2020

Add Health Wave IV Documentation

Add Health
The National Longitudinal Study of Adolescent to Adult Health

Polygenic Scores (PGSs) in the National Longitudinal Study of Adolescent to Adult Health (Add Health) – Release 2

Report prepared by

David Braudt

Kathleen Mullan Harris

UNC | CAROLINA POPULATION CENTER

CAROLINA POPULATION CENTER | CAROLINA SQUARE - SUITE 210 | 123 WEST FRANKLIN STREET | CHAPEL HILL, NC 27516

Add Health is supported by grant P01-HD31921 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development, with cooperative funding from 23 other federal agencies and foundations.

doi.org/XX.XXXXX/XXXXXXXX

Acknowledgments

This report was funded by grant R01 HD073342 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD). Data for this study come from Add Health, a program project directed by Kathleen Mullan Harris and designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris at the University of North Carolina at Chapel Hill, and funded by grant P01 HD31921 from NICHD, with cooperative funding from 23 other federal agencies and foundations. Information on how to obtain the Add Health data files is available on the Add Health website (http://www.cpc.unc.edu/addhealth). Additional support was received from the Population Research Training grant (T32 HD007168) and the Population Research Infrastructure Program (P2C HD050924) awarded to the Carolina Population Center at The University of North Carolina at Chapel Hill by NICHD, the Genetics and Interdisciplinary Training in Demography and Genetics grant (T32 AG052371) awarded to Jason Boardman at the University of Colorado Boulder, and the Add Health GWAS Data: User Support and Research Tools to Enable Widespread Access (R03 HD097630) awarded to Kathleen Mullan Harris and Christy Avery.

Table of Contents

Table of Contents 3
Overview 6
Data 6
  Genome-wide Data 7
  Ancestry Specific Samples 7
Methods 10
  Polygenic Scores 10
  Ancestry-specific Principal Components 10
Using Add Health PGSs 11
Release 2 Notes 12
Citing this Document and Data 12
References 13
Polygenic Scores 15
  Coronary Artery Disease 15
  Myocardial Infarction 17
  Plasma Cortisol 19
  Low-density Lipoprotein (LDL) Cholesterol 21
  High-density Lipoprotein (HDL) Cholesterol 23
  Total Cholesterol 25
  Triglycerides 27
  Type II Diabetes (2012) 29
  Type II Diabetes (2014) 31
  Body Mass Index (2015) 33
  Body Mass Index (2018) 35
  Waist Circumference 37
  Waist-to-Hip Ratio 39
  Height (2014) 41
  Height (2018) 43
  Alzheimer’s Disease (2013) 45
  Alzheimer’s Disease (2019) 47
  Family History of Alzheimer’s Disease 49
  Educational Attainment (2016) 52

Add Health is a program project directed by Kathleen Mullan Harris and designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris at the University of North Carolina at Chapel Hill, and funded by grant P01-HD31921 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development, with cooperative funding from 23 other federal agencies and foundations. Information on how to obtain the Add Health data files is available on the Add Health website (http://www.cpc.unc.edu/addhealth).

Educational Attainment (2018) 54
Cognitive Function 56
Reaction Time 58
Intelligence 60
Autism Spectrum Disorder 62
Ever Regular Smoker (2010) 64
Ever Regular Smoker (2019) 66
Ever Regular Smoker (2019 - UK Biobank) 68
Cigarettes per day (2010) 70
Cigarettes per day (2019) 72
Age of Smoking Initiation 74
Smoking Cessation 76
Drinks per Week (2019 – GSCAN) 78
Drinks per Week (2019 – UK Biobank) 80
Alcohol Dependence 82
Ever-used Cannabis 85
Menarche 87
Menopause 89
Number of Children Ever Born 91
Age at First Birth 95
Post-Traumatic Stress Disorder 99
Attention-Deficit/Hyperactivity Disorder (2010) 102
Attention-Deficit/Hyperactivity Disorder (2019) 104
Bipolar Disorder 106
Schizophrenia 108
Major Depressive Disorder (2013) 110
Major Depressive Disorder (2018) 112
Major Depressive Disorder (2019) 114
Depressive symptoms (2018) 116
Depressive Symptoms (2019) 118
Depressive affect 121
Worry 123
Life Satisfaction 125

Add Health is a program project directed by Kathleen Mullan Harris and designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris at the University of North Carolina at Chapel Hill, and funded by grant P01-HD31921 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development, with cooperative funding from 23 other federal agencies and foundations. Information on how to obtain the Add Health data files is available on the Add Health website (http://www.cpc.unc.edu/addhealth).

Positive Affect 128
Subjective Wellbeing 131
Extraversion 133
Neuroticism (2018) 135
Neuroticism (2019) 137
Anorexia Nervosa 140
Mental Health Cross Disorder 142
Social Isolation 144
Lonely 146
Going to Pubs/Social Clubs Weekly 148
Going to a Gym/Sports Club Weekly 150
Attending Religious Meetings/Groups 152
Risk Tolerance 154
Propensity for Speeding while Driving 156
Lifetime Number of Sexual Partners 158
Insomnia 160
Daytime Sleepiness (Dozing) 162
Ease of getting up in the morning 164
Morning person 166
Daytime Napping 168
Sleep Duration 170
Snoring 172

Overview

Research has shown that many outcomes of interest in the health, behavioral, and social sciences are influenced by genetics (Domingue et al. 2016; Plomin et al. 2016; Turkheimer 2000). For most human traits/behaviors, commonly referred to as phenotypes in genetic literature, it appears that the genetic influence on the phenotype is highly polygenic; i.e., there is no single gene that can account for the association between genetic variance and the outcome. Instead, the influence of genetics on the phenotype is due to many small associations spread across single-nucleotide polymorphisms (SNPs, pronounced snips) (Chabris et al. 2015). Polygenic Scores allow researchers to avoid the methodological complexities of including hundreds, thousands, and possibly millions, of covariates in their analyses by condensing, into a single measure, the associations between individual SNPs and the phenotype of interest (Plomin, Haworth, and Davis 2009).

Polygenic Scores (PGSs), sometimes referred to as polygenic risk scores or genetic risk scores, represent a general measure of the influence of additive genetics on a specific phenotype. The calculation of PGSs relies on summary statistics from genome-wide association studies (GWASs) to create a weighted sum of the associations between allele frequencies at individual SNPs and the associated phenotype. The estimated associations (beta-coefficients, odds-ratios, etc.) for each SNP from a GWAS, conducted on an independent sample, are multiplied by the allele frequencies of the same SNPs for individuals in the sample for which the PGS is being created. This process yields a hypothesis free measure of the cumulative additive genetic influences on the phenotype being studied. PGSs are hypothesis free because the GWAS from which they are based did not impose any a priori hypotheses concerning biological pathways and because PGSs aggregate the individual associations between SNPs and the phenotype, thus removing the possibility of investigating links between specific biological/genetic pathways and the phenotype. While the hypothesis free nature of PGSs removes the ability to investigate specific biological pathways, it allows researchers to capture the additive influence of genetics on a phenotype in a single measure (Belsky and Israel 2014; Dudbridge 2016).

Since PGSs represent the associations between SNPs across the entire genome and a phenotype in a single measure, they can easily be incorporated into many of the quantitative analyses common in the social, behavioral, and health sciences (Benjamin et al. 2012; Braudt 2018; Conley 2016). The PGSs described in this document are meant to facilitate the use of PGSs among users of the National Longitudinal Study of Adolescent to Adult Health (Add Health). While there are many uses for PGS, for users interested in investigating potential gene-environment interactions (GxEs) we highly recommend following the suggestions provided in Domingue et al. (2020).

Data

Add Health is an ongoing nationally representative longitudinal study of adolescents in the U.S. who were in grades 7-12 in 1994-5. Wave I (1994-5, 79% response rate) included a sample of 80 high schools and 52 middle schools chosen from a stratified sample according to region, urbanicity, school size, school type, and racial and ethnic composition with probability of selection proportional to size. With five waves of data—Wave II (1996, 89% response rate), Wave III (2001-2, 77% response rate), Wave IV (2008, 80% response rate), and Wave V (72% response rate)—and information on adolescents' fellow students, school administrators, parents, siblings, friends, and romantic pairs, as well as extensive longitudinal geospatial data on neighborhood measures such

Add Health is a program project directed by Kathleen Mullan Harris and designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris at the University of North Carolina at Chapel Hill, and funded by grant P01-HD31921 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development, with cooperative funding from 23 other federal agencies and foundations. Information on how to obtain the Add Health data files is available on the Add Health website (http://www.cpc.unc.edu/addhealth).

as income, poverty, unemployment, the availability and use of health services, crime, religious membership, and social programs. As such, Add Health represents one of the richest longitudinal studies of health and behavior in the U.S. available today. For more information on the Add Health study design see (Harris et al. 2019).

Genome-wide Data

As part of the Wave IV data collection, saliva samples were obtained from consenting participants (96% of Wave IV respondents). Approximately 12,200, or 80% of those participants, consented to long-term archiving and were consequently eligible for genome-wide genotyping. Genotyping was done on two Illumina platforms, with approximately 80% of the sample genotyping performed with the Illumina Omni1-Quad BeadChip and 20% genotyped with the Illumina Omni2.5-Quad BeadChip. After quality control procedures, genotyped data are available for 9,974 individuals (7,917 from the Omni1 chip and 2,057 from the Omni2 chip) on 609,130 SNPs common across both genotyping platforms (Highland, Heather M.; Avery, Christy L.; Duan, Qing; Li, Yun; Mullan Harris, Kathleen 2018). For more information on the genotyping and quality control procedures see the Add Health GWAS QC report online at: http://www.cpc.unc.edu/projects/addhealth/documentation/guides/copy_of_AH_GWAS_QC.pdf.

Ancestry Specific Samples

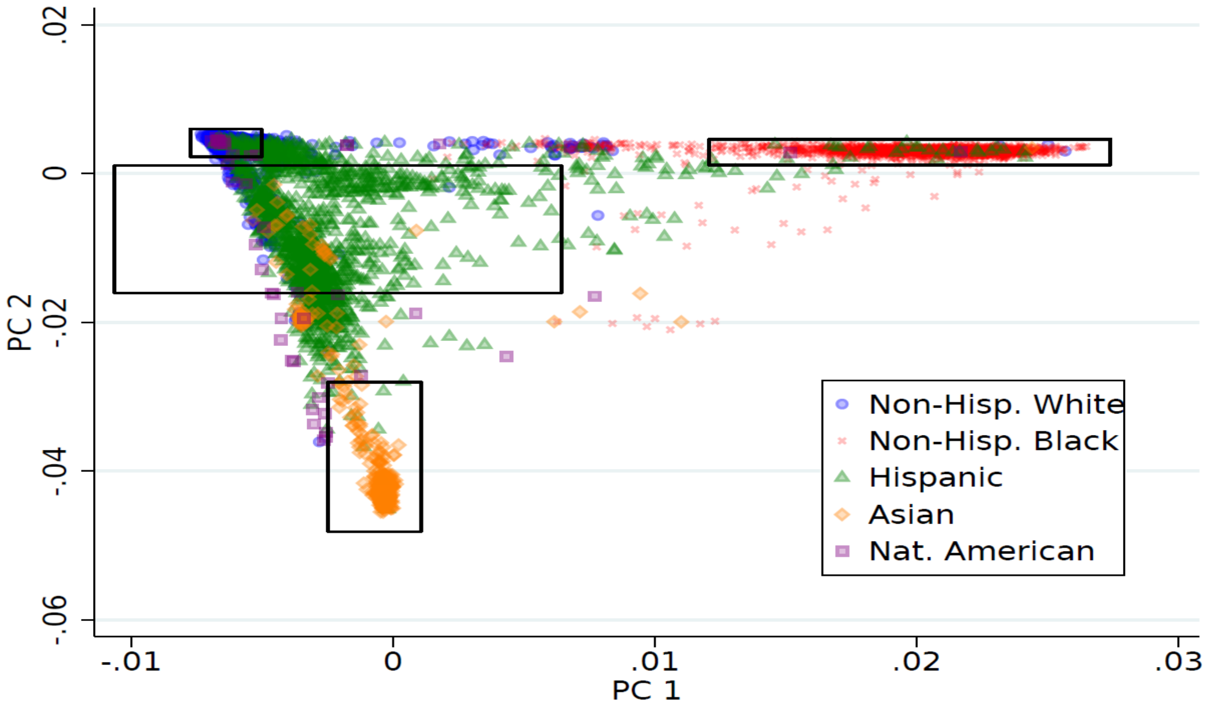
To account for population stratification,[1] we follow the current best practices and restrict the Add Health genotyped sample to four genetic ancestry groups: (1) European ancestry, (2) African ancestry, (3) Hispanic ancestry, and (4) East Asian ancestry. To identify respondents in these four genetic ancestry groups we use principal component analysis on all unrelated members of the full Add Health genotyped sample and project those estimates onto the small remainder of related individuals. Figure 1 provides a visual depiction of this process with the respective cut-offs for inclusion in the genetic ancestry groups represented by the black rectangles.

Each genetic ancestry group is defined by distance from the mean of the first two principal components of the genetic data as shown in Figure 1. To be included in the Hispanic, East Asian, and European ancestry groups individuals must be within +/- 1 standard deviation of the mean of the first two principal components of the genetic data estimated from all individuals in the Add Health genome-wide data who self-identified as Hispanic, Asian, and non-Hispanic White respectively. To be included in the African ancestry group individuals must be within +/- 2 standard deviation of the mean of the first principal component and +/- 1 standard deviation of the mean of the second principal component estimated from all individuals in the genome-wide data who self-identified as non-Hispanic Black. While genetic ancestry and race/ethnicity are correlated (0.89), they are two separate constructs and attempts to conflate the two is not supported by scientific evidence (see footnote 1). Consequently, not all individuals included in a given genetic ancestry may self-identify or be classified by others as the same race and/or ethnicity as other members of their genetic

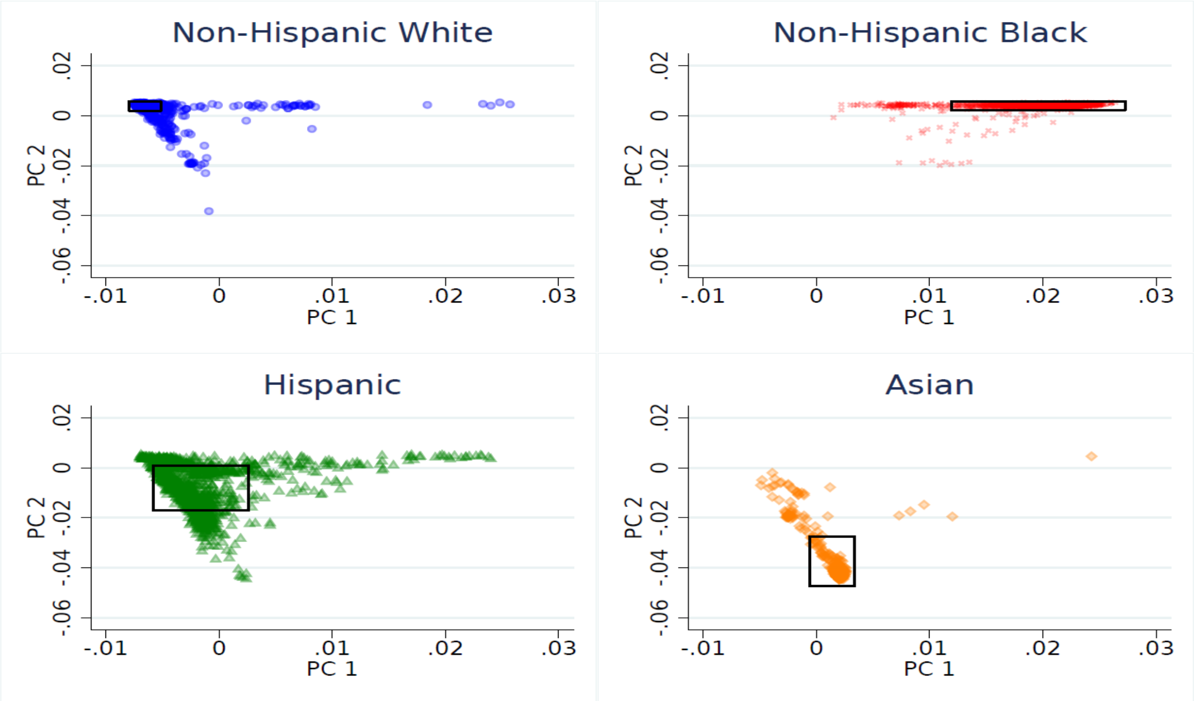
---

[1] Population stratification refers to differences in genetic variation between geographical ancestry groups. Due to the genetic bottle neck created by the small number of humans (roughly 2,000 individuals) who migrated out of Africa early in human history, and the tendency for people to procreate with individuals from the same or nearby geographic regions, genetic variance across the entire genome is highly correlated with geography (for a more detailed discussion see Conley and Fletcher 2017:84–112). Importantly, genetic ancestry should not be confused with race or ethnicity. Race and ethnicity are social constructs based on a multitude of factors, of which ancestry may be included depending on historical and societal differences in racialization (Omi and Winant 1994).

Figure 1: Principal Components and Ancestry-Specific Samples



Add Health is a program project directed by Kathleen Mullan Harris and designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris at the University of North Carolina at Chapel Hill, and funded by grant P01-HD31921 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development, with cooperative funding from 23 other federal agencies and foundations. Information on how to obtain the Add Health data files is available on the Add Health website (http://www.cpc.unc.edu/addhealth).



Notes: (1) The principal components are estimated based on the allele frequencies of the 609,130 genotyped SNPs in the Add Health genome-wide data. (2) Black rectangles represent within Add Health genetic ancestry group boundaries. East Asian, Hispanic, and European ancestry groups are defined as: +/- 1 standard deviation from the within group mean of the first and second principal components estimated from individuals who self-identified as Asian, Hispanic, and non-Hispanic White respectively. African ancestry is defined as +/- 2 standard deviations from the mean of the first principal component and +/- 1 standard deviation from the mean of the second principal component estimated from all individuals in the genotyped sample who self-identified as non-Hispanic Black.

Table 1: Add Health Polygenic Score Sample Sizes by Genetic Ancestry*

	Genetic Ancestry				
Self-Identified Race/Ethnicity	European	African	East Asian	Hispanic	Total
Non-Hispanic White	5,645	5	0	105	5,755
Non-Hispanic Black	0	1,938	0	1	1,939
Native American	14	2	0	6	22
Asian	0	1	422	26	449
Hispanic	72	27	15	849	963
Missing	0	2	0	0	0
Total Sample Size	5,731	1,975	437	987	9,130

* Respondent genetic ancestry classification were updated and harmonized with the Add Health dbGaP data to ensure replicability. The update resulted in marginal differences (3 respondents were added to European ancestry and 1 respondent was removed from African and Hispanic ancestry groups from classifications in Release #1 of the Add Health PGSs).

ancestry group. Table 1 provides a depiction of the resulting sample sizes for each genetic ancestry group and their correspondence to self-identified race/ethnicity.

As a sensitivity analysis, we repeated the above process using group specific medians of the first two principal components of the genetic data as centroids instead of group means. While sample sizes for the four ancestry groups using the median as the centroid were comparable to those using the means of the principal components as centroids, using the mean as the centroid resulted in larger sample size. Given the slight increase in sample size, we define within Add Health genetic ancestry groups as distance from the mean of the first two principal components of the genotyped data.

Methods

Polygenic Scores

We calculate polygenic scores (PGSs) following the procedure outlined in Dudbridge (2013). PGSs are a weighted sum of the reported coefficient for each SNP from an independent GWAS for each phenotype and allele frequencies for the same SNPs in the Add Health genome-wide data. For example, the raw PGS of educational attainment for an individual, i, is calculated as:

$$PGS_{EDU_i} = \sum_{j=1}^{k} \beta_j SNP_{ij} \tag{1}$$

where, SNP_{ij} is the allele frequency of the j^{th} SNP for the i^{th} individual and β_j is the estimated association between SNP j and the number of years of education completed as reported in the GWAS summary statistics based on an independent sample (i.e. if a GWAS included Add Health, the GWAS is re-estimated excluding the Add Health sample and coefficients from the re-estimated GWAS are used in Equation 1). To automate this process, we use a modified version of the PRSice wrapper for R within the PLINK software package (Chang et al. 2015). Once calculated, the raw PGSs are standardized ($\mu_{PGS} = 0 \wedge \sigma_{PGS} = 1$) within ancestry groups.

Standardization of PGSs within ancestry groups is done to account for between group population stratification. To control for within group population stratification, we recommend that researchers include at least the first five ancestry-specific principal components of the genome-wide data in all analyses using PGSs. It should be noted that these are imperfect controls, and PGSs should be re-calculated using GWAS weights for specific ancestry groups when such GWAS become available.

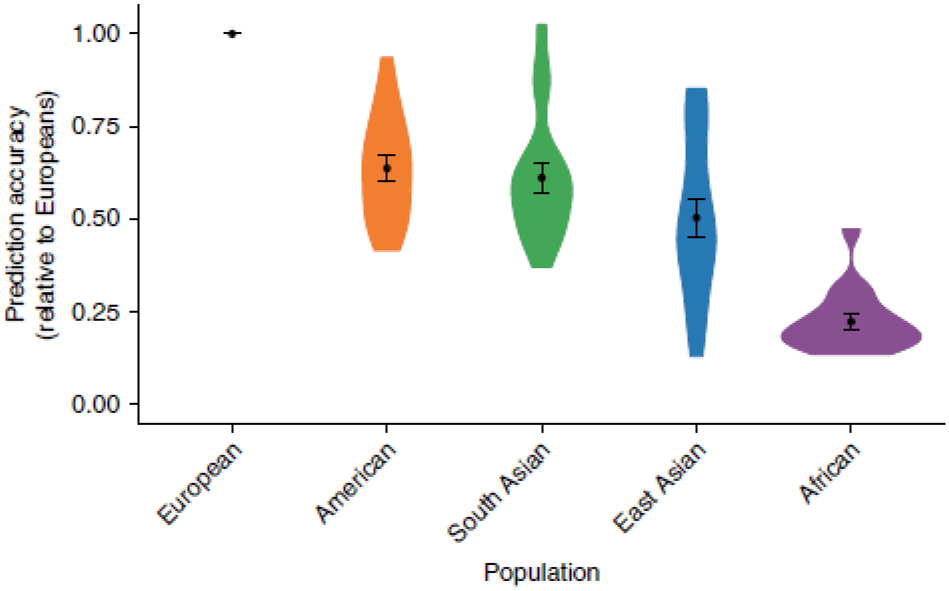
Ancestry-specific Principal Components

Ancestry-specific principal components are estimated following a similar procedure to the full sample principal components described above, with the sample restricted to individuals of the respective genetic ancestry group. The process begins with the sample of individuals in the genetic ancestry group as defined above, randomly removes one sibling of any sibling pairs, estimates the first 20 principal components for all unrelated individuals in the ancestry-specific sample, and then projects those principal components onto the small number of related individuals within each ancestry group.

Using Add Health PGSs

Add Health releases PGSs for European, African, Hispanic, and East Asian genetic ancestry groups. However, researchers should be aware that **PGSs for individuals not of the same ancestry group(s) as the GWAS from which summary statistics are retrieved are less predictive** (Martin et al. 2017, 2019); Figure 2 depicts this graphical. We recommend that researchers use caution when comparing PGSs and/or their estimated associations in analyses that included individuals from multiple genetic-ancestry groups and consult the latest peer-reviewed evidence on potential causes of parameter differences when interpreting estimates.

Figure 2: Prediction Accuracy Relative to Individuals of European Ancestry across 17 Traits and 5 Continental Populations in the UK Biobank



Source: Martin et al. 2019.
Note: Violin plots show distributions of relative prediction accuracies, points show mean values, and error bars show standard error values.

To help account for potential bias due to population stratification and/or differences in genetic structure within ancestry groups we include the first ten ancestry-specific principal components of the genetic data with the PGSs. **It is strongly recommended that researchers perform sensitivity analyses separately by ancestral groups and/or include at least the first five ancestry-specific principal components as covariates in analyses using these PGSs** (Price et al. 2006).

In order to minimize the risk of deductive disclosure, the order of the ancestry-specific principal components (PCS) of the genetic data are randomized in sets of five. Therefore, PCs must be included as sets: PC1-PC5, PC6-PC10, PC11-PC15, PC16-PC20 if any of the PCs of a set are included in analyses. For example, if a research wishes to include the first two ancestry-specific PCs as covariates in their analyses (i.e. PC1 and PC2) they must also include PC3-PC5.

Release 2 Notes

The PGSs for Diastolic Blood Pressure, Systolic Blood Pressure, Pulse Pressure, and Mean Arterial Pressure have been removed from the second release of the PGSs due to a corruption of the summary statistics obtained from the IGAP consortia. The data corruption was discovered after disseminating Release 1; consequently, Add Health users who have not replaced their Release 1 Add Health PGS data with Release 2 PGS data should not use these PGSs from Release 1.

Citing this Document and Data

Please include the following citation in any report, publication, and/or presentation based on the data in this release of the Add Health PGSs as well as the citation for the reference GWAS:

Braudt, David B. and Kathleen Mullan Harris. 2020. "Polygenic Scores (PGSs) in the National Longitudinal Study of Adolescent to Adult Health (Add Health) – Release 2." Chapel Hill, NC: Carolina Population Center, University of North Carolina at Chapel Hill.

References

Belsky, Daniel W., and Salomon Israel. 2014. "Integrating Genetics and Social Science: Genetic Risk Scores." *Biodemography and Social Biology* 60(2):137–55.

Benjamin, Daniel J., David Cesarini, Christopher F. Chabris, Edward L. Glaeser, David I. Laibson, Gene/Environment Susceptibility-Reykjavik Study: Age, Vilmundur Guðnason, Tamara B. Harris, Lenore J. Launer, Shaun Purcell, Albert Vernon Smith, Swedish Twin Registry, Magnus Johannesson, Patrik K. E. Magnusson, Framingham Heart Study, Jonathan P. Beauchamp, Nicholas A. Christakis, Wisconsin Longitudinal Study, Craig S. Atwood, Benjamin Hebert, Jeremy Freese, Robert M. Hauser, Taissa S. Hauser, Swedish Large Schizophrenia Study, Alexander Grankvist, Christina M. Hultman, and Paul Lichtenstein. 2012. "The Promises and Pitfalls of Genoeconomics." *Annual Review of Economics* 4(1):627–62.

Bolton, Jennifer L., Caroline Hayward, Nese Direk, John G. Lewis, Geoffrey L. Hammond, Lesley A. Hill, Anna Anderson, Jennifer Huffman, James F. Wilson, Harry Campbell, Igor Rudan, Alan Wright, Nicholas Hastie, Sarah H. Wild, Fleur P. Velders, Albert Hofman, Andre G. Uitterlinden, Jari Lahti, Katri Räikkönen, Eero Kajantie, Elisabeth Widen, Aarno Palotie, Johan G. Eriksson, Marika Kaakinen, Marjo-Riitta Järvelin, Nicholas J. Timpson, George Davey Smith, Susan M. Ring, David M. Evans, Beate St Pourcain, Toshiko Tanaka, Yuri Milaneschi, Stefania Bandinelli, Luigi Ferrucci, Pim van der Harst, Judith G. M. Rosmalen, Stephen J. L. Bakker, Niek Verweij, Robin P. F. Dullaart, Anubha Mahajan, Cecilia M. Lindgren, Andrew Morris, Lars Lind, Erik Ingelsson, Laura N. Anderson, Craig E. Pennell, Stephen J. Lye, Stephen G. Matthews, Joel Eriksson, Dan Mellstrom, Claes Ohlsson, Jackie F. Price, Mark W. J. Strachan, Rebecca M. Reynolds, Henning Tiemeier, Brian R. Walker, and on behalf of the CORtisol NETwork (CORNET) Consortium. 2014. "Genome Wide Association Identifies Common Variants at the SERPINA6/SERPINA1 Locus Influencing Plasma Cortisol and Corticosteroid Binding Globulin." *PLOS Genetics* 10(7):e1004474.

Braudt, David B. 2018. "Sociogenomics in the 21st Century: An Introduction to the History and Potential of Genetically Informed Social Science." *Sociology Compass* 12(10).

Chabris, Christopher F., James J. Lee, David Cesarini, Daniel J. Benjamin, and David I. Laibson. 2015. "The Fourth Law of Behavior Genetics." *Current Directions in Psychological Science* 24(4):304–312.

Chang, Christopher C., Carson C. Chow, Laurent CAM Tellier, Shashaank Vattikuti, Shaun M. Purcell, and James J. Lee. 2015. "Second-Generation PLINK: Rising to the Challenge of Larger and Richer Datasets." *GigaScience* 4:7.

Conley, Dalton. 2016. "Socio-Genomic Research Using Genome-Wide Molecular Data." *Annual Review of Sociology* 42(1):275–99.

Conley, Dalton, and Jason Fletcher. 2017. *The Genome Factor : What the Social Genomics Revolution Reveals about Ourselves, Our History, and the Future*. Princeton: Princeton University Press.

Domingue, Benjamin, Sam Trejo, Emma Armstrong-Carter, and Elliot M. Tucker-Drob. 2020. *Interactions between Polygenic Scores and Environments: Methodological and Conceptual Challenges*. *preprint*. SocArXiv.

Domingue, Benjamin W., Robbee Wedow, Dalton Conley, Matt McQueen, Thomas J. Hoffmann, and Jason D. Boardman. 2016. “Genome-Wide Estimates of Heritability for Social Demographic Outcomes.” *Biodemography and Social Biology* 62(1):1–18.

Dudbridge, Frank. 2013. “Power and Predictive Accuracy of Polygenic Risk Scores” edited by N. R. Wray. *PLoS Genetics* 9(3):e1003348.

Dudbridge, Frank. 2016. “Polygenic Epidemiology: Polygenic Epidemiology.” *Genetic Epidemiology* 40(4):268–72.

Harris, Kathleen Mullan, Carolyn Tucker Halpern, Eric A. Whitsel, Jon M. Hussey, Ley A. Killeya-Jones, Joyce Tabor, and Sarah C. Dean. 2019. “Cohort Profile: The National Longitudinal Study of Adolescent to Adult Health (Add Health).” *International Journal of Epidemiology*.

Highland, Heather M.; Avery, Christy L.; Duan, Qing; Li, Yun; Mullan Harris, Kathleen. 2018. “Quality Control Analysis of Add Health GWAS Data.”

Martin, Alicia R., Christopher R. Gignoux, Raymond K. Walters, Genevieve L. Wojcik, Benjamin M. Neale, Simon Gravel, Mark J. Daly, Carlos D. Bustamante, and Eimear E. Kenny. 2017. “Human Demographic History Impacts Genetic Risk Prediction across Diverse Populations.” *The American Journal of Human Genetics*.

Martin, Alicia R., Masahiro Kanai, Yoichiro Kamatani, Yukinori Okada, Benjamin M. Neale, and Mark J. Daly. 2019. “Clinical Use of Current Polygenic Risk Scores May Exacerbate Health Disparities.” *Nature Genetics* 51(4):584–91.

Omi, Michael, and Howard |. Winant. 1994. *Racial Formation in the United States : From the 1960s to the 1990s*. 2nd ed. New York: Routledge.

Plomin, Robert, John C. DeFries, Valerie S. Knopik, and Jenae M. Neiderhiser. 2016. “Top 10 Replicated Findings From Behavioral Genetics.” *Perspectives on Psychological Science* 11(1):3–23.

Plomin, Robert, Claire MA Haworth, and Oliver SP Davis. 2009. “Common Disorders Are Quantitative Traits.” *Nature Reviews Genetics* 10(12):872.

Price, Alkes L., Nick J. Patterson, Robert M. Plenge, Michael E. Weinblatt, Nancy A. Shadick, and David Reich. 2006. “Principal Components Analysis Corrects for Stratification in Genome-Wide Association Studies.” *Nature Genetics* 38(8):904–9.

Turkheimer, Eric. 2000. “Three Laws of Behavior Genetics and What They Mean.” *Current Directions in Psychological Science* 9(5):160–164.

Polygenic Scores

Coronary Artery Disease

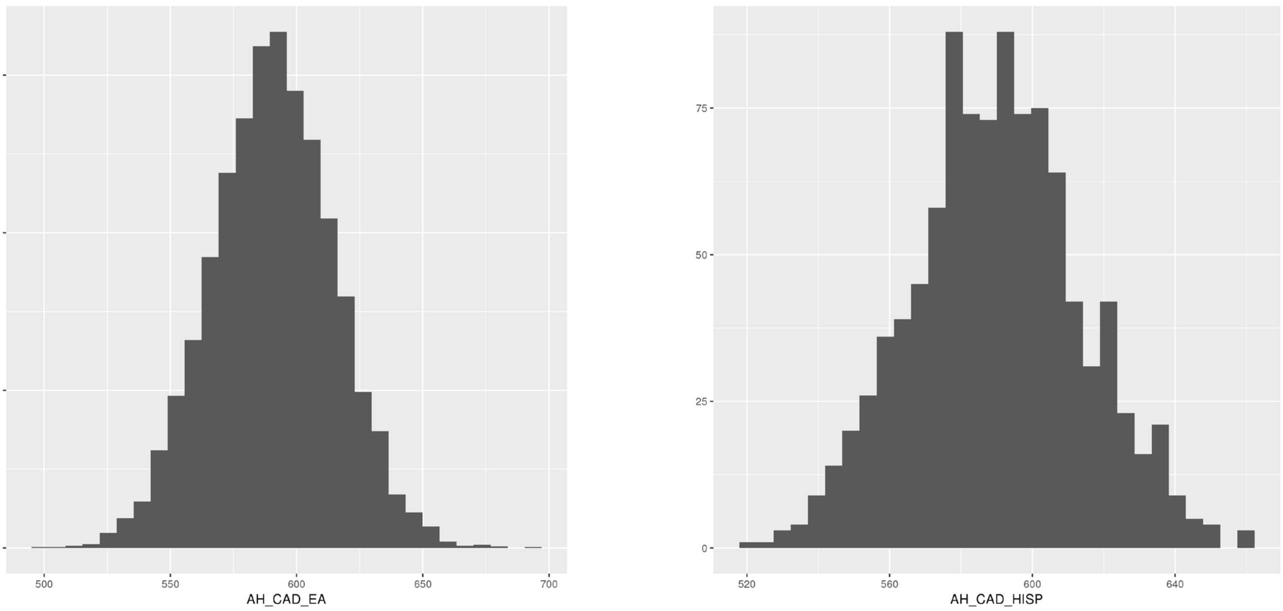
GWAS Summary Statistic Source: Schunkert, Heribert et al. 2011. “Large-Scale Association Analysis Identifies 13 New Susceptibility Loci for Coronary Artery Disease.” *Nature Genetics* 43(4):333–38.

GWAS Ancestry Group(s): European

The PGSs for coronary artery disease (CAD) were created using results from a 2011 study conducted by the Coronary ARtery DIsease Genome-wide Replication and Meta-analysis (CARDIoGRAM) consortium. The GWAS meta-analysis files are publicly available and can be downloaded from: http://www.cardiogramplusc4d.org.

The coronary artery disease GWAS meta-analysis consisted of 14 studies with a total of 22,233 individuals with CAD (cases) and 64,762 without CAD (controls) of European descent imputed to the HapMap3 CEU panel. Replication was performed in a sample of 56,682 individuals (approximately half cases and half controls). Study-specific GWAS adjusted for age of onset of CAD (cases) or age of recruitment (controls), gender, and genetic principal components.

The Add Health PGSs for CAD are standardized (mean = 0, standard deviation = 1) within ancestry groups to control for population stratification. To give researchers an idea of the underlying raw PGS, we present density plots for the unstandardized PGSs below. **Please read the section entitled “Using Add Health PGSs” in the introductory portion of this document prior to conducting any analyses using the provided PGSs.**

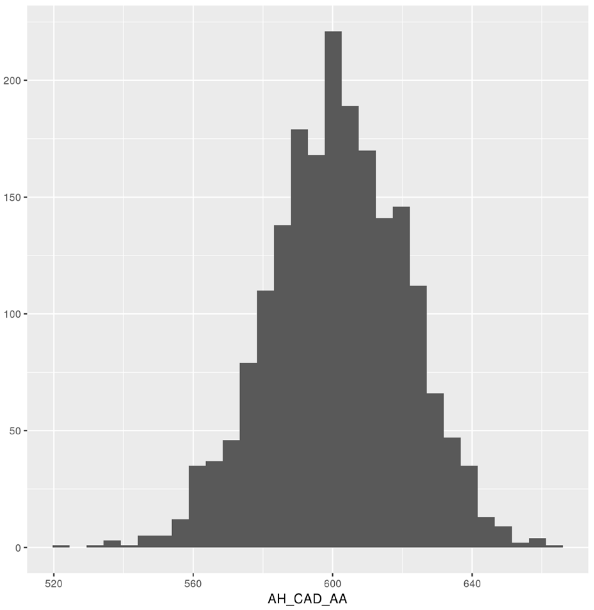


Coronary Artery Disease

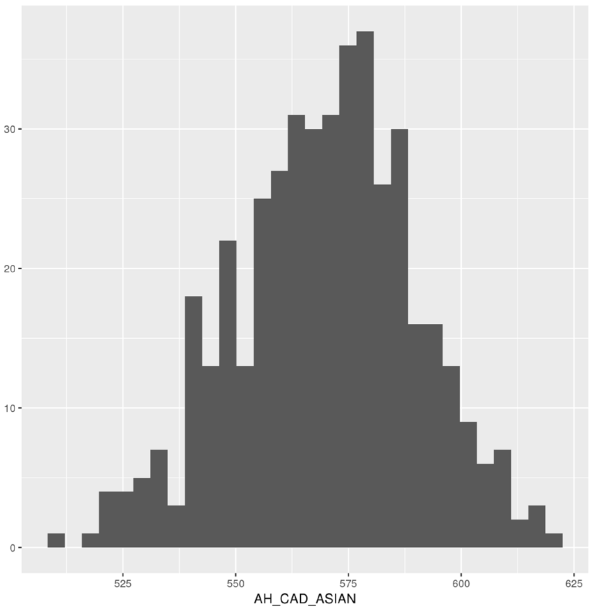
European ancestry raw PGS distribution

Hispanic ancestry raw PGS distribution

African ancestry raw PGS distribution



Asian ancestry raw PGS distribution



Myocardial Infarction

GWAS Summary Statistic Source: Nikpay, Majid et al. 2015. “A Comprehensive 1,000 Genomes-Based Genome-Wide Association Meta-Analysis of Coronary Artery Disease.” Nature Genetics 47(10):1121–30.

GWAS Ancestry Group(s): European (77%), South Asian (13% - India and Pakistan), East Asian (6% - China and Korea), Hispanic and/or African American (4%).

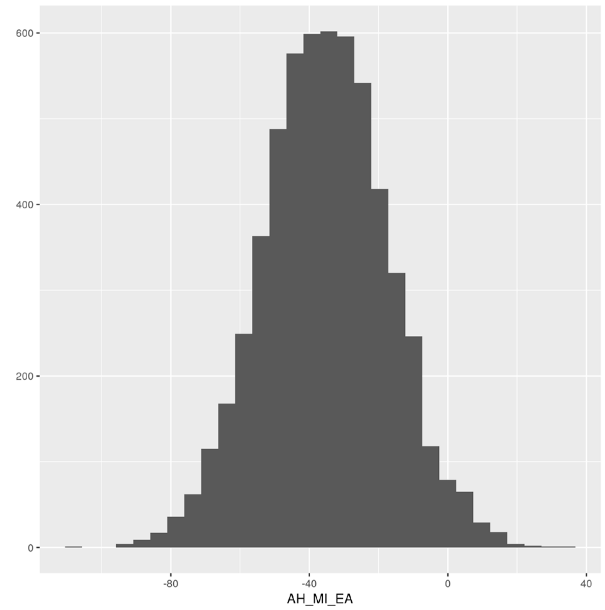
The PGSs for myocardial infarction (MI) were created using 2015 results from a subgroup analysis of coronary artery disease conducted by the Coronary ARtery DIsease Genome wide Replication and Meta-analysis (CARDIoGRAM) consortium. The GWAS meta-analysis files are publicly available and can be downloaded from: http://www.cardiogramplusc4d.org.

The GWAS is a meta-analysis of 48 studies of mainly European, South Asian, and East Asian, descent (described above) imputed using the 1000 Genomes phase 1 v3 training set with 38 million variants. The study included 60,801 individuals with coronary artery disease (cases) and 123,504 without (controls). Case were defined as a reported history of MI (approximately 70% of the total number of cases).status was defined by an inclusive coronary artery disease diagnosis (for example, myocardial infarction, acute coronary syndrome, chronic stable angina or coronary stenosis of >50%). MI subgroup analysis was performed in cases with

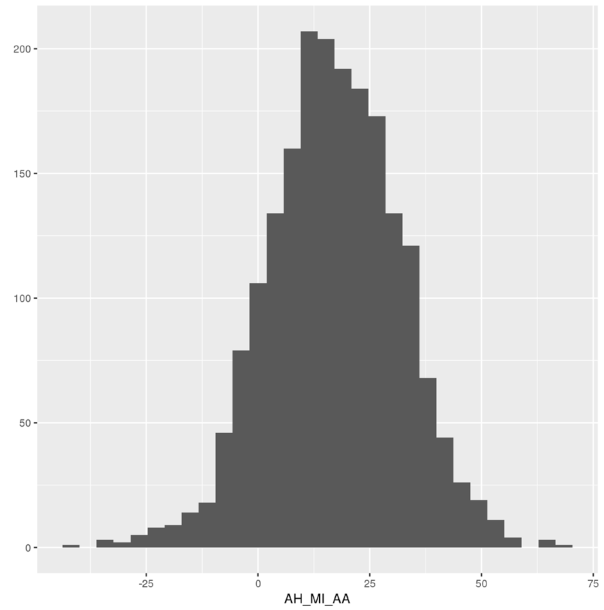
The Add Health PGSs for myocardial infarction are standardized (mean = 0, standard deviation = 1) within ancestry groups to control for population stratification. To give researchers an idea of the underlying raw PGS, we present density plots for the unstandardized PGSs below. **Please read the section entitled “Using Add Health PGSs” in the introductory portion of this document prior to conducting any analyses using the provided PGSs.**

Myocardial Infarction

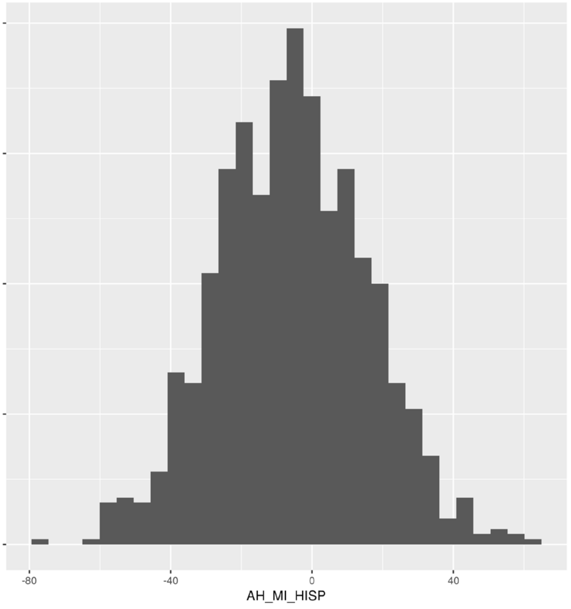
European ancestry raw PGS distribution



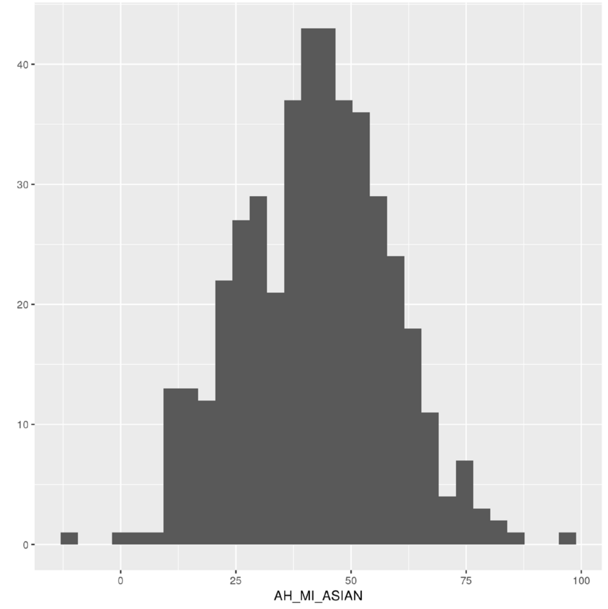
African ancestry raw PGS distribution



Hispanic ancestry raw PGS distribution



Asian ancestry raw PGS distribution



Add Health is a program project directed by Kathleen Mullan Harris and designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris at the University of North Carolina at Chapel Hill, and funded by grant P01-HD31921 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development, with cooperative funding from 23 other federal agencies and foundations. Information on how to obtain the Add Health data files is available on the Add Health website (http://www.cpc.unc.edu/addhealth).

Plasma Cortisol

GWAS Summary Statistic Source: Bolton, Jennifer L. et al. 2014. “Genome Wide Association Identifies Common Variants at the SERPINA6/SERPINA1 Locus Influencing Plasma Cortisol and Corticosteroid Binding Globulin.” PLOS Genetics 10(7):e1004474.

GWAS Ancestry Group(s): Western European

The PGSs for plasma cortisol were created using results from a 2014 GWAS conducted by the CORtisol NETwork (CORNET). The GWAS meta-analysis files are available through CORNET by request only.

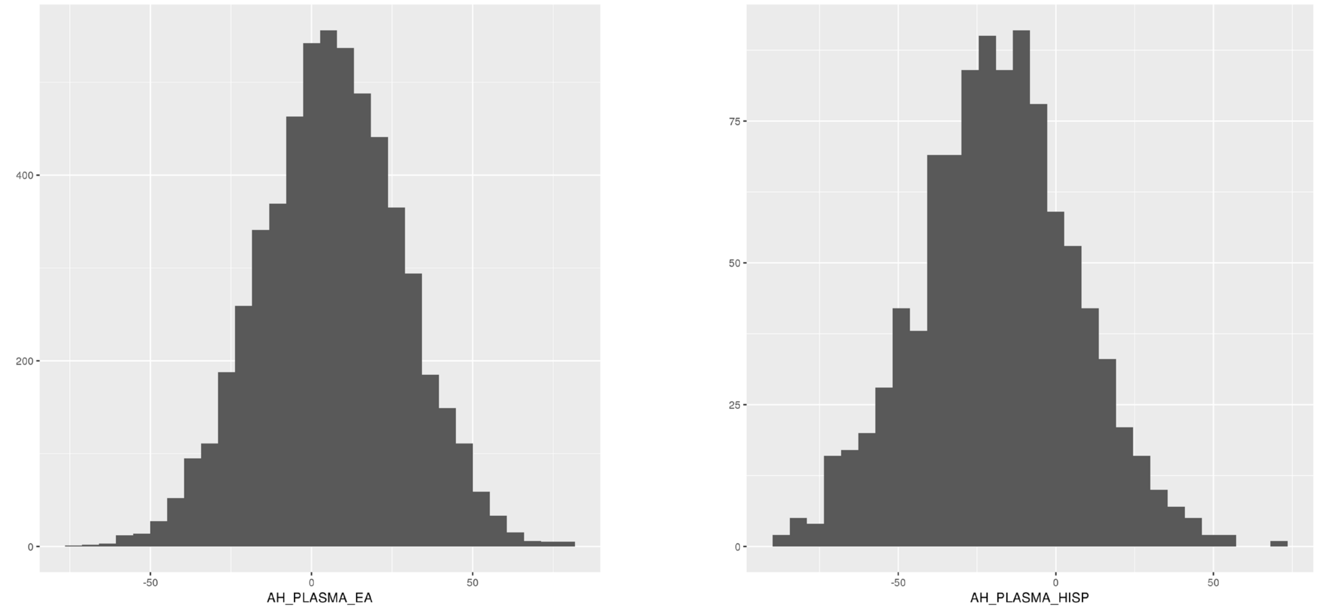
The discovery phase of the meta-analysis included 12,597 participants aged 14-102 of western European descent from 11 studies. The replication sample consisted of 2,795 participants from three independent cohorts: Raine Study (n=797), ET2DS (n=1,069), and MrOS-Sweden (n=929) (Bolton et al. 2014:8). Analyses were adjusted for age and sex. Additional sensitivity analyses showed no difference when time of sampling was included as a covariate.

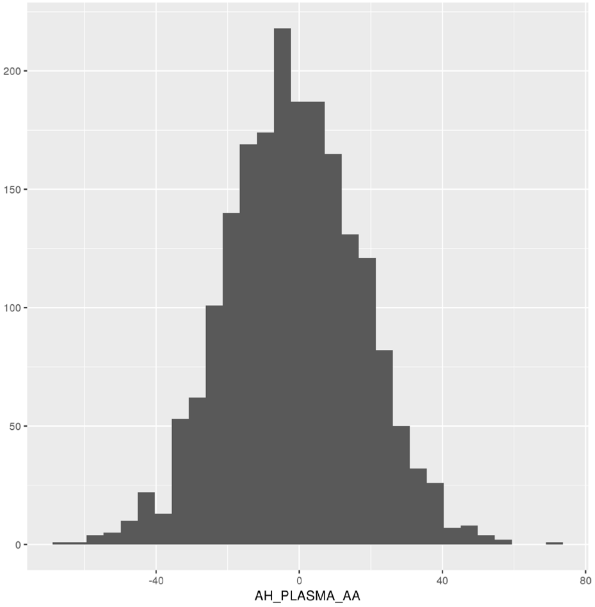
The Add Health PGSs for plasma cortisol are standardized (mean = 0, standard deviation = 1) within ancestry groups to control for population stratification. To give researchers an idea of the underlying raw PGS, we present density plots for the unstandardized PGSs below. **Please read the section entitled “Using Add Health PGSs” in the introductory portion of this document prior to conducting any analyses using the provided PGSs.**

Plasma Cortisol

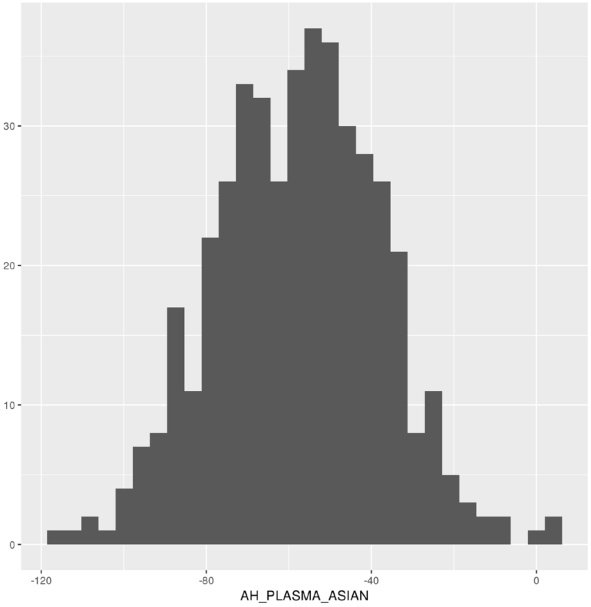
European ancestry raw PGS distribution

Hispanic ancestry raw PGS distribution





African ancestry raw PGS distribution



Asian ancestry raw PGS distribution

Low-density Lipoprotein (LDL) Cholesterol

GWAS Summary Statistic Source: Willer, Cristen J. et al. 2013. “Discovery and Refinement of Loci Associated with Lipid Levels.” *Nature Genetics* 45(11):1274–83.

GWAS Ancestry Group(s): European

The PGSs for low-density lipoprotein (LDL)cholesterol were created using results from a 2013 GWAS conducted by the Global Lipids Genetics Consortium (GLGC). The GWAS meta-analysis files are publicly available from: http://csg.sph.umich.edu/willer/public/lipids2013/.

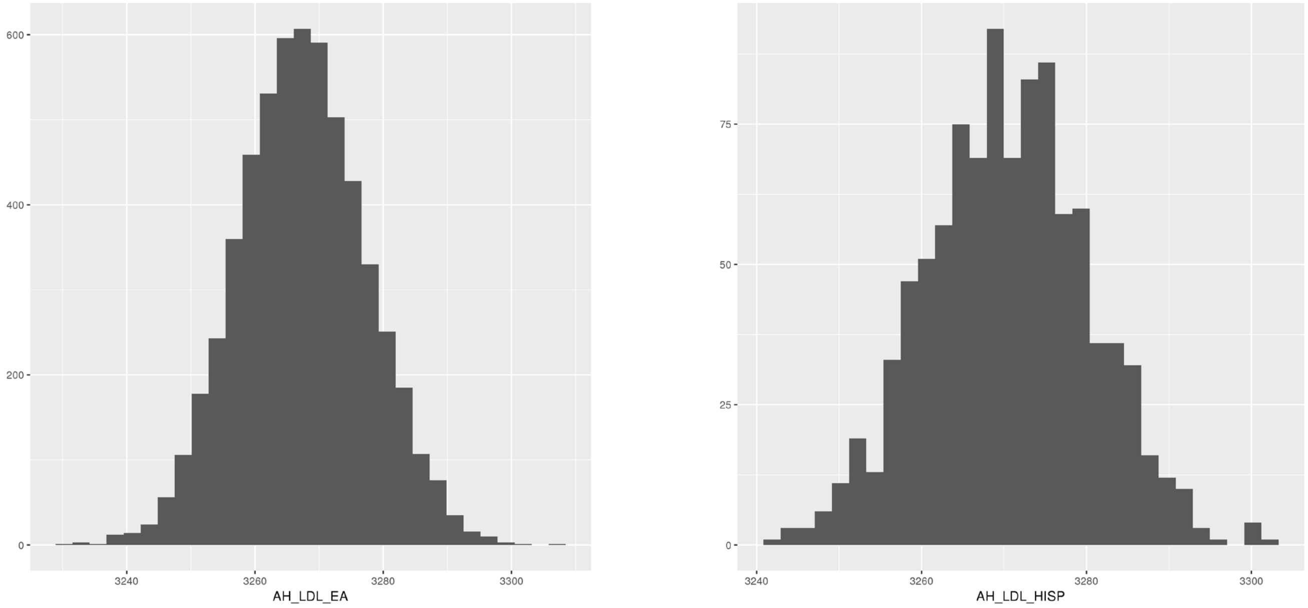
The discovery phase of the meta-analysis included 188,577 individuals of European ancestry from 45 studies, while the replication sample consisted of 7,898 individuals of non-European descent. Lipid levels were measured after eight or more hours of fasting with individuals taking lipid lowering medication excluded from the study. Study-specific GWAS adjusted for age, age-squared, sex, and genetic principal components.

