domingue et al. (In press).

In Domingue et al. (2015), the score was standardized within the black and white subsamples using the AH_RACE variable. This data file contains the raw score returned by Plink and the AH_RACE variable.

AH_RACE

AH_RACE is a single race variable that was constructed using race and ethnicity information available in the Wave I data.

Data Dictionary

AID   Respondent Identifier

GRS_EDU   Weighted sum of education associated alleles, Wave IV.

AH_RACE   5 Categories Constructed Race, Wave I.

Number of observations: 1,886

AID		Char	Respondent identifier NOTE: Smallest 5 and largest 5 values are displayed.
Frequency	Percent	Value	Label
1	0%	10316654	10316654
1	0%	11574211	11574211
1	0%	11718821	11718821
1	0%	12571478	12571478
1	0%	12714332	12714332
1876	99%	Values omitted	NOTE: Range of values omitted from display
1	0%	99719278	99719278
1	0%	99719934	99719934
1	0%	99719938	99719938
1	0%	99884905	99884905
1	0%	99884906	99884906

GRS_EDU		Num	Weighted sum of education associated alleles, Wave IV NOTE: Smallest 5 and largest 5 values are displayed.
Frequency	Percent	Value	Label
1	0%	-0.00002177100000	-0.00002177100000
1	0%	-0.00001979720000	-0.00001979720000
1	0%	-0.00001968290000	-0.00001968290000
1	0%	-0.00001892590000	-0.00001892590000

1	0%	-0.00001889960000	-0.00001889960000
1876	99%	-.000018261-.0000190661	NOTE: Range of values omitted from display
1	0%	0.00002115150000	0.00002115150000
1	0%	0.00002251180000	0.00002251180000
1	0%	0.00002265620000	0.00002265620000
1	0%	0.00002383590000	0.00002383590000
1	0%	0.00002392390000	0.00002392390000

AH_RACE		Num	5 Categories Constructed Race, Wave I
Frequency	Percent	Value	Label
917	49%	1	White
677	36%	2	Black
8	0%	3	Native American
73	4%	4	Asian
209	11%	5	Hispanic
2	0%	.	missing

References

Domingue BW, Belsky DW, Conley D, Harris KM, Boardman JD. Polygenic Influence on Educational Attainment: New Evidence from the National Longitudinal Study of Adolescent to Adult Health. AERA Open 2015; 1(3):1-13.

McQueen, M. B., Boardman, J. D., Domingue, B. W., Smolen, A., Tabor, J., Killeya-Jones, L., ... & Harris, K. M. (2014). The National Longitudinal Study of Adolescent to Adult Health (Add Health) Sibling Pairs Genome-Wide Data. Behavior genetics, 1-12.

Rietveld, C. A., Medland, S. E., Derringer, J., Yang, J., Esko, T., Martin, N. W. & McMahon, G. (2013). GWAS of 126,559 individuals identifies genetic variants associated with educational attainment. Science, 340(6139), 1467-1471.