The Add Health PGSs for LDL are standardized (mean = 0, standard deviation = 1) within ancestry groups to control for population stratification. To give researchers an idea of the underlying raw PGS, we present density plots for the unstandardized PGSs below. **Please read the section entitled “Using Add Health PGSs” in the introductory portion of this document prior to conducting any analyses using the provided PGSs.**

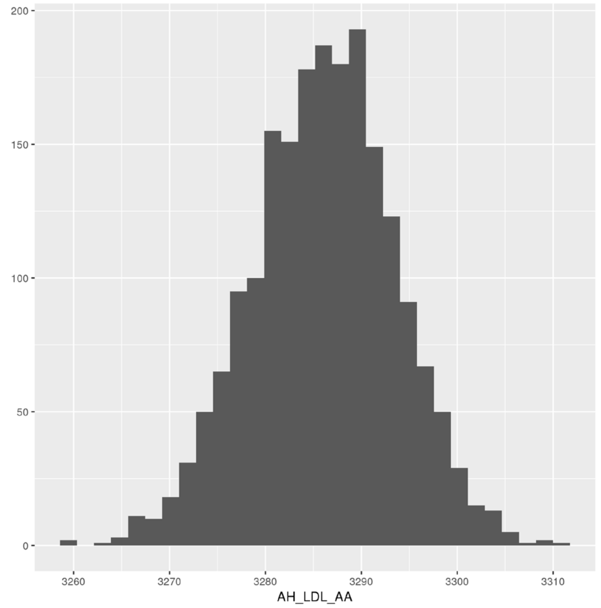
Low-density Lipoprotein (LDL) Cholesterol

European ancestry raw PGS distribution

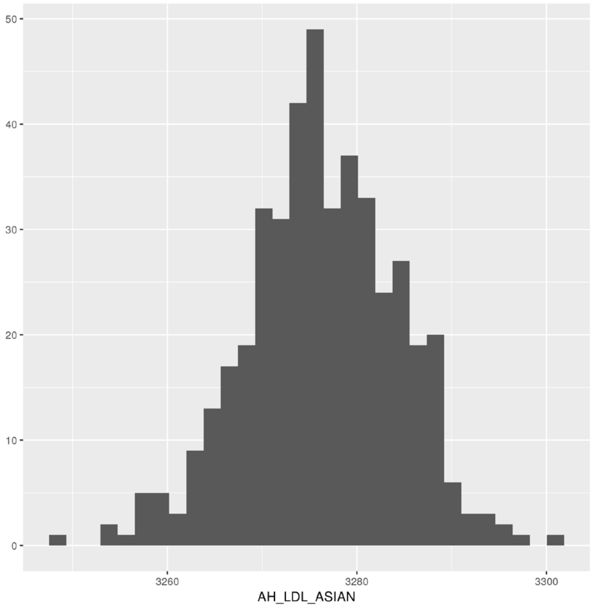
Hispanic ancestry raw PGS distribution



African ancestry raw PGS distribution



Asian ancestry raw PGS distribution



Add Health is a program project directed by Kathleen Mullan Harris and designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris at the University of North Carolina at Chapel Hill, and funded by grant P01-HD31921 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development, with cooperative funding from 23 other federal agencies and foundations. Information on how to obtain the Add Health data files is available on the Add Health website (http://www.cpc.unc.edu/addhealth).

High-density Lipoprotein (HDL) Cholesterol

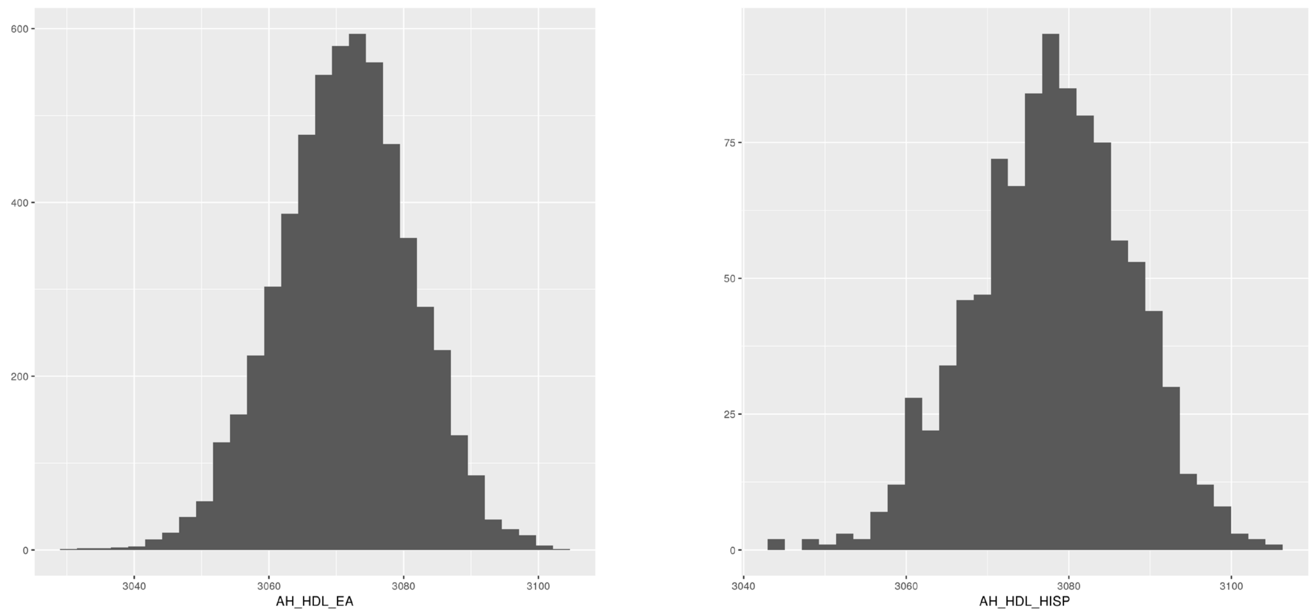
GWAS Summary Statistic Source: Willer, Cristen J. et al. 2013. “Discovery and Refinement of Loci Associated with Lipid Levels.” *Nature Genetics* 45(11):1274–83.

GWAS Ancestry Group(s): European

The PGSs for high-density lipoprotein (HDL) cholesterol were created using results from a 2013 GWAS conducted by the Global Lipids Genetics Consortium (GLGC). The GWAS meta-analysis files are publicly available from: http://csg.sph.umich.edu/willer/public/lipids2013/.

The discovery phase of the meta-analysis included 188,577 individuals of European ancestry from 45 studies, while the replication sample consisted of 7,898 individuals of non-European descent. Lipid levels were measured after eight or more hours of fasting with individuals taking lipid lowering medication excluded from the study. Study-specific GWAS adjusted for age, age-squared, sex, and genetic principal components.

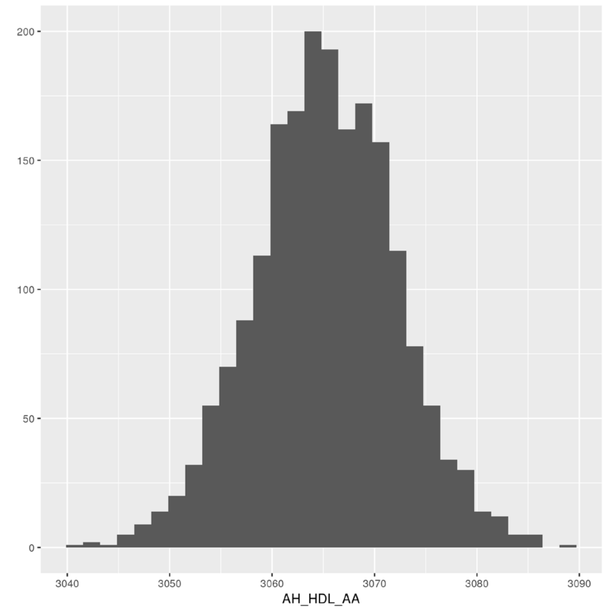
The Add Health PGSs for HDL are standardized (mean = 0, standard deviation = 1) within ancestry groups to control for population stratification. To give researchers an idea of the underlying raw PGS, we present density plots for the unstandardized PGSs below. **Please read the section entitled “Using Add Health PGSs” in the introductory portion of this document prior to conducting any analyses using the provided PGSs.**


High-density Lipoprotein (HDL) Cholesterol

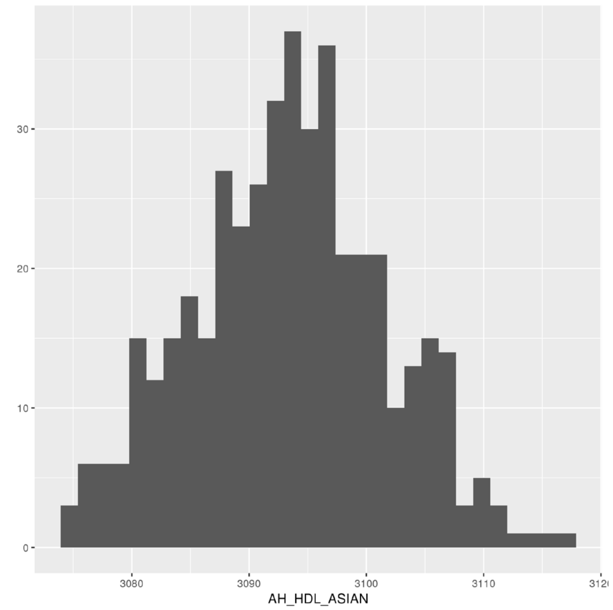
European ancestry raw PGS distribution

Hispanic ancestry raw PGS distribution

African ancestry raw PGS distribution



Asian ancestry raw PGS distribution



Total Cholesterol

GWAS Summary Statistic Source: Willer, Cristen J. et al. 2013. “Discovery and Refinement of Loci Associated with Lipid Levels.” *Nature Genetics* 45(11):1274–83.

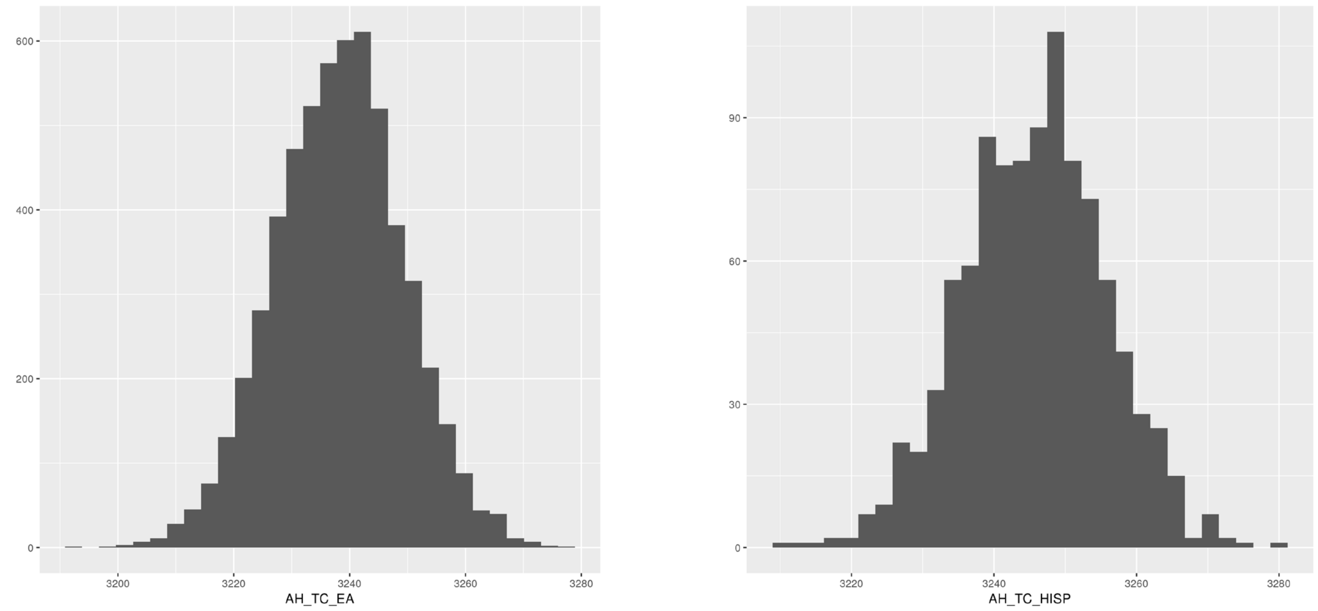
GWAS Ancestry Group(s): European

The PGSs for total cholesterol (TC) were created using results from a 2013 GWAS conducted by the Global Lipids Genetics Consortium (GLGC). The GWAS meta-analysis files are publicly available from: http://csg.sph.umich.edu/willer/public/lipids2013/.

The discovery phase of the meta-analysis included 188,577 individuals of European ancestry from 45 studies, while the replication sample consisted of 7,898 individuals of non-European descent. Lipid levels were measured after eight or more hours of fasting with individuals taking lipid lowering medication excluded from the study. Study-specific GWAS adjusted for age, age-squared, sex, and genetic principal components.

The Add Health PGSs for total cholesterol are standardized (mean = 0, standard deviation = 1) within ancestry groups to control for population stratification. To give researchers an idea of the underlying raw PGS, we present density plots for the unstandardized PGSs below. **Please read the section entitled “Using Add Health PGSs” in the introductory portion of this document prior to conducting any analyses using the provided PGSs.**

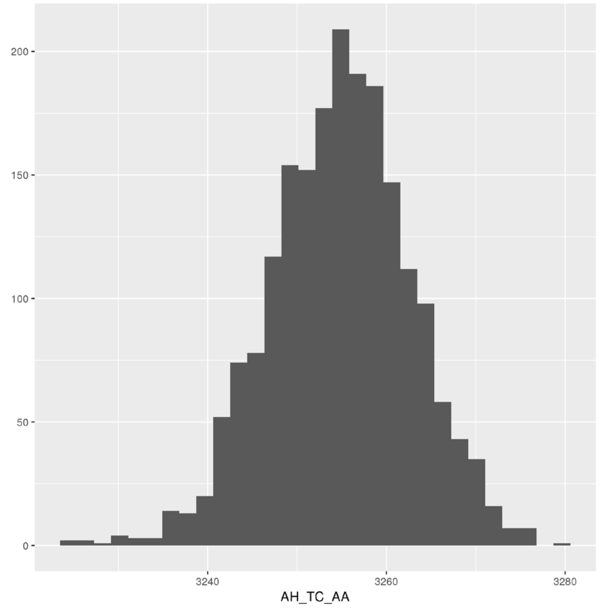
Total Cholesterol (TC)



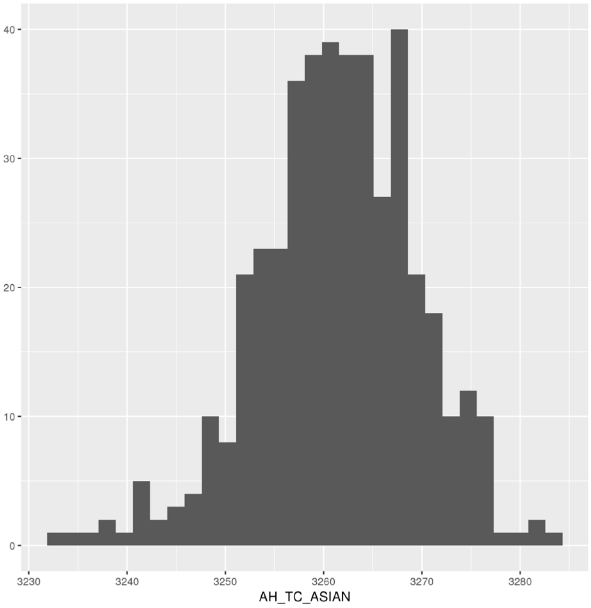
European ancestry raw PGS distribution

Hispanic ancestry raw PGS distribution

African ancestry raw PGS distribution



Asian ancestry raw PGS distribution



Triglycerides

GWAS Summary Statistic Source: Willer, Cristen J. et al. 2013. “Discovery and Refinement of Loci Associated with Lipid Levels.” *Nature Genetics* 45(11):1274–83.

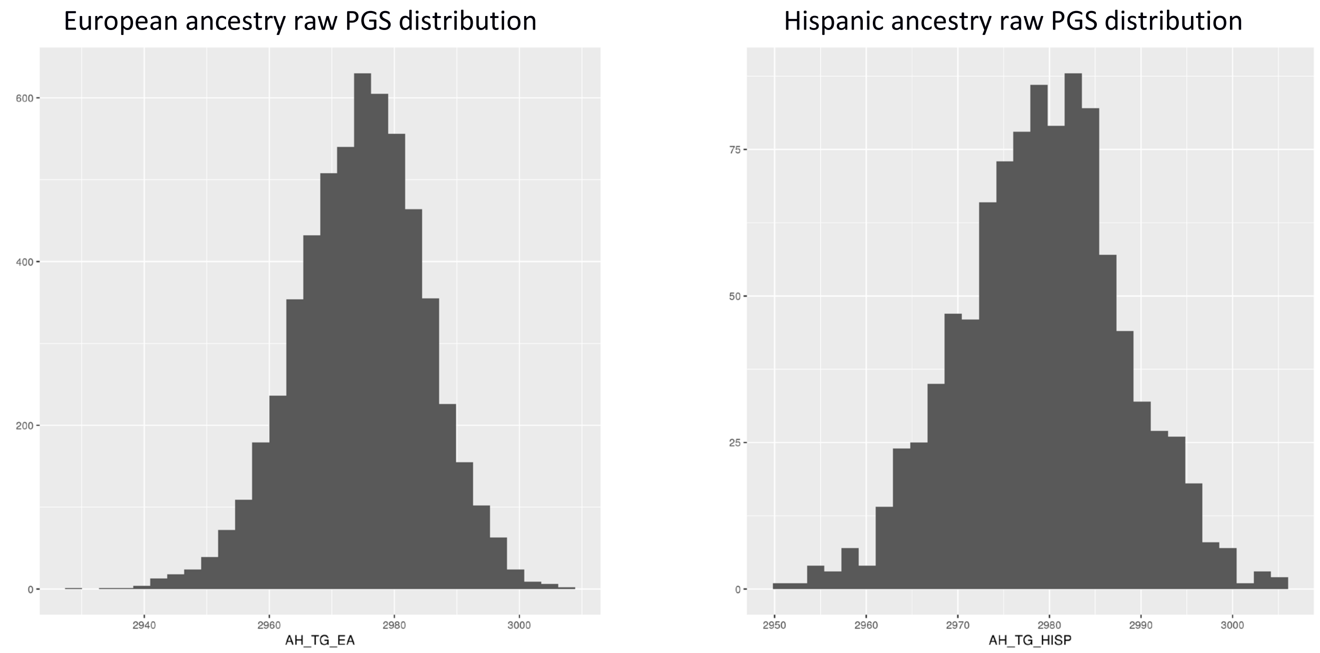
GWAS Ancestry Group(s): European

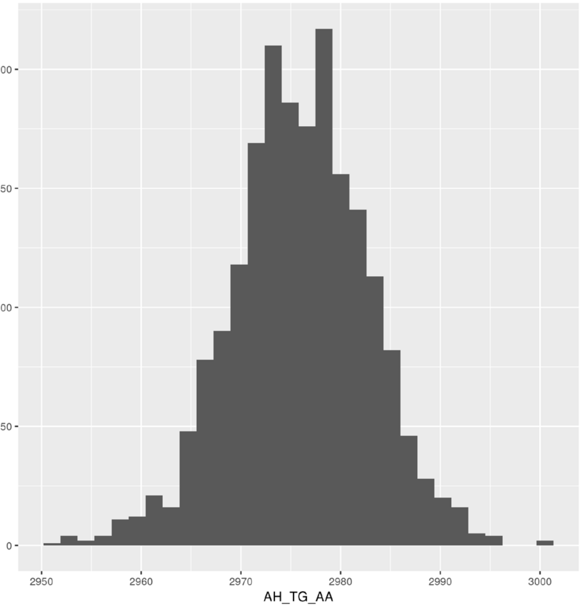
The PGSs for triglycerides (TG) were created using results from a 2013 GWAS conducted by the Global Lipids Genetics Consortium (GLGC). The GWAS meta-analysis files are publicly available and can be downloaded from: http://csg.sph.umich.edu/willer/public/lipids2013/.

The discovery phase of the meta-analysis included 188,577 individuals of European ancestry from 45 studies, while the replication sample consisted of 7,898 individuals of non-European descent. Triglycerides were measured after eight or more hours of fasting with individuals taking lipid lowering medication excluded from the study. Study-specific GWAS adjusted for age, age-squared, sex, and genetic principal components.

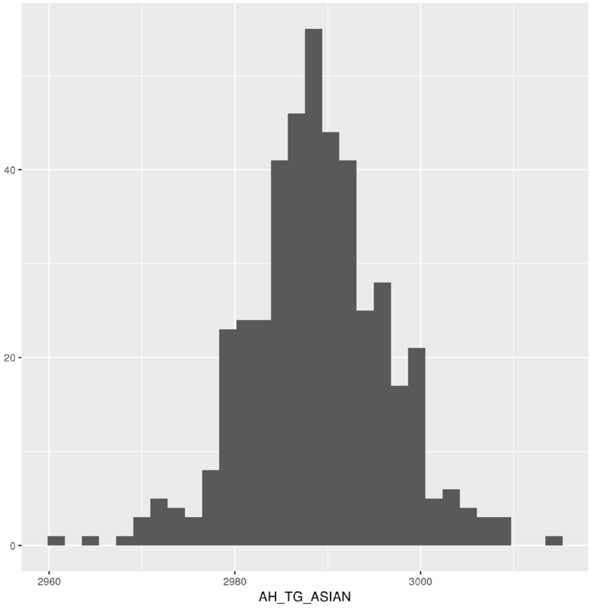
The Add Health PGSs for triglycerides are standardized (mean = 0, standard deviation = 1) within ancestry groups to control for population stratification. To give researchers an idea of the underlying raw PGS, we present density plots for the unstandardized PGSs below. **Please read the section entitled “Using Add Health PGSs” in the introductory portion of this document prior to conducting any analyses using the provided PGSs.**

Triglycerides (TG)





African ancestry raw PGS distribution



Asian ancestry raw PGS distribution

Type II Diabetes (2012)

GWAS Summary Statistic Source: Morris, Andrew P. et al. 2012. "Large-Scale Association Analysis Provides Insights into the Genetic Architecture and Pathophysiology of Type 2 Diabetes." Nature Genetics 44(9):981–90.

GWAS Ancestry Group(s): European

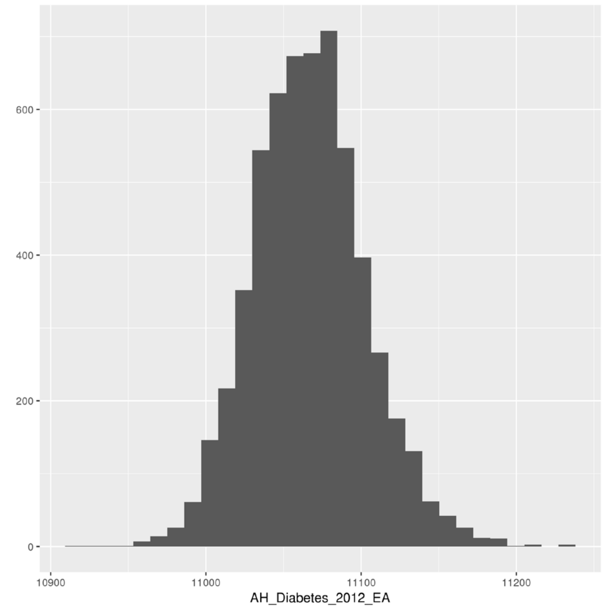
The PGSs for Type II Diabetes 2012 were created using GWAS meta-analysis results from a 2012 study conducted by the DIAbetes Genetics Replication and Meta-analysis (DIAGRAM) Consortium. The GWAS meta-analysis files are available from: http://www.diagram-consortium.org/downloads.html.

The discovery meta-analysis consists of 12,171 T2D cases and 56,862 controls across 12 studies of populations of European descent. The replication meta-analysis consisted of 22,669 cases and 58,119 controls from studies of European populations and 1,178 cases and 2,472 controls from a study of Pakistani populations (Pakistan Risk of Myocardial Infarction Study). The authors note that they found little evidence of heterogeneity in SNP effects between European and Pakistani ancestry groups (p.981). Study-specific GWAS adjusted for age of onset (cases) or age of recruitment (controls), gender, and genetic principal components. Additionally, the results of each GWAS were corrected for residual population structure using the genomic control inflation factor, as reported in Supplementary Table 1 of Morris et al. (2012).

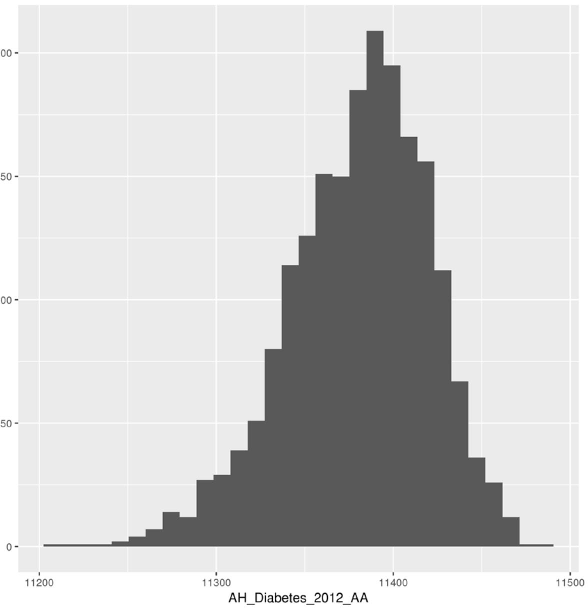
The Add Health PGSs for Type II Diabetes 2012 are standardized (mean = 0, standard deviation = 1) within ancestry groups to control for population stratification. To give researchers an idea of the underlying raw PGS, we present density plots for the unstandardized PGSs below. **Please read the section entitled "Using Add Health PGSs" in the introductory portion of this document prior to conducting any analyses using the provided PGSs.**

Type II Diabetes 2012

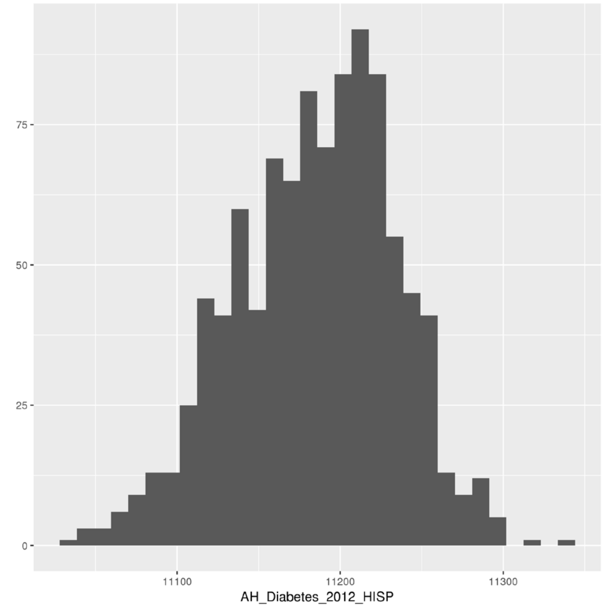
European ancestry raw PGS distribution



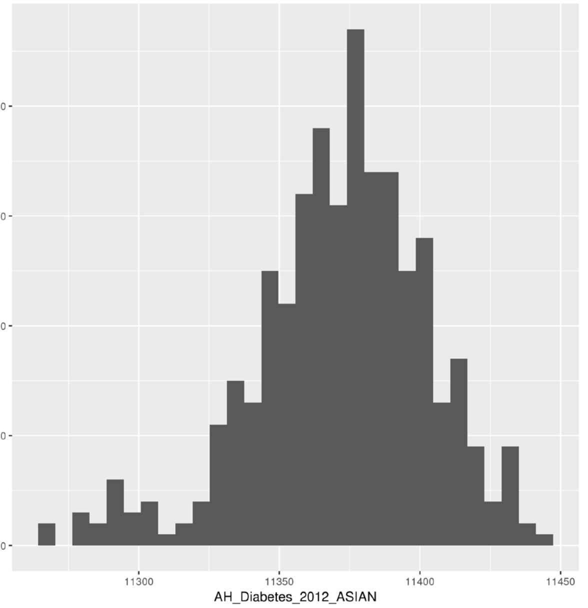
African ancestry raw PGS distribution



Hispanic ancestry raw PGS distribution



Asian ancestry raw PGS distribution



Add Health is a program project directed by Kathleen Mullan Harris and designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris at the University of North Carolina at Chapel Hill, and funded by grant P01-HD31921 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development, with cooperative funding from 23 other federal agencies and foundations. Information on how to obtain the Add Health data files is available on the Add Health website (http://www.cpc.unc.edu/addhealth).

Type II Diabetes (2014)

GWAS Summary Statistic Source: Mahajan, Anubha et al. 2014. “Genome-Wide Trans-Ancestry Meta-Analysis Provides Insight into the Genetic Architecture of Type 2 Diabetes Susceptibility.” Nature Genetics 46(3):234–44.

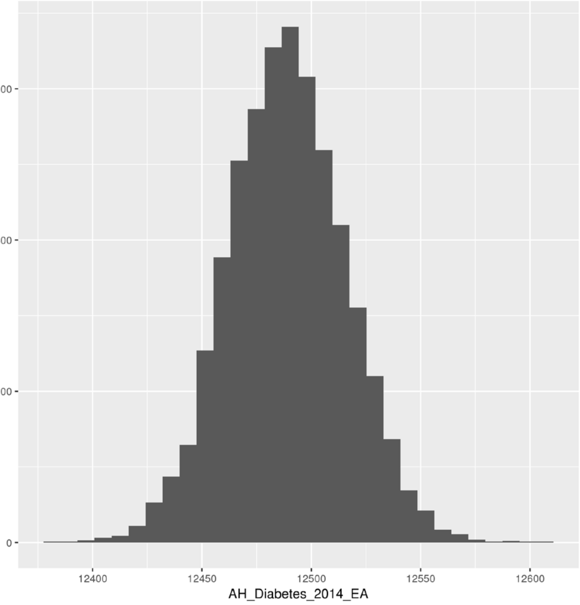
GWAS Ancestry Group(s): European (Ncase = 12,171, Ncontrol = 56,862), East Asian (Ncase = 6,952, Ncontrol = 11,865), South Asian (Ncase = 5,561, Ncontrol = 14,458), Mexican/Mexican American (Ncase = 1,804, Ncontrol = 779).

The PGSs for Type II Diabetes 2014 were created using GWAS meta-analysis results from a 2014 study conducted by the DIAbetes Genetics Replication and Meta-analysis (DIAGRAM) Consortium, the Asian Genetic Epidemiology Network Type 2 Diabetes (AGEN-T2D) Consortium, the South Asian Type 2 Diabetes (SAT2D) Consortium, the Mexican American Type 2 Diabetes (MAT2D) Consortium, and the Type 2 Diabetes Genetic Exploration by Next-generation sequencing in multi-Ethnic Samples (T2D-GENES) Consortium. The GWAS meta-analysis files are publicly available from: http://www.diagram-consortium.org/downloads.html.

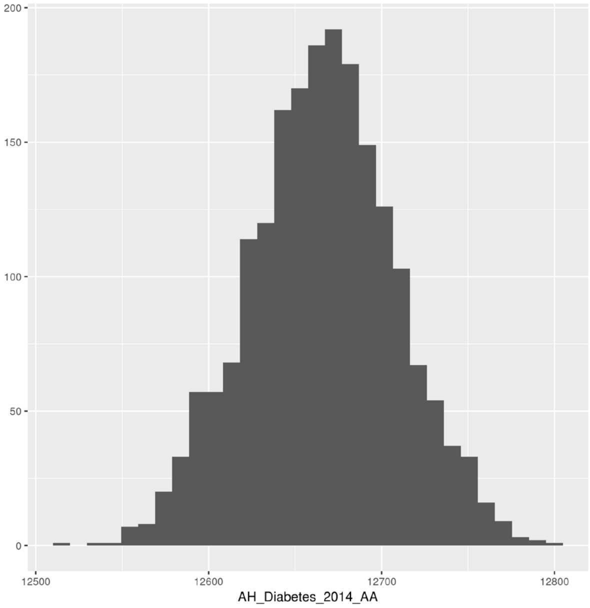
The discovery meta-analysis consists of 26,488 T2D cases and 83,964 controls in populations of European, East Asian, south Asian, and Mexican and Mexican American descent as described above. The replication meta-analysis consisted of 21,491 cases and 55,647 controls in populations of European descent. Study-specific GWAS adjustments varied, but largely included: age, sex, and cohort.

The Add Health PGSs for Type II Diabetes 2014 are standardized (mean = 0, standard deviation = 1) within ancestry groups to control for population stratification. To give researchers an idea of the underlying raw PGS, we present density plots for the unstandardized PGSs below. **Please read the section entitled “Using Add Health PGSs” in the introductory portion of this document prior to conducting any analyses using the provided PGSs.**

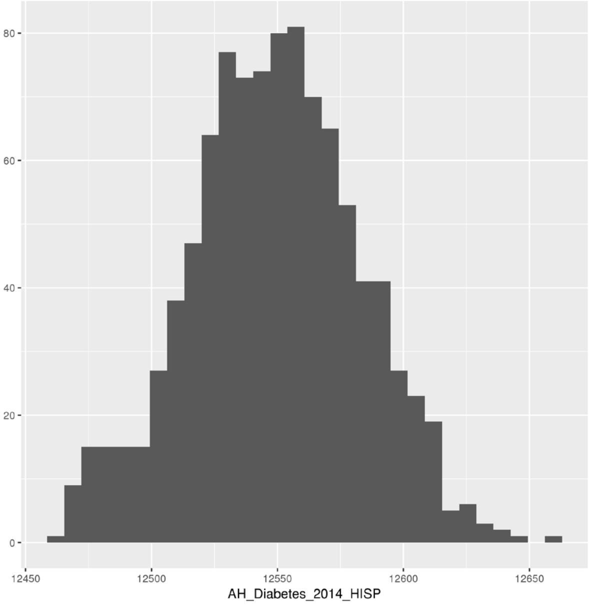
Type II Diabetes 2014



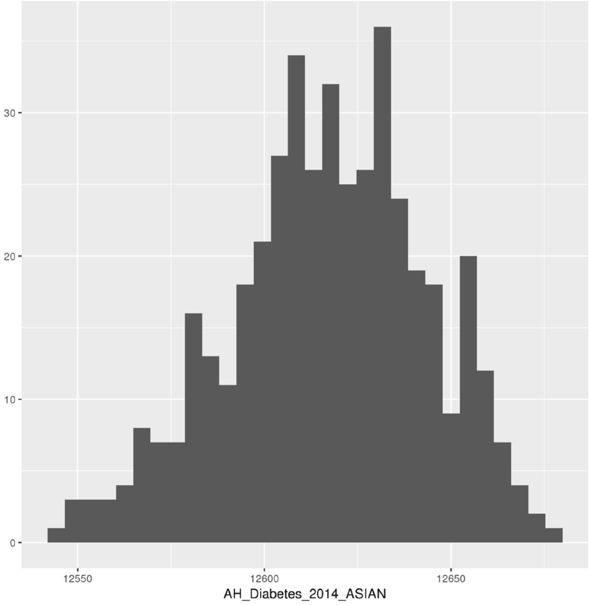
European ancestry raw PGS distribution



African ancestry raw PGS distribution



Hispanic ancestry raw PGS distribution



Asian ancestry raw PGS distribution

Add Health is a program project directed by Kathleen Mullan Harris and designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris at the University of North Carolina at Chapel Hill, and funded by grant P01-HD31921 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development, with cooperative funding from 23 other federal agencies and foundations. Information on how to obtain the Add Health data files is available on the Add Health website (http://www.cpc.unc.edu/addhealth).

Body Mass Index (2015)

GWAS Summary Statistic Source: Locke, Adam E. et al. 2015. “Genetic Studies of Body Mass Index Yield New Insights for Obesity Biology.” *Nature* 518(7538):197–206.

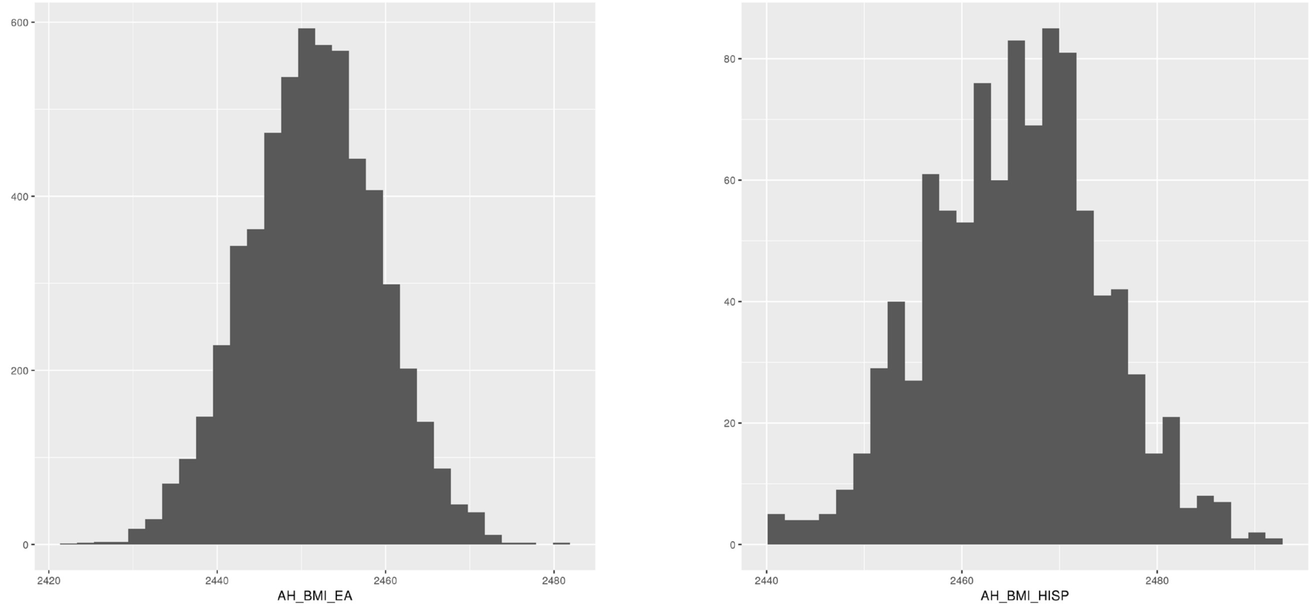
GWAS Ancestry Group(s): European

PGSs for body mass index (BMI) were created using results from a 2015 GWAS conducted by the Genetic Investigation of ANthropometric Traits (GIANT) consortium. The GWAS meta-analysis files are publicly available from: http://portals.broadinstitute.org/collaboration/giant/index.php/GIANT_consortium_data_files.

Joint GWAS and Metabochip meta-analysis was performed on a sample of 332,206 (Ngwas = 234,069; Nmeta = 88,137) individuals of European descent from 114 studies. Additional covariates controlled for during the meta-analysis include: age, the second order polynomial of age, sex, and genetic principal components. The measures of BMI include clinical measurements as well as self-reported BMI.

The Add Health BMI are standardized (mean = 0, standard deviation = 1) within ancestry groups to control for population stratification. To give researchers an idea of the underlying raw PGS, we present density plots for the unstandardized PGSs below. **Please read the section entitled “Using Add Health PGSs” in the introductory portion of this document prior to conducting any analyses using the provided PGSs.**

Body Mass Index (2015)

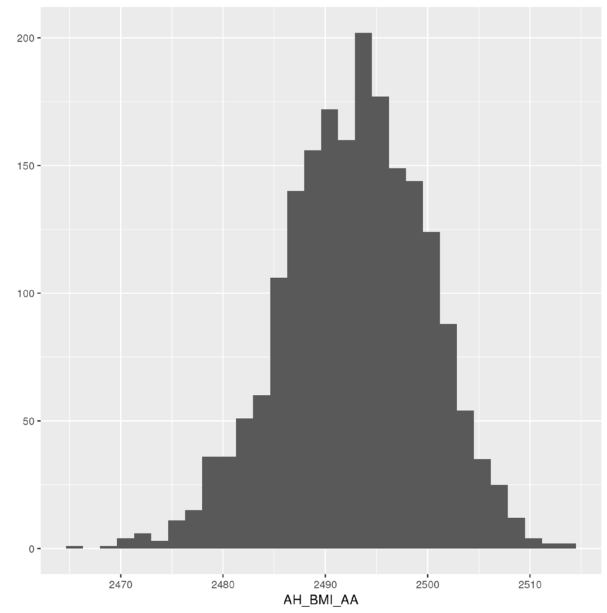


European ancestry raw PGS distribution

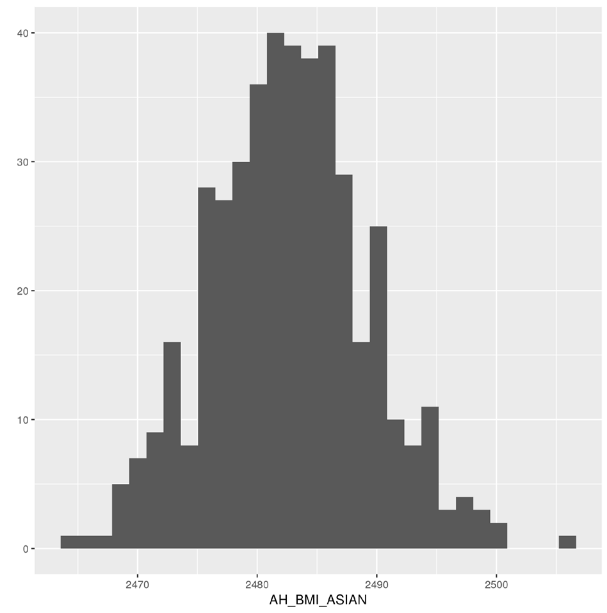
Hispanic ancestry raw PGS distribution

Add Health is a program project directed by Kathleen Mullan Harris and designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris at the University of North Carolina at Chapel Hill, and funded by grant P01-HD31921 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development, with cooperative funding from 23 other federal agencies and foundations. Information on how to obtain the Add Health data files is available on the Add Health website (http://www.cpc.unc.edu/addhealth).

African ancestry raw PGS distribution



Asian ancestry raw PGS distribution



Add Health is a program project directed by Kathleen Mullan Harris and designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris at the University of North Carolina at Chapel Hill, and funded by grant P01-HD31921 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development, with cooperative funding from 23 other federal agencies and foundations. Information on how to obtain the Add Health data files is available on the Add Health website (http://www.cpc.unc.edu/addhealth).

Body Mass Index (2018)

GWAS Summary Statistic Source: Yengo et al. 2018. “Meta-Analysis of Genome-Wide Association Studies for Height and Body Mass Index in ~700,000 Individuals of European Ancestry.” *Human Molecular Genetics* 27(20):3641–49.

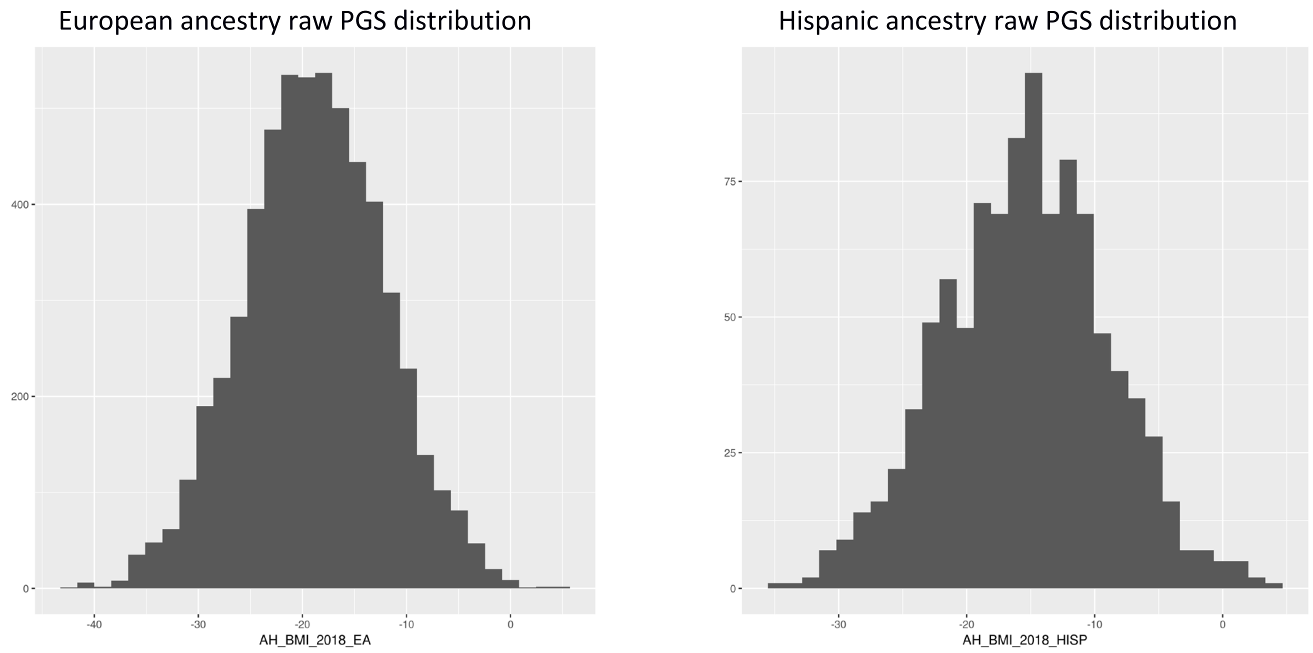
GWAS Ancestry Group(s): European

The PGSs for body mass index (BMI) were created using results from a 2018 study conducted by the Loic Yengo and colleagues. The GWAS meta-analysis files are publicly available and can be downloaded from: https://portals.broadinstitute.org/collaboration/giant/index.php/GIANT_consortium_data_files.

The BMI GWAS meta-analyses are based on the analysis of 456,426 individuals of European genetic ancestry from the UK Biobank and summary statistics from Adam Locke, the Genetic Investigation of ANthropometric Traits (GIANT) consortium, and colleagues (2015). GWAS analyses adjusted for age, sex, recruitment centers, genotyping batches, and the first 10 principal components of the genetic data (Yengo et al. 2018 for more details).

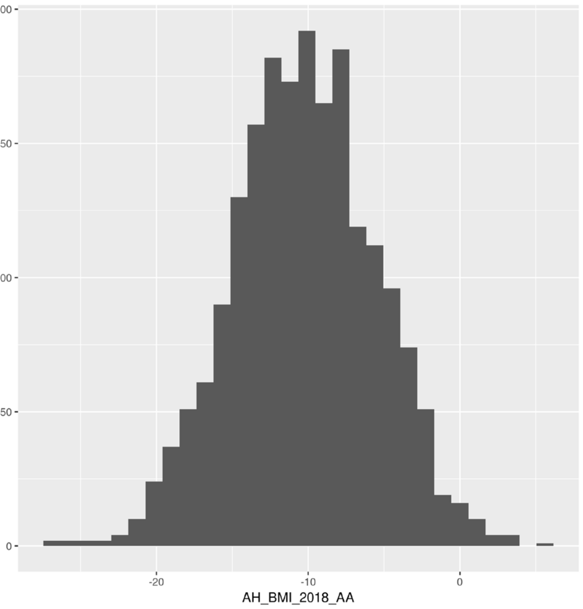
The Add Health PGSs for BMI are standardized (mean = 0, standard deviation = 1) within ancestry groups to control for population stratification. To give researchers an idea of the underlying raw PGS, we present density plots for the unstandardized PGSs below. **Please read the section entitled “Using Add Health PGSs” in the introductory portion of this document prior to conducting any analyses using the provided PGSs.**

Body Mass Index (2018)

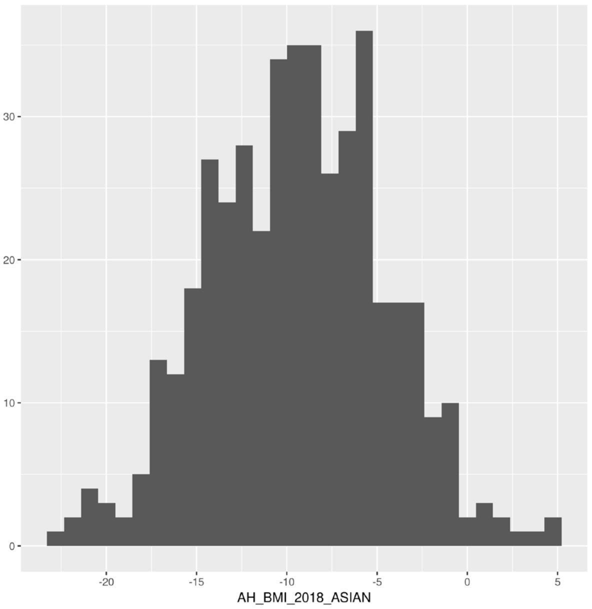


European ancestry raw PGS distribution

Hispanic ancestry raw PGS distribution



African ancestry raw PGS distribution



Asian ancestry raw PGS distribution

Add Health is a program project directed by Kathleen Mullan Harris and designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris at the University of North Carolina at Chapel Hill, and funded by grant P01-HD31921 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development, with cooperative funding from 23 other federal agencies and foundations. Information on how to obtain the Add Health data files is available on the Add Health website (http://www.cpc.unc.edu/addhealth).

Waist Circumference

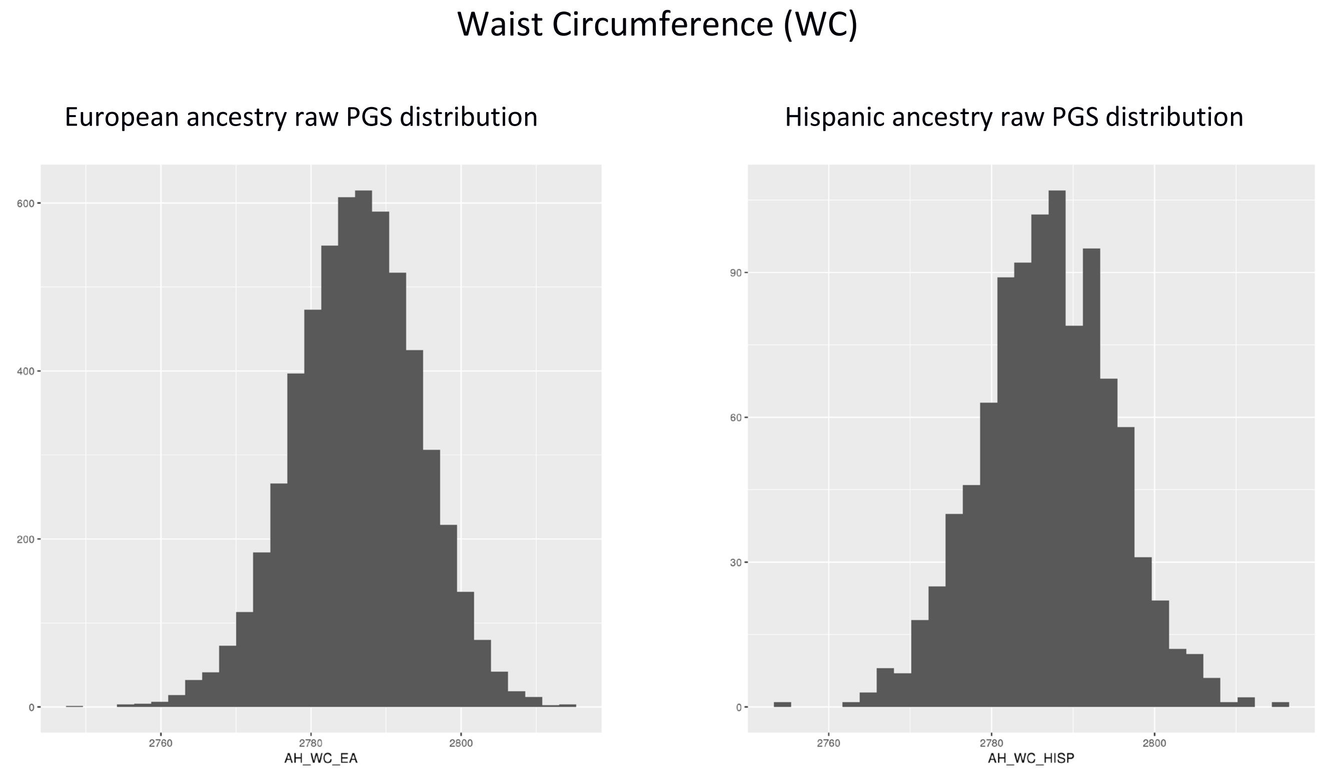
GWAS Summary Statistic Source: Shungin, Dmitry et al. 2015. “New Genetic Loci Link Adipose and Insulin Biology to Body Fat Distribution.” *Nature* 518(7538):187–96.

GWAS Ancestry Group(s): European

PGSs for waist circumference were created using results from a 2015 study conducted by the Genetic Investigation of ANthropometric Traits (GIANT) consortium. The GWAS meta-analysis files are publicly available from: http://portals.broadinstitute.org/collaboration/giant/index.php/GIANT_consortium_data_files.

A GWAS meta-analysis was performed on a sample of 142,762 individuals of European descent from 57 studies, and separately in a Metabochip (MC) meta-analysis on a sample of 67,326 individuals of European descent from 44 studies. A joint GWAS and MC meta-analysis was then conducted on 210,088 individuals of European descent. Additional covariates in both GWAS and MC analyses included: age, the second order polynomial of age, study-specific covariates as necessary, and BMI.

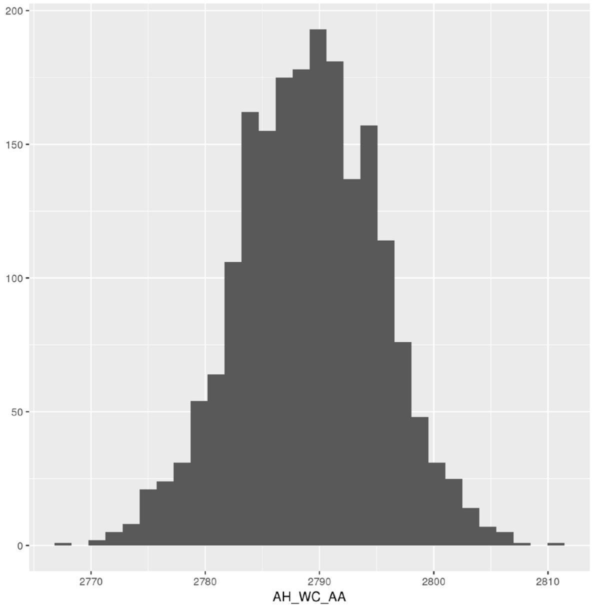
The Add Health PGSs for waist circumference are standardized (mean = 0, standard deviation = 1) within ancestry groups to control for population stratification. To give researchers an idea of the underlying raw PGS, we present density plots for the unstandardized PGSs below. **Please read the section entitled “Using Add Health PGSs” in the introductory portion of this document prior to conducting any analyses using the provided PGSs.**



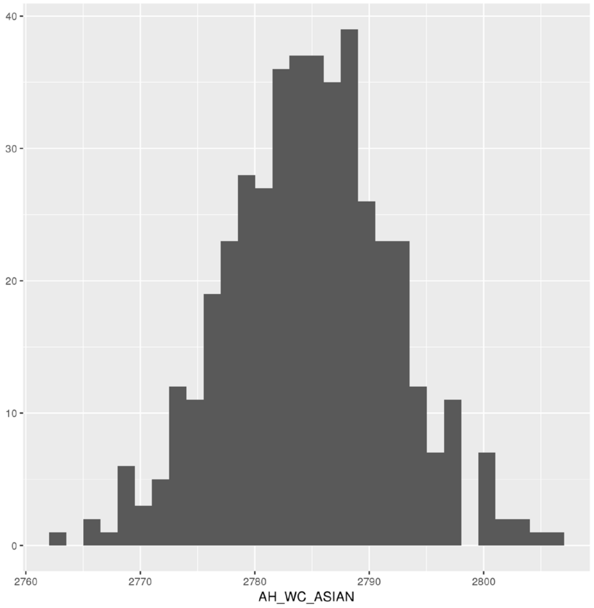
Waist Circumference (WC)

European ancestry raw PGS distribution

Hispanic ancestry raw PGS distribution



African ancestry raw PGS distribution



Asian ancestry raw PGS distribution

Waist-to-Hip Ratio

GWAS Summary Statistic Source: Shungin, Dmitry et al. 2015. "New Genetic Loci Link Adipose and Insulin Biology to Body Fat Distribution." *Nature* 518(7538):187–96.

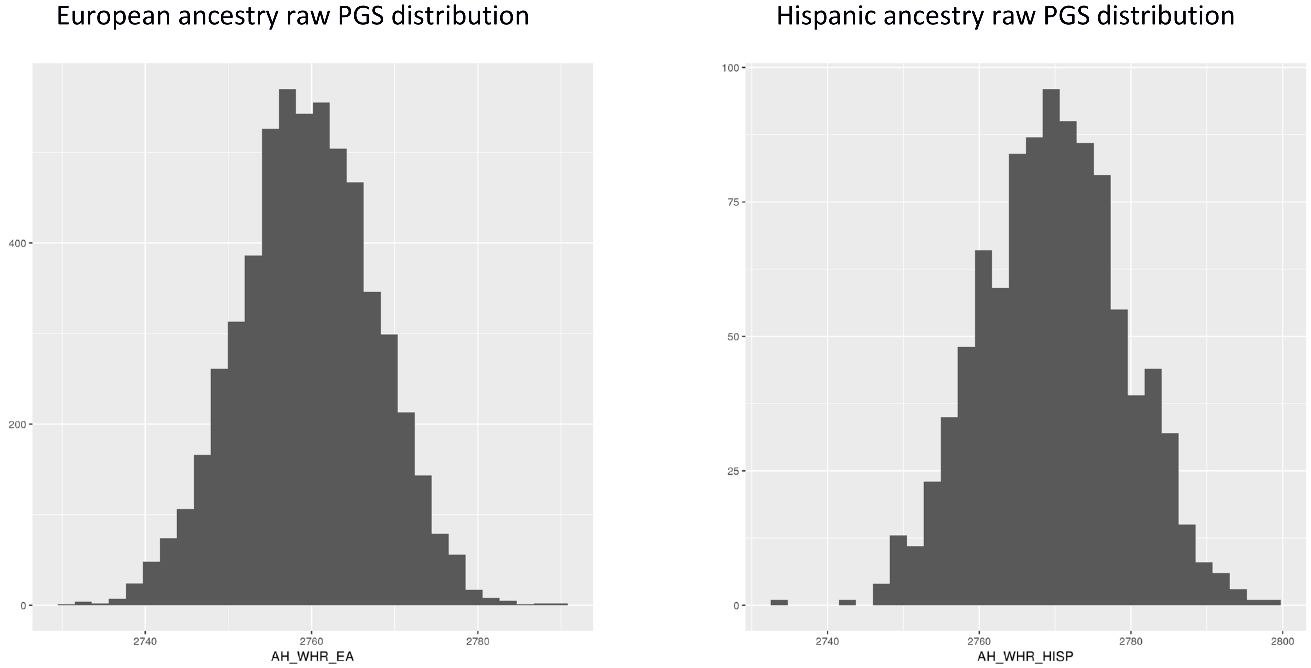
GWAS Ancestry Group(s): European

PGSs for waist-to-hip ratio were created using results from a 2015 study conducted by the Genetic Investigation of ANthropometric Traits (GIANT) consortium. The GWAS meta-analysis files are publicly available from: http://portals.broadinstitute.org/collaboration/giant/index.php/GIANT_consortium_data_files.

GWAS meta-analysis was performed on a sample of 142,762 individuals of European descent from 57 studies, and separately in a Metabochip (MC) meta-analysis on a sample of 67,326 individuals of European descent from 44 studies. A joint GWAS and MC meta-analysis was then conducted on 210,088 individuals of European descent. Additional covariates in both GWAS and MC analyses included: age, the second order polynomial of age, study-specific covariates as necessary, and BMI.

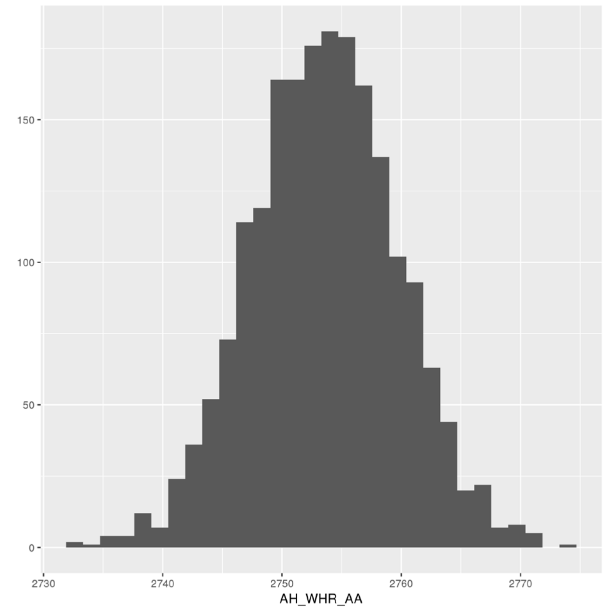
The Add Health PGSs for waist-to-hip ratio are standardized (mean = 0, standard deviation = 1) within ancestry groups to control for population stratification. To give researchers an idea of the underlying raw PGS, we present density plots for the unstandardized PGSs below. **Please read the section entitled "Using Add Health PGSs" in the introductory portion of this document prior to conducting any analyses using the provided PGSs.**

Waist-to-Hip Ratio (WHR)

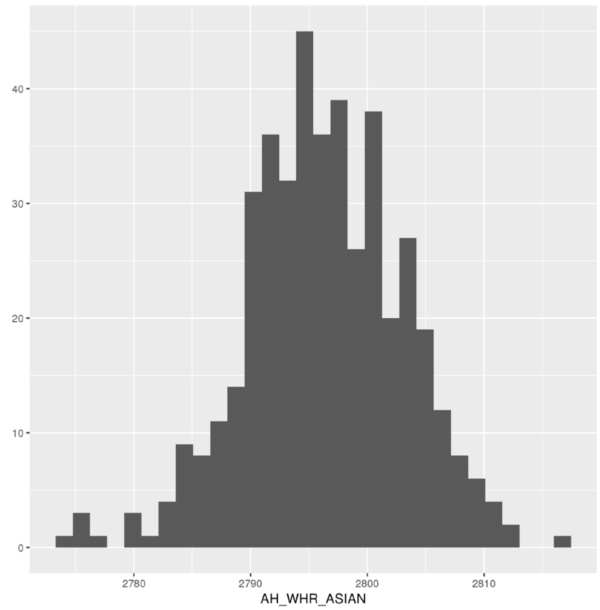

European ancestry raw PGS distribution

Hispanic ancestry raw PGS distribution

African ancestry raw PGS distribution



Asian ancestry raw PGS distribution



Add Health is a program project directed by Kathleen Mullan Harris and designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris at the University of North Carolina at Chapel Hill, and funded by grant P01-HD31921 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development, with cooperative funding from 23 other federal agencies and foundations. Information on how to obtain the Add Health data files is available on the Add Health website (http://www.cpc.unc.edu/addhealth).

Height (2014)

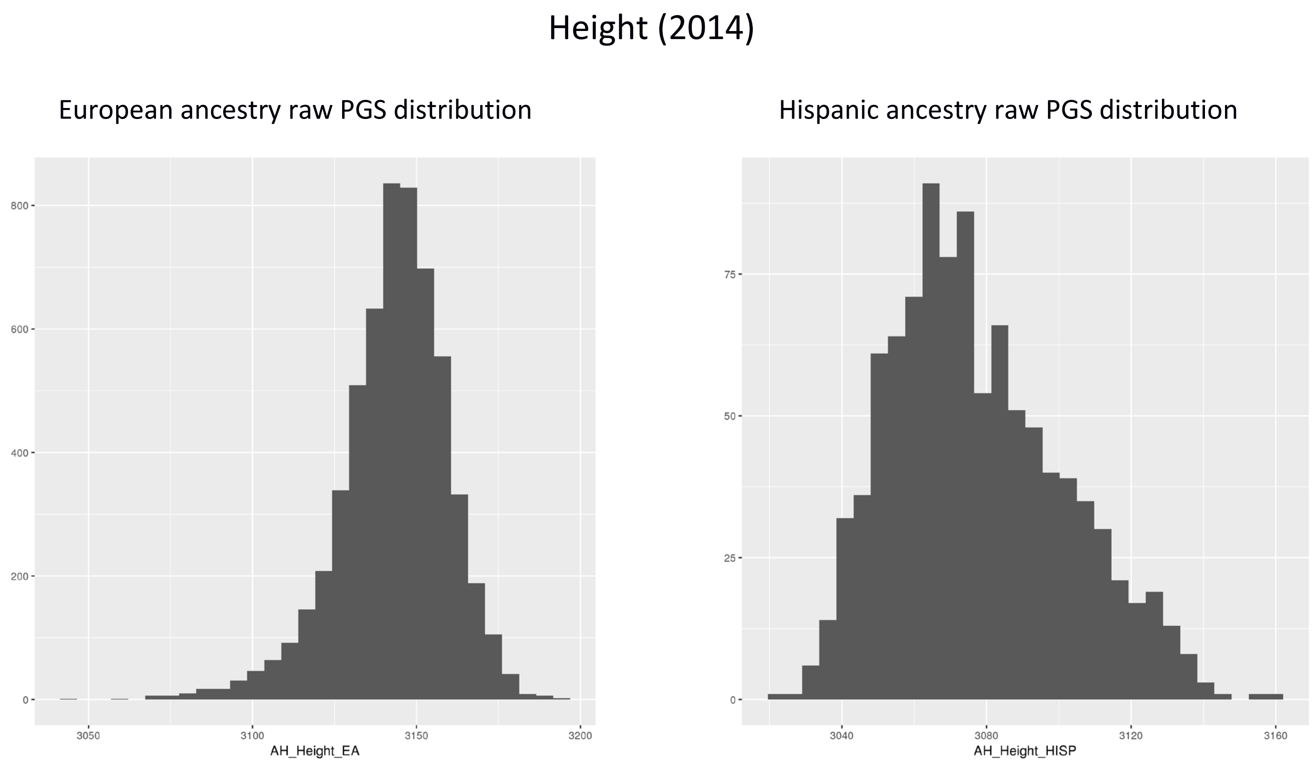
GWAS Summary Statistic Source: Wood, Andrew R. et al. 2014. “Defining the Role of Common Variation in the Genomic and Biological Architecture of Adult Human Height.” *Nature Genetics* 46(11):1173–86.

GWAS Ancestry Group(s): European

PGSs for height were created using results from a 2014 study conducted by the Genetic Investigation of ANthropometric Traits (GIANT) consortium. The GWAS meta-analysis files are publicly available from: http://portals.broadinstitute.org/collaboration/giant/index.php/GIANT_consortium_data_files.

The meta-analysis included 253,288 individuals of European descent from 79 studies imputed to HapMap II. Replication was performed in a sample of 80,067 individuals of European descent. Height was measured as sex standardized height. Additional covariates in GWAS analyses included age and genetic principal components.

The Add Health PGSs for height are standardized (mean = 0, standard deviation = 1) within ancestry groups to control for population stratification. To give researchers an idea of the underlying raw PGS, we present density plots for the unstandardized PGSs below. **Please read the section entitled “Using Add Health PGSs” in the introductory portion of this document prior to conducting any analyses using the provided PGSs.**

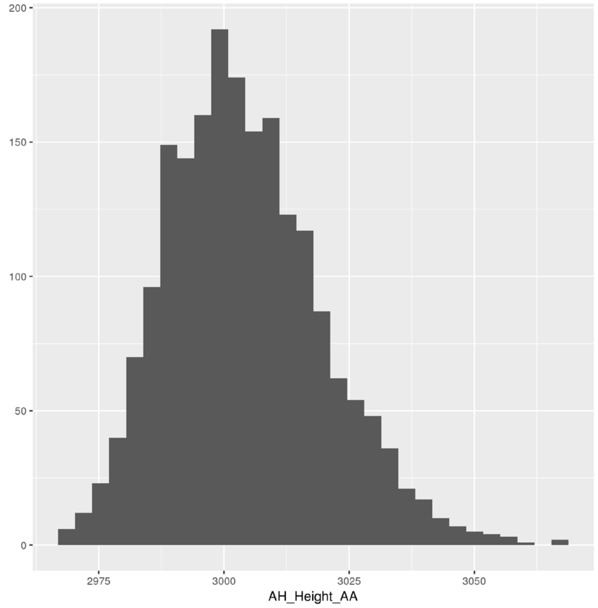

Height (2014)

European ancestry raw PGS distribution

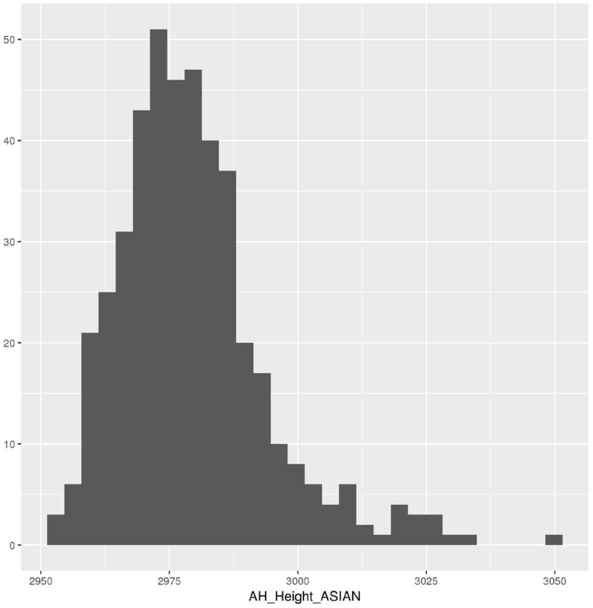
Hispanic ancestry raw PGS distribution

Add Health is a program project directed by Kathleen Mullan Harris and designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris at the University of North Carolina at Chapel Hill, and funded by grant P01-HD31921 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development, with cooperative funding from 23 other federal agencies and foundations. Information on how to obtain the Add Health data files is available on the Add Health website (http://www.cpc.unc.edu/addhealth).

African ancestry raw PGS distribution



Asian ancestry raw PGS distribution



Height (2018)

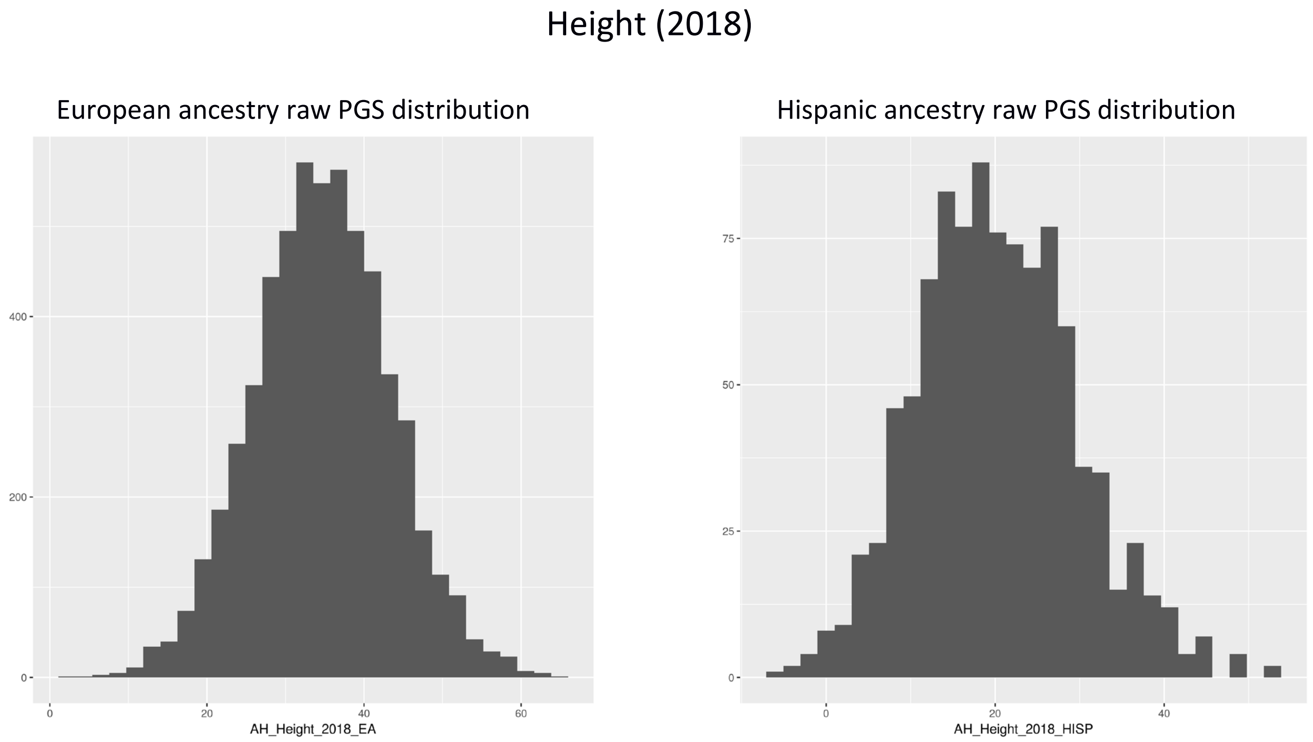
GWAS Summary Statistic Source: Yengo et al. 2018. “Meta-Analysis of Genome-Wide Association Studies for Height and Body Mass Index in ~700,000 Individuals of European Ancestry.” *Human Molecular Genetics* 27(20):3641–49.

GWAS Ancestry Group(s): European

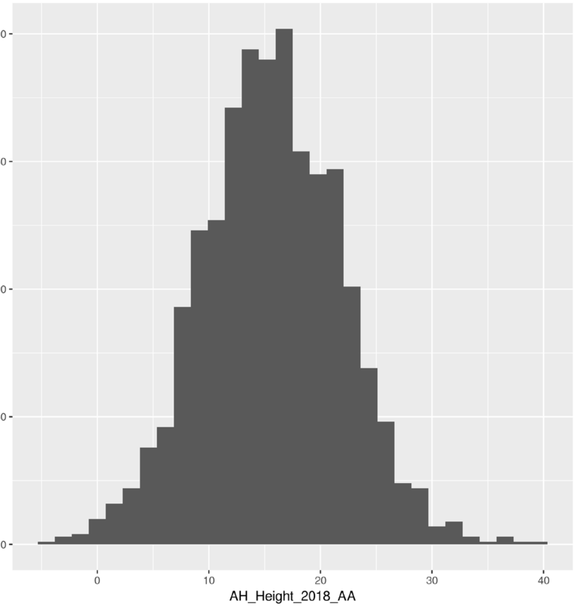
The PGSs for height were created using results from a 2018 study conducted by the Loic Yengo and colleagues. The GWAS meta-analysis files are publicly available and can be downloaded from: https://portals.broadinstitute.org/collaboration/giant/index.php/GIANT_consortium_data_files.

The Height GWAS meta-analyses are based on the analysis of 456,426 individuals of European genetic ancestry from the UK Biobank and summary statistics from Andrew Wood, the Genetic Investigation of ANthropometric Traits (GIANT) consortium, and others (2014). GWAS analyses adjusted for age, sex, recruitment centers, genotyping batches, and the first 10 principal components of the genetic data (Yengo et al. 2018 for more details).

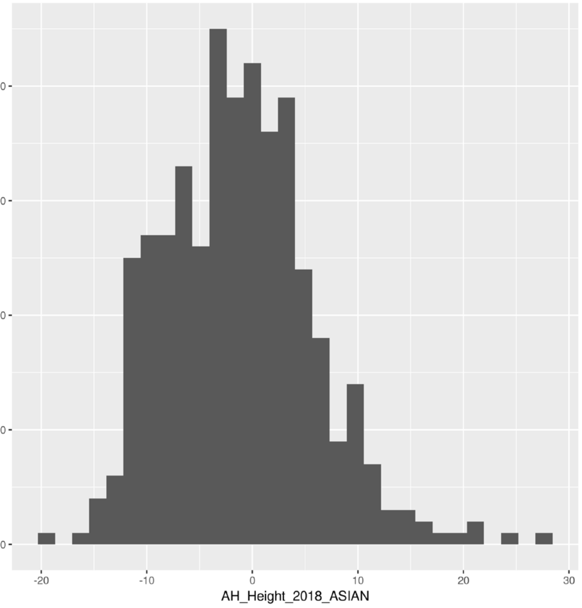
The Add Health PGSs for Height are standardized (mean = 0, standard deviation = 1) within ancestry groups to control for population stratification. To give researchers an idea of the underlying raw PGS, we present density plots for the unstandardized PGSs below. **Please read the section entitled “Using Add Health PGSs” in the introductory portion of this document prior to conducting any analyses using the provided PGSs.**



African ancestry raw PGS distribution



Asian ancestry raw PGS distribution



Alzheimer’s Disease (2013)

GWAS Summary Statistic Source: Lambert, Jean-Charles et al. 2013. “Meta-Analysis of 74,046 Individuals Identifies 11 New Susceptibility Loci for Alzheimer’s Disease.” *Nature Genetics* 45(12):1452–58.

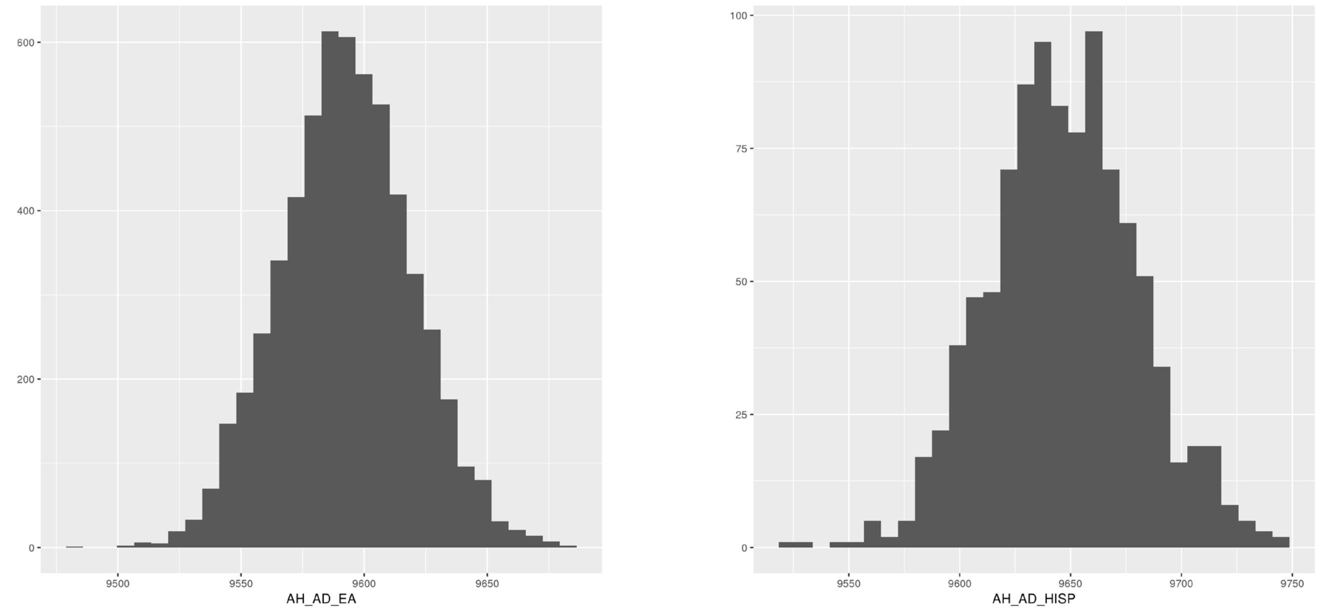
GWAS Ancestry Group(s): European

The PGSs for Alzheimer’s disease were created using results from a 2013 GWAS conducted by the International Genomics of Alzheimer’s Project (IGAP). The GWAS meta-analysis files are publicly available from: http://web.pasteur-lille.fr/en/recherche/u744/igap/igap_download.php.

The meta-analysis of Alzheimer’s disease included 54,162 participants (cases=17,008 and controls=37,154) of European decent imputed to 1000 Genomes 2010 release. The replication sample included 19,884 participants of European ancestry (cases=8,572 and controls=11,312). Additional covariates included in the GWAS for each contributing cohort included age, sex, and genetic principal components.

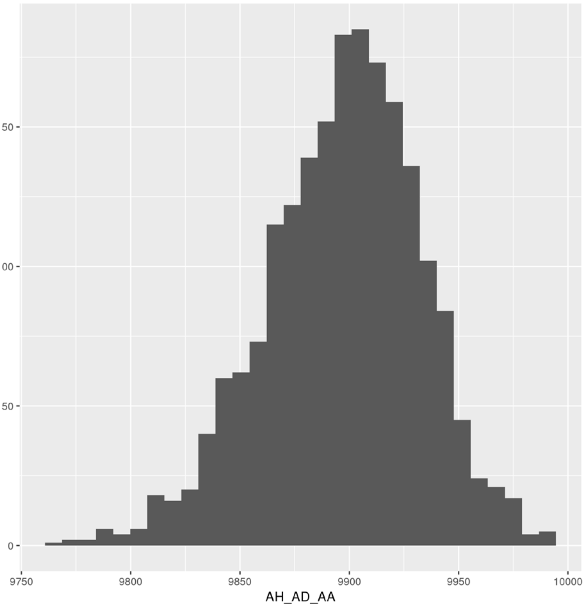
The Add Health PGSs for Alzheimer’s disease are standardized (mean = 0, standard deviation = 1) within ancestry groups to control for population stratification. To give researchers an idea of the underlying raw PGS, we present density plots for the unstandardized PGSs below. **Please read the section entitled “Using Add Health PGSs” in the introductory portion of this document prior to conducting any analyses using the provided PGSs.**

Alzheimer’s Disease (2013)

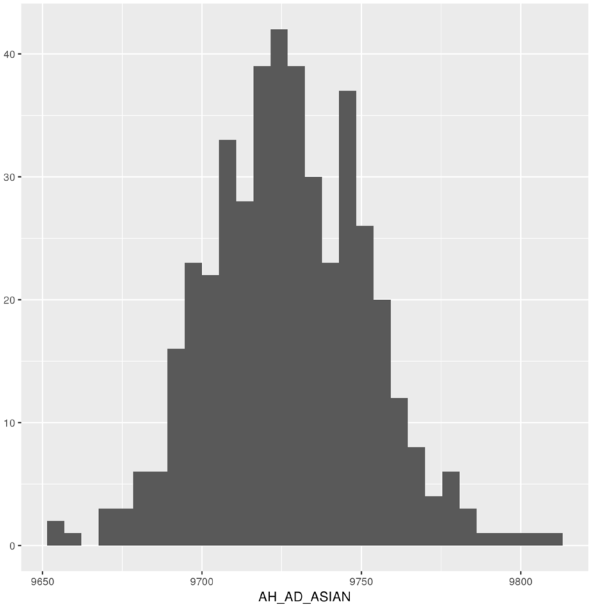

European ancestry raw PGS distribution

Hispanic ancestry raw PGS distribution

African ancestry raw PGS distribution



Asian ancestry raw PGS distribution



Alzheimer’s Disease (2019)

GWAS Summary Statistic Source: Jansen et al. 2019. “Genome-Wide Meta-Analysis Identifies New Loci and Functional Pathways Influencing Alzheimer’s Disease Risk.” *Nature Genetics* 51:404-413.

GWAS Ancestry Group(s): European

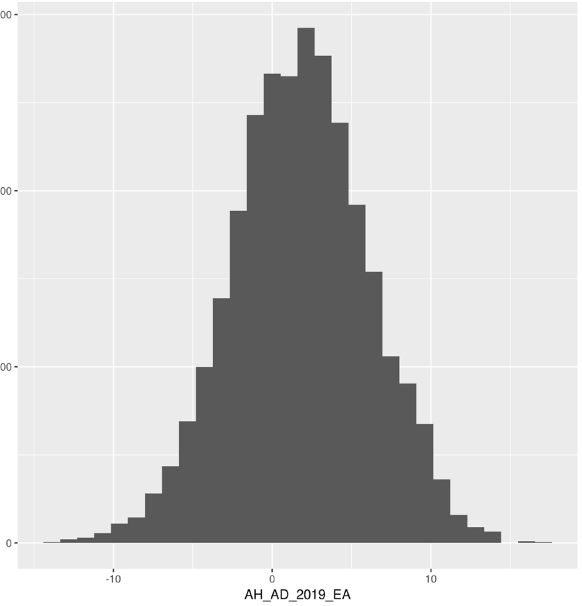
The PGSs for Alzheimer’s Disease (AD) were created using results from a 2019 study conducted by the Iris Jansen and colleagues. The GWAS meta-analysis files are publicly available and can be downloaded from: https://ctg.cncr.nl/software/summary_statistics.

The AD GWAS meta-analysis consisted of 71,880 individuals with either a clinical diagnosis of AD or a parent with a clinical diagnosis for AD (cases) and 383,378 without AD (controls) of European genetic ancestry. The author’s note that parental diagnosis of AD has a strong genetic correlation with individual-level diagnosis (rG = 0.81). GWAS meta-analyses consisted of two stages: 1) a GWAS meta-analysis of clinical diagnosis of AD (24,087 cases and 55,058 controls) spanning cohorts in three AD related consortia (the Psychiatric Genomics Consortium, the International Genomics of Alzheimer’s Projects, and the Alzheimer’s Disease Sequencing Project) and 2) a GWAS of parental diagnosis of AD (47,793 cases and 328,320 controls) from the UK Biobank with cases in which both parents had clinical diagnoses of AD being more heavily weighted than those with a single parent with an AD diagnosis (see the article and supplemental documents for information). GWAS analyses generally accounted for sex, genotyping batch, and the first 4 principal components of the genetic data while some analyses included age (or an age proxy) and additional principal components of the genetic data as covariates. GWAS analysis of the UK Biobank included age, sex, genotyping array, assessment center, and the first 12 principal components of the genetic data.

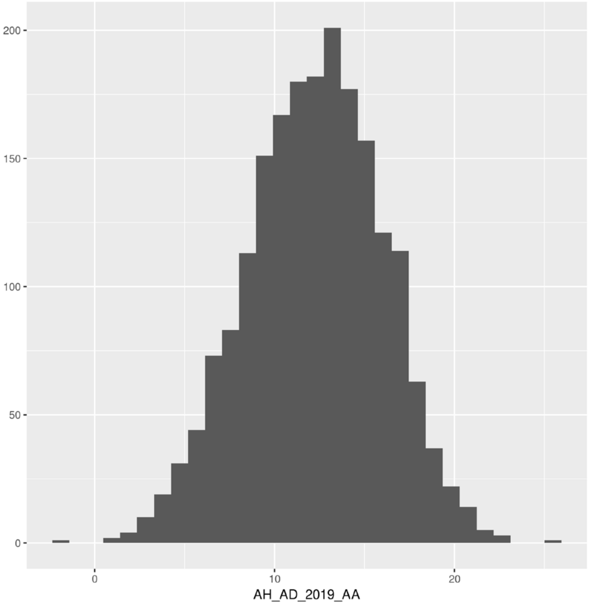
The Add Health PGSs for AD are standardized (mean = 0, standard deviation = 1) within ancestry groups to control for population stratification. To give researchers an idea of the underlying raw PGS, we present density plots for the unstandardized PGSs below. **Please read the section entitled “Using Add Health PGSs” in the introductory portion of this document prior to conducting any analyses using the provided PGSs.**

Alzheimer’s Disease (2019)

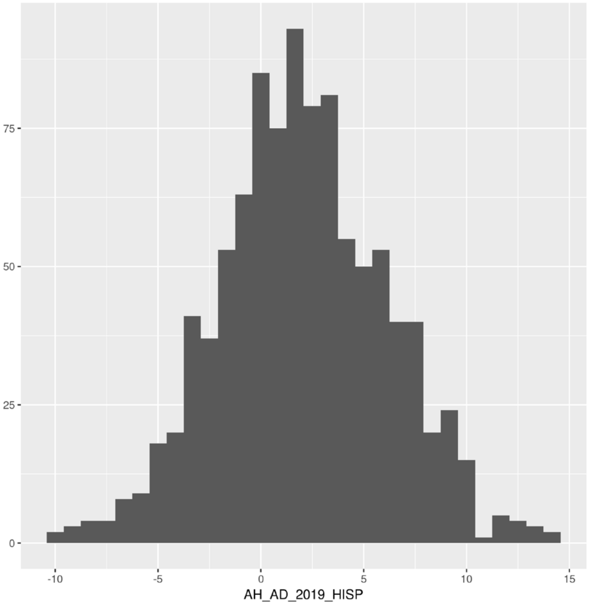
European ancestry raw PGS distribution



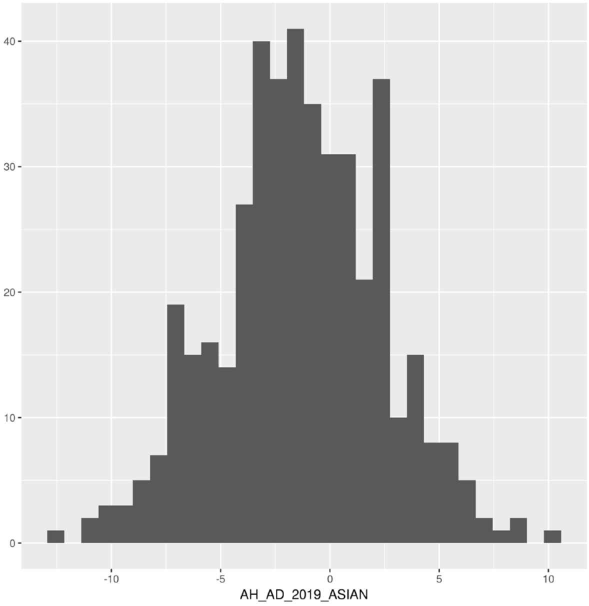
African ancestry raw PGS distribution



Hispanic ancestry raw PGS distribution



Asian ancestry raw PGS distribution



Add Health is a program project directed by Kathleen Mullan Harris and designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris at the University of North Carolina at Chapel Hill, and funded by grant P01-HD31921 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development, with cooperative funding from 23 other federal agencies and foundations. Information on how to obtain the Add Health data files is available on the Add Health website (http://www.cpc.unc.edu/addhealth).

Family History of Alzheimer's Disease

GWAS Summary Statistic Source: Maroni et al. 2018. "GWAS on Family History of Alzheimer's Disease." *Translational Psychiatry* 8(1):1–7.

GWAS Ancestry Group(s): European

The PGSs for a family history of Alzheimer's Disease were created using results from a 2018 study conducted by Riccardo Marioni and colleagues. The GWAS meta-analysis files are publicly available and can be downloaded from: http://www.ccace.ed.ac.uk/node/335.

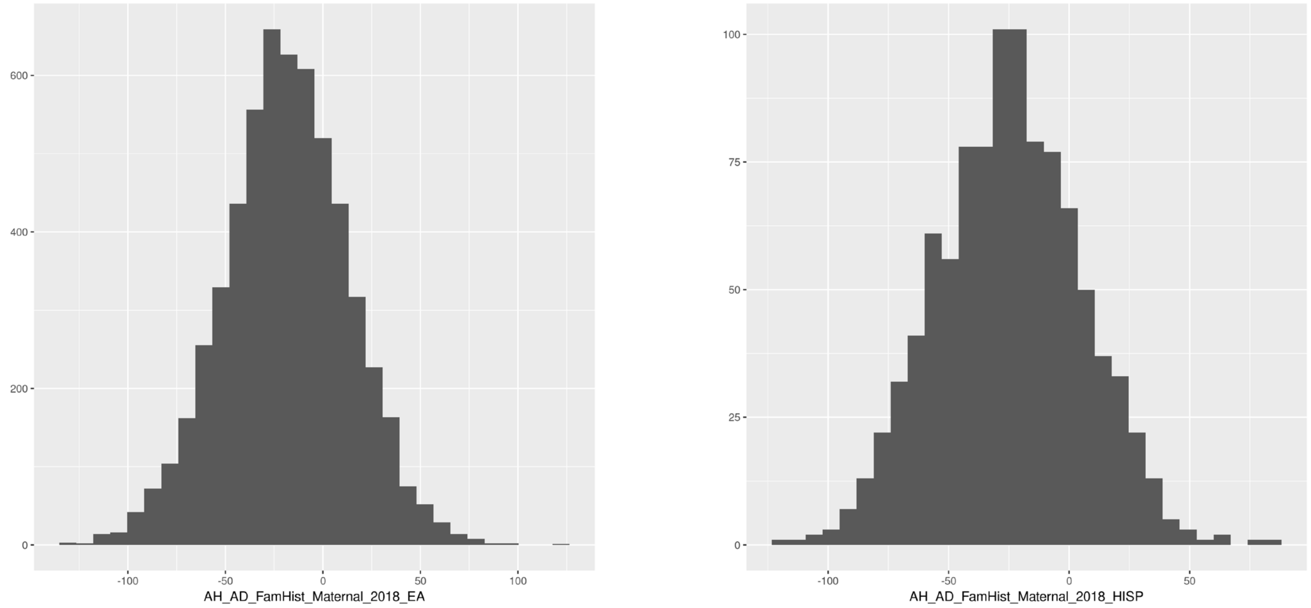
The family history of Alzheimer's Disease GWAS meta-analysis consisted of 314,278 individuals of European genetic ancestry from the UK Biobank with 27,696 maternal cases and 14,338 paternal cases meta-analyses with published consortium data. Family history of Alzheimer's Disease was ascertained by asking respondents if either their mother and/or father had Alzheimer's Disease or dementia. The GWAS was conducted on residuals from a linear regression of self-reports of mother or father Alzheimer's Disease or dementia on age of parent at death or time of interview [whichever occurred first], assessment center, a fixed effect for genotyping batch, genotyping array, and the first 40 principal components of the genome-wide data.

The Add Health PGSs for Family History of Alzheimer's Disease (FAD) are standardized (mean = 0, standard deviation = 1) within ancestry groups to control for population stratification. To give researchers an idea of the underlying raw PGS, we present density plots for the unstandardized PGSs below. **Please read the section entitled "Using Add Health PGSs" in the introductory portion of this document prior to conducting any analyses using the provided PGSs.**

Family History (Maternal) of Alzheimer's Disease

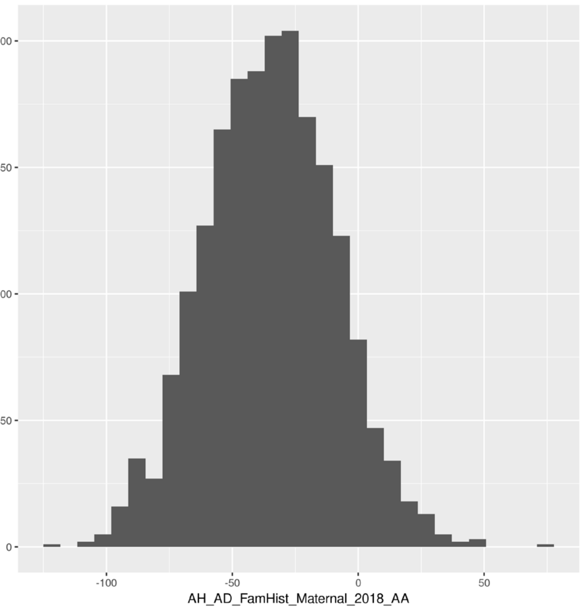
European ancestry raw PGS distribution

Hispanic ancestry raw PGS distribution

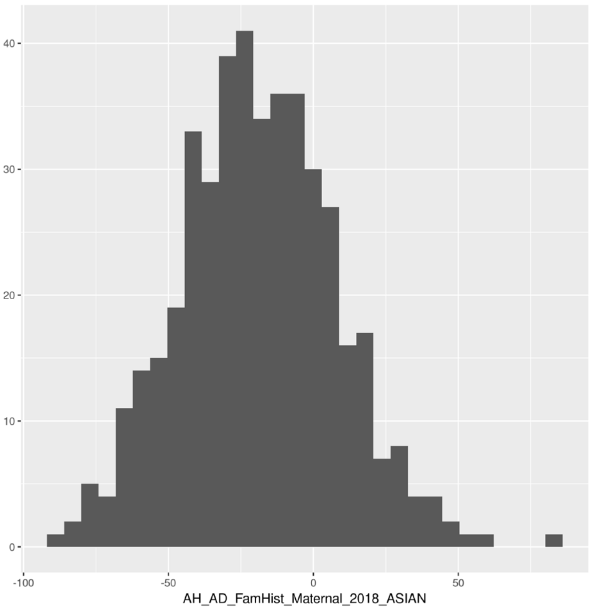


Add Health is a program project directed by Kathleen Mullan Harris and designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris at the University of North Carolina at Chapel Hill, and funded by grant P01-HD31921 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development, with cooperative funding from 23 other federal agencies and foundations. Information on how to obtain the Add Health data files is available on the Add Health website (http://www.cpc.unc.edu/addhealth).

African ancestry raw PGS distribution



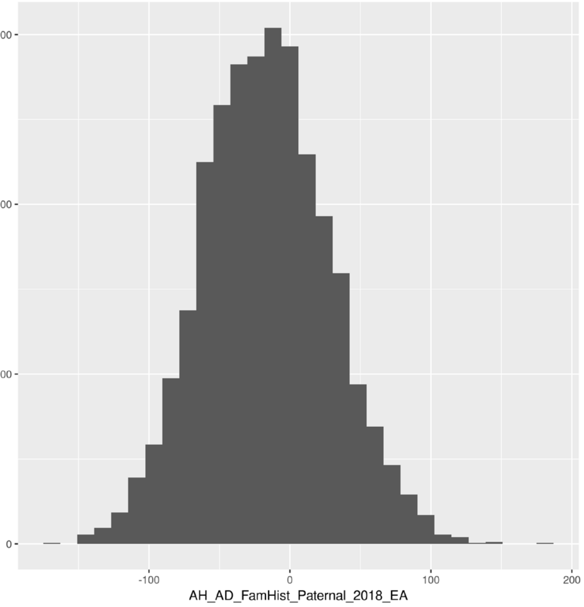
Asian ancestry raw PGS distribution



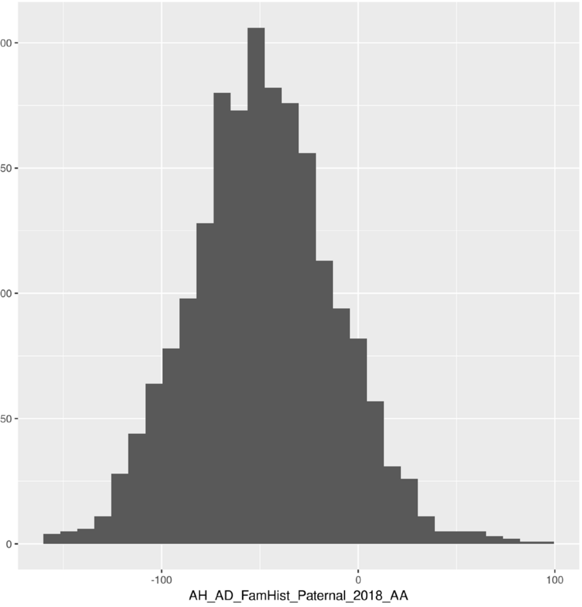
Add Health is a program project directed by Kathleen Mullan Harris and designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris at the University of North Carolina at Chapel Hill, and funded by grant P01-HD31921 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development, with cooperative funding from 23 other federal agencies and foundations. Information on how to obtain the Add Health data files is available on the Add Health website (http://www.cpc.unc.edu/addhealth).

Family History (Paternal) of Alzheimer's Disease

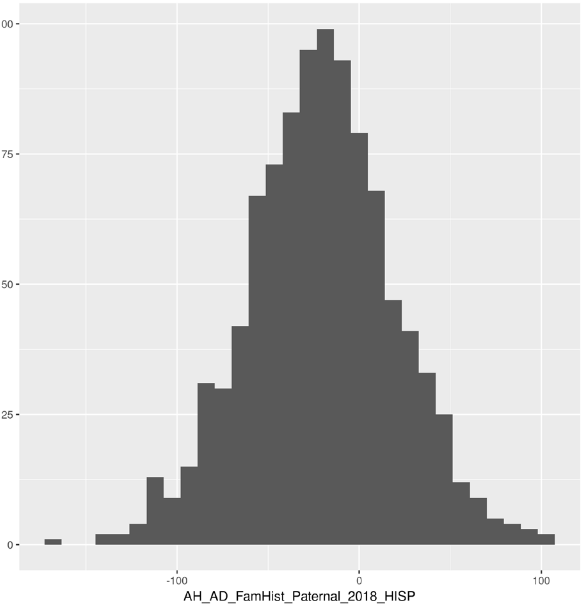
European ancestry raw PGS distribution



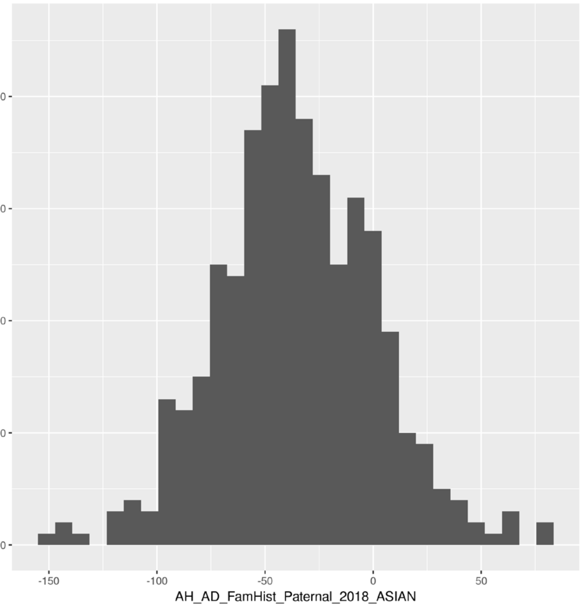
African ancestry raw PGS distribution



Hispanic ancestry raw PGS distribution



Asian ancestry raw PGS distribution



Add Health is a program project directed by Kathleen Mullan Harris and designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris at the University of North Carolina at Chapel Hill, and funded by grant P01-HD31921 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development, with cooperative funding from 23 other federal agencies and foundations. Information on how to obtain the Add Health data files is available on the Add Health website (http://www.cpc.unc.edu/addhealth).

Educational Attainment (2016)

GWAS Summary Statistic Source: Okbay, Aysu et al. 2016. “Genome-Wide Association Study Identifies 74 Loci Associated with Educational Attainment.” *Nature* 533(7604):539–42.

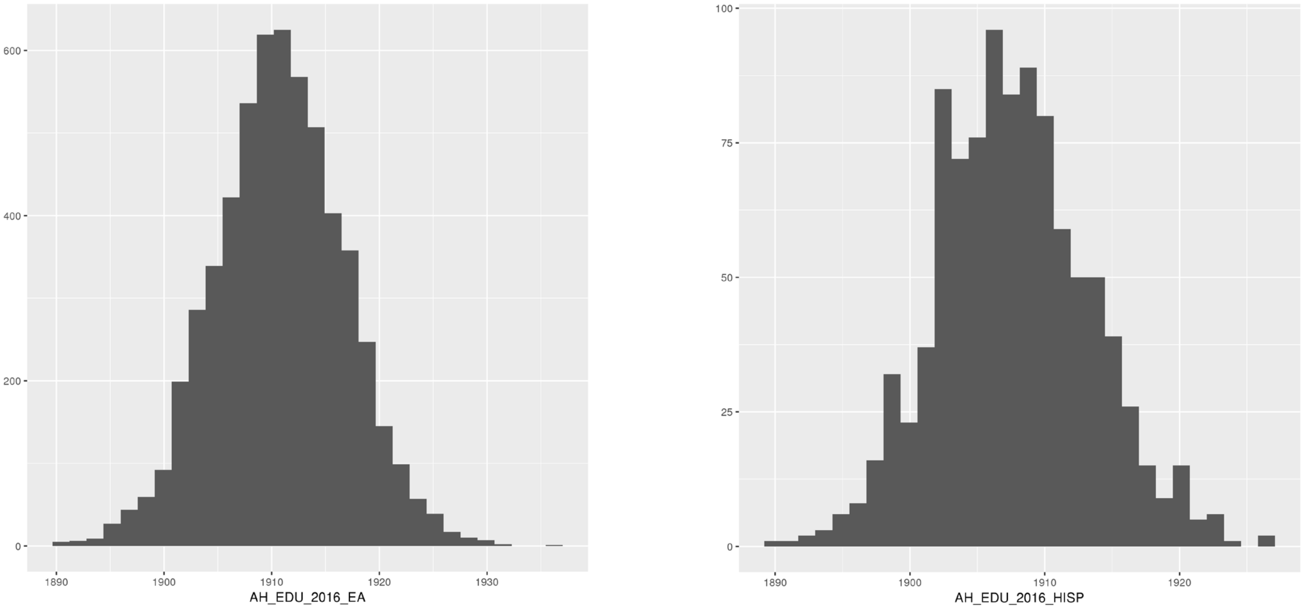
GWAS Ancestry Group(s): European

The educational attainment 2016 PGSs were created using results from a 2016 study by the Social Science Genetic Association Consortium (SSGAC). The GWAS meta-analysis files are publicly available from: https://www.thessgac.org/data.

Due to data sharing restrictions from 23andMe, the PGSs are based on summary statistics that exclude the 23andMe cohort. Study-specific GWASs controlled for the first ten principal components of the genotypic data, age, a second-order polynomial of age, a third-order polynomial in age, an indicator for being female, interactions between age and female, and study-specific controls, including dummy variables for major events such as wars or policy changes that may have affected access to education in their specific sample.

The Add Health PGSs for educational attainment 2016 are standardized (mean = 0, standard deviation = 1) within ancestry groups to control for population stratification. To give researchers an idea of the underlying raw PGS, we present density plots for the unstandardized PGSs below. **Please read the section entitled “Using Add Health PGSs” in the introductory portion of this document prior to conducting any analyses using the provided PGSs.**

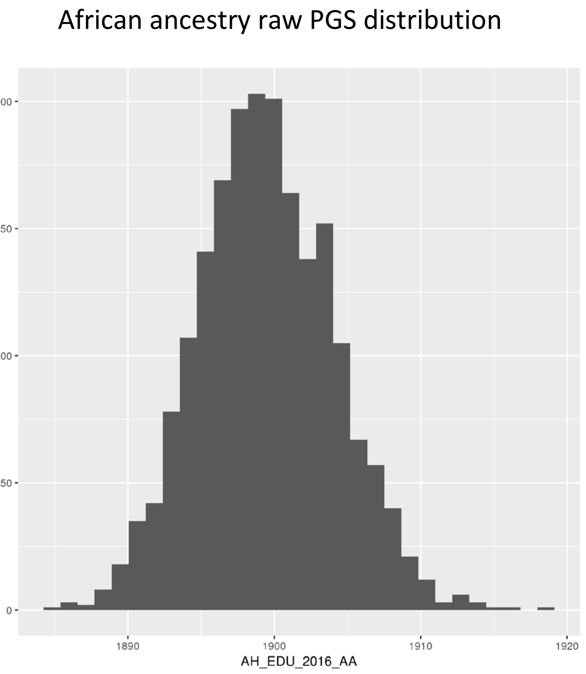
Educational Attainment 2016

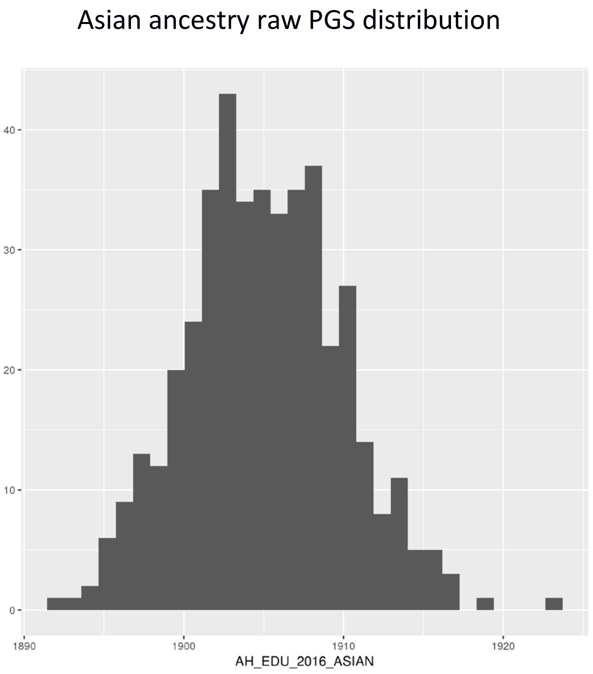


European ancestry raw PGS distribution

Hispanic ancestry raw PGS distribution

Add Health is a program project directed by Kathleen Mullan Harris and designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris at the University of North Carolina at Chapel Hill, and funded by grant P01-HD31921 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development, with cooperative funding from 23 other federal agencies and foundations. Information on how to obtain the Add Health data files is available on the Add Health website (http://www.cpc.unc.edu/addhealth).


African ancestry raw PGS distribution


Asian ancestry raw PGS distribution

Educational Attainment (2018)

GWAS Summary Statistic Source: Lee et al. 2018. "Gene discovery and polygenic prediction from a 1.1-million-person GWAS of educational attainment." Nature Genetics 50(8):1112-1121.

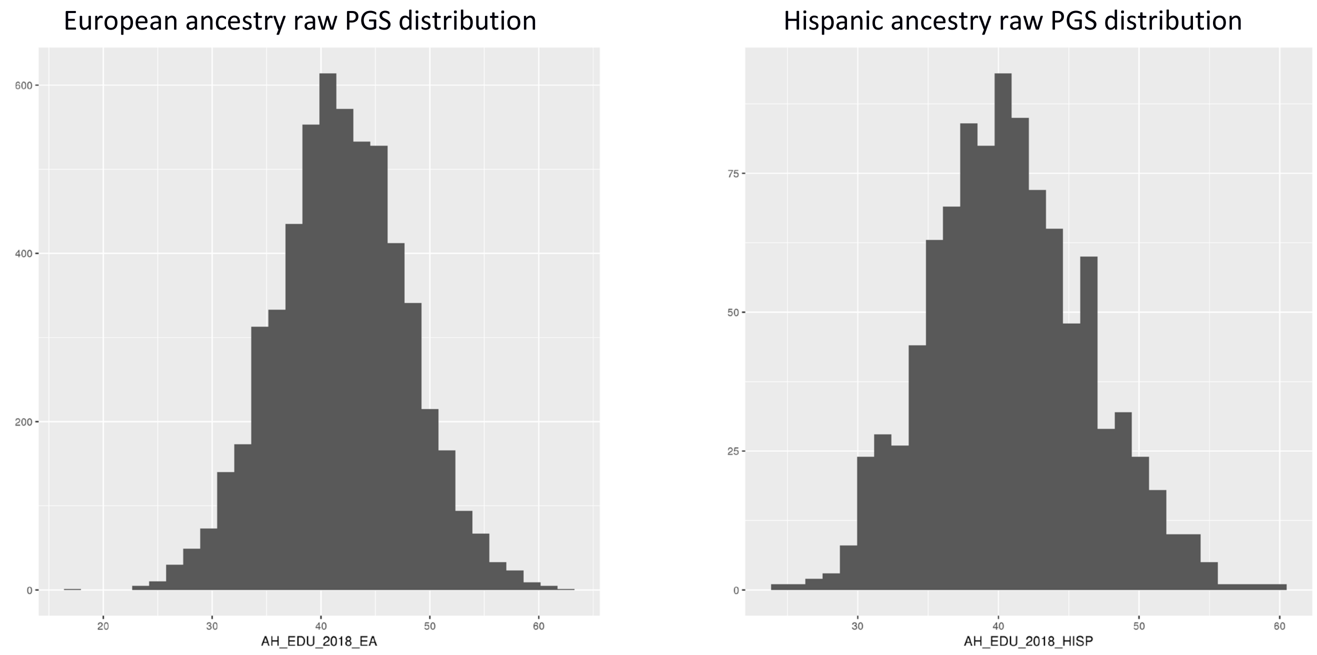
GWAS Ancestry Group(s): European

The educational attainment 2018 PGSs were created using results from a forthcoming 2018 study by the Social Science Genetic Association Consortium (SSGAC). The GWAS meta-analysis files are publicly available from: https://www.thessgac.org/data.

The discovery meta-analysis of educational attainment 2018 included 1,131,881 individuals of European decent imputed to 1000 Genomes 2010 release. However, due to data use restrictions imposed by 23andMe, the summary statistics used to create the Add Health PGSs exclude the 23andMe sample resulting in a sample size of 766,345.

The Add Health PGSs for educational attainment 2018 are standardized (mean = 0, standard deviation = 1) within ancestry groups to control for population stratification. To give researchers an idea of the underlying raw PGS, we present density plots for the unstandardized PGSs below. **Please read the section entitled "Using Add Health PGSs" in the introductory portion of this document prior to conducting any analyses using the provided PGSs.**

Educational Attainment 2018

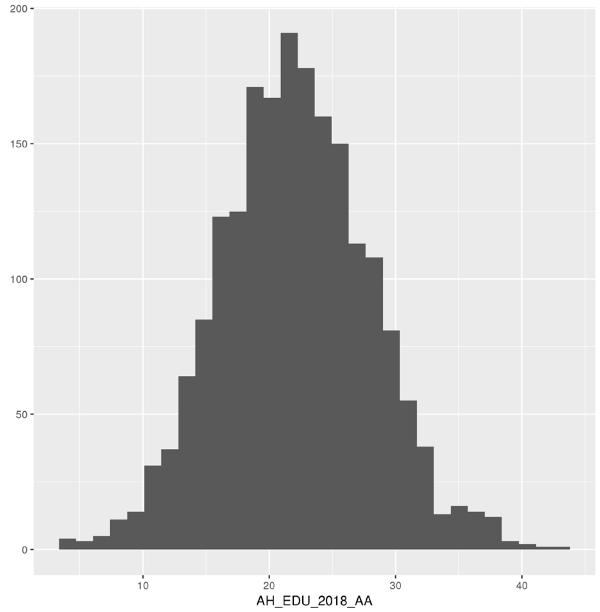


European ancestry raw PGS distribution

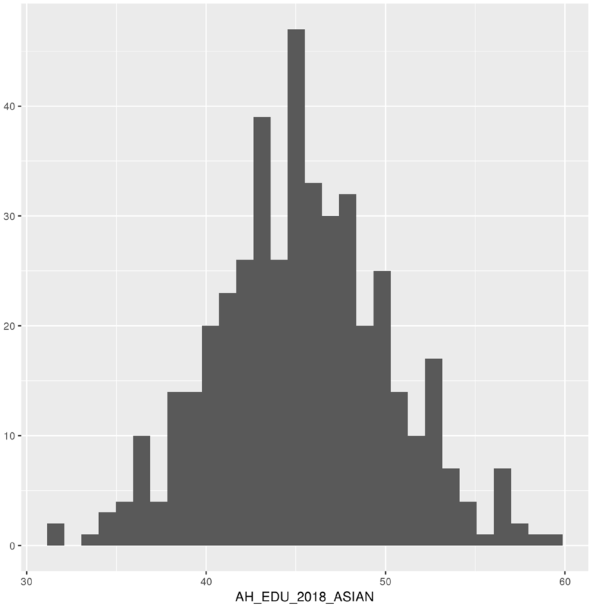
Hispanic ancestry raw PGS distribution

Add Health is a program project directed by Kathleen Mullan Harris and designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris at the University of North Carolina at Chapel Hill, and funded by grant P01-HD31921 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development, with cooperative funding from 23 other federal agencies and foundations. Information on how to obtain the Add Health data files is available on the Add Health website (http://www.cpc.unc.edu/addhealth).

African ancestry raw PGS distribution



Asian ancestry raw PGS distribution



Cognitive Function

GWAS Summary Statistic Source: Davies et al. 2018. “Study of 300,486 Individuals Identifies 148 Independent Genetic Loci Influencing General Cognitive Function.” *Nature Communications* 9(1):2098.

GWAS Ancestry Group(s): European

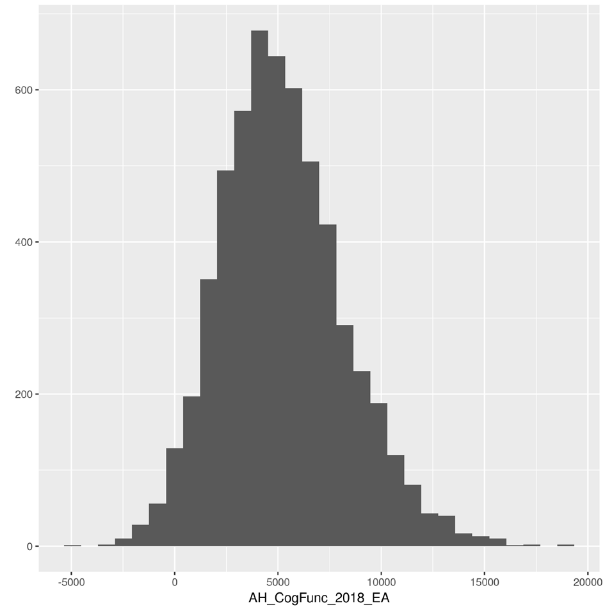
The PGSs for general cognitive function were created using results from a 2018 study conducted by the Gail Davies, Max Lam, and colleagues in collaboration with the CHARGE and COGENT consortia. The GWAS meta-analysis files are publicly available and can be downloaded from: http://www.ccace.ed.ac.uk/node/335.

The general cognitive function GWAS meta-analysis consisted of a combination of 57 population-based cohorts resulting in 300,486 individuals of European genetic ancestry (CHARGE and COGENT cohorts were combined with the UK Biobank). General cognitive function was measured as the first principal component of a series of indicators of cognitive domains which are in-turn based on individual-level cognitive tests (see the Supplemental Information from Davies et al. 2018 for more details). Individuals were excluded if they had had a stroke (including self-reported stroke) or dementia. Study-specific GWAS adjusted for age, sex, test site, familial relationships, and genetic principal components (more details can be found in the supplemental materials).

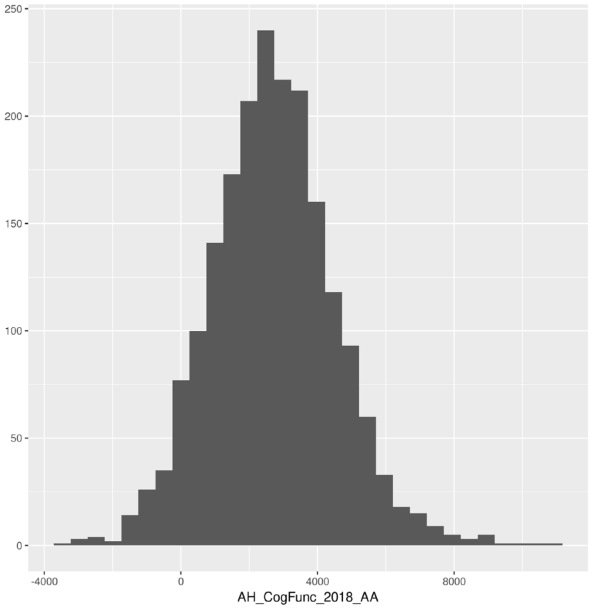
The Add Health PGSs for general cognitive function are standardized (mean = 0, standard deviation = 1) within ancestry groups to control for population stratification. To give researchers an idea of the underlying raw PGS, we present density plots for the unstandardized PGSs below. **Please read the section entitled “Using Add Health PGSs” in the introductory portion of this document prior to conducting any analyses using the provided PGSs.**

Cognitive Function

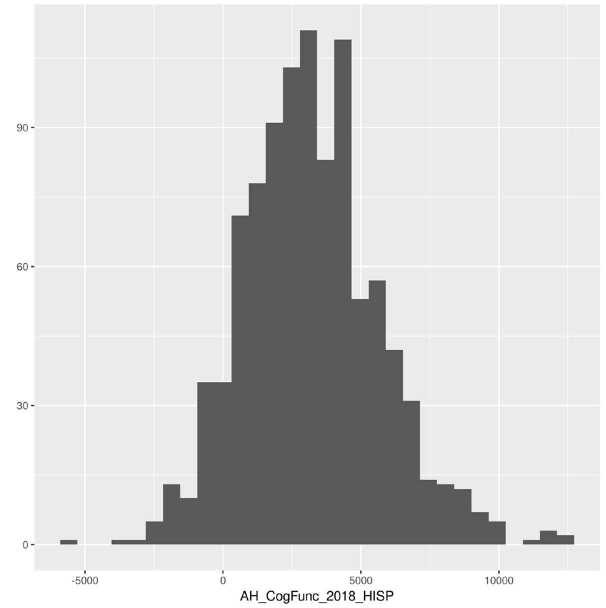
European ancestry raw PGS distribution



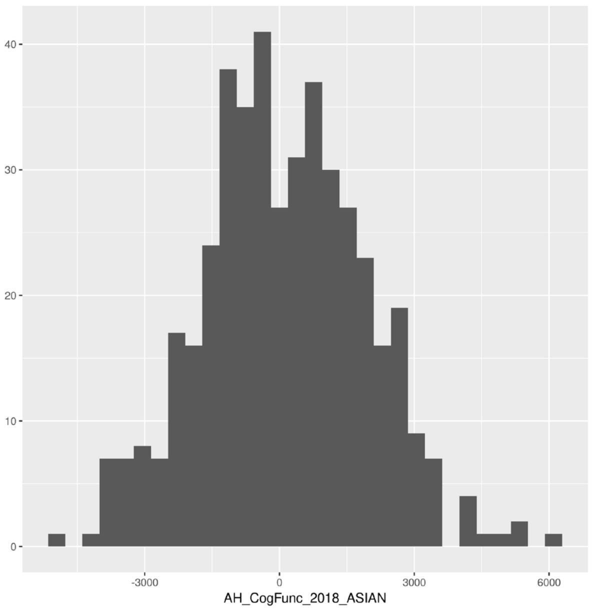
African ancestry raw PGS distribution



Hispanic ancestry raw PGS distribution



Asian ancestry raw PGS distribution



Add Health is a program project directed by Kathleen Mullan Harris and designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris at the University of North Carolina at Chapel Hill, and funded by grant P01-HD31921 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development, with cooperative funding from 23 other federal agencies and foundations. Information on how to obtain the Add Health data files is available on the Add Health website (http://www.cpc.unc.edu/addhealth).

Reaction Time

GWAS Summary Statistic Source: Davis et al. 2018. “Study of 300,486 Individuals Identifies 148 Independent Genetic Loci Influencing General Cognitive Function.” *Nature Communications* 9(1):2098.

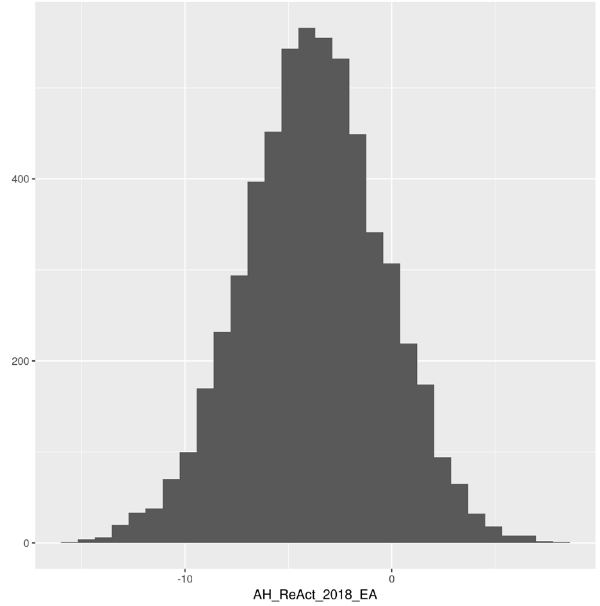
GWAS Ancestry Group(s): European

The PGSs for reaction time were created using results from a 2018 study conducted by the Gail Davies, Max Lam, and colleagues in collaboration with the CHARGE and COGENT consortia. The GWAS summary files are publicly available and can be downloaded from: http://www.ccace.ed.ac.uk/node/335.

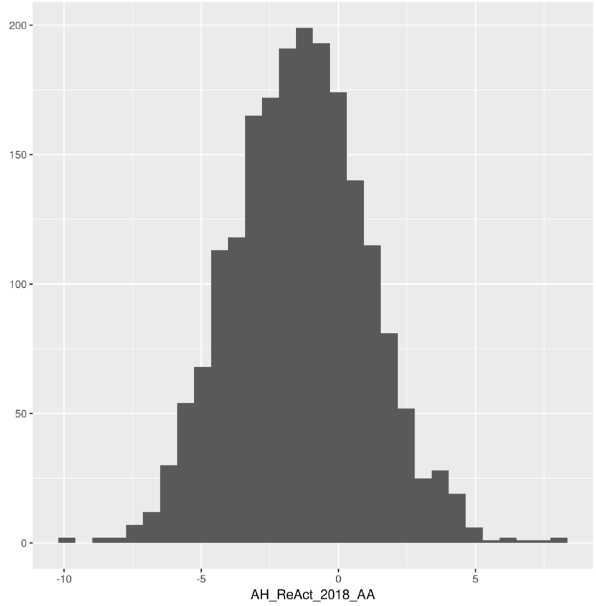
The reaction time GWAS consisted of 330,069 individuals of European genetic ancestry from the UK Biobank. Participants of the UK Biobank were presented with pairs of cards on a computer screen. The two cards could either be the same or different. If the two cards were identical, participants were to push a button box as quickly as possible. There were 12 trials in total. The first five were used as a practice. Of the remaining seven trials, four presented matching cards. The measure of reaction time used in the GWAS is the mean time, in milliseconds, to correctly identify the matching cards in these four trials (see the supplemental information in Davies et al. 2018 for more details). Individuals were excluded if they had had a stroke (including self-reported stroke) or dementia. GWAS analyses adjusted for age, sex, test site, genotyping batch, and the first 40 genetic principal components (more details can be found in the supplemental materials).

The Add Health PGSs for reaction time are standardized (mean = 0, standard deviation = 1) within ancestry groups to control for population stratification. To give researchers an idea of the underlying raw PGS, we present density plots for the unstandardized PGSs below. **Please read the section entitled “Using Add Health PGSs” in the introductory portion of this document prior to conducting any analyses using the provided PGSs.**

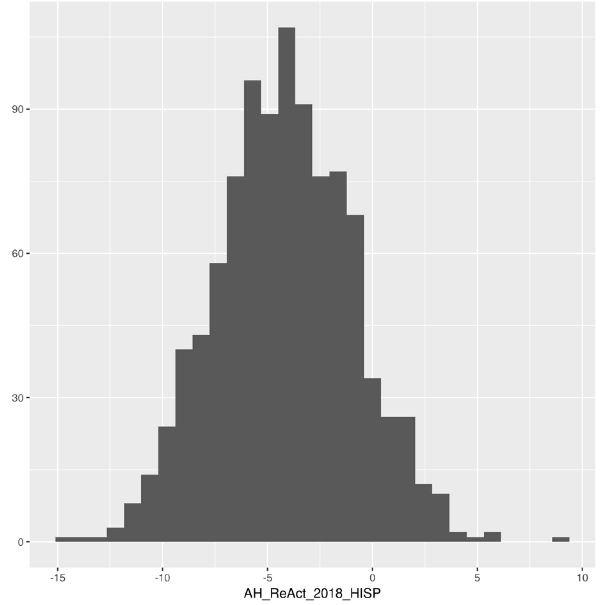
Reaction Time



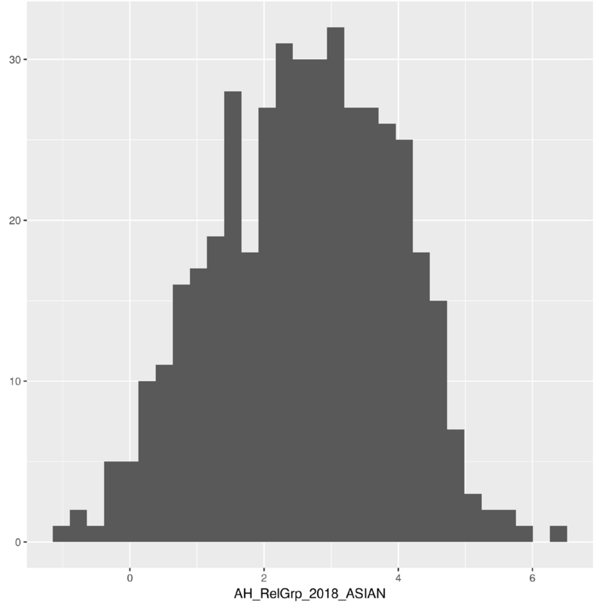
European ancestry raw PGS distribution



African ancestry raw PGS distribution



Hispanic ancestry raw PGS distribution



Asian ancestry raw PGS distribution

Intelligence

GWAS Summary Statistic Source: Savage, Stringer, et al. 2018. "Genome-Wide Association Meta-Analysis in 269,867 Individuals Identifies New Genetic and Functional Links to Intelligence." *Nature Genetics* 50(7):912-919.

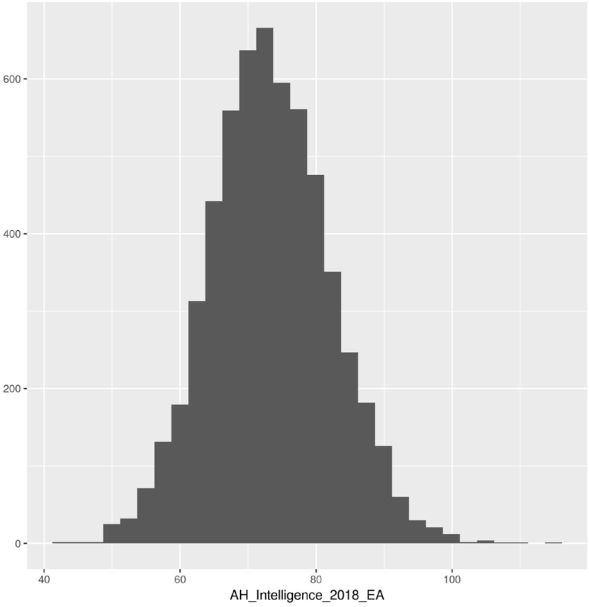
GWAS Ancestry Group(s): European

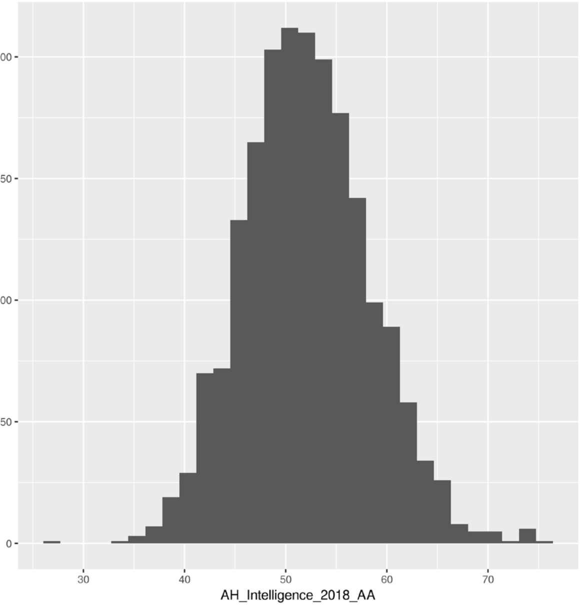
The PGSs for intelligence were created using results from a 2018 study conducted by the Jeanne Savage and colleagues. The GWAS meta-analysis files are publicly available and can be downloaded from: https://ctg.cncr.nl/software/summary_statistics.

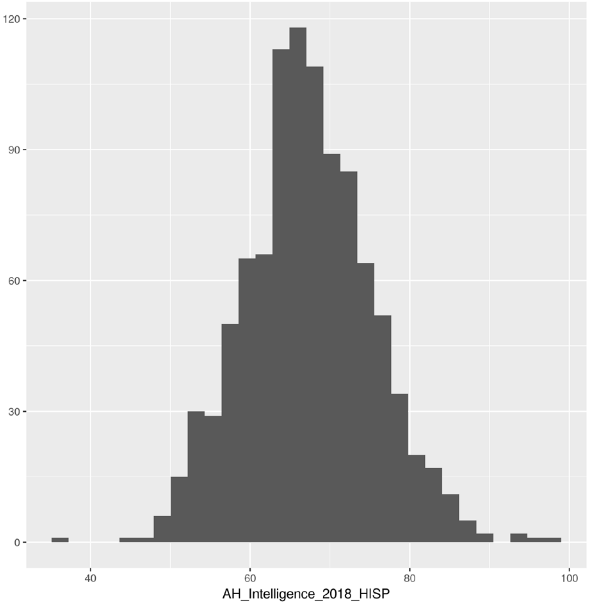
The intelligence GWAS is based on the analysis of 269,867 individuals of European genetic ancestry from 14 independent cohorts. Intelligence was measured as a latent variable based on a neuropsychiatric battery of test in 10 of the 14 cohorts, fluid intelligence in the UK Biobank, IQ scales in the Generation R Study (GENR), SAT scores in the Split for Science (S4S) cohort, and high IQ case controls in the High IQ (HIQ) and Health and Retirement Study (HRS). GWAS analyses adjusted for different sets of covariates in the various cohorts, but generally included age, sex, and different numbers of principal components of the genetic data. More details are available in the Supplementary Information provided by Savage et al. 2018.

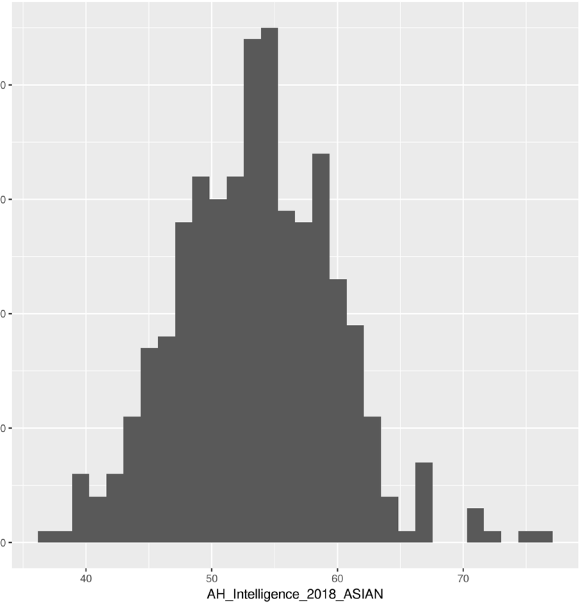
The Add Health PGSs for intelligence are standardized (mean = 0, standard deviation = 1) within ancestry groups to control for population stratification. To give researchers an idea of the underlying raw PGS, we present density plots for the unstandardized PGSs below. **Please read the section entitled "Using Add Health PGSs" in the introductory portion of this document prior to conducting any analyses using the provided PGSs.**

Intelligence


European ancestry raw PGS distribution


African ancestry raw PGS distribution


Hispanic ancestry raw PGS distribution


Asian ancestry raw PGS distribution

Autism Spectrum Disorder

GWAS Summary Statistic Source: Grove et al. 2019. “Identification of Common Genetic Risk Variants for Autism Spectrum Disorder.” *Nature Genetics* 51: 431-444.

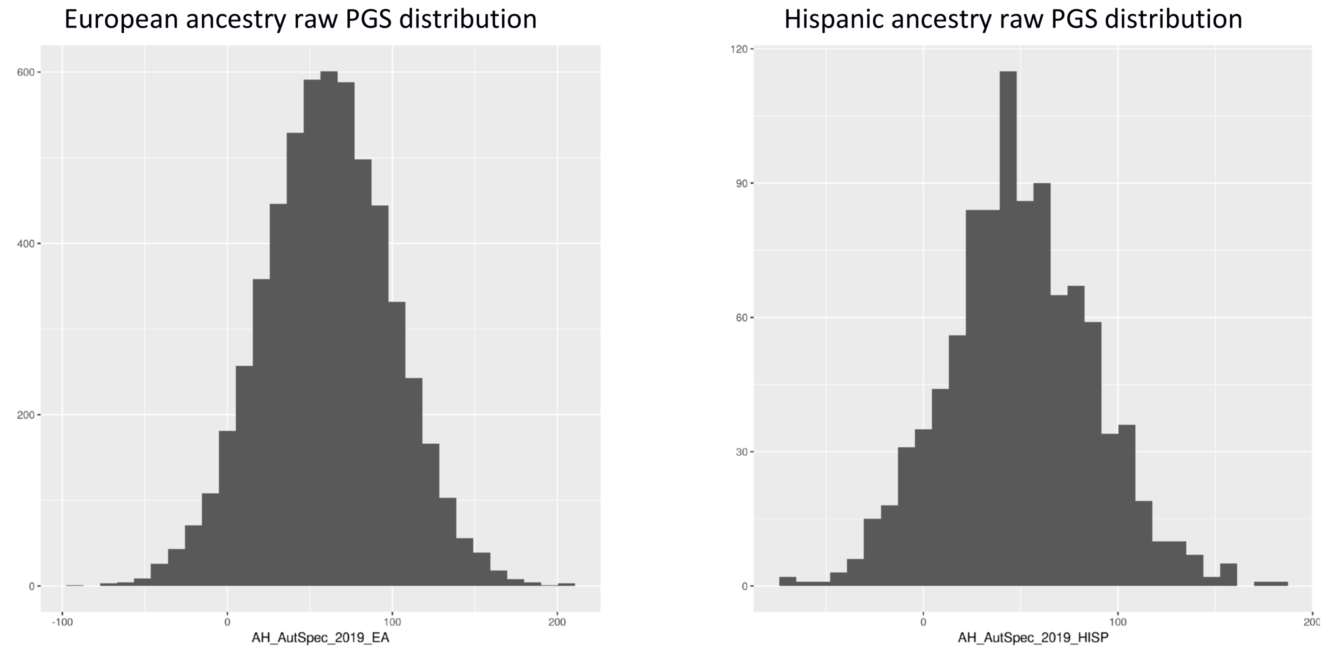
GWAS Ancestry Group(s): European

The PGSs for autism spectrum disorder were created using results from a 2019 study conducted by Jakob Grove and colleagues. The GWAS meta-analysis files are publicly available and can be downloaded from: https://www.med.unc.edu/pgc/results-and-downloads.

The autism spectrum disorder (ASD) GWAS meta-analyses are based on the analysis of 46,351 (18,382 cases, 27,969 controls) individuals of European genetic ancestry. Cases are defined as ever being diagnosed with ASD according to the ICD-10. GWAS analyses adjusted for different numbers of PCs, typically selected as the maximum number of PCs with a statistically significant association with ASD (see the methods section for in the original publication for more details).

The Add Health PGSs for ASD are standardized (mean = 0, standard deviation = 1) within ancestry groups to control for population stratification. To give researchers an idea of the underlying raw PGS, we present density plots for the unstandardized PGSs below. **Please read the section entitled “Using Add Health PGSs” in the introductory portion of this document prior to conducting any analyses using the provided PGSs.**

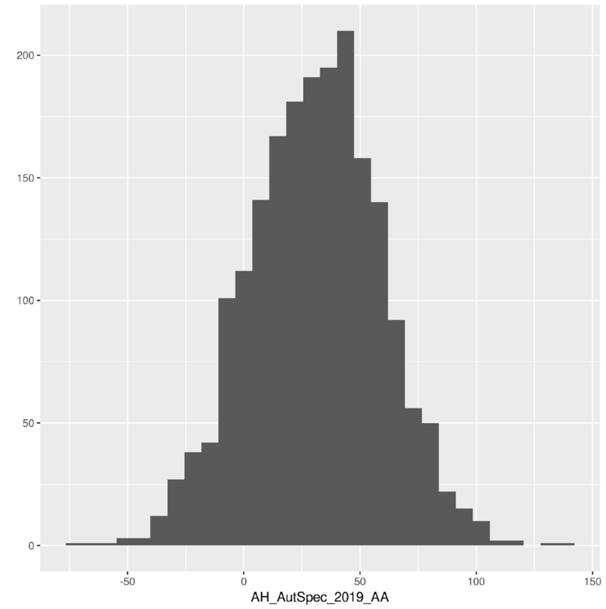
Autism Spectrum Disorder



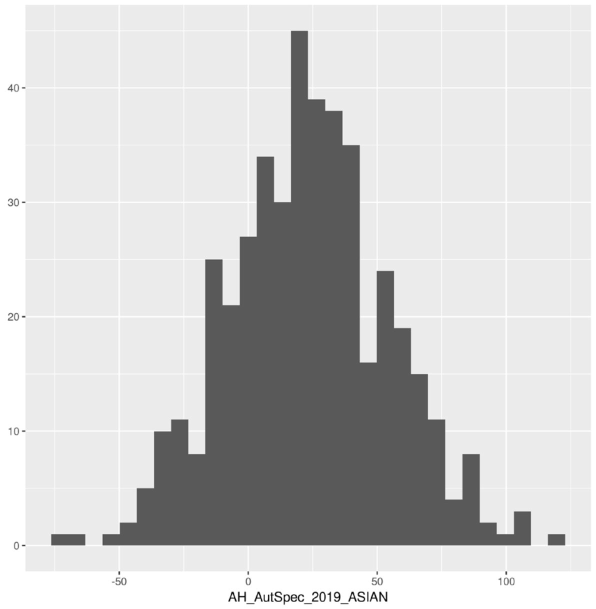
European ancestry raw PGS distribution

Hispanic ancestry raw PGS distribution

African ancestry raw PGS distribution



Asian ancestry raw PGS distribution



Ever Regular Smoker (2010)

GWAS Summary Statistic Source: Furberg, Helena et al. 2010. “Genome-Wide Meta-Analyses Identify Multiple Loci Associated with Smoking Behavior.” *Nature Genetics* 42(5):441–47.

GWAS Ancestry Group(s): European

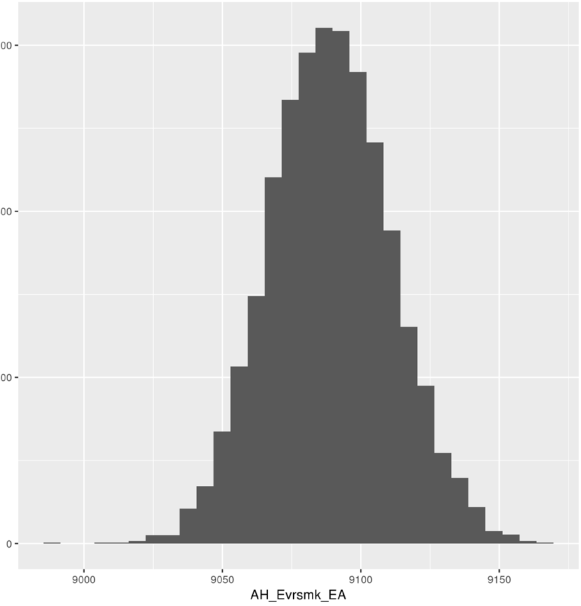
The PGSs for current/ever smoker were created using results from a 2010 GWAS conducted by the Tobacco and Genetics (TAG) Consortium. The GWAS meta-analysis files are publicly available from: https://www.med.unc.edu/pgc/results-and-downloads.

The discovery phase of analysis included 74,053 individuals of European descent. The replication of the 15 most significant genetic regions included 143,023 individuals. Individuals who were recorded as having ever been regular smokers were defined as those who reported having smoked at least 100 cigarettes during their lifetime, and never regular smokers were defined as those who reported having smoked between 0 and 99 cigarettes during their lifetime. GWAS analyses controlled for imputed allele dosage for a SNP and if the individual was classified as a case in the primary study. If the primary study was case-control in design and the phenotype being studied was known to be associated with smoking, the GWAS adjusted for case status to reduce potential confounding. Analyses were run and meta-analyzed separately for males and females.

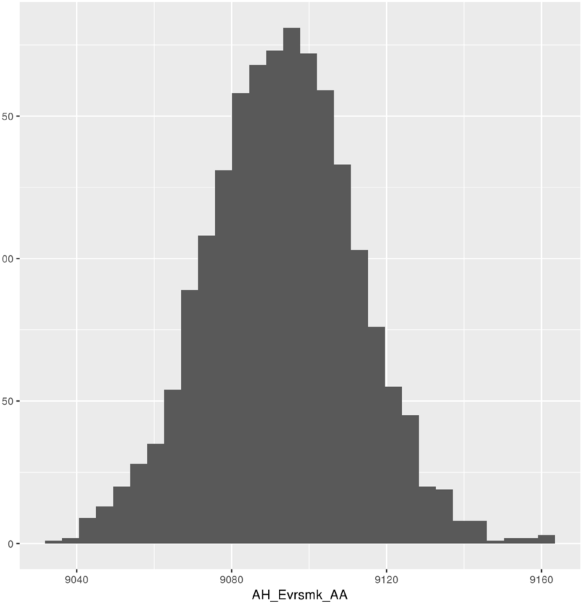
The Add Health current/ever smoker are standardized (mean = 0, standard deviation = 1) within ancestry groups to control for population stratification, but to give researchers an idea of the underlying raw PGS, we present density plots for the unstandardized PGSs below. **Please read the section entitled “Using Add Health PGSs” in the introductory portion of this document prior to conducting any analyses using the provided PGSs.**

Ever Regular Smoker (2010)

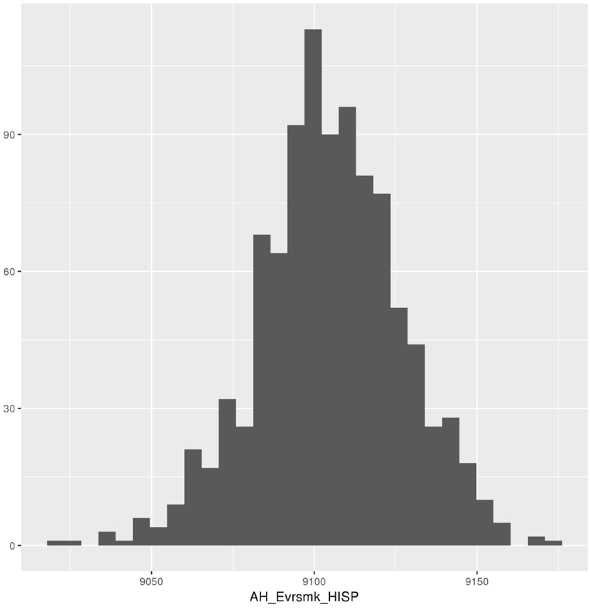
European ancestry raw PGS distribution



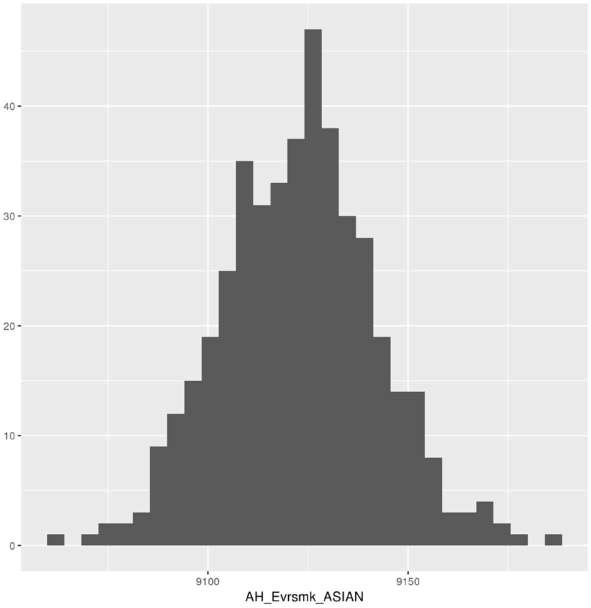
African ancestry raw PGS distribution



Hispanic ancestry raw PGS distribution



Asian ancestry raw PGS distribution



Add Health is a program project directed by Kathleen Mullan Harris and designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris at the University of North Carolina at Chapel Hill, and funded by grant P01-HD31921 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development, with cooperative funding from 23 other federal agencies and foundations. Information on how to obtain the Add Health data files is available on the Add Health website (http://www.cpc.unc.edu/addhealth).

Ever Regular Smoker (2019)

GWAS Summary Statistic Source: Liu et al. 2019. “Association Studies of up to 1.2 Million Individuals Yield New Insights into the Genetic Etiology of Tobacco and Alcohol Use.” Nature Genetics 51(2):237-244.

GWAS Ancestry Group(s): European

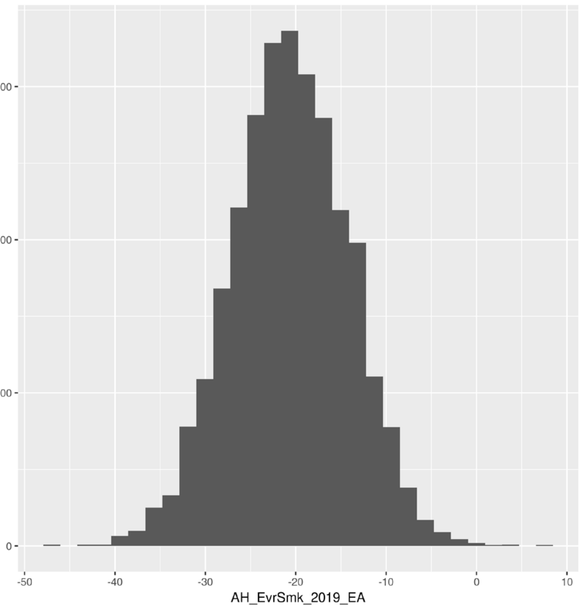
The PGSs for ever smoked were created using results from a 2019 study conducted by the Mengzhen Liu and the other members of the GWAS and Sequencing Consortium of Alcohol and Nicotine use (GSCAN). The GWAS meta-analysis files are publicly available and can be downloaded from: https://genome.psych.umn.edu/index.php/GSCAN.

The ever smoked GWAS meta-analyses are based on the analysis of 1,232,091 individuals of European genetic ancestry from the UK Biobank and 23andMe. However, due to data use restrictions imposed by 23andMe, the summary statistics used to create the Add Health PGSs exclude the 23andMe sample resulting in a sample size of 632,802. The phenotype was assessed as a dichotomous measure based on responses to the question: “Have you ever smoked?” GWAS analyses adjusted for an undisclosed number of principal components of the genetic data and other covariates. See Liu et al. 2019 Supplemental Materials for more information.

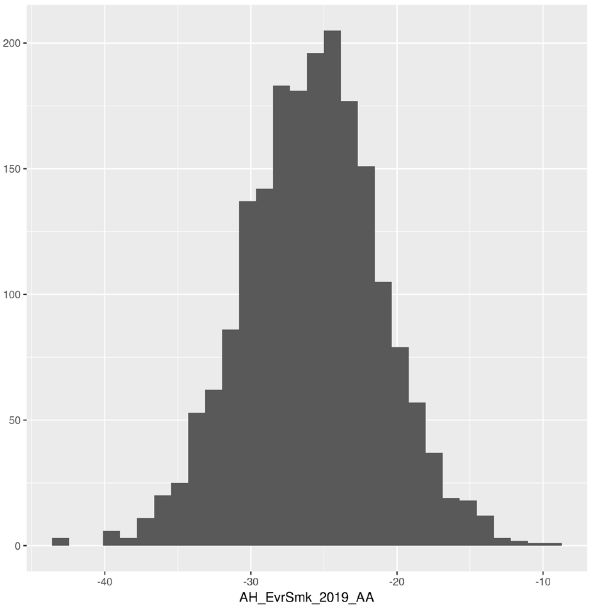
The Add Health PGSs for ever smoked are standardized (mean = 0, standard deviation = 1) within ancestry groups to control for population stratification. To give researchers an idea of the underlying raw PGS, we present density plots for the unstandardized PGSs below. **Please read the section entitled “Using Add Health PGSs” in the introductory portion of this document prior to conducting any analyses using the provided PGSs.**

Ever Regular Smoker (2019)

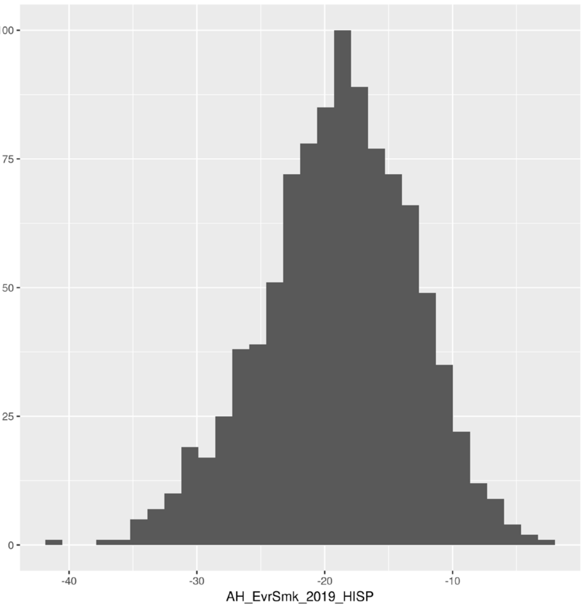
European ancestry raw PGS distribution



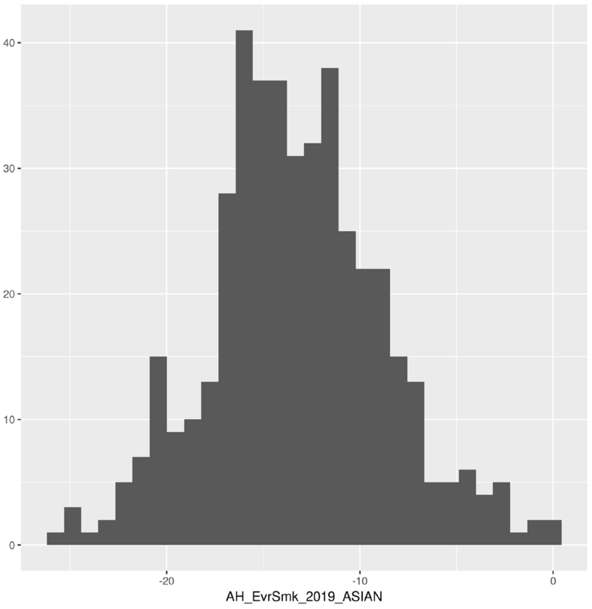
African ancestry raw PGS distribution



Hispanic ancestry raw PGS distribution



Asian ancestry raw PGS distribution



Ever Regular Smoker (2019 - UK Biobank)

GWAS Summary Statistic Source: Karlsson Linnér et al. 2019. "Genome-wide association analysis of risk tolerance and risky behaviors in over 1 million individuals identify hundreds of loci and shared genetic influences." *Nature Genetics* 51: 245-257.

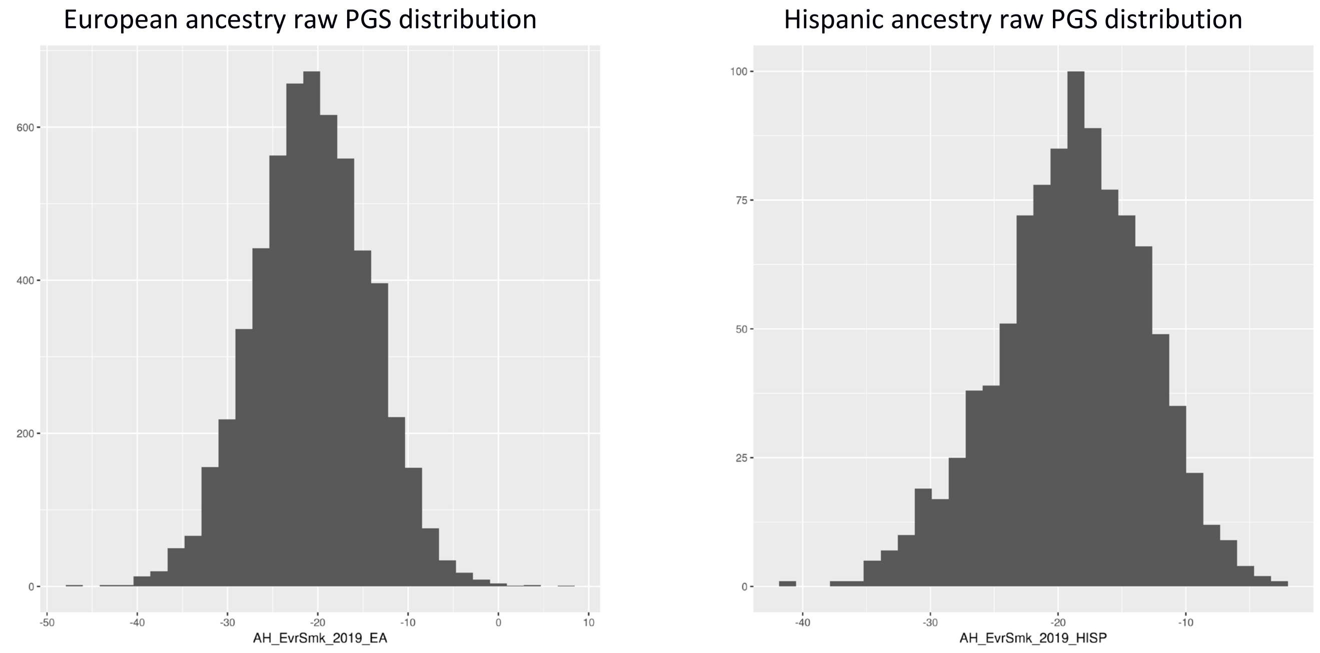
GWAS Ancestry Group(s): European

The PGSs for ever smoking regularly were created using results from a 2018 study conducted by Richard Karlsson Linnér and colleagues. The GWAS meta-analysis files are publicly available and can be downloaded from: https://www.thessgac.org/data.

The ever smoking regularly GWAS meta-analyses are based on the analysis of 518,633 individuals of European genetic ancestry from the UK Biobank. The measure of ever being a regular smoker is a dichotomous measure equal to one if a respondent reported being a current or previous smoker and zero if the respondent reported never smoking or only smoking once or twice in their lifetime. See the supplemental materials from Linnér et al. (2019) for more details.

The Add Health PGSs for ever smoking regularly are standardized (mean = 0, standard deviation = 1) within ancestry groups to control for population stratification. To give researchers an idea of the underlying raw PGS, we present density plots for the unstandardized PGSs below. **Please read the section entitled "Using Add Health PGSs" in the introductory portion of this document prior to conducting any analyses using the provided PGSs.**

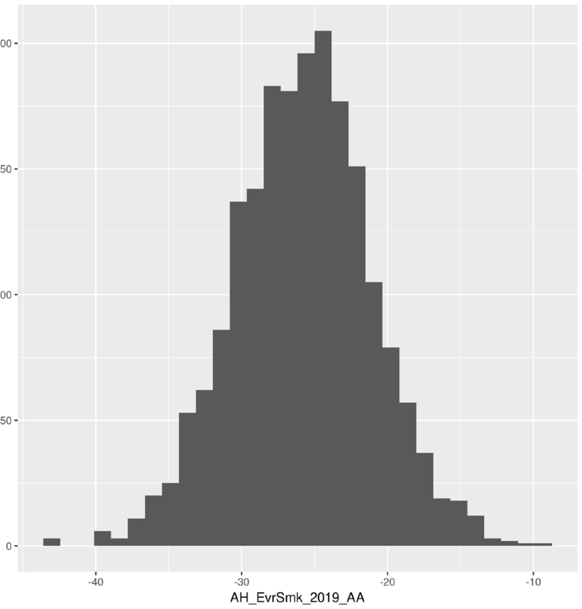
Ever a Regular Smoker (UK Biobank)


European ancestry raw PGS distribution

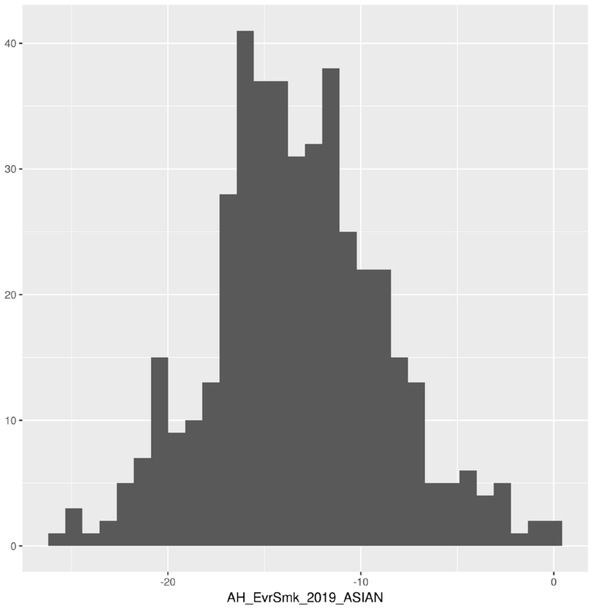
Hispanic ancestry raw PGS distribution

Add Health is a program project directed by Kathleen Mullan Harris and designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris at the University of North Carolina at Chapel Hill, and funded by grant P01-HD31921 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development, with cooperative funding from 23 other federal agencies and foundations. Information on how to obtain the Add Health data files is available on the Add Health website (http://www.cpc.unc.edu/addhealth).

African ancestry raw PGS distribution



Asian ancestry raw PGS distribution



Cigarettes per day (2010)

GWAS Summary Statistic Source: Furberg, Helena et al. 2010. “Genome-Wide Meta-Analyses Identify Multiple Loci Associated with Smoking Behavior.” *Nature Genetics* 42(5):441–47.

GWAS Ancestry Group(s): European

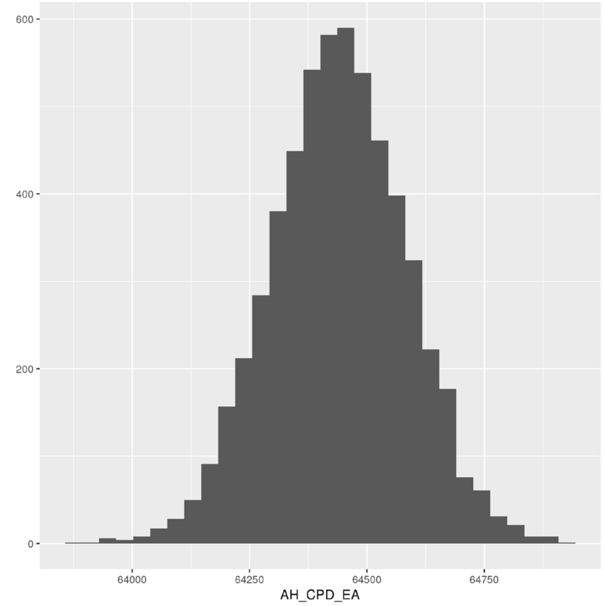
The PGSs for the number of cigarettes smoked per day were created using results from a 2010 GWAS conducted by the Tobacco and Genetics (TAG) Consortium. The GWAS meta-analysis files are publicly available from: https://www.med.unc.edu/pgc/results-and-downloads.

The discovery phase of analysis included 74,053 individuals of European descent. The replication of the 15 most significant genetic regions included 143,023 individuals. The number of cigarettes smoked per day was reported as either the average number of cigarettes smoked per day or the maximum number of cigarettes smoked in a single day (see the online methods). GWAS analyses controlled for imputed allele dosage for a SNP and if the individual was classified as a case in the primary study. If the primary study was case-control in design and the phenotype being studied was known to be associated with smoking, the GWAS adjusted for case status to reduce potential confounding. Analyses were run and meta-analyzed separately for males and females.

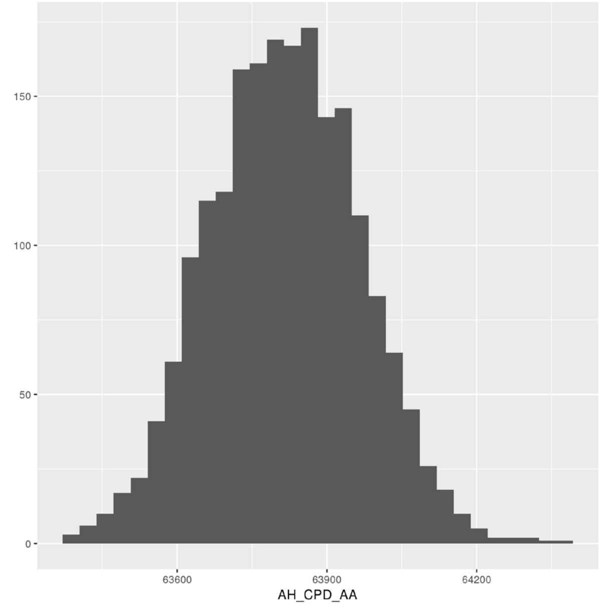
The Add Health cigarettes smoked per day are standardized (mean = 0, standard deviation = 1) within ancestry groups to control for population stratification, but to give researchers an idea of the underlying raw PGS, we present density plots for the unstandardized PGSs below. **Please read the section entitled “Using Add Health PGSs” in the introductory portion of this document prior to conducting any analyses using the provided PGSs.**

Number of Cigarettes per Day

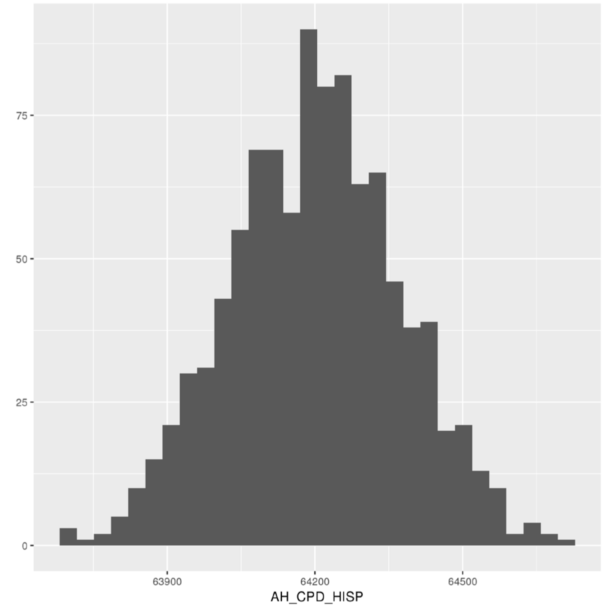
European ancestry raw PGS distribution



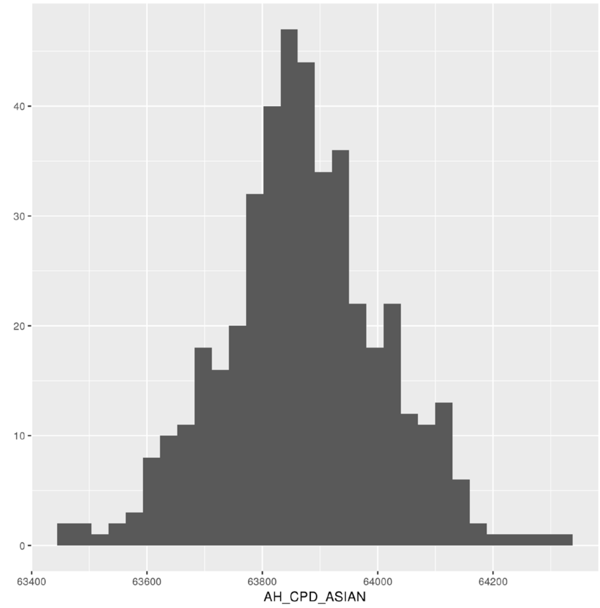
African ancestry raw PGS distribution



Hispanic ancestry raw PGS distribution



Asian ancestry raw PGS distribution



Cigarettes per day (2019)

GWAS Summary Statistic Source: Liu et al. 2019. “Association Studies of up to 1.2 Million Individuals Yield New Insights into the Genetic Etiology of Tobacco and Alcohol Use.” Nature Genetics 51(2):237-244.

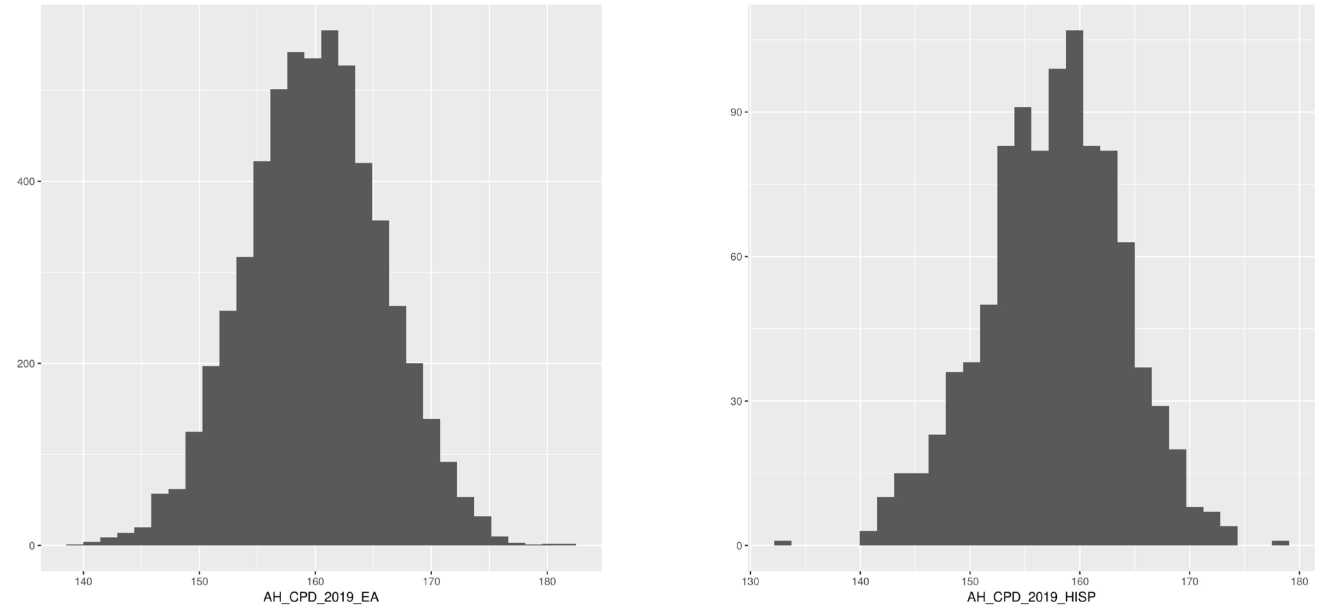
GWAS Ancestry Group(s): European

The PGSs for Cigarettes per day (CPD) were created using results from a 2019 study conducted by the Mengzhen Liu and the other members of the GWAS and Sequencing Consortium of Alcohol and Nicotine use (GSCAN). The GWAS meta-analysis files are publicly available and can be downloaded from: https://genome.psych.umn.edu/index.php/GSCAN.

The CPD GWAS meta-analyses are based on the analysis of 337,334individuals of European genetic ancestry from the UK Biobank and 23andMe. However, due to data use restrictions imposed by 23andMe, the summary statistics used to create the Add Health PGSs exclude the 23andMe sample resulting in a sample size of 263,954. GWAS analyses adjusted for an undisclosed number of principal components of the genetic data and other covariates (see Liu et al. 2019 Supplemental Materials for more information).

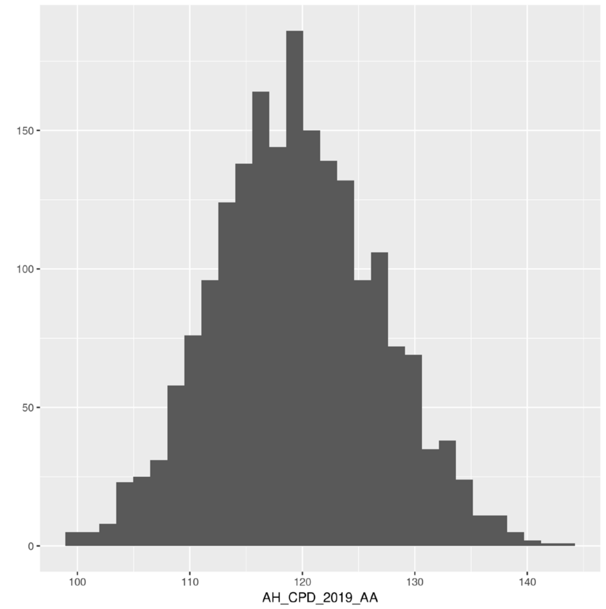
The Add Health PGSs for CPD are standardized (mean = 0, standard deviation = 1) within ancestry groups to control for population stratification. To give researchers an idea of the underlying raw PGS, we present density plots for the unstandardized PGSs below. **Please read the section entitled “Using Add Health PGSs” in the introductory portion of this document prior to conducting any analyses using the provided PGSs.**

Cigarettes per Day (2019)

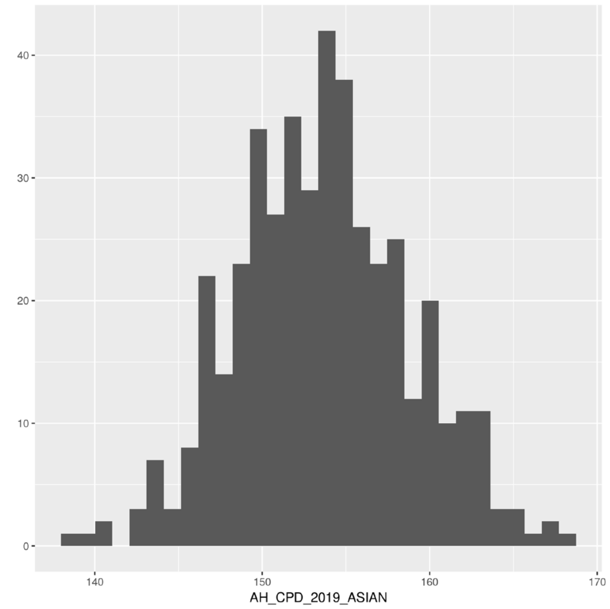


European ancestry raw PGS distribution

Hispanic ancestry raw PGS distribution



African ancestry raw PGS distribution



Asian ancestry raw PGS distribution

Age of Smoking Initiation

GWAS Summary Statistic Source: Liu et al. 2019. “Association Studies of up to 1.2 Million Individuals Yield New Insights into the Genetic Etiology of Tobacco and Alcohol Use.” Nature Genetics 51(2):237-244.

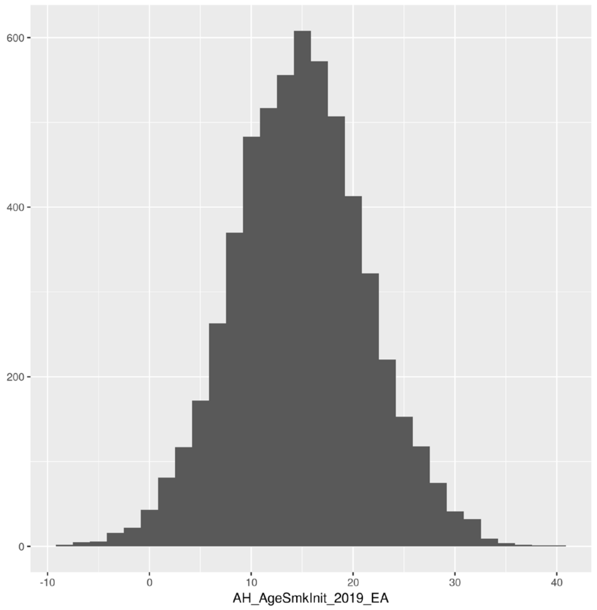
GWAS Ancestry Group(s): European

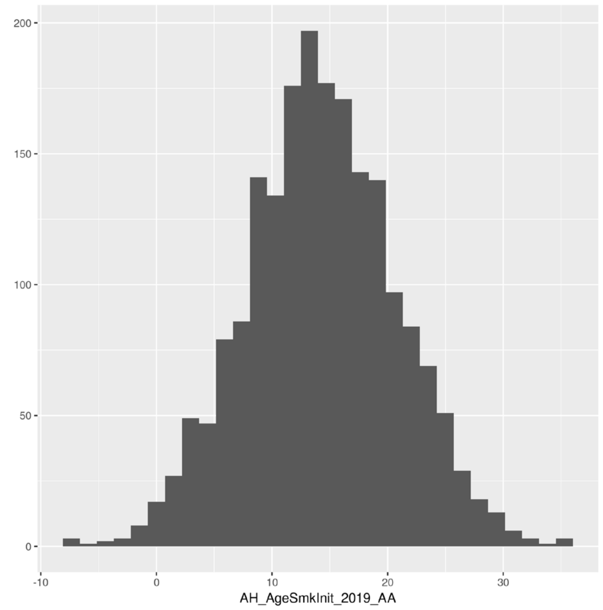
The PGSs for age of smoking initiation were created using results from a 2019 study conducted by the Mengzhen Liu and the other members of the GWAS and Sequencing Consortium of Alcohol and Nicotine use (GSCAN). The GWAS meta-analysis files are publicly available and can be downloaded from: https://genome.psych.umn.edu/index.php/GSCAN.

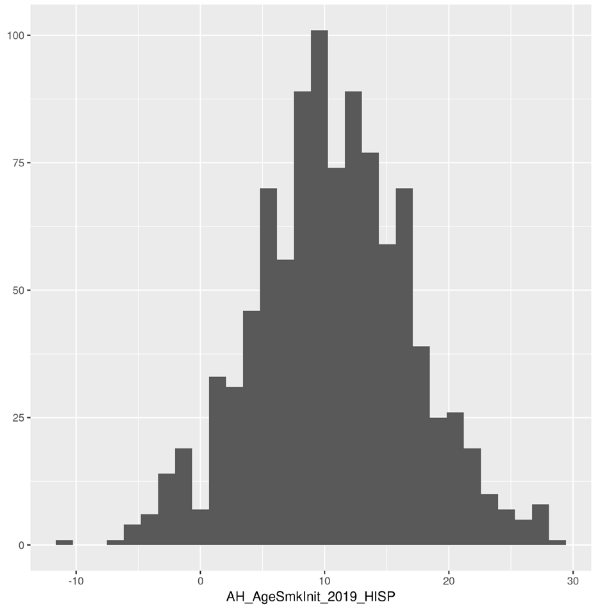
The age of smoking initiation GWAS meta-analyses are based on the analysis of 341,427 individuals of European genetic ancestry from the UK Biobank and 23andMe. However, due to data use restrictions imposed by 23andMe, the summary statistics used to create the Add Health PGSs exclude the 23andMe sample resulting in a sample size of 262,990. GWAS analyses adjusted for an undisclosed number of principal components of the genetic data and other covariates (see Liu et al. 2019 Supplemental Materials for more information).

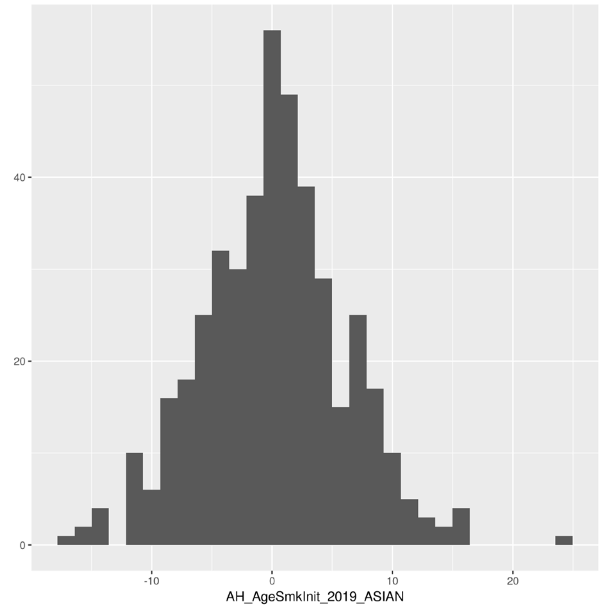
The Add Health PGSs for age of smoking initiation are standardized (mean = 0, standard deviation = 1) within ancestry groups to control for population stratification. To give researchers an idea of the underlying raw PGS, we present density plots for the unstandardized PGSs below. **Please read the section entitled “Using Add Health PGSs” in the introductory portion of this document prior to conducting any analyses using the provided PGSs.**

Age of Smoking Initiation 2019


European ancestry raw PGS distribution


African ancestry raw PGS distribution


Hispanic ancestry raw PGS distribution


Asian ancestry raw PGS distribution

Add Health is a program project directed by Kathleen Mullan Harris and designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris at the University of North Carolina at Chapel Hill, and funded by grant P01-HD31921 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development, with cooperative funding from 23 other federal agencies and foundations. Information on how to obtain the Add Health data files is available on the Add Health website (http://www.cpc.unc.edu/addhealth).

Smoking Cessation

GWAS Summary Statistic Source: Liu et al. 2019. “Association Studies of up to 1.2 Million Individuals Yield New Insights into the Genetic Etiology of Tobacco and Alcohol Use.” Nature Genetics 51(2):237-244.

GWAS Ancestry Group(s): European

The PGSs for smoking cessation were created using results from a 2019 study conducted by the Mengzhen Liu and the other members of the GWAS and Sequencing Consortium of Alcohol and Nicotine use (GSCAN). The GWAS meta-analysis files are publicly available and can be downloaded from: https://genome.psych.umn.edu/index.php/GSCAN.

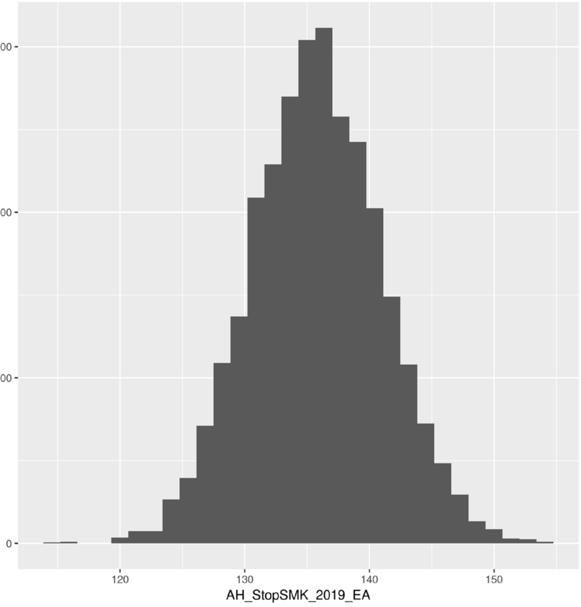
The smoking cessation GWAS meta-analyses are based on the analysis of 547,219 individuals of European genetic ancestry from the UK Biobank and 23andMe. However, due to data use restrictions imposed by 23andMe, the summary statistics used to create the Add Health PGSs exclude the 23andMe sample resulting in a sample size of 312,821. Smoking cessation was assessed as a dichotomous measure based on responses to the question: “Have you ever stopped smoking?” GWAS analyses adjusted for an undisclosed number of principal components of the genetic data and other covariates (see Liu et al. 2019 Supplemental Materials for more information).

The Add Health PGSs for smoking cessation are standardized (mean = 0, standard deviation = 1) within ancestry groups to control for population stratification. To give researchers an idea of the underlying raw PGS, we present density plots for the unstandardized PGSs below. **Please read the section entitled “Using Add Health PGSs” in the introductory portion of this document prior to conducting any analyses using the provided PGSs.**

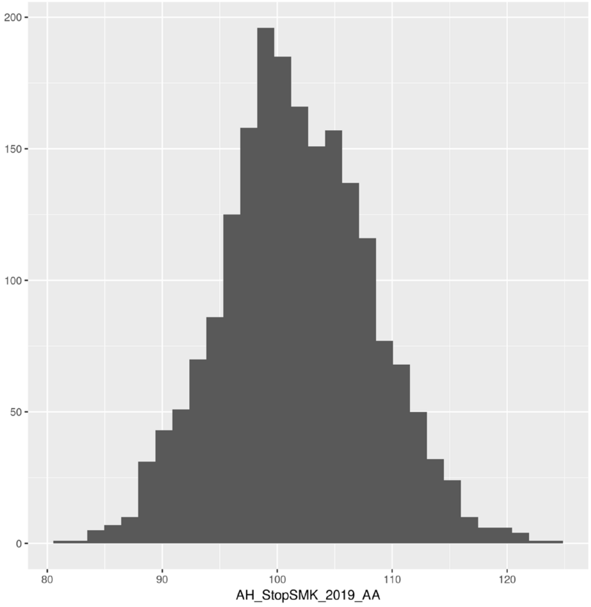
Add Health is a program project directed by Kathleen Mullan Harris and designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris at the University of North Carolina at Chapel Hill, and funded by grant P01-HD31921 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development, with cooperative funding from 23 other federal agencies and foundations. Information on how to obtain the Add Health data files is available on the Add Health website (http://www.cpc.unc.edu/addhealth).

Smoking Cessation

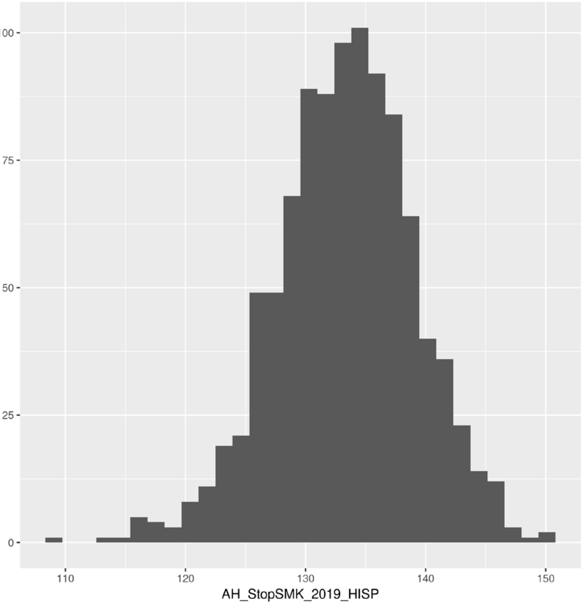
European ancestry raw PGS distribution



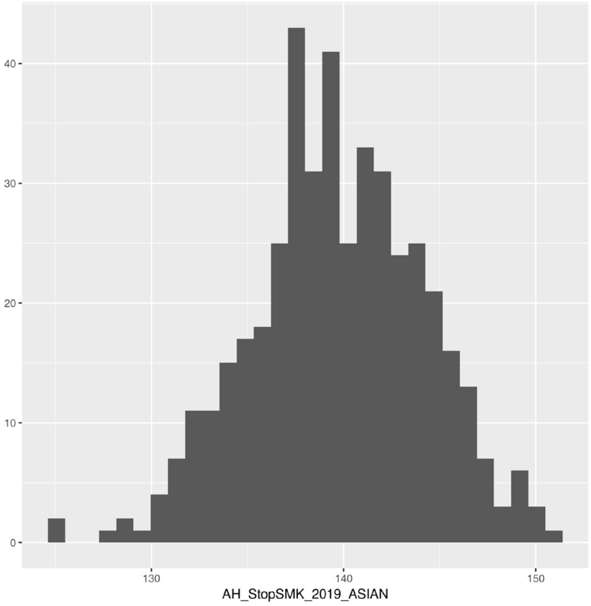
African ancestry raw PGS distribution



Hispanic ancestry raw PGS distribution



Asian ancestry raw PGS distribution



Drinks per Week (2019 – GSCAN)

GWAS Summary Statistic Source: Liu et al. 2019. “Association Studies of up to 1.2 Million Individuals Yield New Insights into the Genetic Etiology of Tobacco and Alcohol Use.” Nature Genetics 51(2):237-244.

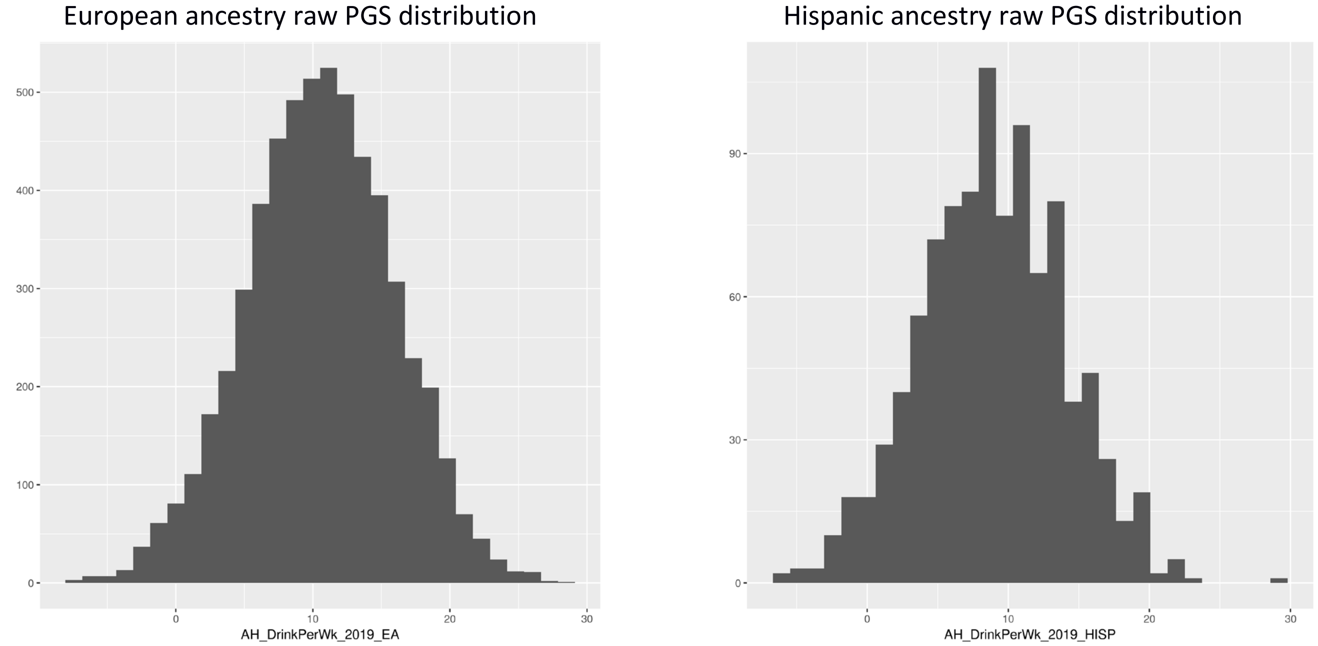
GWAS Ancestry Group(s): European

The PGSs for drinks per week were created using results from a 2019 study conducted by the Mengzhen Liu and the other members of the GWAS and Sequencing Consortium of Alcohol and Nicotine use (GSCAN). The GWAS meta-analysis files are publicly available and can be downloaded from: https://genome.psych.umn.edu/index.php/GSCAN.

The drinks per week GWAS meta-analyses are based on the analysis of 941,280 individuals of European genetic ancestry from the UK Biobank and 23andMe. However, due to data use restrictions imposed by 23andMe, the summary statistics used to create the Add Health PGSs exclude the 23andMe sample resulting in a sample size of 537,349. GWAS analyses adjusted for an undisclosed number of principal components of the genetic data and other covariates (see Liu et al. 2019 Supplemental Materials for more information).

The Add Health PGSs for drinks per week are standardized (mean = 0, standard deviation = 1) within ancestry groups to control for population stratification. To give researchers an idea of the underlying raw PGS, we present density plots for the unstandardized PGSs below. **Please read the section entitled “Using Add Health PGSs” in the introductory portion of this document prior to conducting any analyses using the provided PGSs.**

Drinks per Week (2019 – GSCAN)

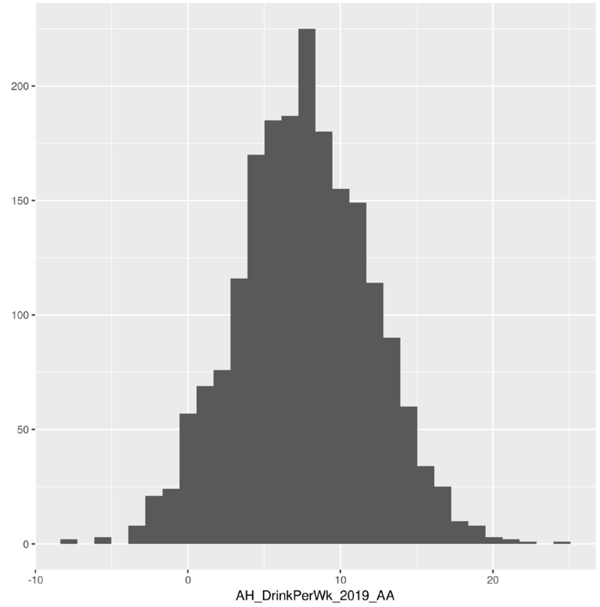


European ancestry raw PGS distribution

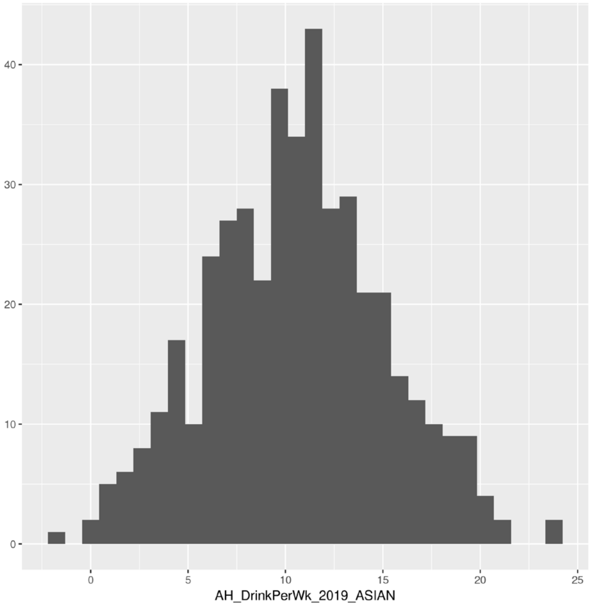
Hispanic ancestry raw PGS distribution

Add Health is a program project directed by Kathleen Mullan Harris and designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris at the University of North Carolina at Chapel Hill, and funded by grant P01-HD31921 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development, with cooperative funding from 23 other federal agencies and foundations. Information on how to obtain the Add Health data files is available on the Add Health website (http://www.cpc.unc.edu/addhealth).

African ancestry raw PGS distribution



Asian ancestry raw PGS distribution



Drinks per Week (2019 – UK Biobank)

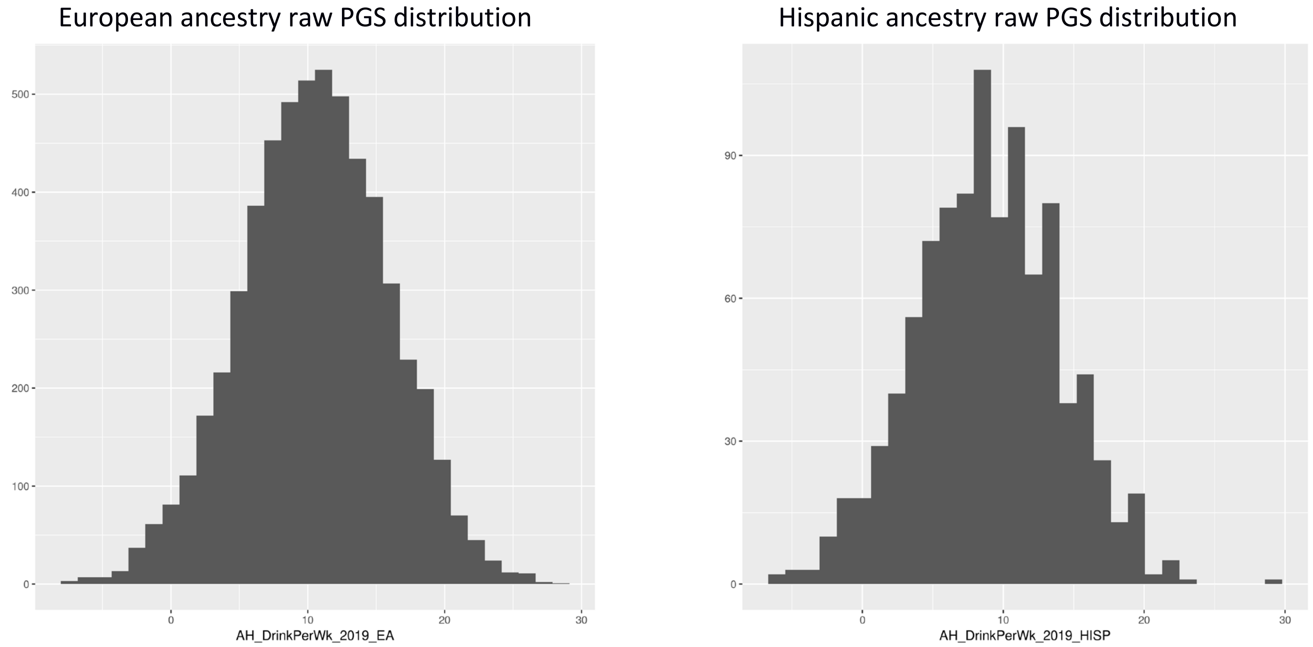
GWAS Summary Statistic Source: Karlsson Linnér et al. 2019. "Genome-wide association analysis of risk tolerance and risky behaviors in over 1 million individuals identify hundreds of loci and shared genetic influences." *Nature Genetics* 51: 245-257.

GWAS Ancestry Group(s): European

The PGSs for number of drinks per week were created using results from a 2018 study conducted by Richard Karlsson Linnér and colleagues. The GWAS meta-analysis files are publicly available and can be downloaded from: https://www.thessgac.org/data.

The number of drinks per week GWAS meta-analyses are based on the analysis of 414,343 individuals of European genetic ancestry from the UK Biobank. The measure of the number of drinks per week is a composite measure based on responses to a series of questions about drinking habits of alcoholic beverages. See the supplemental materials from Linnér et al. (2019) for more details.

The Add Health PGSs for number of drinks per week are standardized (mean = 0, standard deviation = 1) within ancestry groups to control for population stratification. To give researchers an idea of the underlying raw PGS, we present density plots for the unstandardized PGSs below. **Please read the section entitled "Using Add Health PGSs" in the introductory portion of this document prior to conducting any analyses using the provided PGSs.**

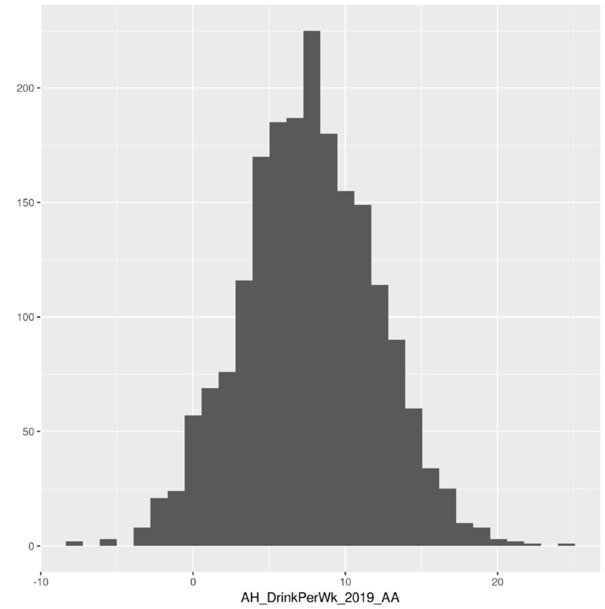


Drinks per Week (2019 – UK Biobank)

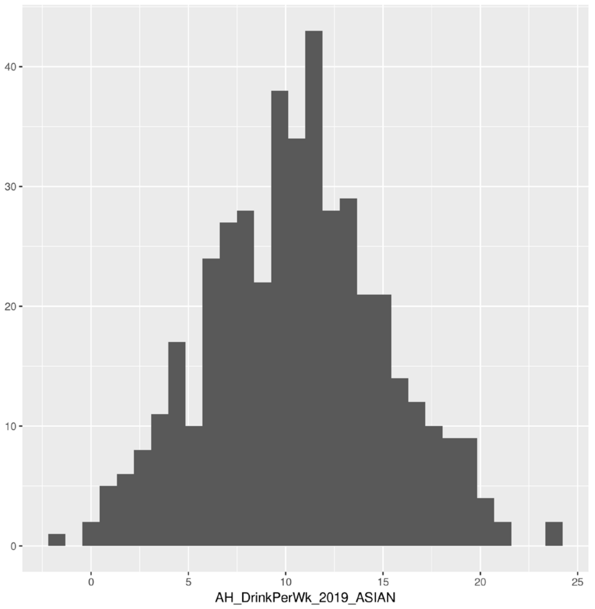
European ancestry raw PGS distribution

Hispanic ancestry raw PGS distribution

African ancestry raw PGS distribution



Asian ancestry raw PGS distribution



Alcohol Dependence

GWAS Summary Statistic Source: Walters et al. 2018. "Transancestral GWAS of Alcohol Dependence Reveals Common Genetic Underpinnings with Psychiatric Disorders." *Nature Neuroscience* 21(12):1656.

GWAS Ancestry Group(s): European and African

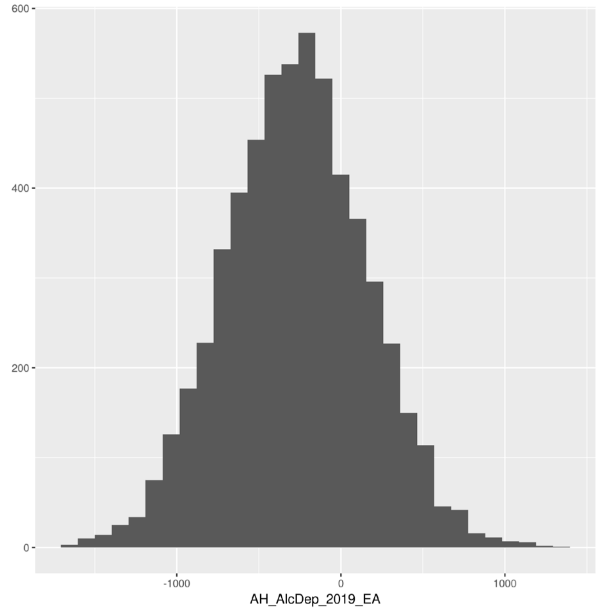
There are five PGSs for alcohol dependence created using results from a 2018 study conducted by Raymond K. Walters and colleagues. The first PGS is based on summary statistics from the combined sample GWAS, used to create a PGS for all four genetic ancestry groups. The other four PGSs use summary statistics from the GWASs limited to African and European ancestry respectively and are only made available for Add Health respondents of the corresponding genetic ancestry (see the introductory material for the methodology used to determine genetic ancestry of Add Health respondents). The first PGS within both African and European genetic ancestry groups uses summary statistics from the full ancestry-specific sample GWAS while the second PGS uses summary statistics from a GWAS limited to unrelated individuals within each genetic ancestry group. The GWAS meta-analysis files are publicly available and can be downloaded from: https://www.med.unc.edu/pgc/results-and-downloads/alcohol-dependence/?choice=AlcoholAlcohol+Dependence+%28ALCDEP%29.

The alcohol dependence GWAS meta-analyses are based on the analysis of 52,848 individuals of European (EA) and African (AA) genetic ancestry (N_{EA} = 46,568 [N_{EA} unrelated = 38,686], N_{AA} = 6,280 [N_{AA} unrelated = 5,799]. The measure of alcohol dependence is based on clinical diagnoses of alcohol dependence and/or semi-structured interviews following the DSM-IV criteria. GWAS analyses adjusted for sex and 5-10 principal components of the genome-wide data depending on the cohort sample size.

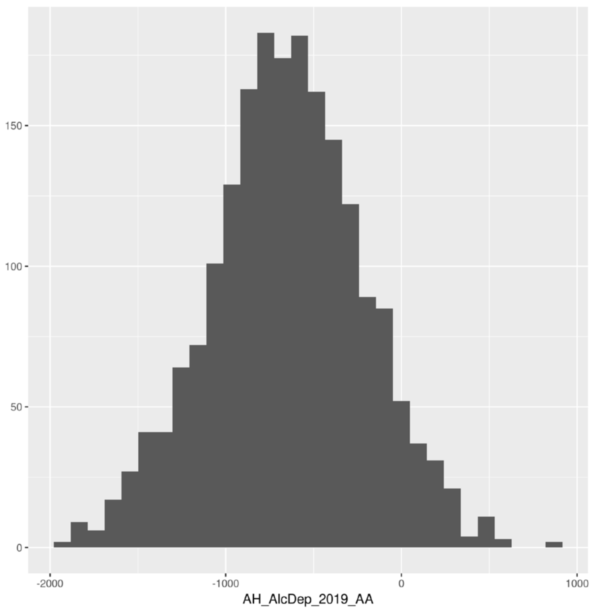
The Add Health PGSs for alcohol dependence are standardized (mean = 0, standard deviation = 1) within ancestry groups to control for population stratification. To give researchers an idea of the underlying raw PGS, we present density plots for the unstandardized PGSs below. **Please read the section entitled "Using Add Health PGSs" in the introductory portion of this document prior to conducting any analyses using the provided PGSs.**

Alcohol Dependence (Combined Sample)

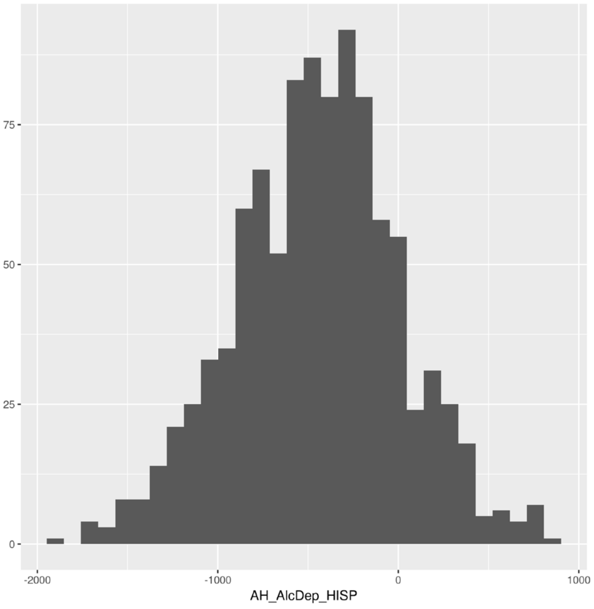
European ancestry raw PGS distribution



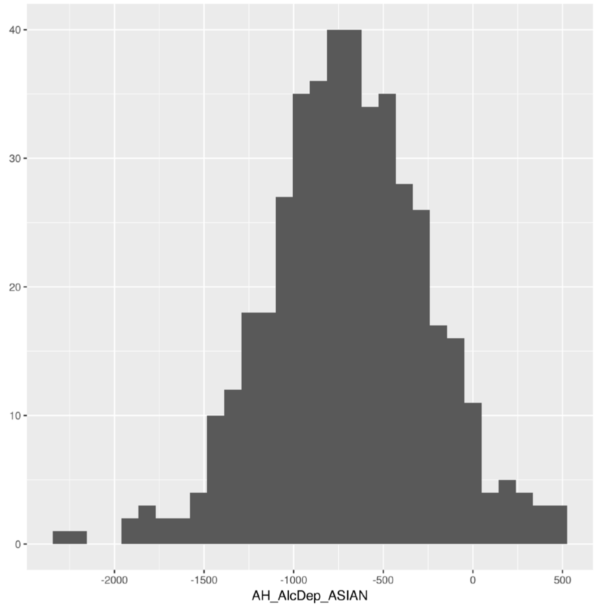
African ancestry raw PGS distribution



Hispanic ancestry raw PGS distribution

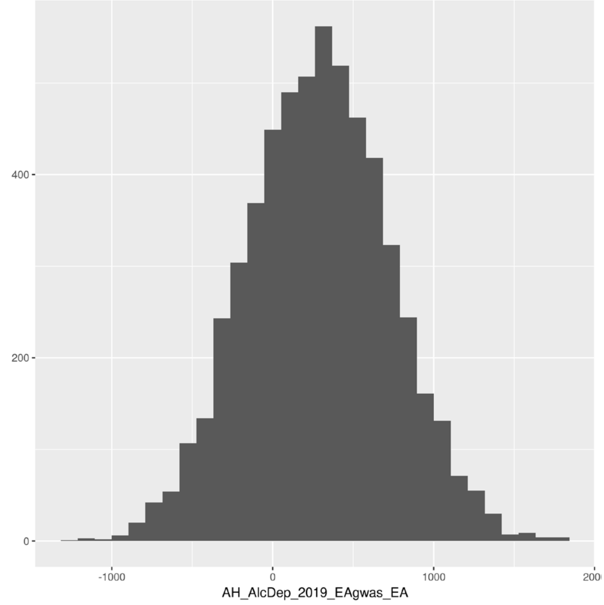


Asian ancestry raw PGS distribution

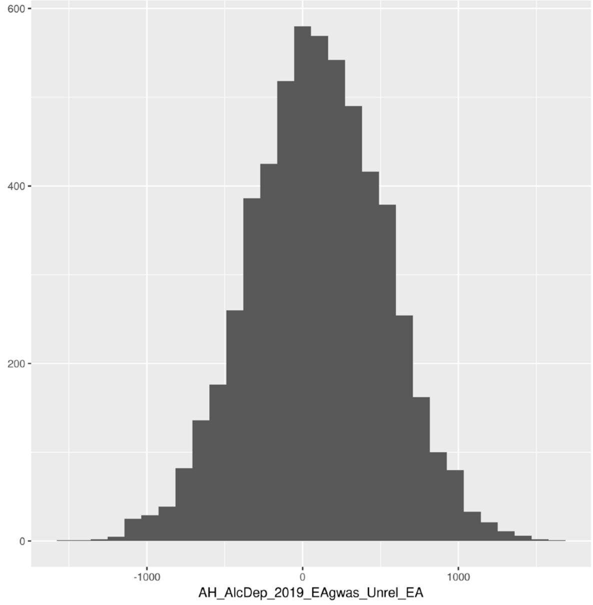


Alcohol Dependence (European Ancestry GWASs)

European ancestry Full Sample raw PGS distribution

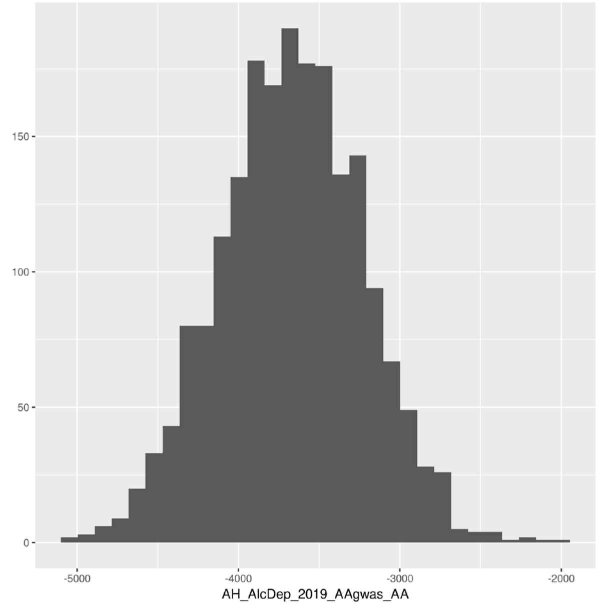


European ancestry Unrelated Sample raw PGS distribution

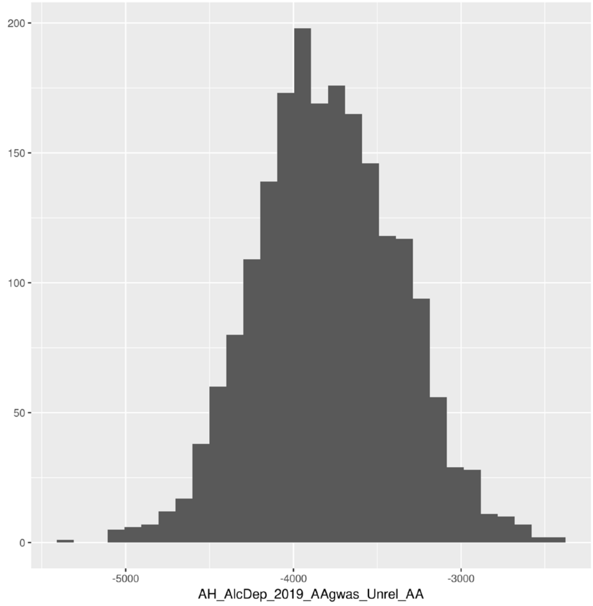


Alcohol Dependence (African Ancestry GWASs)

African ancestry Full Sample raw PGS distribution



African ancestry Unrelated Sample raw PGS distribution



Add Health is a program project directed by Kathleen Mullan Harris and designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris at the University of North Carolina at Chapel Hill, and funded by grant P01-HD31921 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development, with cooperative funding from 23 other federal agencies and foundations. Information on how to obtain the Add Health data files is available on the Add Health website (http://www.cpc.unc.edu/addhealth).

Ever-used Cannabis

GWAS Summary Statistic Source: Pasman et al. 2018. "GWAS of Lifetime Cannabis Use Reveals New Risk Loci, Genetic Overlap with Psychiatric Traits, and a Causal Influence of Schizophrenia." *Nature Neuroscience* 21(9):1161.

GWAS Ancestry Group(s): European

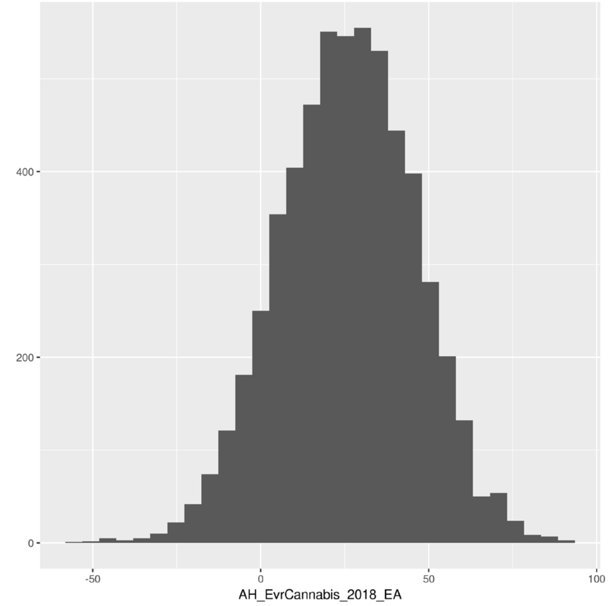
The PGSs for ever-used cannabis were created using results from a 2018 study conducted by Joelle Pasman and colleagues. The GWAS meta-analysis files are publicly available and can be downloaded from: https://www.ru.nl/bsi/research/group-pages/substance-use-addiction-food-saf/vm-saf/genetics/international-cannabis-consortium-icc/.

The ever-used cannabis GWAS meta-analyses are based on the analysis of 184,765 individuals of European genetic ancestry from the International Cannabis Consortium, the UK Biobank, and 23andMe. However, due to data use restrictions imposed by 23andMe, the summary statistics used to create the Add Health PGSs exclude the 23andMe sample resulting in a sample size of 162,082. The measure of cannabis use is a dichotomous indicator of if a person had ever used cannabis in their lifetime. The exact survey instrument varied between cohorts. GWAS analyses adjusted for sex, age (and age-squared in the UK Biobank), genotype batch, and an undisclosed number of principal components of the genome-wide data.

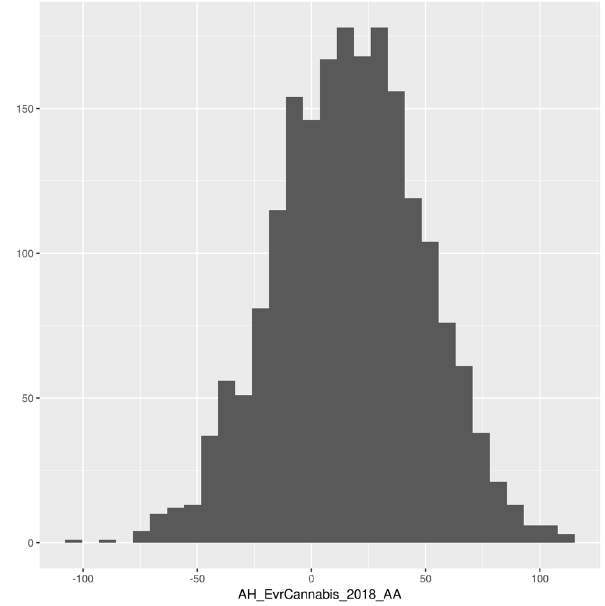
The Add Health PGSs for ever-used cannabis are standardized (mean = 0, standard deviation = 1) within ancestry groups to control for population stratification. To give researchers an idea of the underlying raw PGS, we present density plots for the unstandardized PGSs below. **Please read the section entitled "Using Add Health PGSs" in the introductory portion of this document prior to conducting any analyses using the provided PGSs.**

Ever-used Cannabis

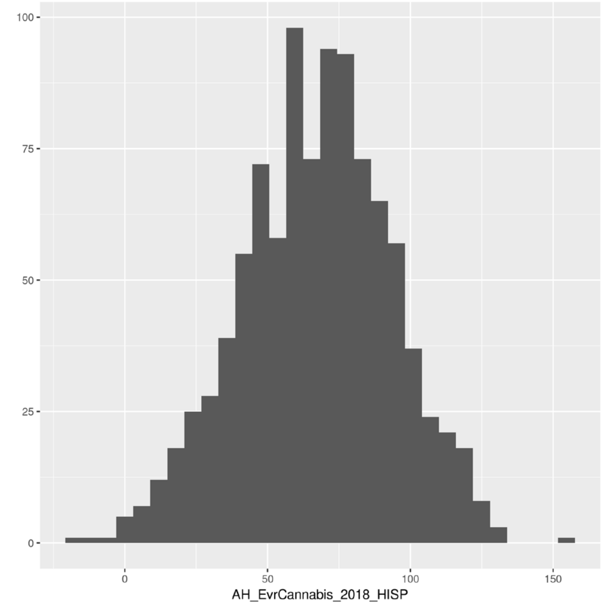
European ancestry raw PGS distribution



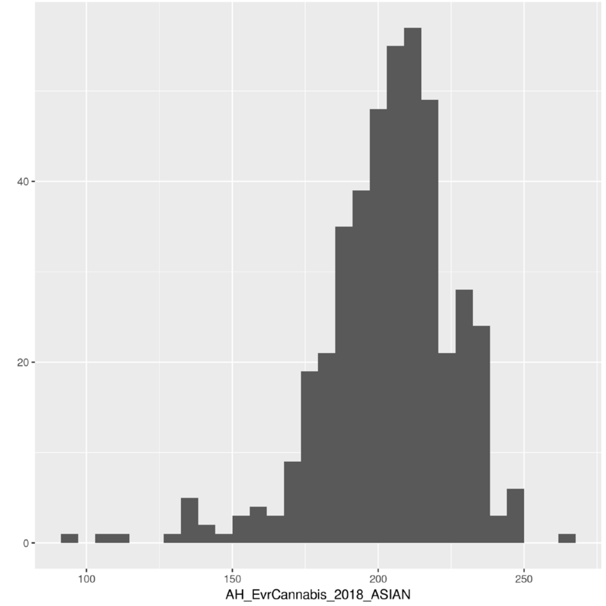
African ancestry raw PGS distribution



Hispanic ancestry raw PGS distribution



Asian ancestry raw PGS distribution



Menarche

GWAS Summary Statistic Source: Perry, John R. B. et al. 2014. "Parent-of-Origin-Specific Allelic Associations among 106 Genomic Loci for Age at Menarche." *Nature* 514(7520):92–97.

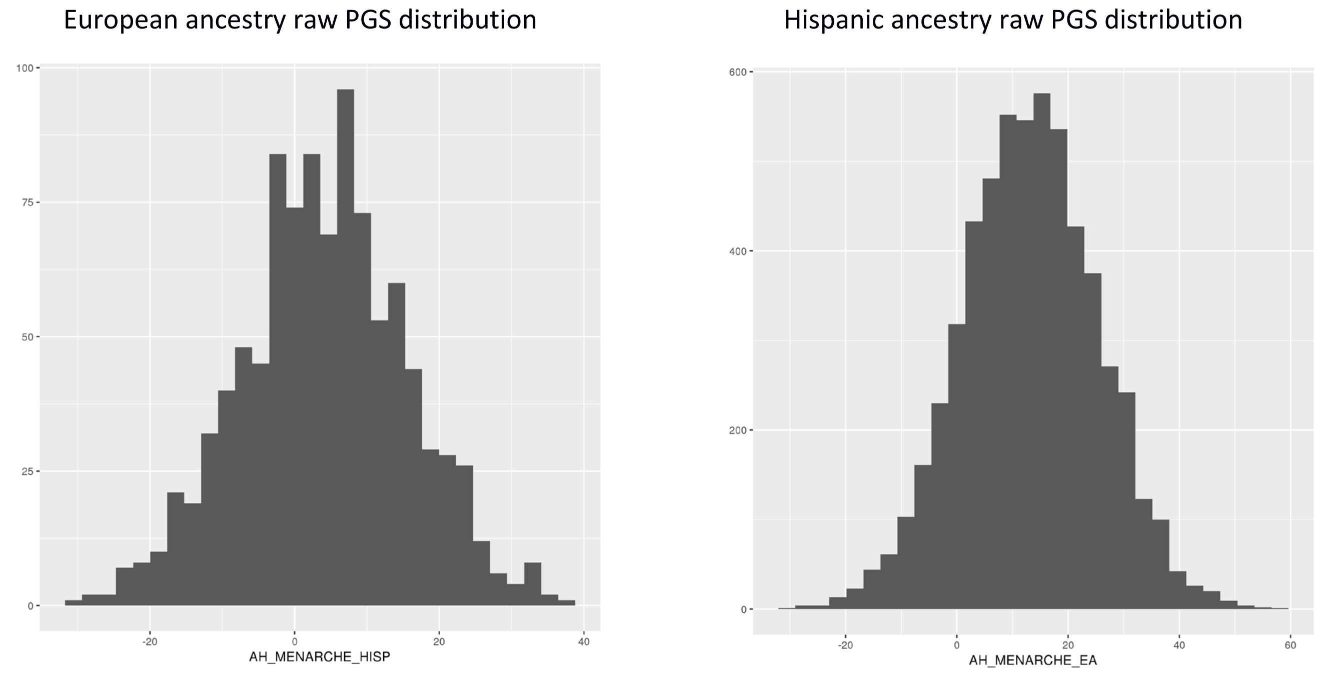
GWAS Ancestry Group(s): European

PGSs for age at menarche were created using results from a 2014 study conducted by the Reproductive Genetics (ReproGen) consortium. The GWAS meta-analysis files are publicly available from: http://www.reprogen.org/data_download.html.

The discovery meta-analysis included 132,989 women of European descent from 58 studies. Women who reported their age at menarche as less than 9 years or greater than 17 years old were excluded from the analysis. Analyses adjusted for birth year in order to account for secular trends in menarche timing, study-specific genomic control inflation factors, and relatedness between individuals when necessary.

The Add Health PGSs for age at menarche are standardized (mean = 0, standard deviation = 1) within ancestry groups to control for population stratification. To give researchers an idea of the underlying raw PGS, we present density plots for the unstandardized PGSs below. **Please read the section entitled "Using Add Health PGSs" in the introductory portion of this document prior to conducting any analyses using the provided PGSs.**

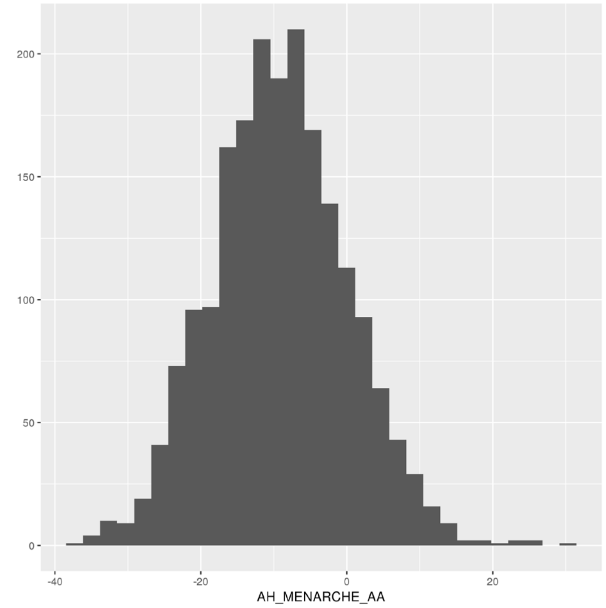
Menarche


European ancestry raw PGS distribution

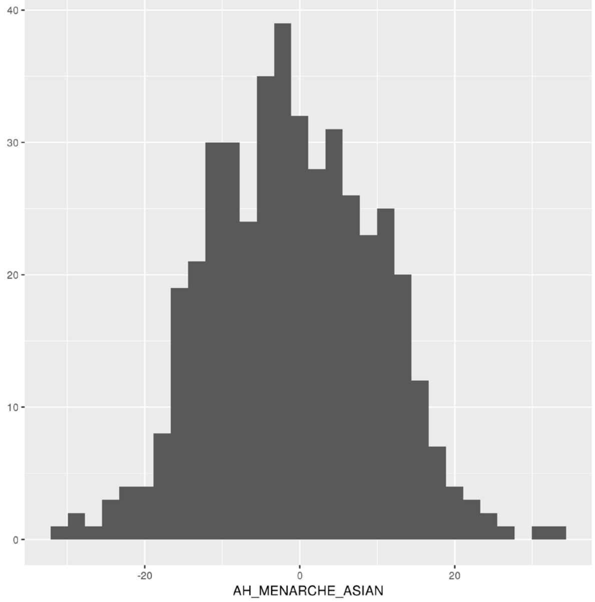
Hispanic ancestry raw PGS distribution

Add Health is a program project directed by Kathleen Mullan Harris and designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris at the University of North Carolina at Chapel Hill, and funded by grant P01-HD31921 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development, with cooperative funding from 23 other federal agencies and foundations. Information on how to obtain the Add Health data files is available on the Add Health website (http://www.cpc.unc.edu/addhealth).

African ancestry raw PGS distribution



Asian ancestry raw PGS distribution



Add Health is a program project directed by Kathleen Mullan Harris and designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris at the University of North Carolina at Chapel Hill, and funded by grant P01-HD31921 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development, with cooperative funding from 23 other federal agencies and foundations. Information on how to obtain the Add Health data files is available on the Add Health website (http://www.cpc.unc.edu/addhealth).

Menopause

GWAS Summary Statistic Source: Day, Felix R. et al. 2015. “Large-Scale Genomic Analyses Link Reproductive Aging to Hypothalamic Signaling, Breast Cancer Susceptibility and BRCA1-Mediated DNA Repair.” *Nature Genetics* 47(11):1294–1303.

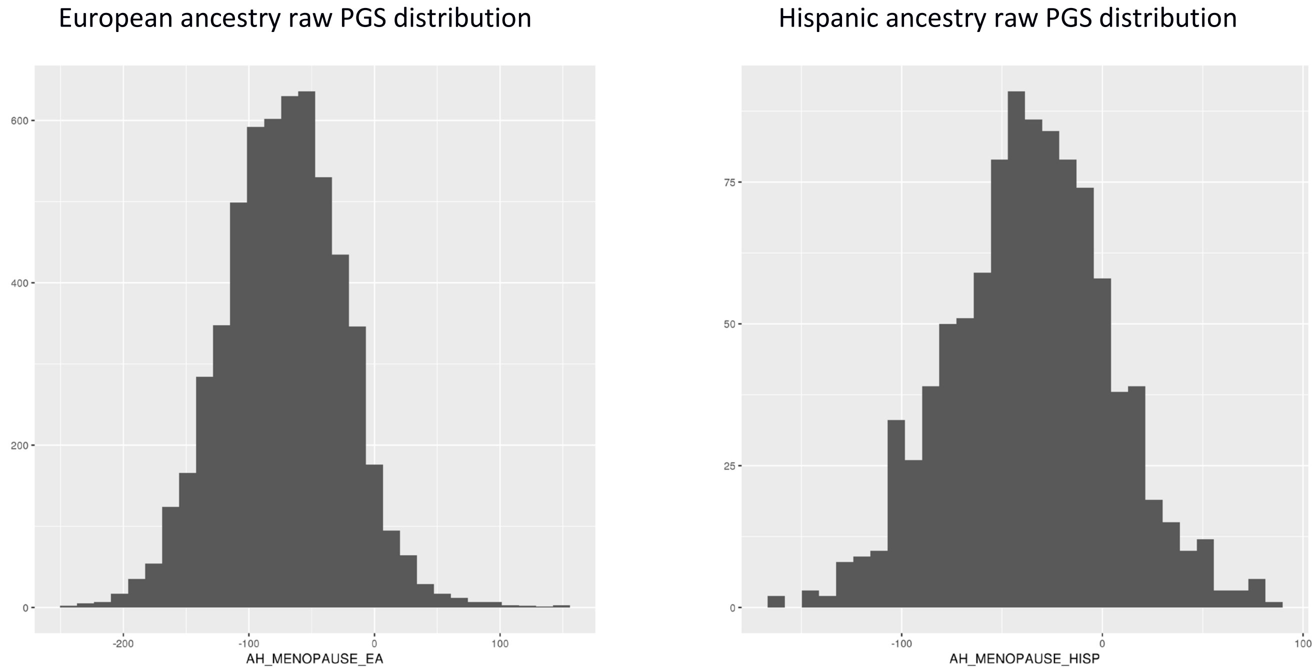
GWAS Ancestry Group(s): European

PGSs for age at menopause were created using results from a 2014 study conducted by the Reproductive Genetics (ReproGen) consortium. The GWAS meta-analysis files are publicly available from: http://www.reprogen.org/data_download.html.

The ReproGen meta-analysis included 69,360 women of European descent from 50 studies imputed to HapMap Phase 2. GWAS analyses included study specific controls and the top principal components of the genetic data. In all studies, age at naturally occurring menopause is self-reported. Age at menopause is defined as the age at last occurring menstrual period followed by 12 consecutive months of amenorrhea.

The Add Health PGSs for age at menopause are standardized (mean = 0, standard deviation = 1) within ancestry groups to control for population stratification. To give researchers an idea of the underlying raw PGS, we present density plots for the unstandardized PGSs below. **Please read the section entitled “Using Add Health PGSs” in the introductory portion of this document prior to conducting any analyses using the provided PGSs.**

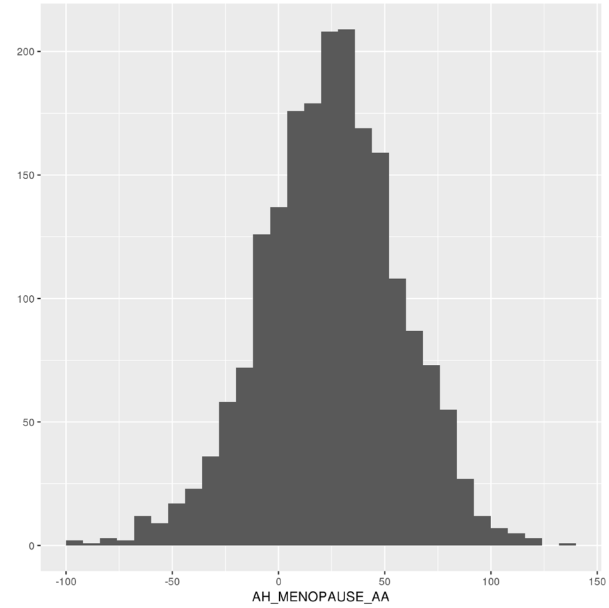
Menopause


European ancestry raw PGS distribution

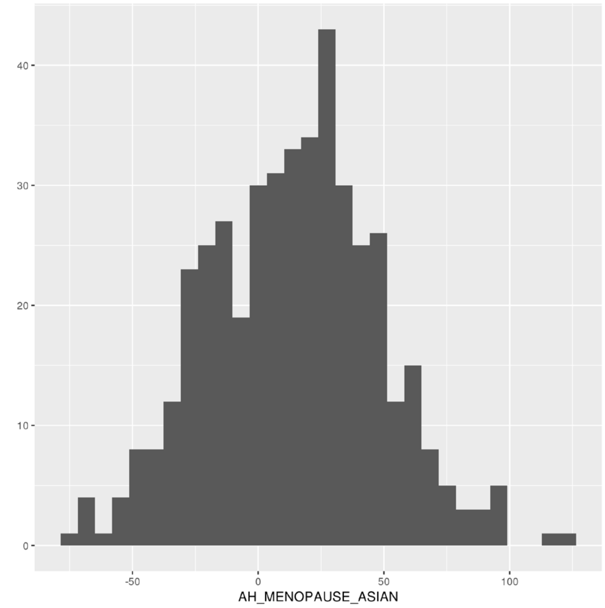
Hispanic ancestry raw PGS distribution

Add Health is a program project directed by Kathleen Mullan Harris and designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris at the University of North Carolina at Chapel Hill, and funded by grant P01-HD31921 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development, with cooperative funding from 23 other federal agencies and foundations. Information on how to obtain the Add Health data files is available on the Add Health website (http://www.cpc.unc.edu/addhealth).

African ancestry raw PGS distribution



Asian ancestry raw PGS distribution



Number of Children Ever Born

GWAS Summary Statistic Source: Barban, Nicola et al. 2016. “Genome-Wide Analysis Identifies 12 Loci Influencing Human Reproductive Behavior.” *Nature Genetics* 48(12):1462–72.

GWAS Ancestry Group(s): European

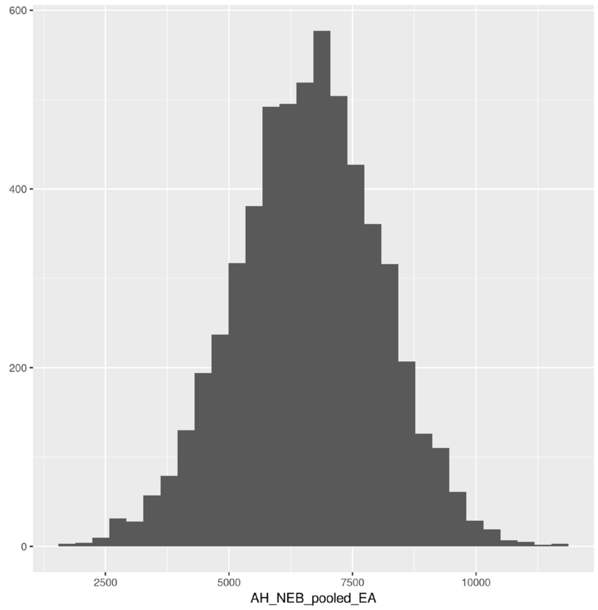
There are three PGSs for the number of children ever born using results from a 2016 study conducted as part of the Sociogenome project (http://www.sociogenome.com/). The first is based on summary statistics from the pooled sample of men and women while the other two use summary statistics from sex-specific GWASs and are only made available for Add Health respondents of the corresponding sex. The GWAS meta-analysis files are publicly available from: ftp://ftp.ebi.ac.uk/pub/databases/gwas/summary_statistics/BarbanN_27798627_GCST003795.

The meta-analysis included 343,072 individuals of European descent above age 44 for women and 54 for men from 62 studies. Additional covariates include age, the second order polynomial of age, the third order polynomial of age, and genetic principal components.

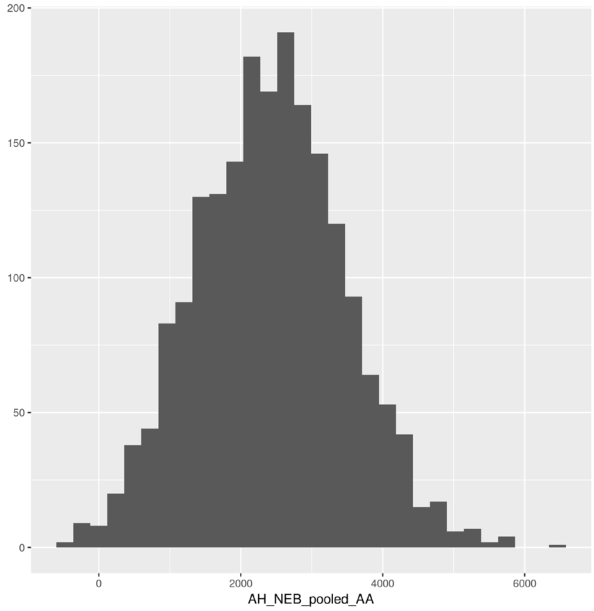
The Add Health PGSs for number of children ever born are standardized (mean = 0, standard deviation = 1) within ancestry groups to control for population stratification. To give researchers an idea of the underlying raw PGS, we present density plots for the unstandardized PGSs below. **Please read the section entitled “Using Add Health PGSs” in the introductory portion of this document prior to conducting any analyses using the provided PGSs.**

Number of Children Ever Born (pooled sample)

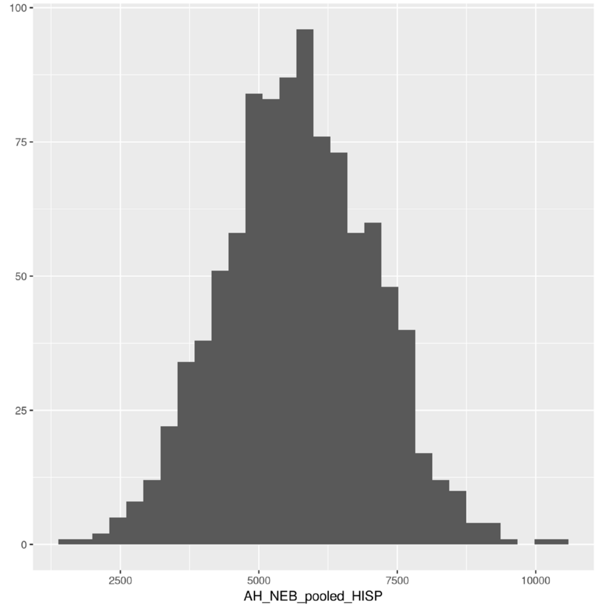
European ancestry raw PGS distribution



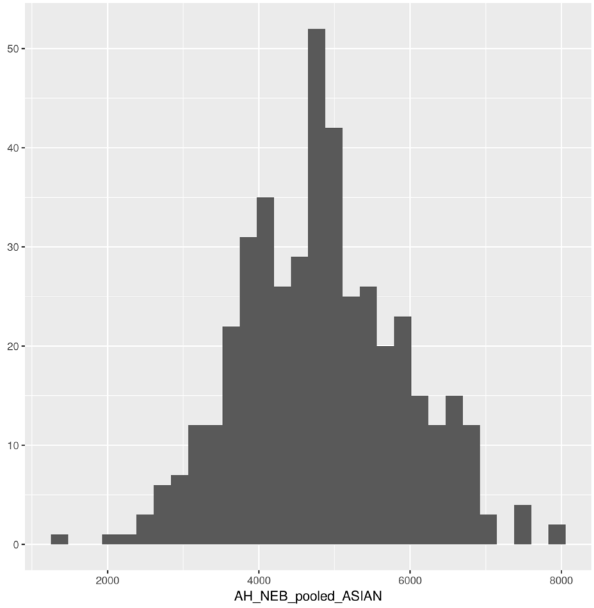
African ancestry raw PGS distribution



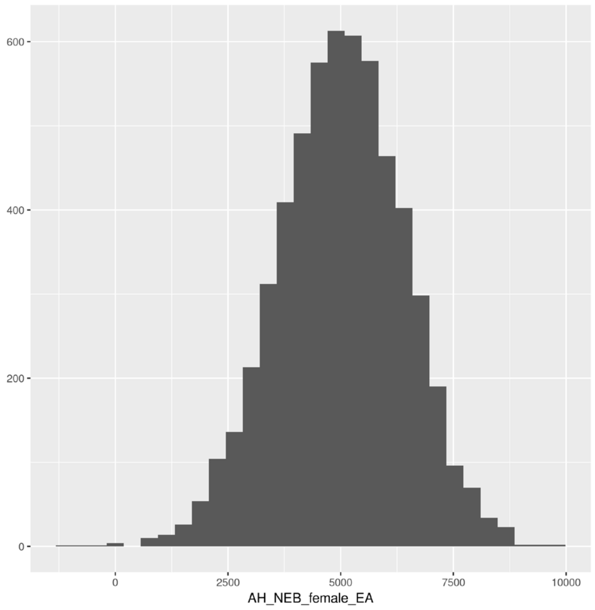
Hispanic ancestry raw PGS distribution



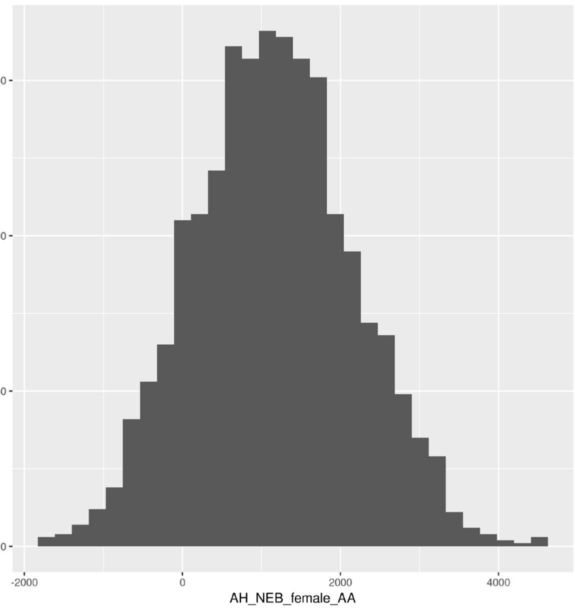
Asian ancestry raw PGS distribution



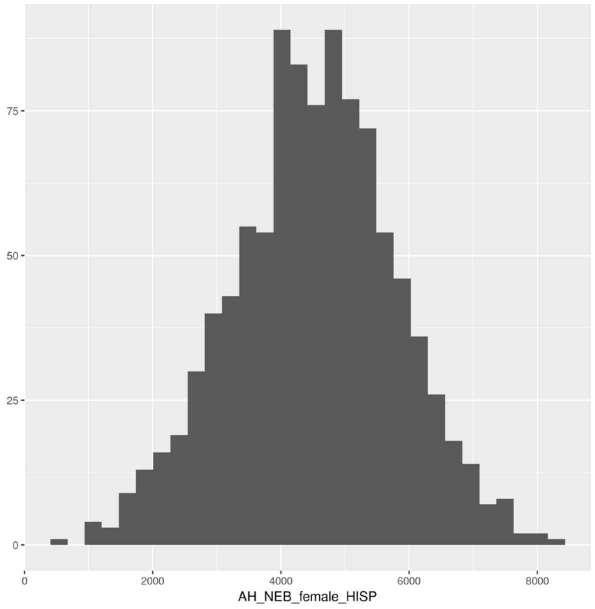
Number of Children Ever Born (female only sample)



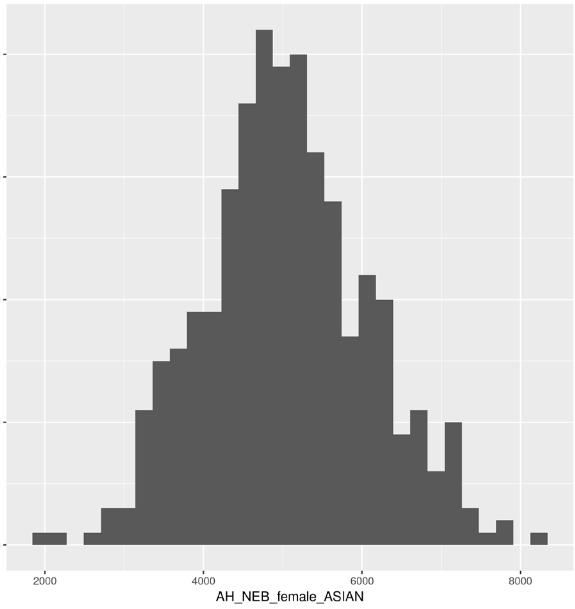
European ancestry raw PGS distribution



African ancestry raw PGS distribution



Hispanic ancestry raw PGS distribution

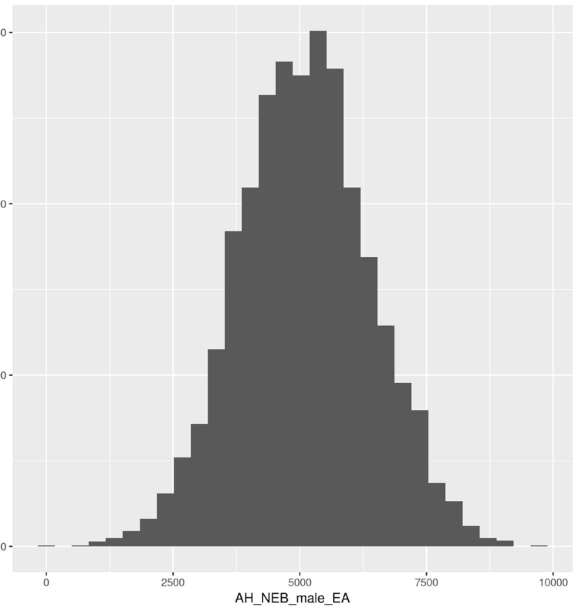


Asian ancestry raw PGS distribution

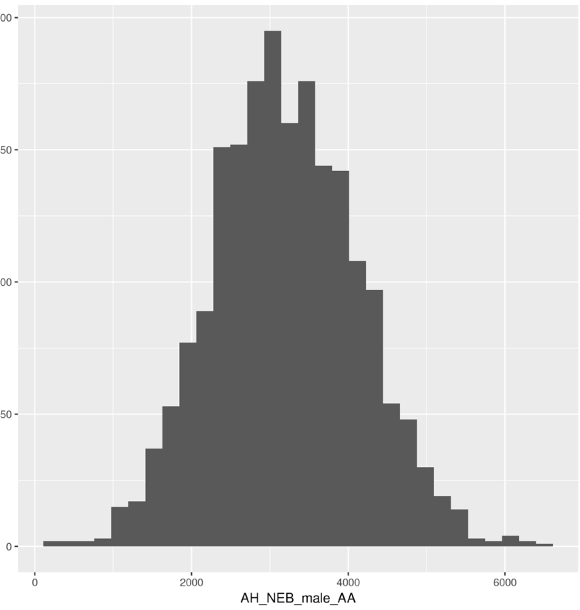
Add Health is a program project directed by Kathleen Mullan Harris and designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris at the University of North Carolina at Chapel Hill, and funded by grant P01-HD31921 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development, with cooperative funding from 23 other federal agencies and foundations. Information on how to obtain the Add Health data files is available on the Add Health website (http://www.cpc.unc.edu/addhealth).

Number of Children Ever Born (male only sample)

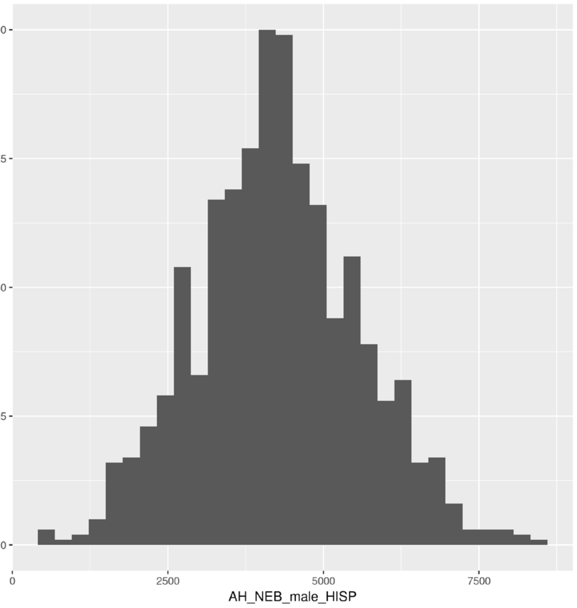
European ancestry raw PGS distribution



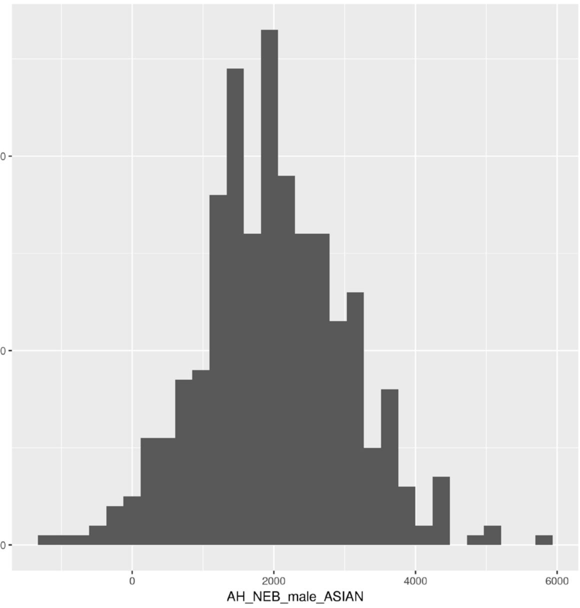
African ancestry raw PGS distribution



Hispanic ancestry raw PGS distribution



Asian ancestry raw PGS distribution



Age at First Birth

GWAS Summary Statistic Source: Barban, Nicola et al. 2016. “Genome-Wide Analysis Identifies 12 Loci Influencing Human Reproductive Behavior.” *Nature Genetics* 48(12):1462–72.

GWAS Ancestry Group(s): European

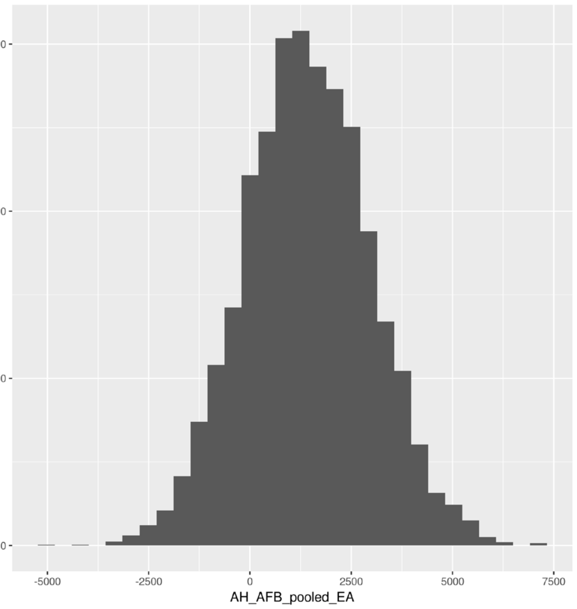
There are three PGSs for the age at first birth using results from a 2016 study conducted as part of the Sociogenome project (http://www.sociogenome.com/). The first is based on summary statistics from the pooled sample of men and women while the other two use summary statistics from sex-specific GWASs and are only made available for Add Health respondents of the corresponding sex. The GWAS meta-analysis files are publicly available from:
ftp://ftp.ebi.ac.uk/pub/databases/gwas/summary_statistics/BarbanN_27798627_GCST003795.

The GWAS meta-analysis included 251,151 individuals of European descent who had a biological child at the time of interview from 62 studies. Additional covariates include age, the second order polynomial of age, the third order polynomial of age, and genetic principal components.

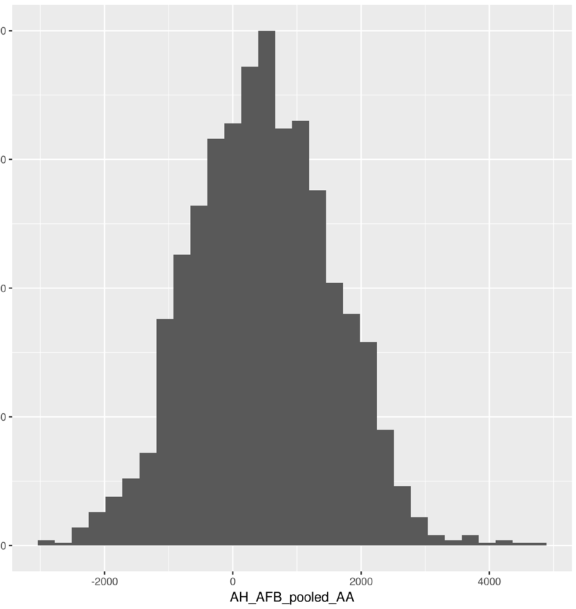
The Add Health PGSs for age at first birth are standardized (mean = 0, standard deviation = 1) within ancestry groups to control for population stratification. To give researchers an idea of the underlying raw PGS, we present density plots for the unstandardized PGSs below. **Please read the section entitled “Using Add Health PGSs” in the introductory portion of this document prior to conducting any analyses using the provided PGSs.**

Age at First Birth (pooled sample)

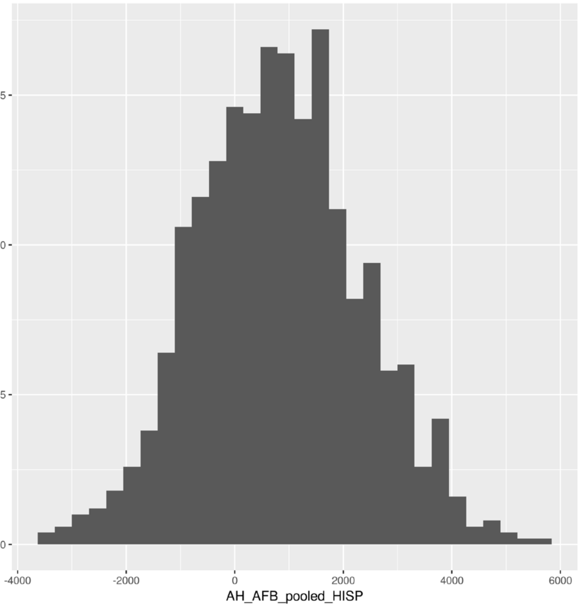
European ancestry raw PGS distribution



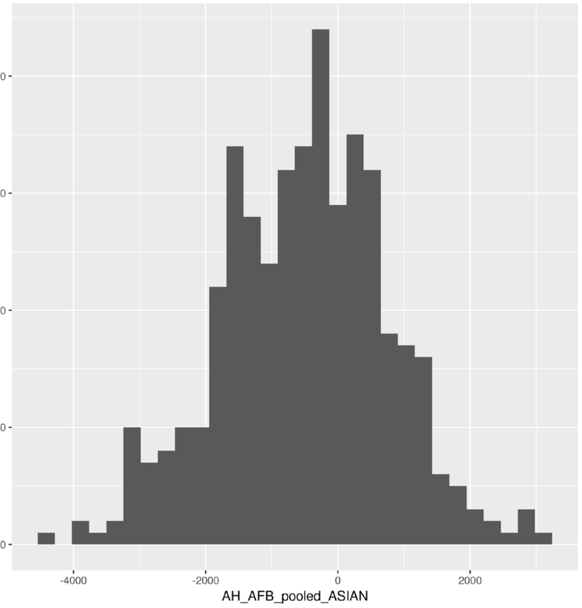
African ancestry raw PGS distribution



Hispanic ancestry raw PGS distribution



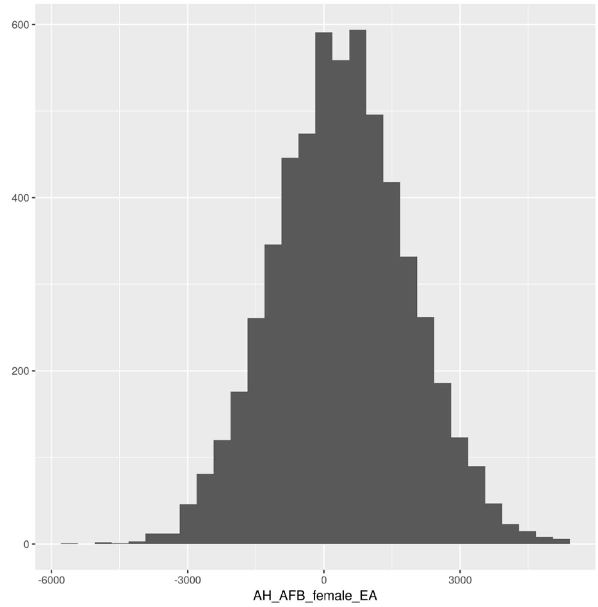
Asian ancestry raw PGS distribution



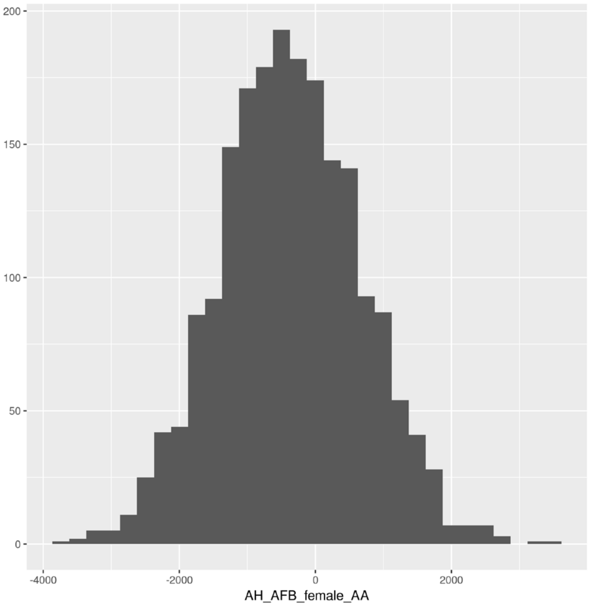
Add Health is a program project directed by Kathleen Mullan Harris and designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris at the University of North Carolina at Chapel Hill, and funded by grant P01-HD31921 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development, with cooperative funding from 23 other federal agencies and foundations. Information on how to obtain the Add Health data files is available on the Add Health website (http://www.cpc.unc.edu/addhealth).

Age at First Birth (female sample)

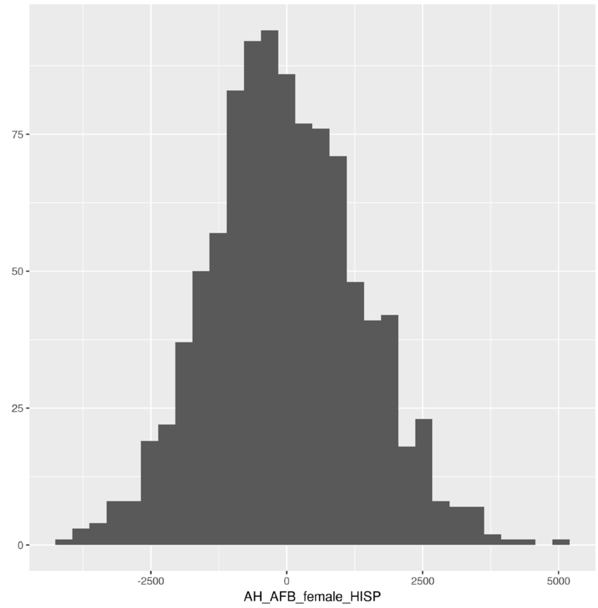
European ancestry raw PGS distribution



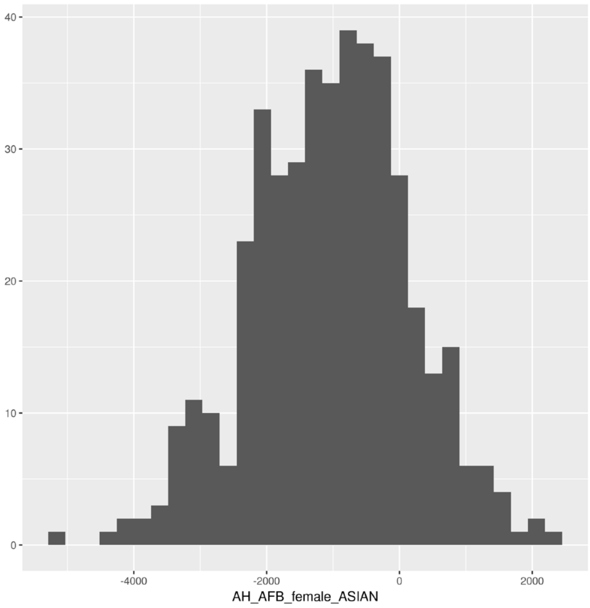
African ancestry raw PGS distribution



Hispanic ancestry raw PGS distribution



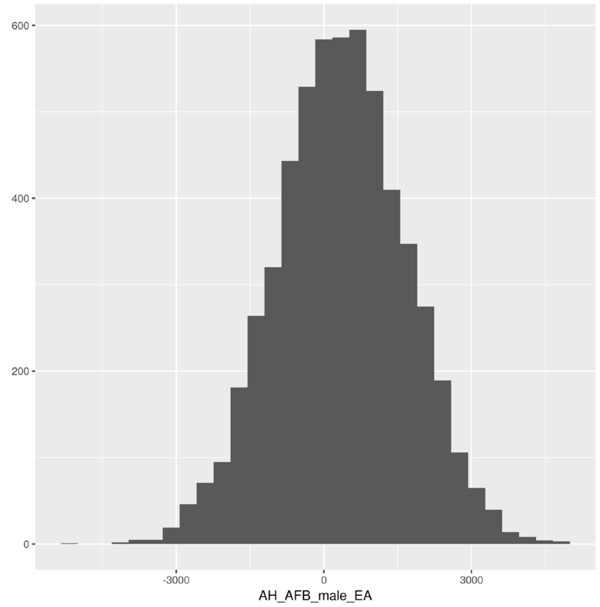
Asian ancestry raw PGS distribution



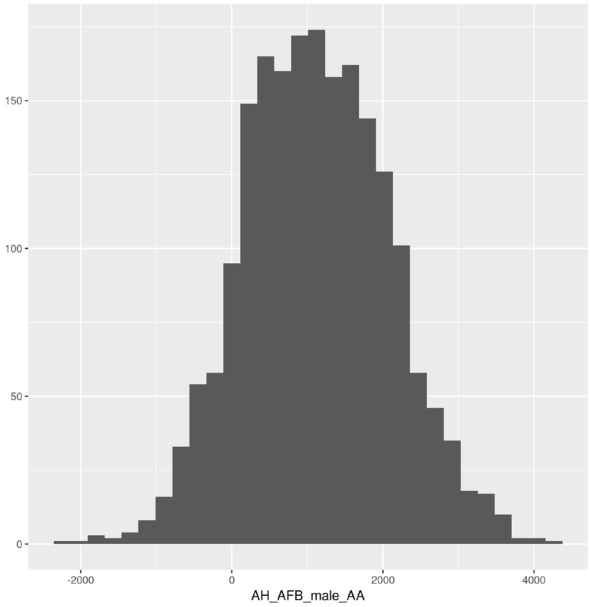
Add Health is a program project directed by Kathleen Mullan Harris and designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris at the University of North Carolina at Chapel Hill, and funded by grant P01-HD31921 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development, with cooperative funding from 23 other federal agencies and foundations. Information on how to obtain the Add Health data files is available on the Add Health website (http://www.cpc.unc.edu/addhealth).

Age at First Birth (male sample)

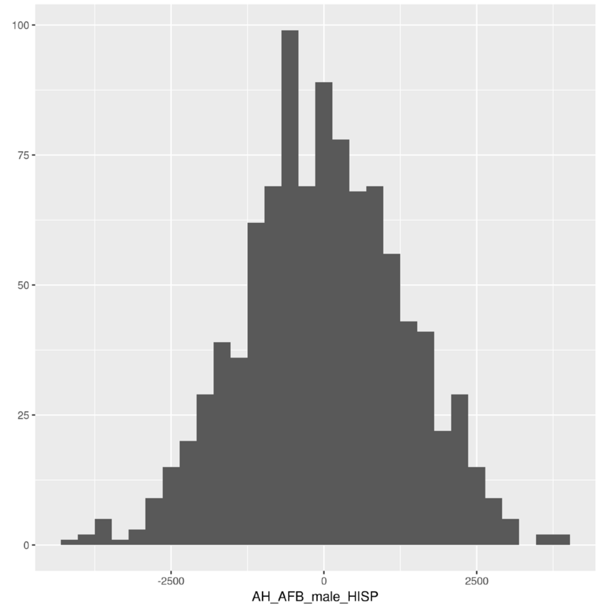
European ancestry raw PGS distribution



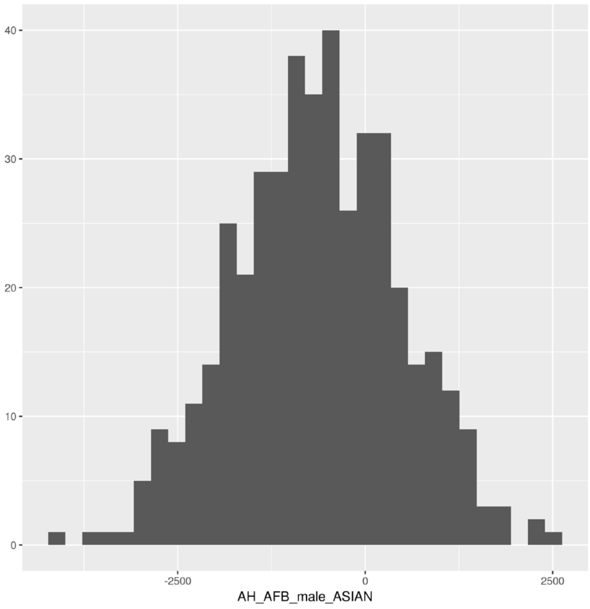
African ancestry raw PGS distribution



Hispanic ancestry raw PGS distribution



Asian ancestry raw PGS distribution



Add Health is a program project directed by Kathleen Mullan Harris and designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris at the University of North Carolina at Chapel Hill, and funded by grant P01-HD31921 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development, with cooperative funding from 23 other federal agencies and foundations. Information on how to obtain the Add Health data files is available on the Add Health website (http://www.cpc.unc.edu/addhealth).

Post-Traumatic Stress Disorder

GWAS Summary Statistic Source: Duncan et al. 2018. "Largest GWAS of PTSD (N=20,070) Yields Genetic Overlap with Schizophrenia and Sex Differences in Heritability." *Molecular Psychiatry* 23(3):666–73.

GWAS Ancestry Group(s): African, South African, Latino, and European

There are three PGSs for post-traumatic stress disorder (PTSD) created using results from the 2018 study conducted by the PGC-PTSD and colleagues. The first PGS is based on summary statistics from the combined sample GWAS, used to create a PGS for all four genetic ancestry groups. The other two PGSs use summary statistics from the GWASs limited to African and European ancestry respectively and are only made available for Add Health respondents of the corresponding genetic ancestry (see the introductory material for the methodology used to determine genetic ancestry of Add Health respondents). The GWAS meta-analysis files are publicly available and can be downloaded from: https://www.med.unc.edu/pgc/results-and-downloads.

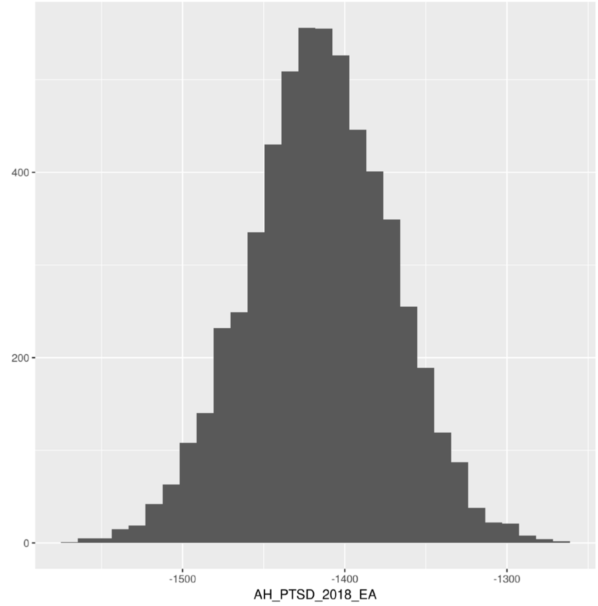
The PTSD GWAS meta-analyses are based on the analysis of 20,730 individuals (9,691 African genetic ancestry, 9,956 European genetic ancestry, 698 Latino genetic ancestry, 387 South African genetic ancestry). Importantly, the vast majority (87%) of controls were trauma-exposed controls. The measurement of PTSD differs between cohorts but generally consisted of a validated assessment (e.g., the PTSD checklist—a 17-item self-report measure of DSM-IV symptoms). GWAS analyses adjusted for different covariates in the various cohorts, but generally included the first 10 genetic principal components as well as ancestry-specific PCs (see the Supplemental Information for more details).

The Add Health PGSs for PTSD are standardized (mean = 0, standard deviation = 1) within ancestry groups to control for population stratification. To give researchers an idea of the underlying raw PGS, we present density plots for the unstandardized PGSs below. **Please read the section entitled "Using Add Health PGSs" in the introductory portion of this document prior to conducting any analyses using the provided PGSs.**

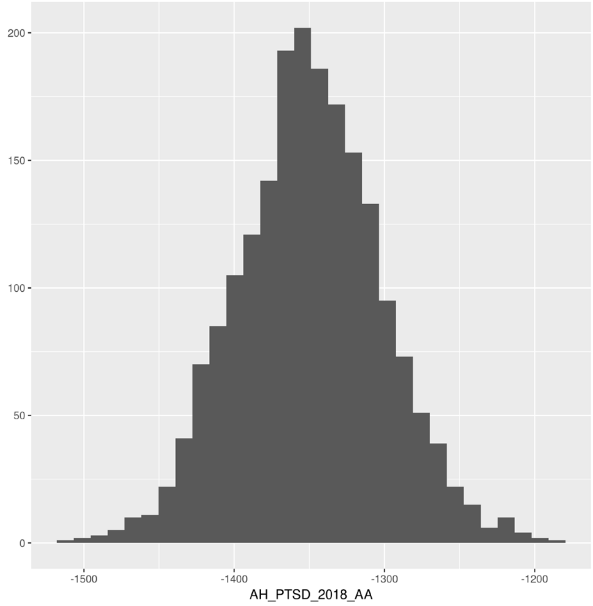
Add Health is a program project directed by Kathleen Mullan Harris and designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris at the University of North Carolina at Chapel Hill, and funded by grant P01-HD31921 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development, with cooperative funding from 23 other federal agencies and foundations. Information on how to obtain the Add Health data files is available on the Add Health website (http://www.cpc.unc.edu/addhealth).

PTSD (Combined Sample)

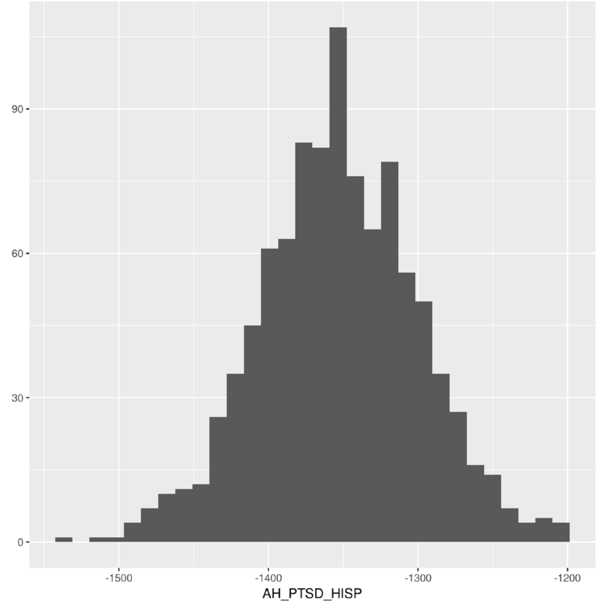
European ancestry raw PGS distribution



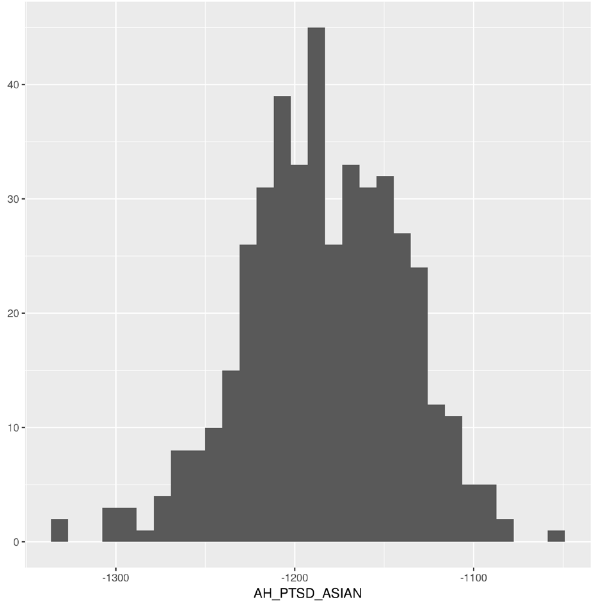
African ancestry raw PGS distribution



Hispanic ancestry raw PGS distribution



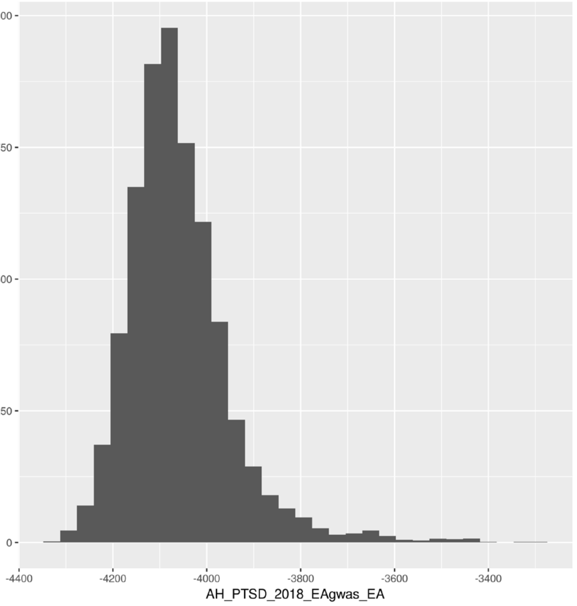
Asian ancestry raw PGS distribution



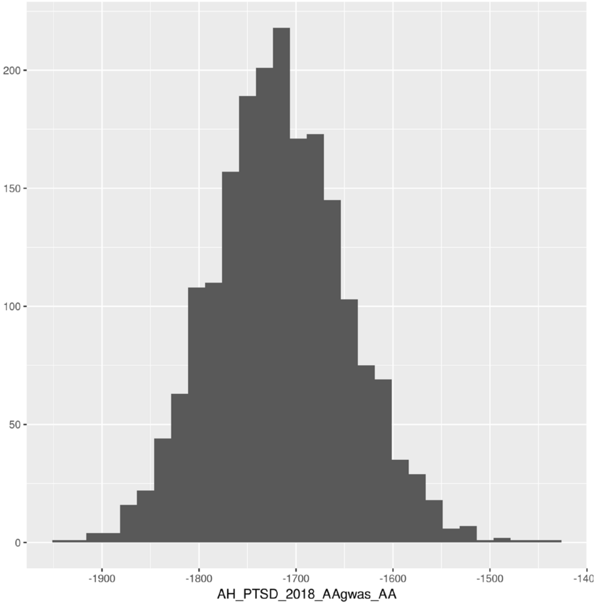
Add Health is a program project directed by Kathleen Mullan Harris and designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris at the University of North Carolina at Chapel Hill, and funded by grant P01-HD31921 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development, with cooperative funding from 23 other federal agencies and foundations. Information on how to obtain the Add Health data files is available on the Add Health website (http://www.cpc.unc.edu/addhealth).

PTSD (Ancestry-Specific GWASs)

European ancestry GWAS raw PGS distribution



African ancestry GWAS raw PGS distribution



Attention-Deficit/Hyperactivity Disorder (2010)

GWAS Summary Statistic Source: Neale, Benjamin M. et al. 2010. "Meta-Analysis of Genome-Wide Association Studies of Attention-Deficit/Hyperactivity Disorder." *Journal of the American Academy of Child & Adolescent Psychiatry* 49(9):884–97.

GWAS Ancestry Group(s): European

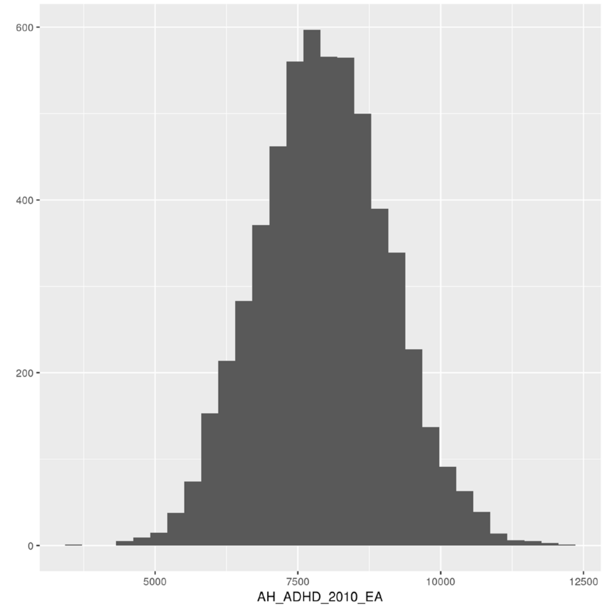
The PGSs for attention-deficit/hyperactivity disorder (ADHD) were created using results from a 2010 GWAS conducted by the ADHD subgroup of the Psychiatric GWAS Consortium. The GWAS meta-analysis files are publicly available from: http://www.med.unc.edu/pgc/results-and-downloads (pgc.adhd.full.2012-10.txt).

The meta-analysis included 2,064 trios, 896 cases, and 2,455 controls from four projects: (1) the Children's Hospital of Philadelphia (CHOP), (2) phase I of the International Multisite ADHD Genetics Project (IMAGE), (3) phase II of IMAGE (IMAGE II), and (4) the Pfizer funded study from the University of California, Los Angeles, Washington University, and the Massachusetts General Hospital (PUWMa). Measures of ADHD were harmonized across samples and were based on a combination of semi-structured interviews and parent and/or teacher report on questionnaires (see Neale et al., p. 3-5, 2010).

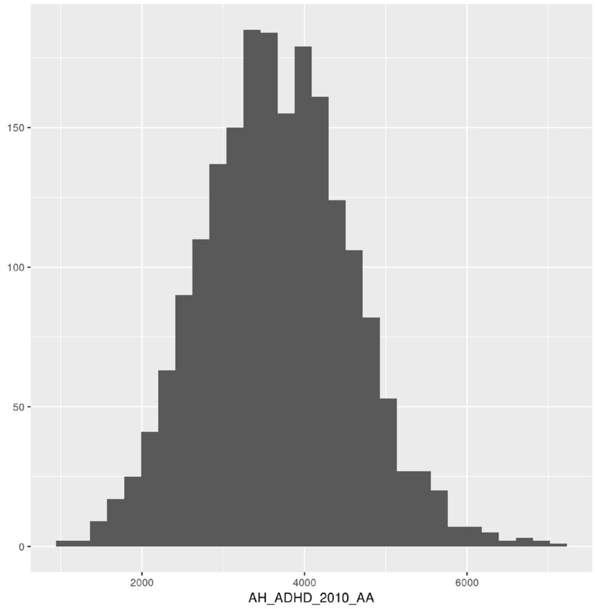
The Add Health PGSs for ADHD (2010) are standardized (mean = 0, standard deviation = 1) within ancestry groups to control for population stratification. To give researchers an idea of the underlying raw PGS, we present density plots for the unstandardized PGSs below. **Please read the section entitled "Using Add Health PGSs" in the introductory portion of this document prior to conducting any analyses using the provided PGSs.**

Attention Deficit Disorder (2010)

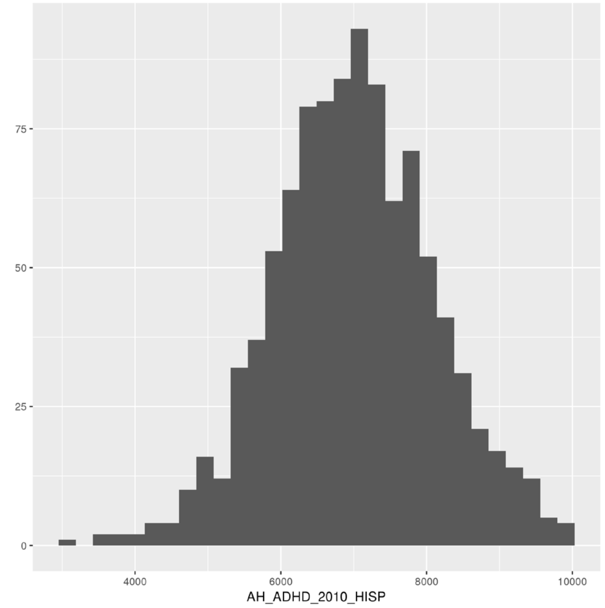
European ancestry raw PGS distribution



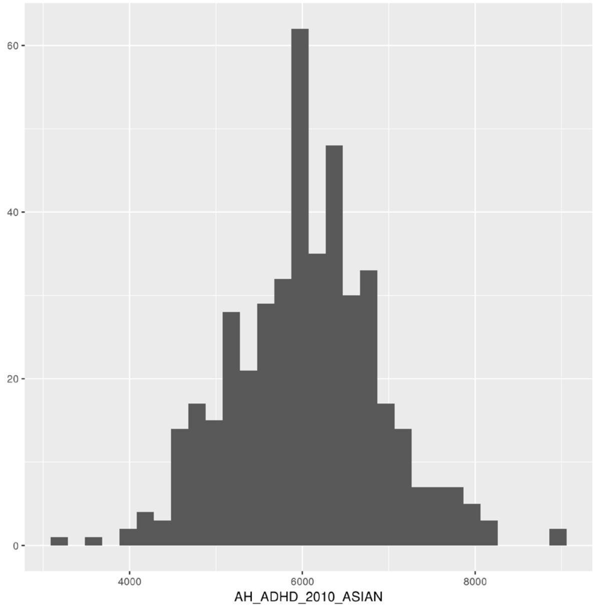
African ancestry raw PGS distribution



Hispanic ancestry raw PGS distribution



Asian ancestry raw PGS distribution



Add Health is a program project directed by Kathleen Mullan Harris and designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris at the University of North Carolina at Chapel Hill, and funded by grant P01-HD31921 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development, with cooperative funding from 23 other federal agencies and foundations. Information on how to obtain the Add Health data files is available on the Add Health website (http://www.cpc.unc.edu/addhealth).

Attention-Deficit/Hyperactivity Disorder (2019)

GWAS Summary Statistic Source: Demontis, Ditte et al. 2019. “Discovery of the First Genome-Wide Significant Risk Loci for Attention Deficit/Hyperactivity Disorder.” Nature Genetics 51(1):63–75.

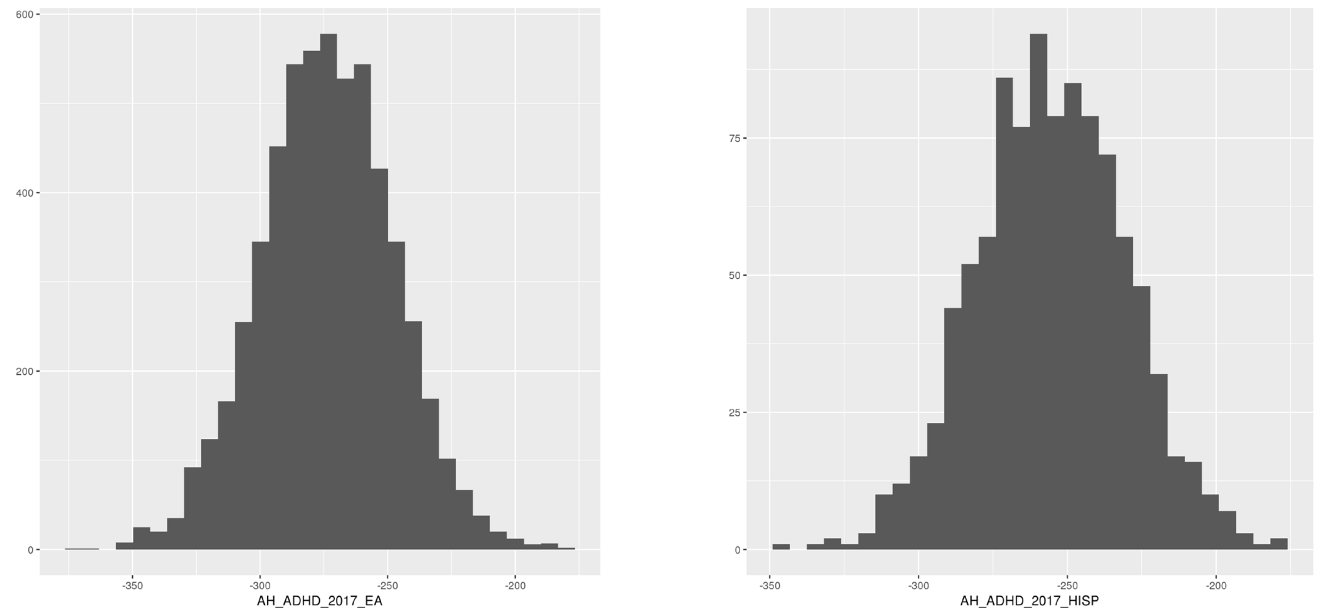
GWAS Ancestry Group(s): European (Ncases = 19,099; Ncontrols = 34,194), Han Chinese (Ncases = 1,040; Ncontrol = 963); Multi-ethnic (PUWMa sample with N = 78)

The PGSs for attention-deficit/hyperactivity disorder (ADHD) PGSs were created using results from a 2017 BioRxiv preprint from the Lundbeck Foundation Initiative for Integrative Psychiatric Research (iPSYCH) and Psychiatric Genetics Consortium (PGC). The GWAS meta-analysis files are publicly available from: http://www.med.unc.edu/pgc/results-and-downloads.

The discovery meta-analysis included 55,374 individuals (20,183 cases and 35,191 controls) from 12 studies of mixed ancestry while the replication sample included 93,916 individuals from two cohorts of multiple ancestry groups. Additional covariates included study specific controls and genetic ancestry principal components.

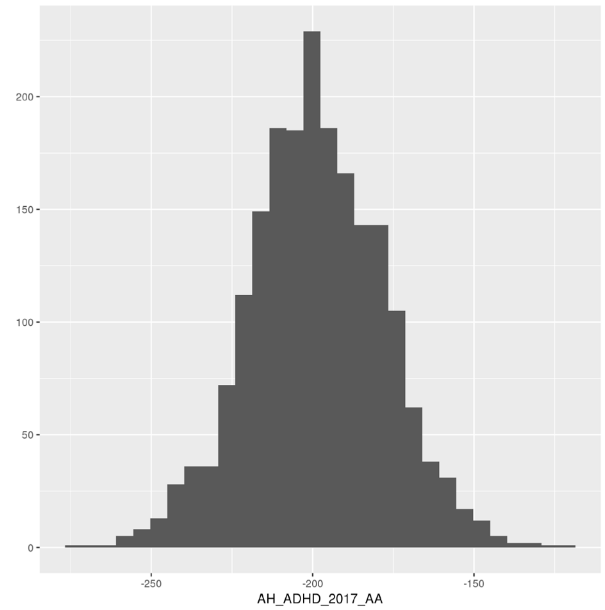
The Add Health PGSs for ADHD (2017) are standardized (mean = 0, standard deviation = 1) within ancestry groups to control for population stratification. To give researchers an idea of the underlying raw PGS, we present density plots for the unstandardized PGSs below. **Please read the section entitled “Using Add Health PGSs” in the introductory portion of this document prior to conducting any analyses using the provided PGSs.**

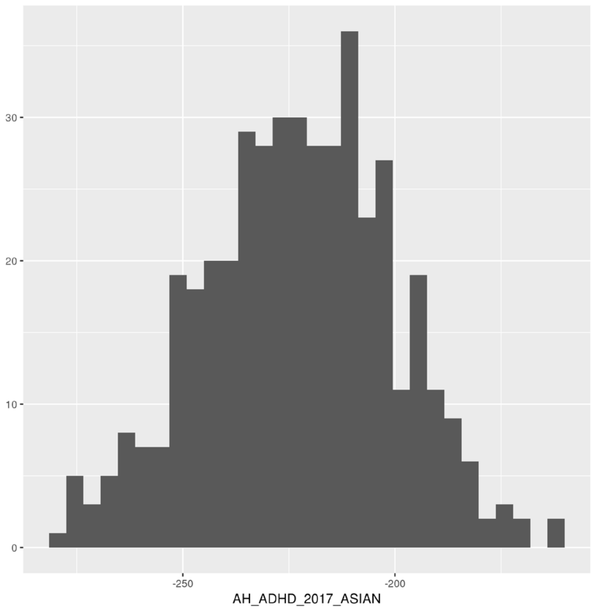
Attention Deficit Disorder (2019)



European ancestry raw PGS distribution

Hispanic ancestry raw PGS distribution


African ancestry raw PGS distribution


Asian ancestry raw PGS distribution

Bipolar Disorder

GWAS Summary Statistic Source: Psychiatric GWAS Consortium Bipolar Disorder Working Group. 2011. “Large-Scale Genome-Wide Association Analysis of Bipolar Disorder Identifies a New Susceptibility Locus near ODZ4.” Nature Genetics 43(10):977–83.

GWAS Ancestry Group(s): European

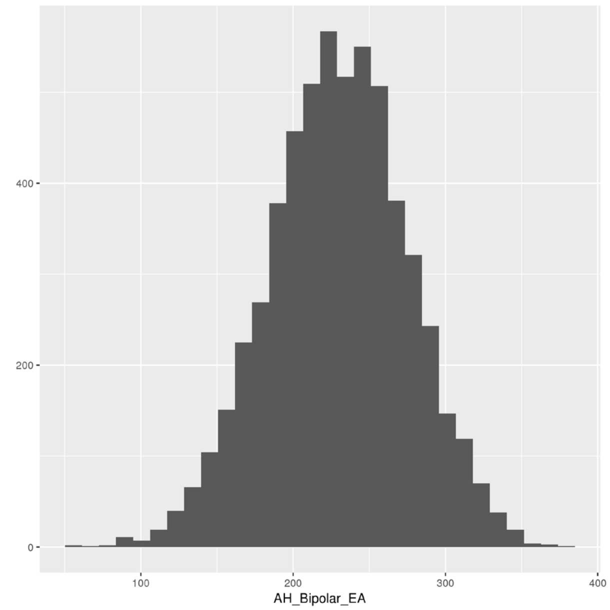
The PGSs for bipolar disorder were created using results from a 2011 GWAS conducted by the Bipolar Disorder working group of the Psychiatric GWAS Consortium. The GWAS meta-analysis files are publicly available from: http://www.med.unc.edu/pgc/results-and-downloads (pgc.bip.2012-04.zip).

The discovery phase of the meta-analysis included 7,481 cases and 9,250 controls. The replication phase of the 34 most significant genetic regions identified in the discovery phase was conducted in a sample of 4,496 cases and 42,422 controls. Samples were drawn from 11 studies (see Table 1 and supplemental materials, from the original publication). All analyses adjusted for the top five principal components of the genetic data and study fixed effects. Bipolar disorder was measured as diagnoses of either: bipolar disorder type 1, bipolar disorder type 2, schizoaffective disorder bipolar, and a few cases with other bipolar diagnoses (p. 997).

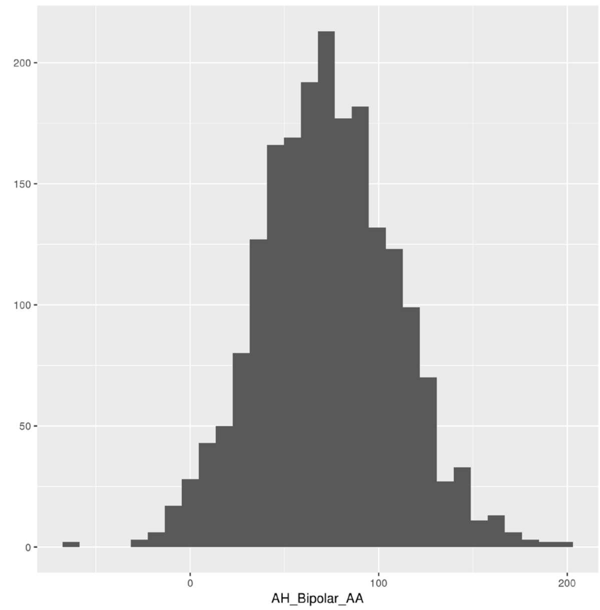
The Add Health PGSs for bipolar disorder are standardized (mean = 0, standard deviation = 1) within ancestry groups to control for population stratification, but to give researchers an idea of the underlying raw PGS, we present density plots for the unstandardized PGSs below. **Please read the section entitled “Using Add Health PGSs” in the introductory portion of this document prior to conducting any analyses using the provided PGSs.**

Bipolar Disorder

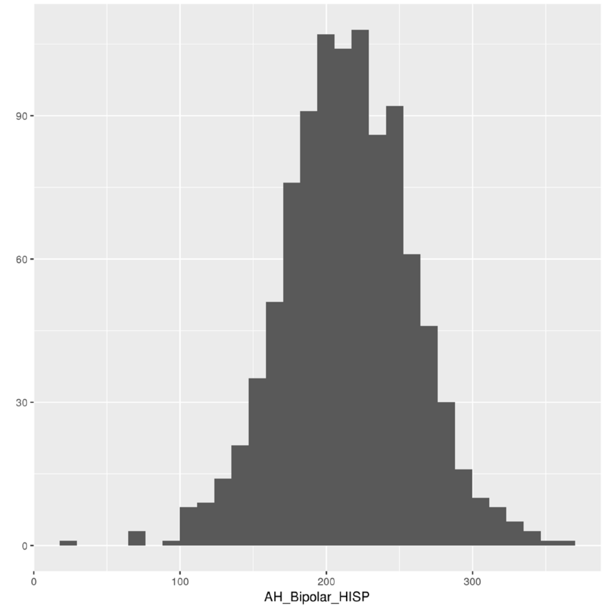
European ancestry raw PGS distribution



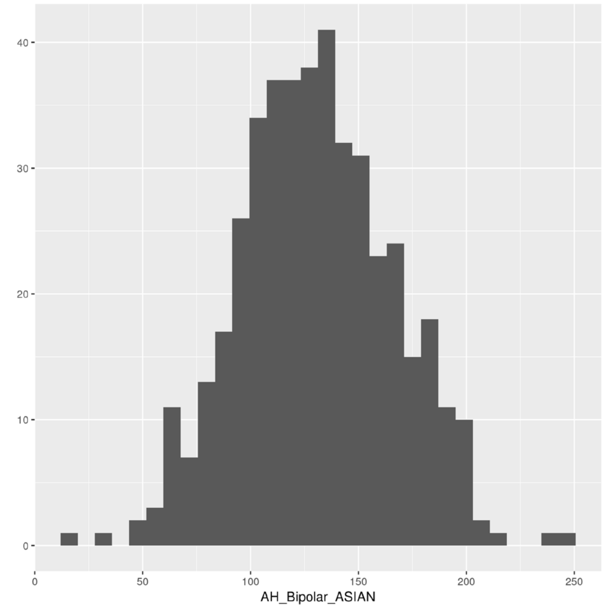
African ancestry raw PGS distribution



Hispanic ancestry raw PGS distribution



Asian ancestry raw PGS distribution



Schizophrenia

GWAS Summary Statistic Source: Ripke, Stephan et al. 2014. “Biological Insights from 108 Schizophrenia-Associated Genetic Loci.” Nature 511(7510):421–27.

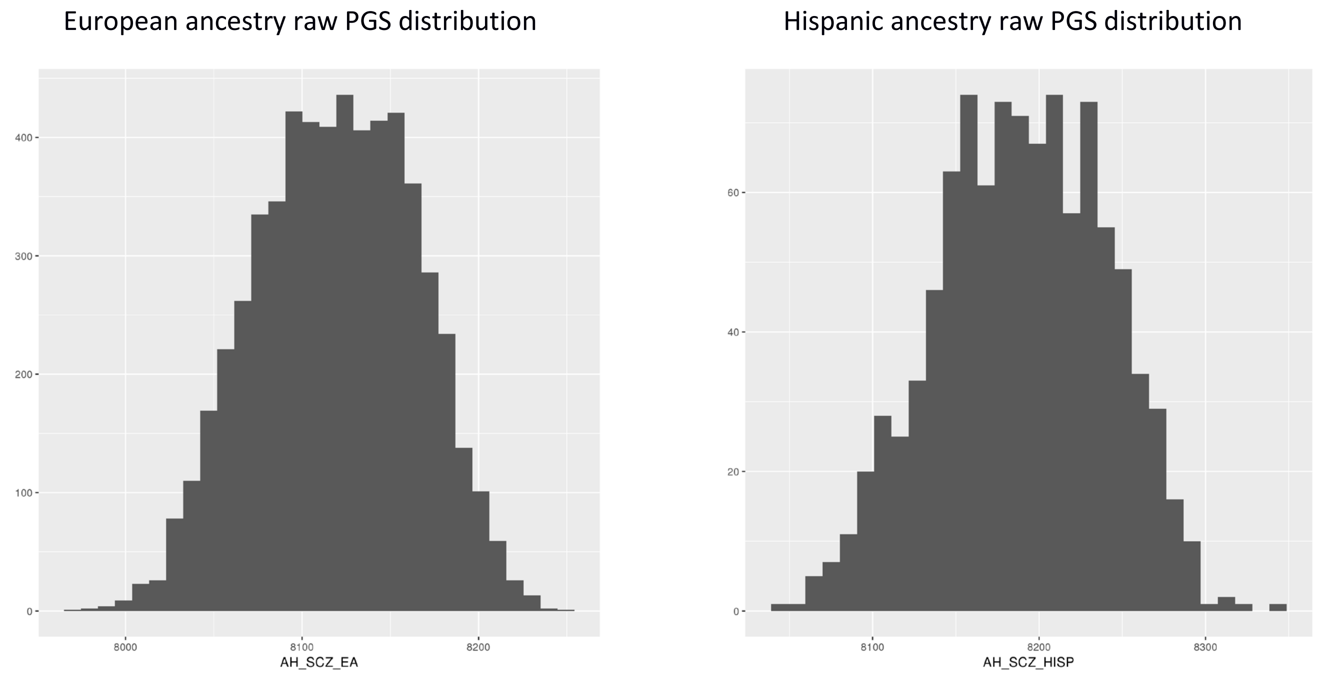
GWAS Ancestry Group(s): European and East Asian

The PGSs for schizophrenia were created using results from a 2014 GWAS conducted by the Schizophrenia Working Group of the Psychiatric Genomics Consortium (PGC). The GWAS meta-analysis files are publicly available from: https://www.med.unc.edu/pgc/results-and-downloads.

The schizophrenia GWAS combined meta-analysis included 36,989 cases and 113,075 controls (N=150,064) from individuals of European and East Asian descent. The replication sample consisted of 1,513 cases and 66,236 controls from individuals of European descent. To enable acquisition of large samples, some of the participating groups ascertained cases via clinician diagnosis rather than a research-based assessment. Genetic principal components and study fixed effects were included as covariates.

The Add Health schizophrenia are standardized (mean = 0, standard deviation = 1) within ancestry groups to control for population stratification. To give researchers an idea of the underlying raw PGS, we present density plots for the unstandardized PGSs below. **Please read the section entitled “Using Add Health PGSs” in the introductory portion of this document prior to conducting any analyses using the provided PGSs.**

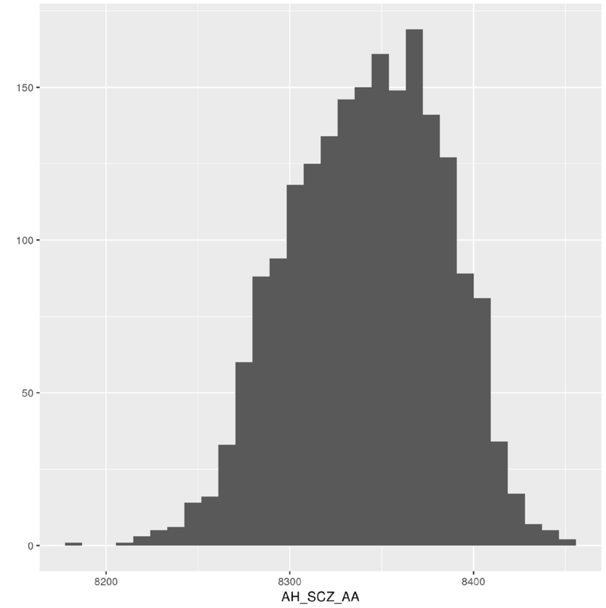
Schizophrenia


European ancestry raw PGS distribution

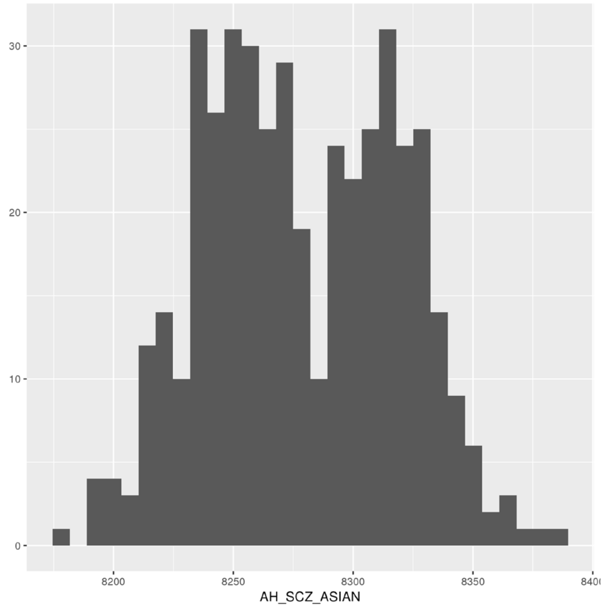
Hispanic ancestry raw PGS distribution

Add Health is a program project directed by Kathleen Mullan Harris and designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris at the University of North Carolina at Chapel Hill, and funded by grant P01-HD31921 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development, with cooperative funding from 23 other federal agencies and foundations. Information on how to obtain the Add Health data files is available on the Add Health website (http://www.cpc.unc.edu/addhealth).

African ancestry raw PGS distribution



Asian ancestry raw PGS distribution



Major Depressive Disorder (2013)

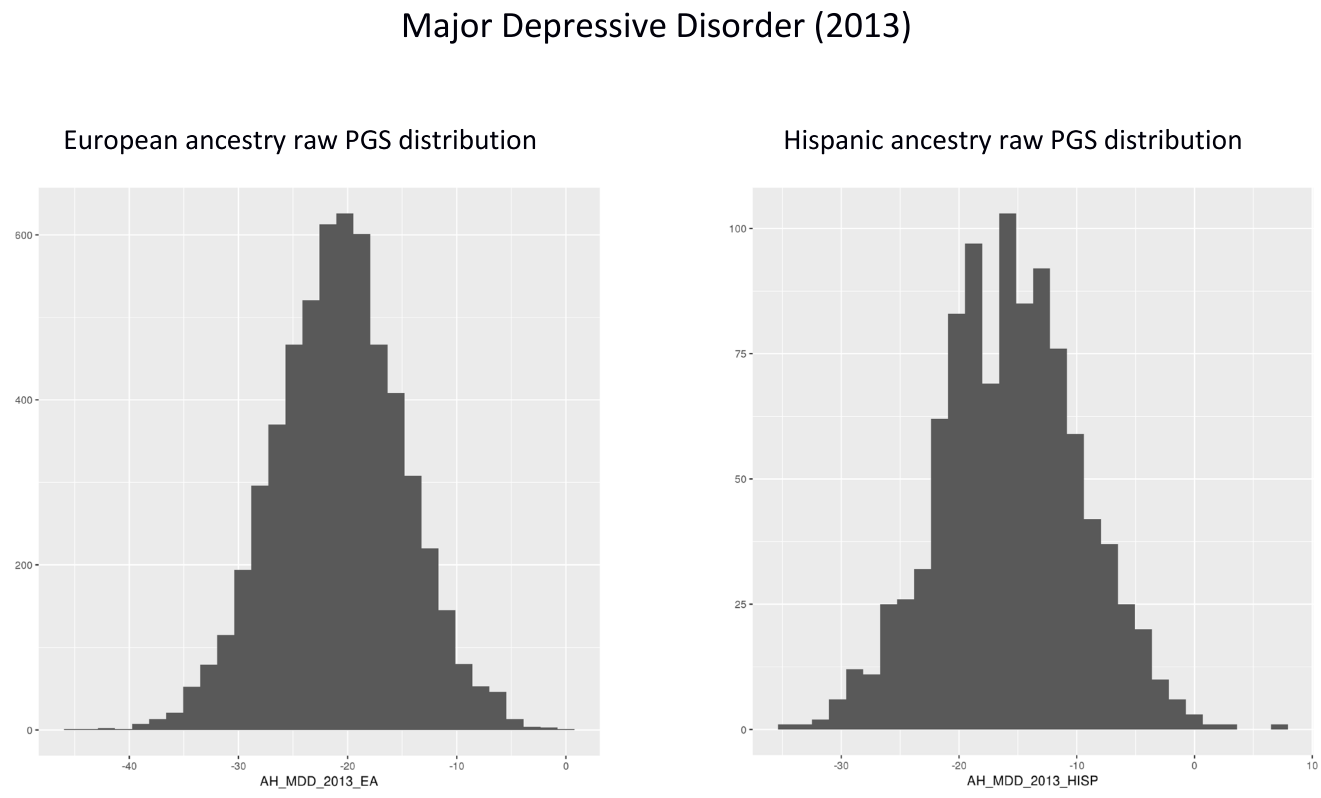
GWAS Summary Statistic Source: Ripke, Stephan et al. 2013. “A Mega-Analysis of Genome-Wide Association Studies for Major Depressive Disorder.” Molecular Psychiatry 18(4):497–511.

GWAS Ancestry Group(s): European

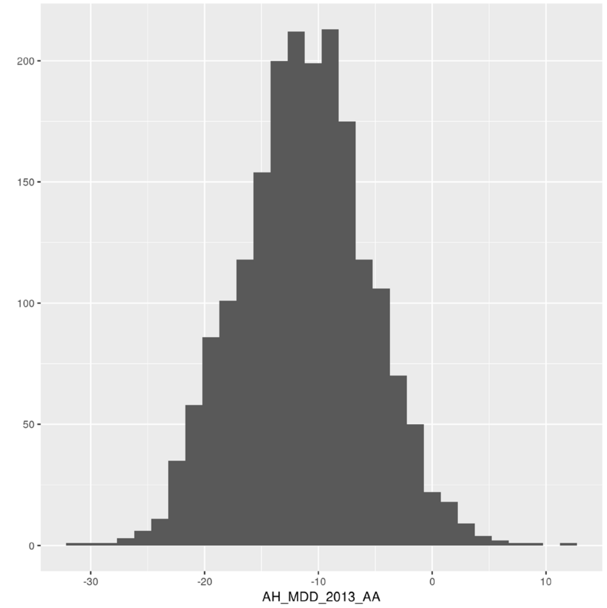
The PGSs for major depressive disorder (MDD) were created using results from a 2013 GWAS conducted by the MDD Working Group of the Psychiatric GWAS Consortium. The GWAS meta-analysis files are publicly available from: https://www.med.unc.edu/pgc/results-and-downloads.

The MDD GWAS combined meta-analysis included 9,240 cases and 9,519 controls (N=18,759). The replication sample consisted of 6,783 cases and 50,695 controls. Study level fixed effects and the first five genetic principal components were included as covariates. Analyses were also stratified by sex, recurrent MDD, recurrent early-onset MDD, age of onset, pre-pubertal onset MDD, and latent classes of MDD criteria.

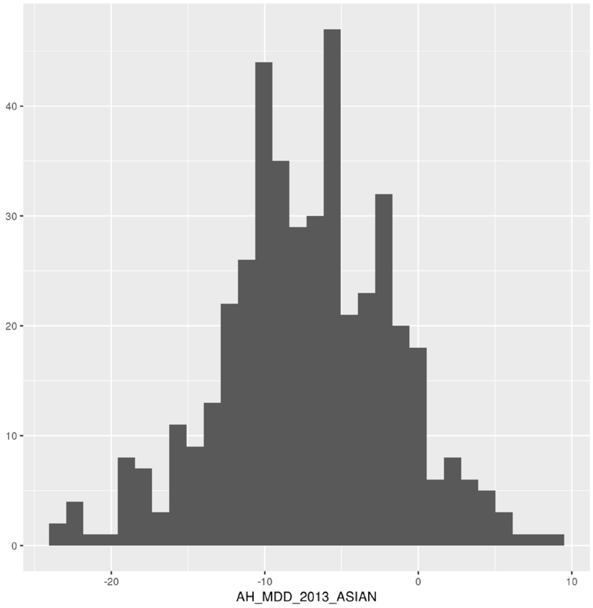
The Add Health MDD (2013) are standardized (mean = 0, standard deviation = 1) within ancestry groups to control for population stratification. To give researchers an idea of the underlying raw PGS, we present density plots for the unstandardized PGSs below. **Please read the section entitled “Using Add Health PGSs” in the introductory portion of this document prior to conducting any analyses using the provided PGSs.**



African ancestry raw PGS distribution



Asian ancestry raw PGS distribution



Major Depressive Disorder (2018)

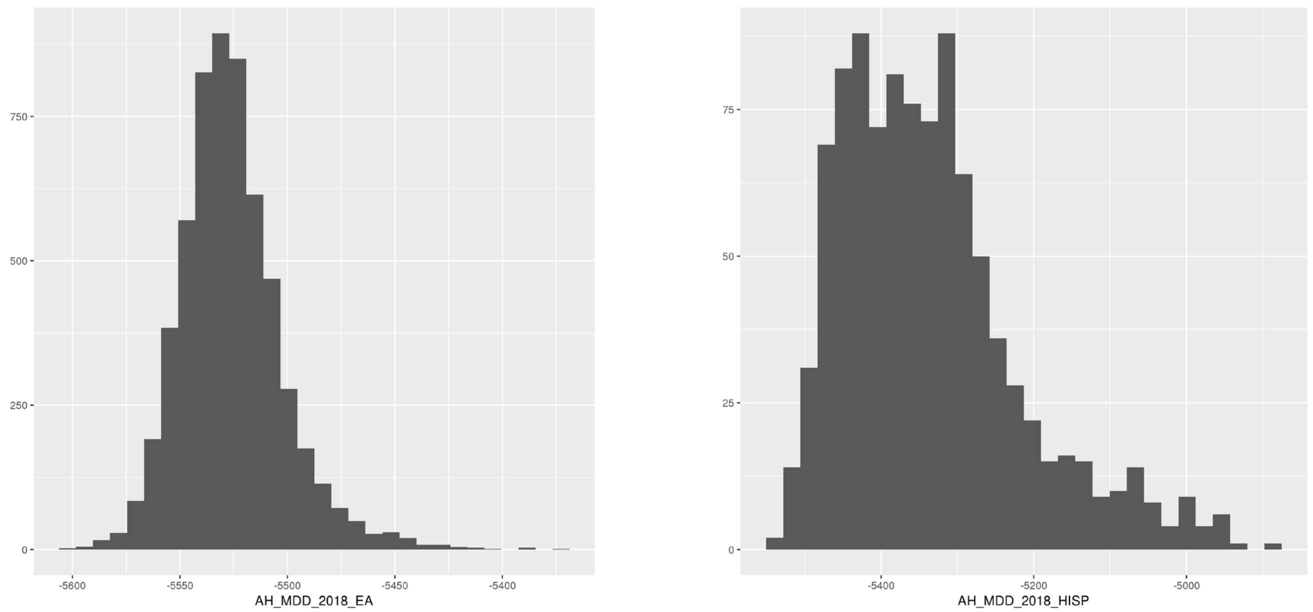
GWAS Summary Statistic Source: Wray, Naomi R. et al. 2018. “Genome-Wide Association Analyses Identify 44 Risk Variants and Refine the Genetic Architecture of Major Depression.” Nature Genetics 50(5):668–81.

GWAS Ancestry Group(s): European

The PGSs for major depressive disorder (MDD) were created using results from a 2018 GWAS conducted by the MDD Working Group of the Psychiatric GWAS Consortium. The GWAS meta-analysis files are publicly available from: https://www.med.unc.edu/pgc/results-and-downloads.

The MDD 2018 GWAS combined meta-analysis included 135,458 cases and 344,901 controls (N = 480,359) from 35 studies of individuals of European descent. Due to data sharing restrictions from 23andMe, the PGSs are based on summary statistics that exclude the 23andMe cohort (Ncase = 75,607; Ncontrols = 231,747). MDD is cases are based on meeting standard criteria for MDD.

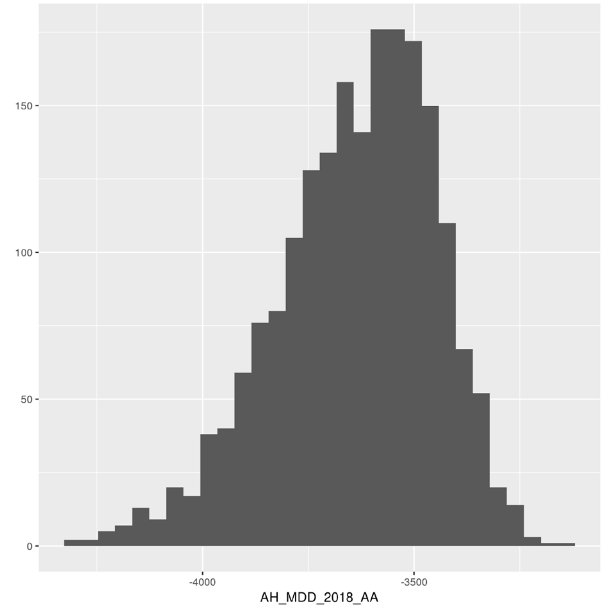
The Add Health PGSs for MDD (2018) are standardized (mean = 0, standard deviation = 1) within ancestry groups to control for population stratification. To give researchers an idea of the underlying raw PGS, we present density plots for the unstandardized PGSs below. **Please read the section entitled “Using Add Health PGSs” in the introductory portion of this document prior to conducting any analyses using the provided PGSs.**


Major Depressive Disorder 2018

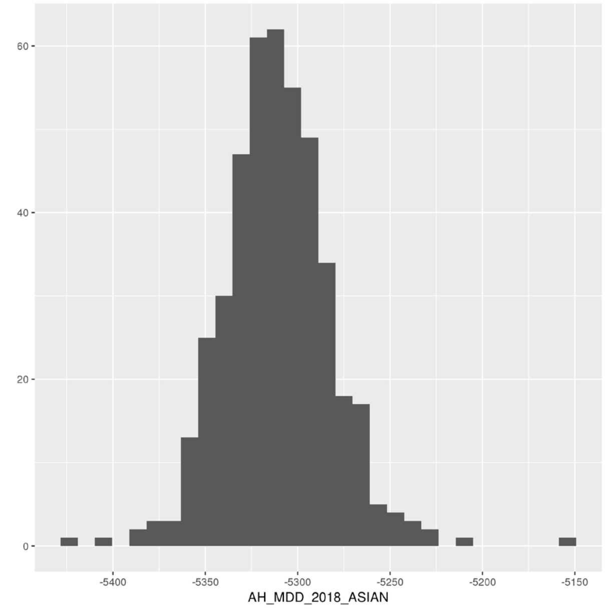
European ancestry raw PGS distribution

Hispanic ancestry raw PGS distribution

African ancestry raw PGS distribution



Asian ancestry raw PGS distribution



Major Depressive Disorder (2019)

GWAS Summary Statistic Source: Howard et al. 2019. “Genome-Wide Meta-Analysis of Depression Identifies 102 Independent Variants and Highlights the Importance of the Prefrontal Brain Regions.” *Nature Neuroscience* 22(3):343.

GWAS Ancestry Group(s): European

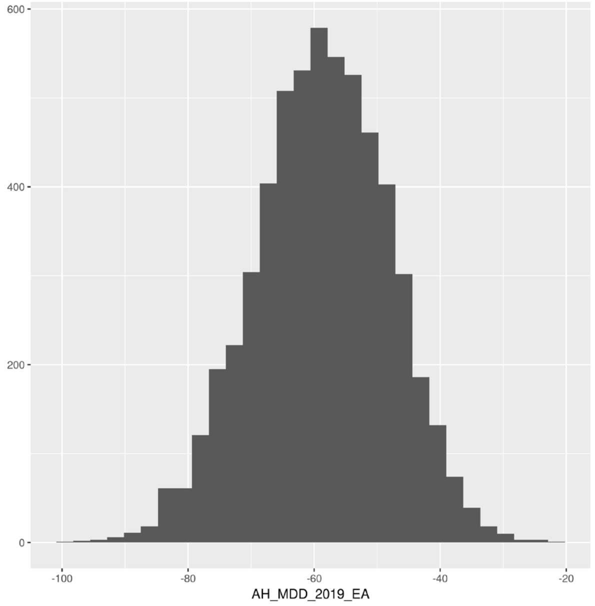
The PGSs for major depressive disorder (MDD) were created using results from a 2019 study conducted by David M. Howard, members of the MDD Working Group of the Psychiatric Genomics Consortium, and other colleagues. The GWAS meta-analysis files are publicly available and can be downloaded from: https://www.med.unc.edu/pgc/results-and-downloads.

The depression GWAS meta-analyses are based on the analysis of 807,553 individuals of European genetic ancestry. However, due to data use restrictions imposed by 23andMe, the summary statistics used to create the Add Health PGSs exclude the 23andMe sample resulting in a sample size of 500,199 (170,756 cases and 329,443 controls). Cases are defined as ever being diagnosed with MDD or “broad depression” based on self-reported help-seeking for problems with nerves, anxiety, tension, or depression. GWAS analyses adjusted for different covariates in the various cohorts, but generally included sex, age, genetic principal components, and genotyping array (in the UK Biobank). See the methods section of Howard et al. (2019) for more details.

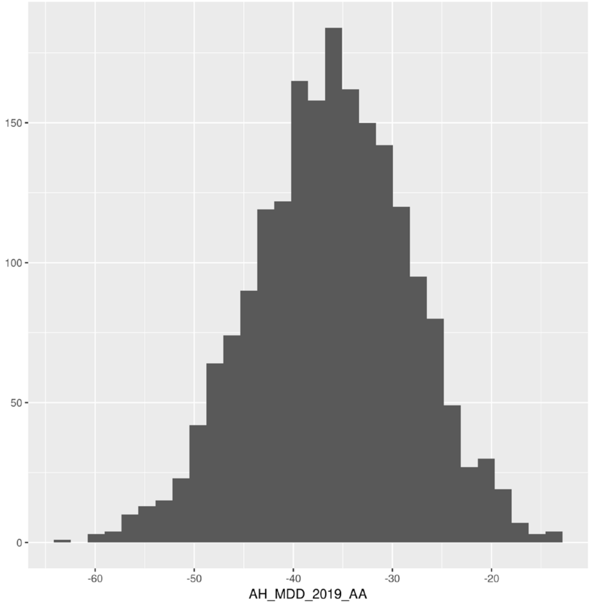
The Add Health PGSs for MDD (2019) are standardized (mean = 0, standard deviation = 1) within ancestry groups to control for population stratification. To give researchers an idea of the underlying raw PGS, we present density plots for the unstandardized PGSs below. **Please read the section entitled “Using Add Health PGSs” in the introductory portion of this document prior to conducting any analyses using the provided PGSs.**

Major Depressive Disorder 2019

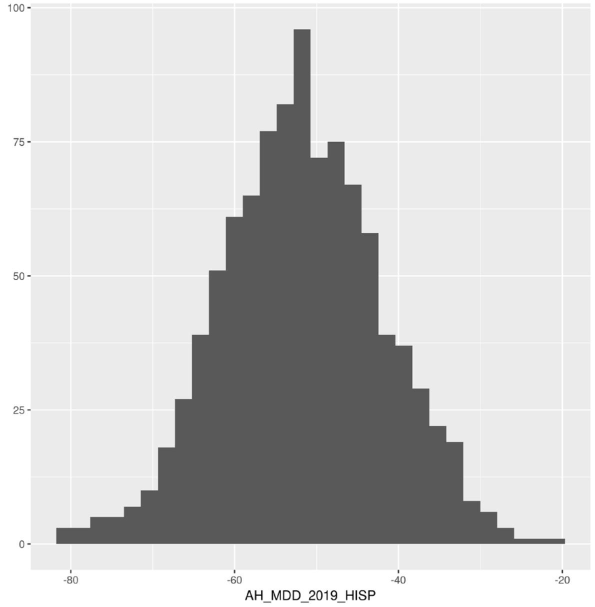
European ancestry raw PGS distribution



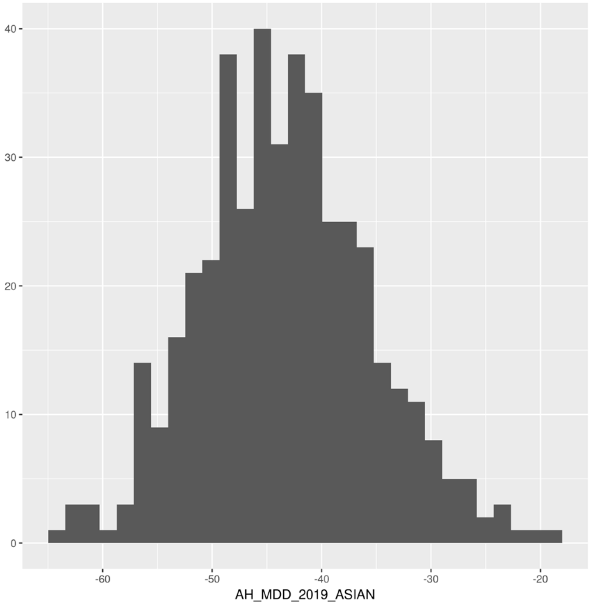
African ancestry raw PGS distribution



Hispanic ancestry raw PGS distribution



Asian ancestry raw PGS distribution



Depressive symptoms (2018)

GWAS Summary Statistic Source: Nagel et al. 2018. “Meta-Analysis of Genome-Wide Association Studies for Neuroticism in 449,484 Individuals Identifies Novel Genetic Loci and Pathways.” *Nature Genetics* 50(7):920–27.

GWAS Ancestry Group(s): European

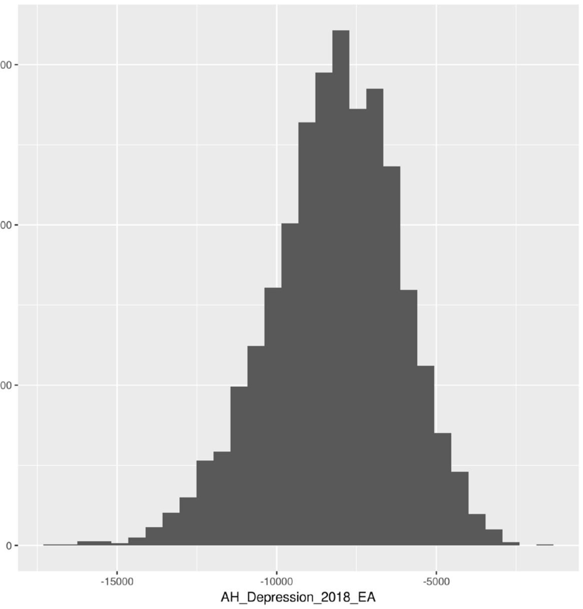
The PGSs for depressive symptoms were created using results from a 2018 study conducted by the Mats Nagel and colleagues. The GWAS meta-analysis files are publicly available and can be downloaded from: https://ctg.cncr.nl/software/summary_statistics.

The depressive symptoms GWAS is based on the analysis of 688,809 individuals of European genetic ancestry from the UK Biobank, 23adnMe, and the Psychiatric Genetics Consortium (PGC). However, due to data use restrictions imposed by 23andMe, the summary statistics used to create the Add Health PGSs are based solely on a GWAS of the UK Biobank and GPC (N = 381,455). In the UK Biobank, depressive symptoms was measured as the sum of two 4-item Likert scales based on the following questions: 1) “Over the past two weeks, how often have you felt down, depressed or hopeless?” and 2) “Over the past two weeks, how often have you had little interest or pleasure in doing things?” In the PGC, depressive symptoms are measured as case/controls of diagnoses of major depressive disorder. GWAS analyses adjusted for different sets of covariates in the various cohorts, but generally included age, sex, and principal components of the genetic data. More details are available in the Supplementary Information provided by Nagel et al. 2018.

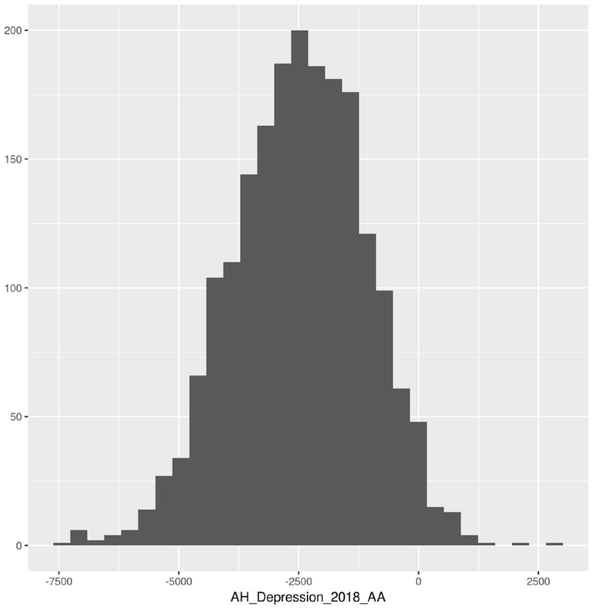
The Add Health PGSs for depressive symptoms are standardized (mean = 0, standard deviation = 1) within ancestry groups to control for population stratification. To give researchers an idea of the underlying raw PGS, we present density plots for the unstandardized PGSs below. **Please read the section entitled “Using Add Health PGSs” in the introductory portion of this document prior to conducting any analyses using the provided PGSs.**

Depressive Symptoms 2018

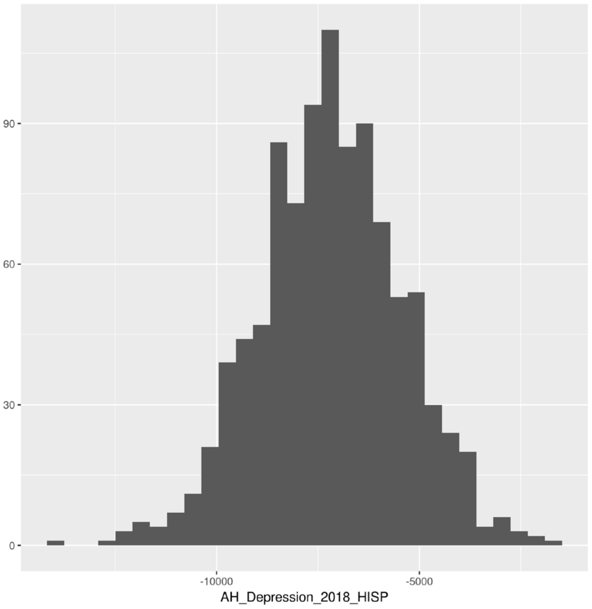
European ancestry raw PGS distribution



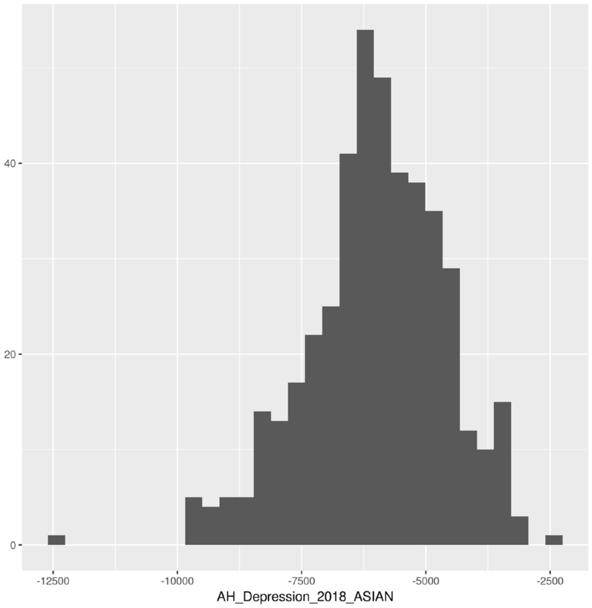
African ancestry raw PGS distribution



Hispanic ancestry raw PGS distribution



Asian ancestry raw PGS distribution



Add Health is a program project directed by Kathleen Mullan Harris and designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris at the University of North Carolina at Chapel Hill, and funded by grant P01-HD31921 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development, with cooperative funding from 23 other federal agencies and foundations. Information on how to obtain the Add Health data files is available on the Add Health website (http://www.cpc.unc.edu/addhealth).

Depressive Symptoms (2019)

GWAS Summary Statistic Source: Baselmans et al. 2019. “Multivariate Genome-Wide Analyses of the Well-Being Spectrum.” *Nature Genetics* 51: 445-451.

GWAS Ancestry Group(s): European

The PGSs for depressive symptoms were created using results from a 2019 study conducted by the Bart Baselmans and colleagues including members of the BIOS Consortium and the Social Science Genetic Association Consortium. The GWAS meta-analysis files are publicly available and can be downloaded from: https://surfdrive.surf.nl/files/index.php/s/Ow1qCDpFT421ZOO.

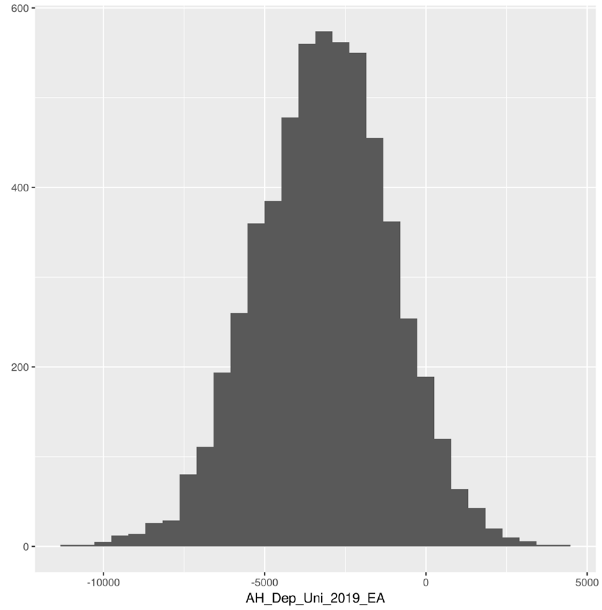
There are two PGSs for depressive symptoms from the Baselmans et al. (2019) study, one based on summary statistics from a traditional univariate GWAS meta-analyses and a second based on summary statistics from a novel statistical approach to genome-wide studies: model-averaging multivariate genome-wide association meta-analysis (MA-GWAMA). Details on the MA-GWAMA method can be found in Baselmans et al. (2019). Results from Baselmans and colleagues (2019) suggest that PGSs based on the MA-GWAMA summary statistics are up to approximately 50% more predictive than PGSs based on the traditional univariate GWAS meta-analysis methods.

The depressive symptoms GWAS meta-analyses and MA-GWAMA are based on the analysis of 1,295,946 individuals of European genetic ancestry. However, due to data use restrictions imposed by 23andMe, the summary statistics used to create the Add Health PGSs exclude the 23andMe sample resulting in a sample size of 1,067,913. Details on the contributing cohorts can be found in Supplementary Figure 2 of the online supplement accompanying the original article.

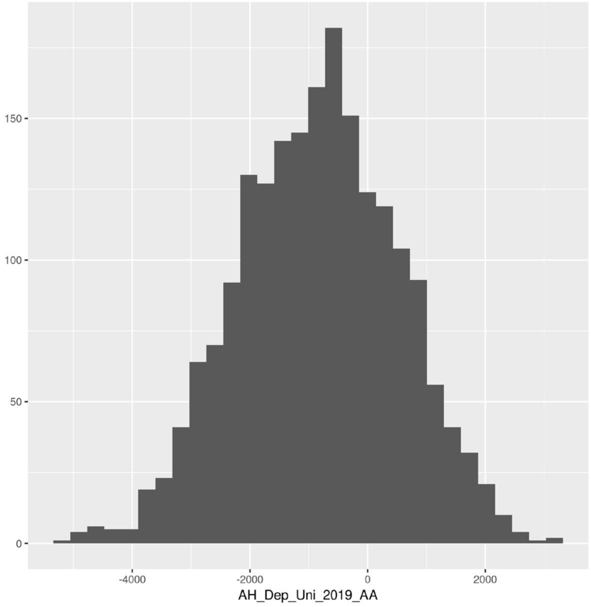
The Add Health PGSs for depressive symptoms are standardized (mean = 0, standard deviation = 1) within ancestry groups to control for population stratification. To give researchers an idea of the underlying raw PGS, we present density plots for the unstandardized PGSs below. **Please read the section entitled “Using Add Health PGSs” in the introductory portion of this document prior to conducting any analyses using the provided PGSs.**

Depressive Symptoms 2019 – Univariate GWAS

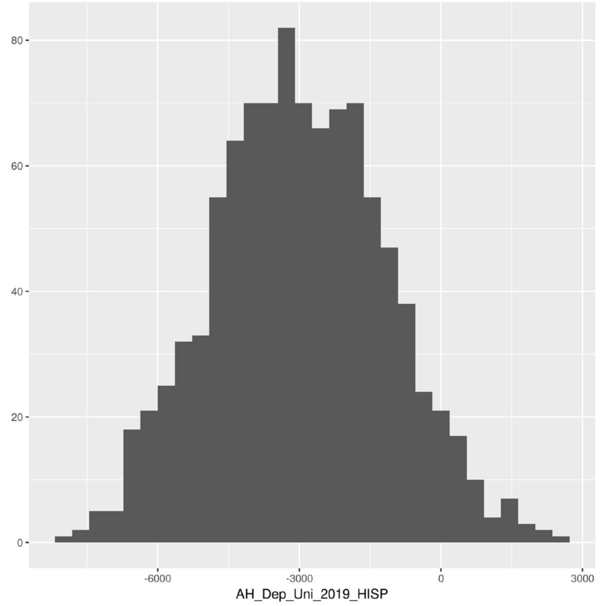
European ancestry raw PGS distribution



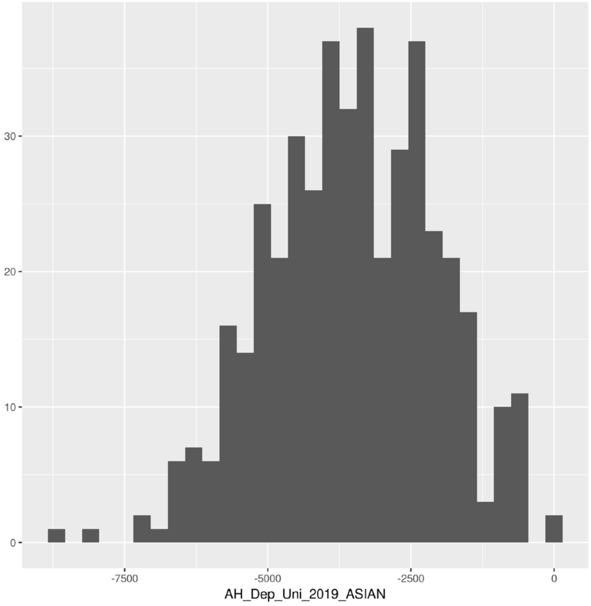
African ancestry raw PGS distribution



Hispanic ancestry raw PGS distribution

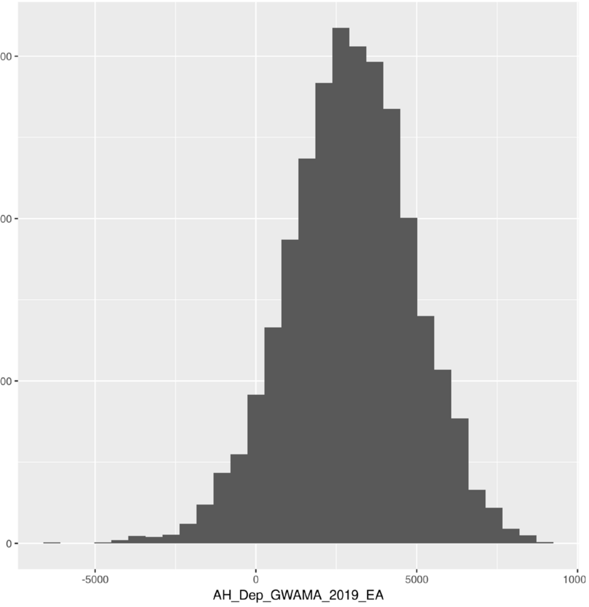


Asian ancestry raw PGS distribution

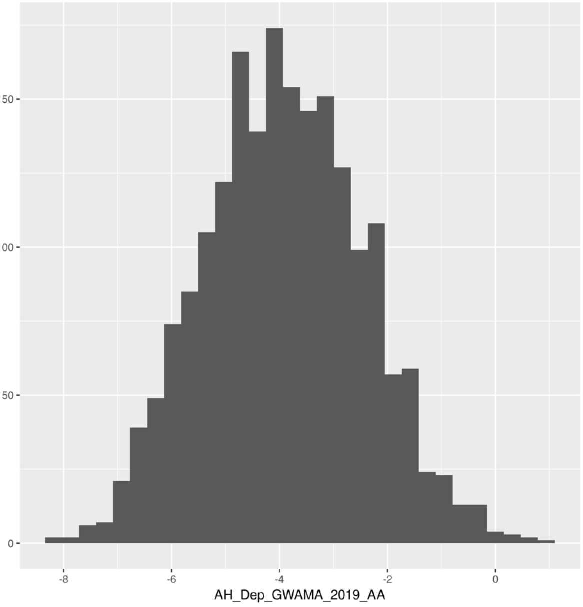


Depressive Symptoms 2019 – MA-GWAMA

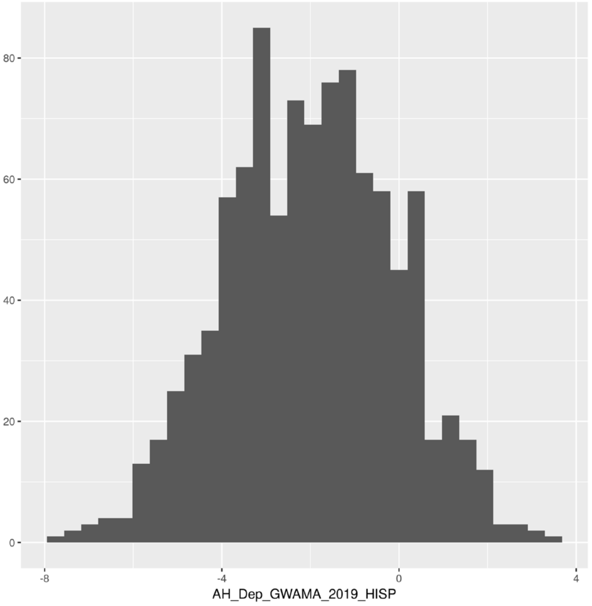
European ancestry raw PGS distribution



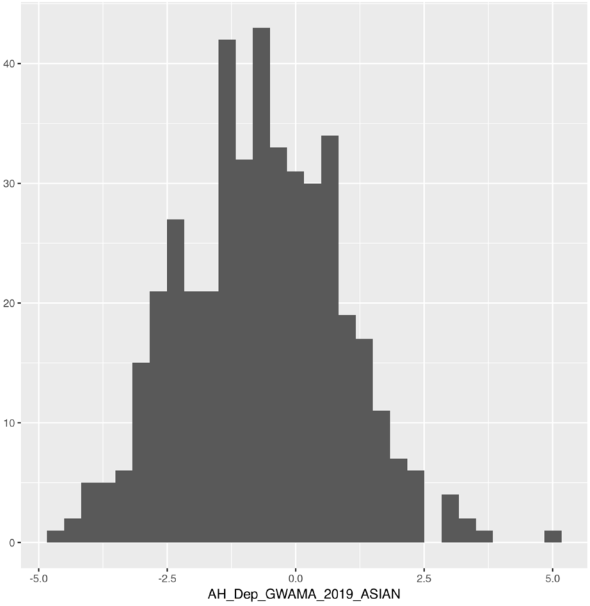
African ancestry raw PGS distribution



Hispanic ancestry raw PGS distribution



Asian ancestry raw PGS distribution



Add Health is a program project directed by Kathleen Mullan Harris and designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris at the University of North Carolina at Chapel Hill, and funded by grant P01-HD31921 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development, with cooperative funding from 23 other federal agencies and foundations. Information on how to obtain the Add Health data files is available on the Add Health website (http://www.cpc.unc.edu/addhealth).

Depressive affect

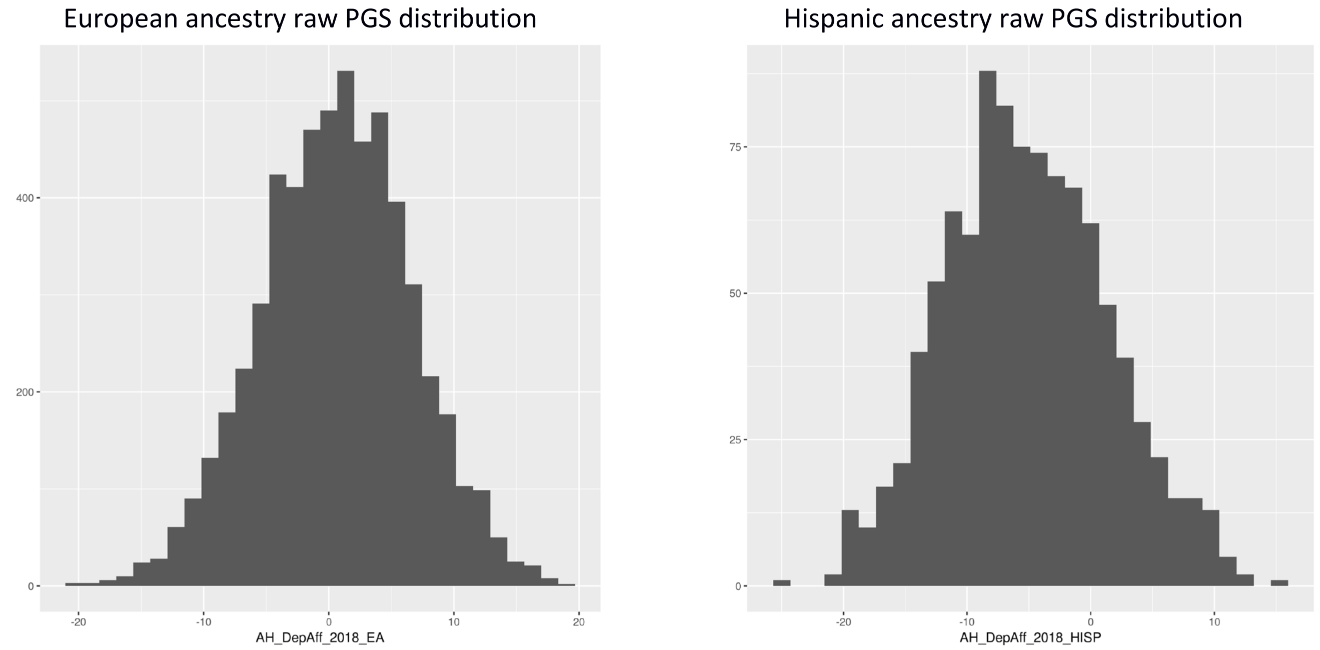
GWAS Summary Statistic Source: Nagel et al. 2018. “Meta-Analysis of Genome-Wide Association Studies for Neuroticism in 449,484 Individuals Identifies Novel Genetic Loci and Pathways.” *Nature Genetics* 50(7):920–27.

GWAS Ancestry Group(s): European

The PGSs for depressive affect were created using results from a 2018 study conducted by the Mats Nagel and colleagues. The GWAS meta-analysis files are publicly available and can be downloaded from: https://ctg.cncr.nl/software/summary_statistics.

The depressive affect GWAS is based solely on the analysis of 357,957 individuals of European genetic ancestry from the UK Biobank. Depressive affect was measured as the sum of four questions: (1) Does your mood often go up and down?; (2) Do you ever feel ‘just miserable’ for no reason?; (3) Do you often feel ‘fed-up’?; and (4) Do you often feel lonely? GWAS analyses adjusted for age, sex, the Townsend deprivation index, genotyping array, and the first 10 genetic principal components (see the Supplementary Information provided by Nagel et al. 2018 for more details).

The Add Health PGSs for depressive affect are standardized (mean = 0, standard deviation = 1) within ancestry groups to control for population stratification. To give researchers an idea of the underlying raw PGS, we present density plots for the unstandardized PGSs below. **Please read the section entitled “Using Add Health PGSs” in the introductory portion of this document prior to conducting any analyses using the provided PGSs.**

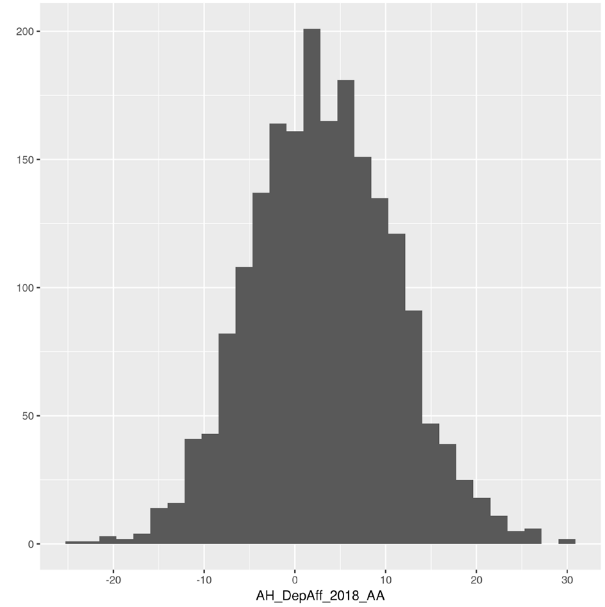


Depressive Affect

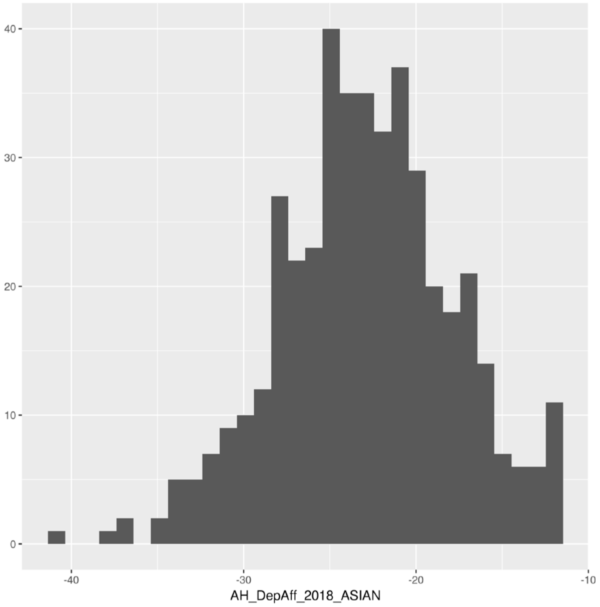
European ancestry raw PGS distribution

Hispanic ancestry raw PGS distribution

African ancestry raw PGS distribution



Asian ancestry raw PGS distribution



Worry

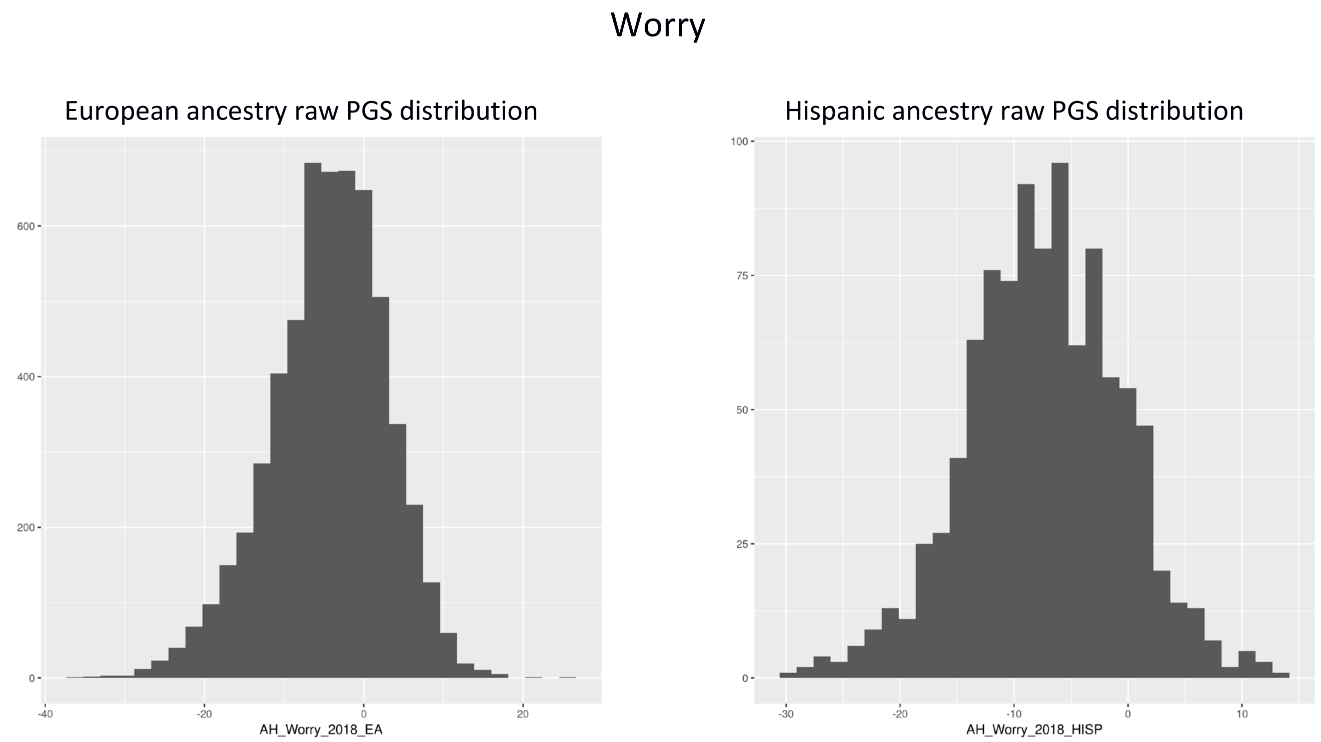
GWAS Summary Statistic Source: Nagel et al. 2018. “Meta-Analysis of Genome-Wide Association Studies for Neuroticism in 449,484 Individuals Identifies Novel Genetic Loci and Pathways.” *Nature Genetics* 50(7):920–27.

GWAS Ancestry Group(s): European

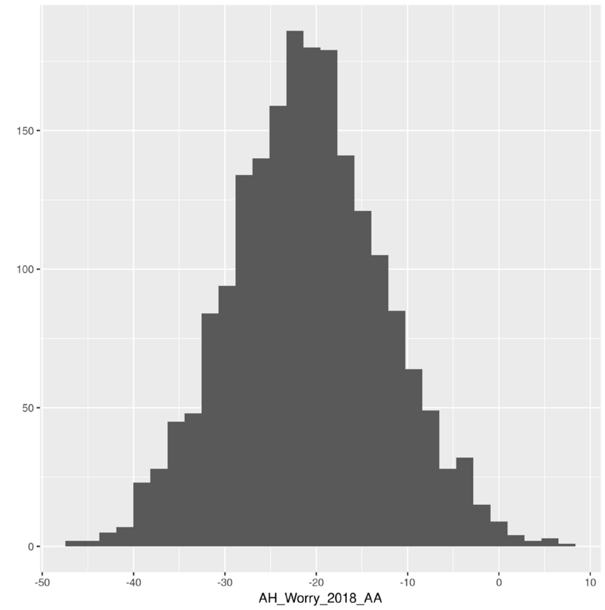
The PGSs for worry were created using results from a 2018 study conducted by the Mats Nagel and colleagues. The GWAS meta-analysis files are publicly available and can be downloaded from: https://ctg.cncr.nl/software/summary_statistics.

The worry GWAS is based solely on the analysis of 348,219 individuals of European genetic ancestry from the UK Biobank. Worry was measured as the sum of three questions: (1) Would you call yourself a nervous person?; (2) Are you a worrier?; and (3) Would you call yourself tense or ‘highly strung’?. GWAS analyses adjusted for age, sex, the Townsend deprivation index, genotyping array, and the first 10 genetic principal components (see the Supplementary Information provided by Nagel et al. 2018 for more details).

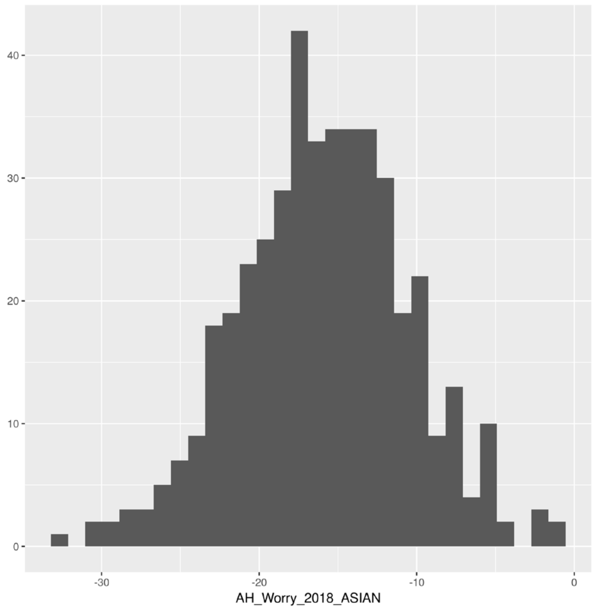
The Add Health PGSs for worry are standardized (mean = 0, standard deviation = 1) within ancestry groups to control for population stratification. To give researchers an idea of the underlying raw PGS, we present density plots for the unstandardized PGSs below. **Please read the section entitled “Using Add Health PGSs” in the introductory portion of this document prior to conducting any analyses using the provided PGSs.**



African ancestry raw PGS distribution



Asian ancestry raw PGS distribution



Life Satisfaction

GWAS Summary Statistic Source: Baselmans et al. 2019. “Multivariate Genome-Wide Analyses of the Well-Being Spectrum.” *Nature Genetics* 51: 445-451.

GWAS Ancestry Group(s): European

The PGSs for life satisfaction were created using results from a 2019 study conducted by the Bart Baselmans and colleagues including members of the BIOS Consortium and the Social Science Genetic Association Consortium. The GWAS meta-analysis files are publicly available and can be downloaded from: https://surfdrive.surf.nl/files/index.php/s/Ow1qCDpFT421ZOO.

There are two PGSs for life satisfaction from the Baselmans et al. (2019) study, one based on summary statistics from a traditional univariate GWAS meta-analyses and a second based on summary statistics from a novel statistical approach to genome-wide studies: model-averaging multivariate genome-wide association meta-analysis (MA-GWAMA). Details on the MA-GWAMA method can be found in the original article. Results from Baselmans and colleagues suggest that PGSs based on the MA-GWAMA summary statistics are up to approximately 50% more predictive than PGSs based on the traditional univariate GWAS meta-analysis methods.

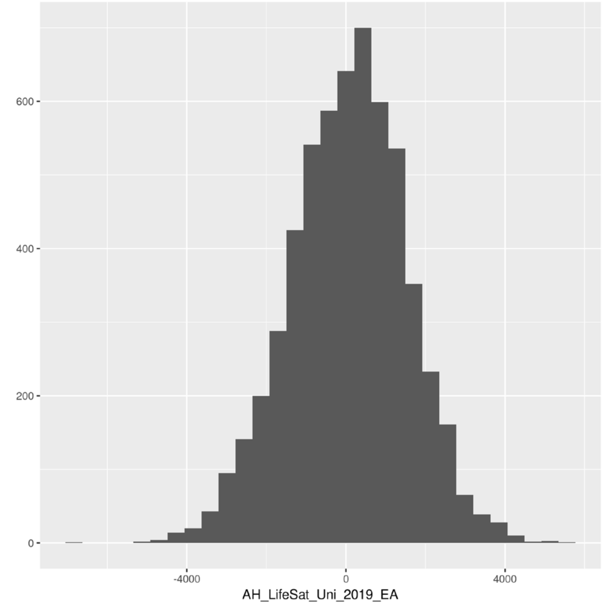
The life satisfaction GWAS meta-analyses and MA-GWAMA are based on the analysis of 80,852 individuals of European genetic ancestry. Details on the contributing cohorts can be found in Supplementary Figure 2 of the online supplement accompanying Baselmans et al. (2019).

The Add Health PGSs for life satisfaction are standardized (mean = 0, standard deviation = 1) within ancestry groups to control for population stratification. To give researchers an idea of the underlying raw PGS, we present density plots for the unstandardized PGSs below. **Please read the section entitled “Using Add Health PGSs” in the introductory portion of this document prior to conducting any analyses using the provided PGSs.**

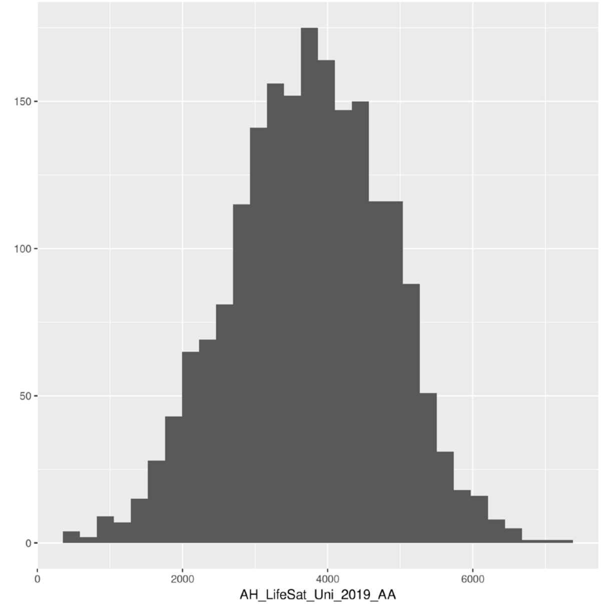
Add Health is a program project directed by Kathleen Mullan Harris and designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris at the University of North Carolina at Chapel Hill, and funded by grant P01-HD31921 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development, with cooperative funding from 23 other federal agencies and foundations. Information on how to obtain the Add Health data files is available on the Add Health website (http://www.cpc.unc.edu/addhealth).

Life Satisfaction 2019 – Univariate GWAS

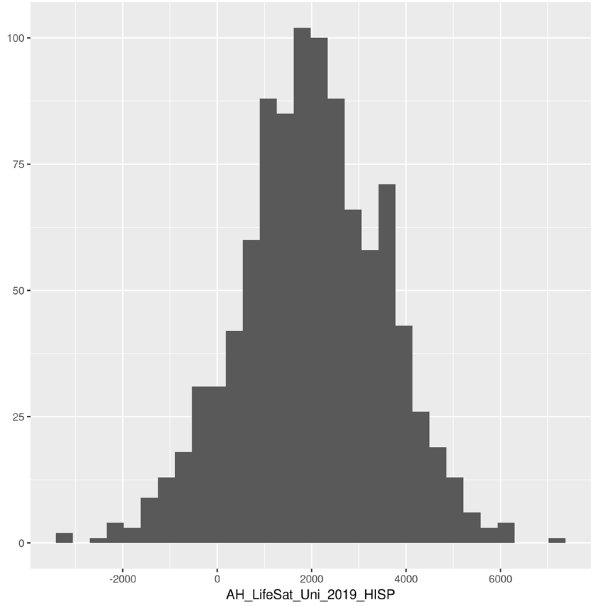
European ancestry raw PGS distribution



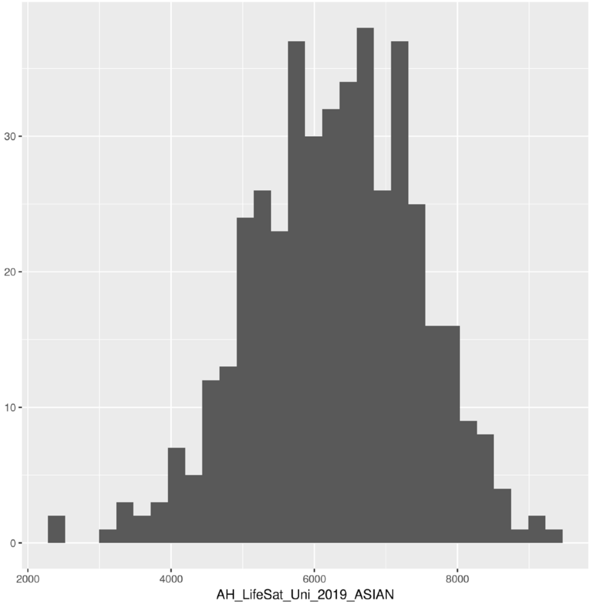
African ancestry raw PGS distribution



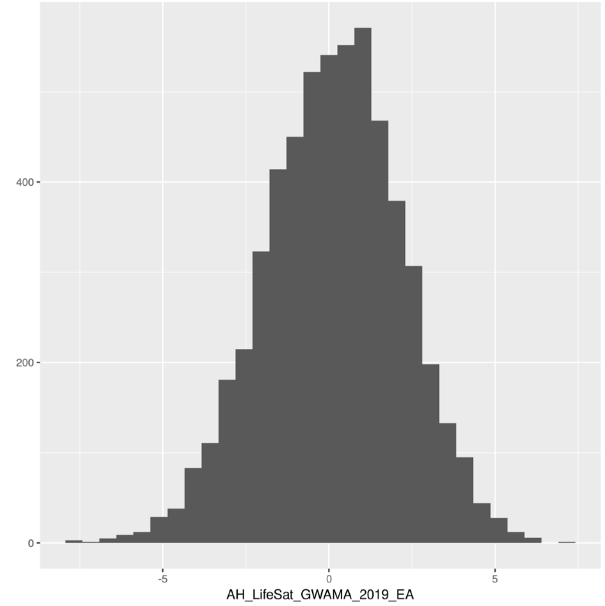
Hispanic ancestry raw PGS distribution



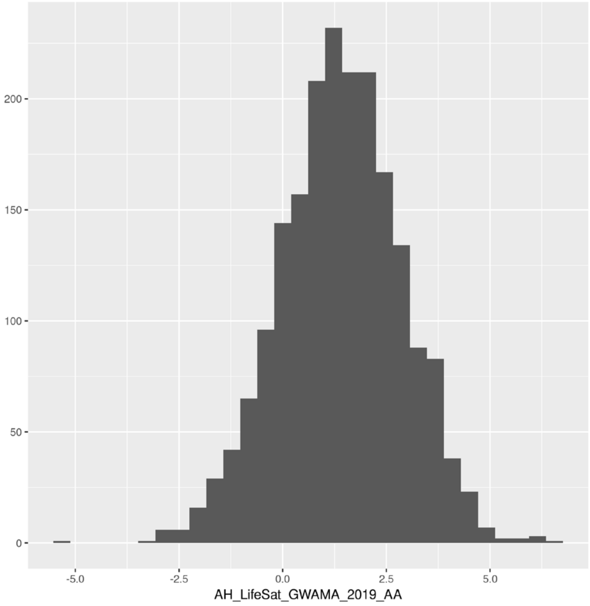
Asian ancestry raw PGS distribution



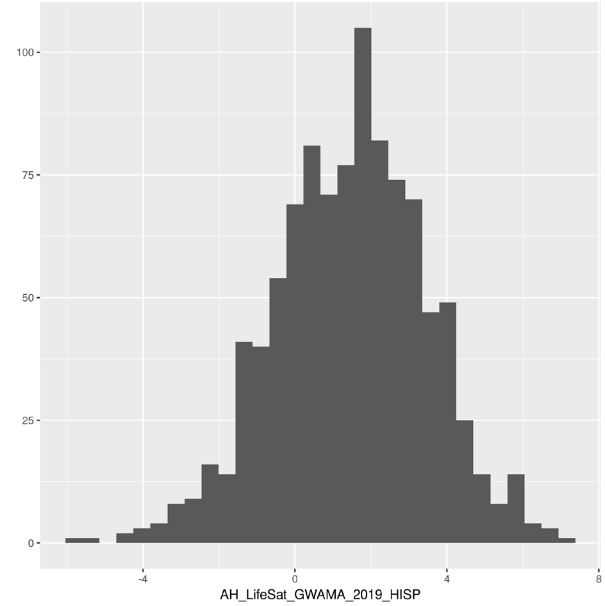
Life Satisfaction 2019 – MA-GWAMA



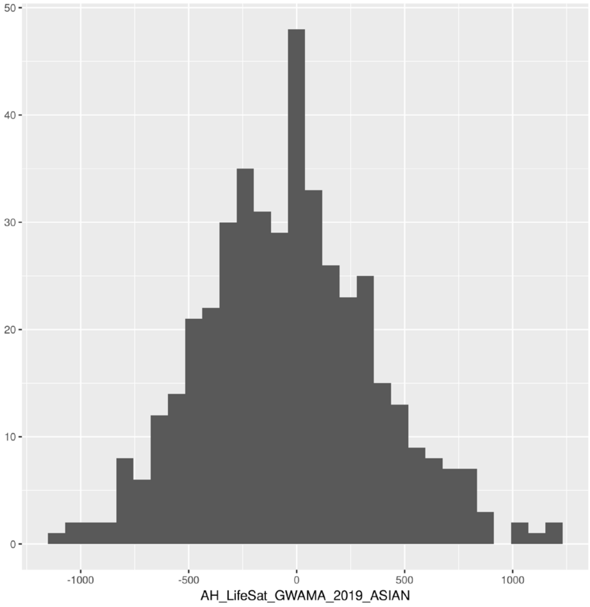
European ancestry raw PGS distribution



African ancestry raw PGS distribution



Hispanic ancestry raw PGS distribution



Asian ancestry raw PGS distribution

Add Health is a program project directed by Kathleen Mullan Harris and designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris at the University of North Carolina at Chapel Hill, and funded by grant P01-HD31921 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development, with cooperative funding from 23 other federal agencies and foundations. Information on how to obtain the Add Health data files is available on the Add Health website (http://www.cpc.unc.edu/addhealth).

Positive Affect

GWAS Summary Statistic Source: Baselmans et al. 2019. “Multivariate Genome-Wide Analyses of the Well-Being Spectrum.” *Nature Genetics* 51: 445-451.

GWAS Ancestry Group(s): European

The PGSs for positive affect were created using results from a 2019 study conducted by the Bart Baselmans and colleagues including members of the BIOS Consortium and the Social Science Genetic Association Consortium. The GWAS meta-analysis files are publicly available and can be downloaded from: https://surfdrive.surf.nl/files/index.php/s/Ow1qCDpFT421ZOO.

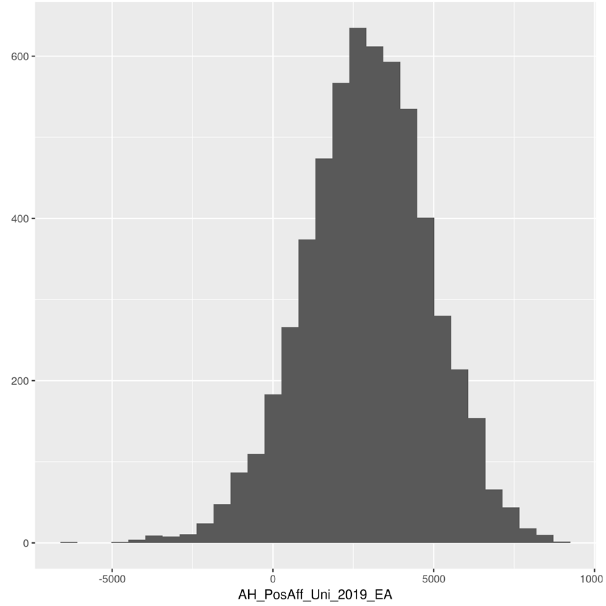
There are two PGSs for positive affect from the Baselmans et al. (2019) study, one based on summary statistics from a traditional univariate GWAS meta-analyses and a second based on summary statistics from a novel statistical approach to genome-wide studies: model-averaging multivariate genome-wide association meta-analysis (MA-GWAMA). Details on the MA-GWAMA method can be found in the original article. Results from Baselmans and colleagues suggest that PGSs based on the MA-GWAMA summary statistics are up to approximately 50% more predictive than PGSs based on the traditional univariate GWAS meta-analysis methods.

The positive affect GWAS meta-analyses and MA-GWAMA are based on the analysis of 410,603 individuals of European genetic ancestry. Details on the contributing cohorts can be found in Supplementary Figure 2 of the online supplement accompanying Baselmans et al. (2019).

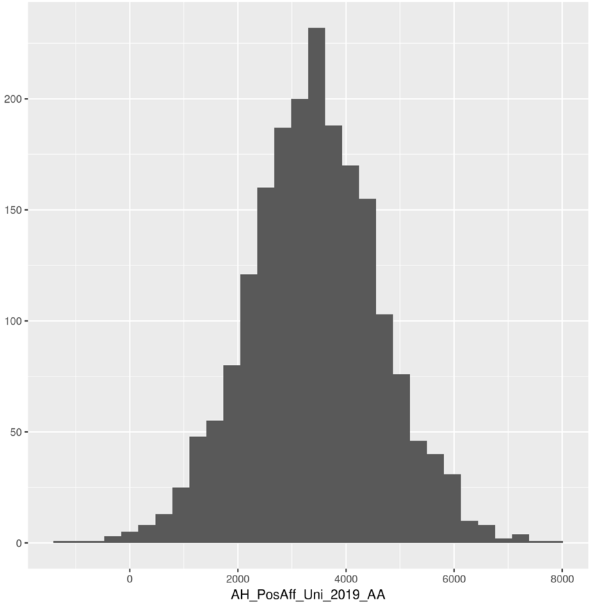
The Add Health PGSs for positive affect are standardized (mean = 0, standard deviation = 1) within ancestry groups to control for population stratification. To give researchers an idea of the underlying raw PGS, we present density plots for the unstandardized PGSs below. **Please read the section entitled “Using Add Health PGSs” in the introductory portion of this document prior to conducting any analyses using the provided PGSs.**

Positive Affect 2019 – Univariate GWAS

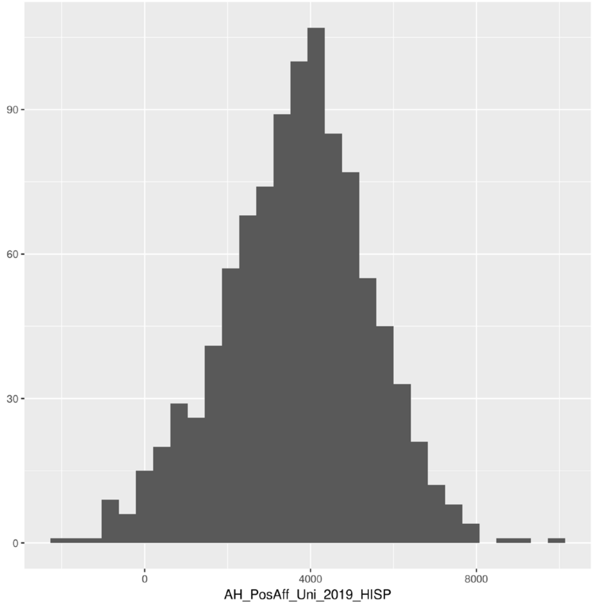
European ancestry raw PGS distribution



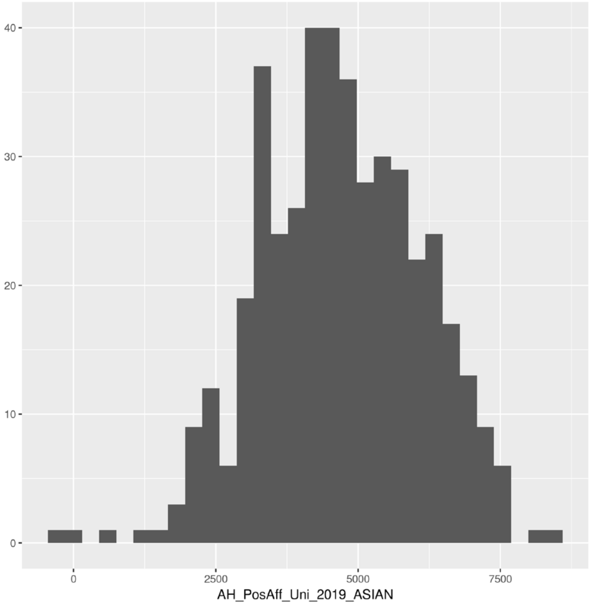
African ancestry raw PGS distribution



Hispanic ancestry raw PGS distribution

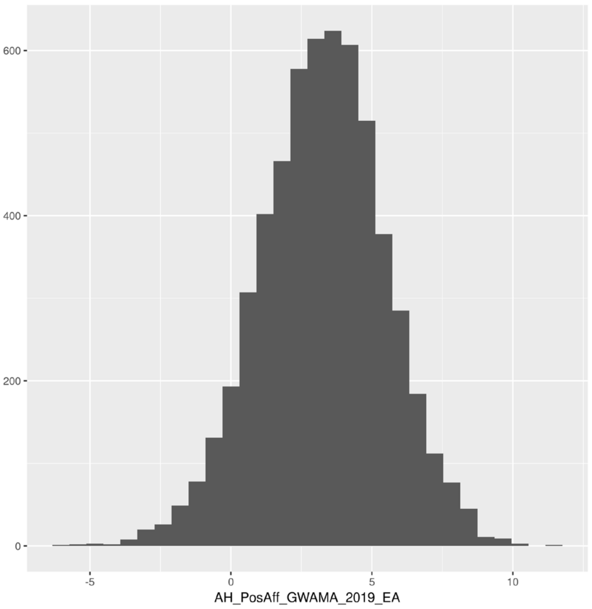


Asian ancestry raw PGS distribution

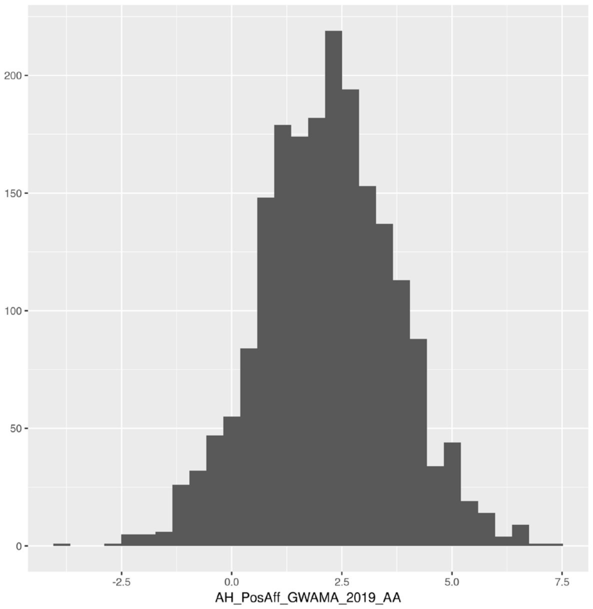


Positive Affect 2019 – MA-GWAMA

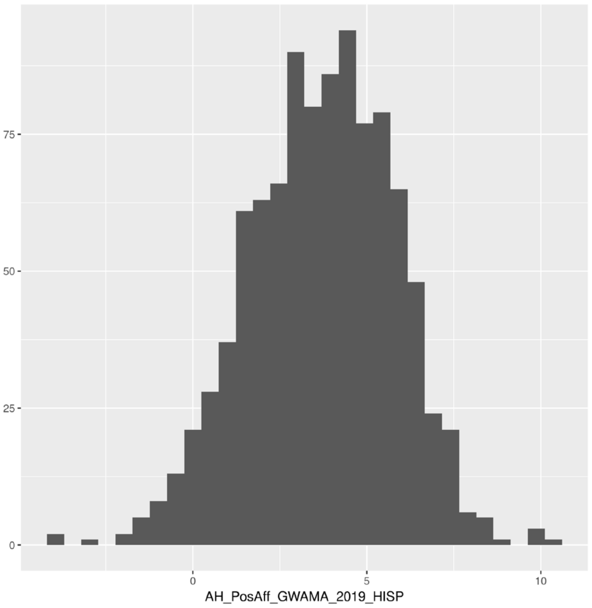
European ancestry raw PGS distribution



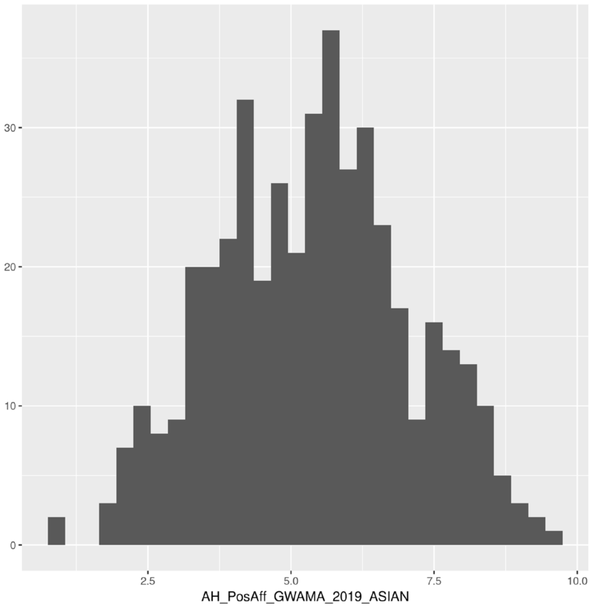
African ancestry raw PGS distribution



Hispanic ancestry raw PGS distribution



Asian ancestry raw PGS distribution



Subjective Wellbeing

GWAS Summary Statistic Source: Baselmans et al. 2019. "Multivariate Genome-Wide Analyses of the Well-Being Spectrum." *Nature Genetics* 51: 445-451.

GWAS Ancestry Group(s): European

The PGSs for subjective wellbeing were created using results from a 2019 study conducted by the Bart Baselmans and colleagues including members of the BIOS Consortium and the Social Science Genetic Association Consortium. The GWAS meta-analysis files are publicly available and can be downloaded from: https://surfdrive.surf.nl/files/index.php/s/Ow1qCDpFT421ZOO.

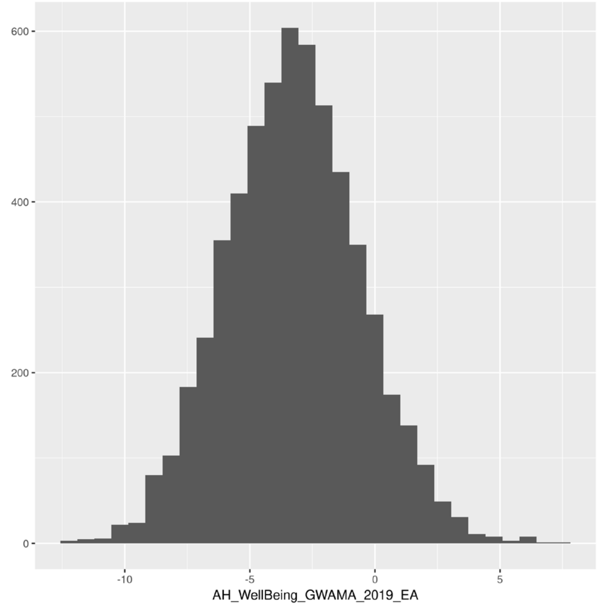
There is only one PGS for subjective wellbeing from the Baselmans et al. (2019) study based on summary statistics from a novel statistical approach to genome-wide studies: N-weighted multivariate genome-wide association meta-analysis (N-GWAMA). Details on the N-GWAMA method can be found in Baselmans et al. (2019). Results from Baselmans and colleagues (2019) suggest that PGSs based on the N -GWAMA summary statistics are up to approximately 50% more predictive than PGSs based on the traditional univariate GWAS meta-analysis methods.

The subjective wellbeing N-GWAMA are based on the analysis of 2,370,390 individuals of European genetic ancestry. Details on the contributing cohorts can be found in Supplementary Figure 2 of the online supplement accompanying Baselmans et al. (2019). However, due to data use restrictions imposed by 23andMe, the summary statistics used to create the Add Health PGSs exclude the 23andMe sample resulting in a sample size of 1,550,297.

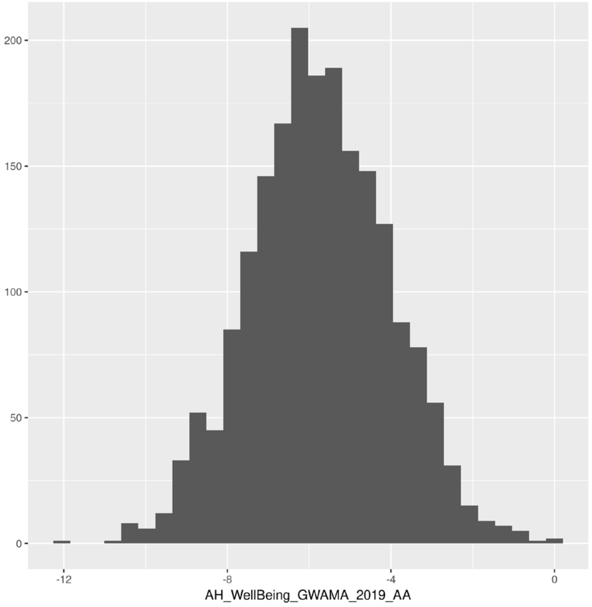
The Add Health PGSs for subjective wellbeing are standardized (mean = 0, standard deviation = 1) within ancestry groups to control for population stratification. To give researchers an idea of the underlying raw PGS, we present density plots for the unstandardized PGSs below. **Please read the section entitled "Using Add Health PGSs" in the introductory portion of this document prior to conducting any analyses using the provided PGSs.**

Subjective Wellbeing – N-GWAMA

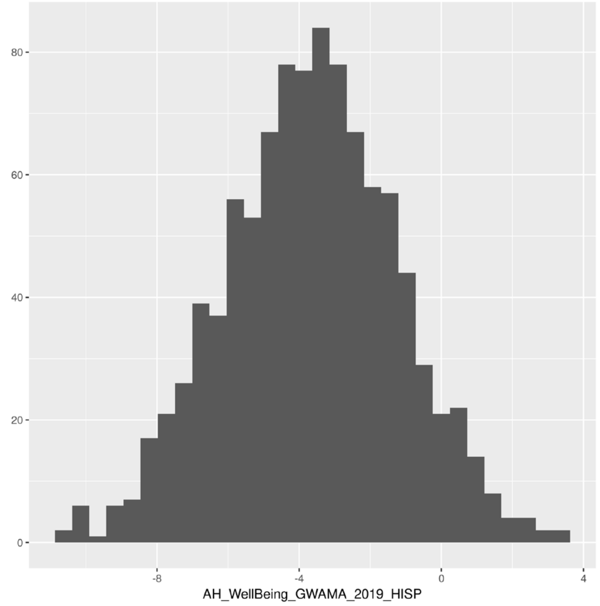
European ancestry raw PGS distribution



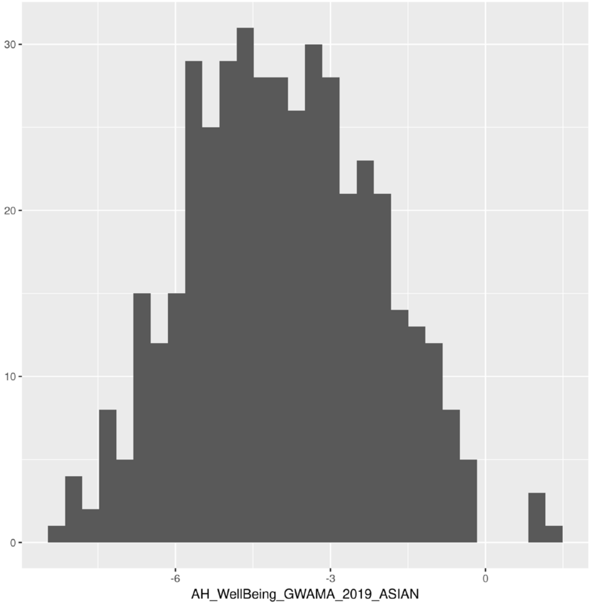
African ancestry raw PGS distribution



Hispanic ancestry raw PGS distribution



Asian ancestry raw PGS distribution



Add Health is a program project directed by Kathleen Mullan Harris and designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris at the University of North Carolina at Chapel Hill, and funded by grant P01-HD31921 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development, with cooperative funding from 23 other federal agencies and foundations. Information on how to obtain the Add Health data files is available on the Add Health website (http://www.cpc.unc.edu/addhealth).

Extraversion

GWAS Summary Statistic Source: van den Berg, Stéphanie M. et al. 2016. "Meta-Analysis of Genome-Wide Association Studies for Extraversion: Findings from the Genetics of Personality Consortium." Behavior Genetics 46(2):170–82.

GWAS Ancestry Group(s): European

The PGSs for extraversion were created using results from a 2016 study by the Genetics of Personality Consortium (GPC). Summary statistics are publicly available from: www.tweelingenregister.org/GPC/.

The discovery meta-analysis included 63,030 individuals of European descent from 29 cohorts. The replication analysis consisted of 9,783 individuals, also of European descent, from a single cohort. All cohorts imputed SNPs based on the 1000 genomes reference panel (1000G). Additional study-specific covariates included sex, age, and ancestry principal components of the genetic data. Extraversion was measured as a continuous latent variable as described in van den Berg et al. (2014).

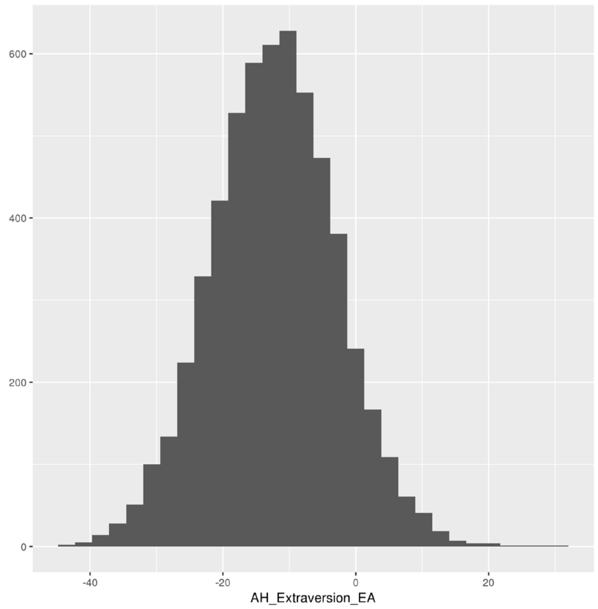
The Add Health PGSs for extraversion are standardized (mean = 0, standard deviation = 1) within ancestry groups to control for population stratification. To give researchers an idea of the underlying raw PGS, we present density plots for the unstandardized PGSs below. **Please read the section entitled "Using Add Health PGSs" in the introductory portion of this document prior to conducting any analyses using the provided PGSs.**

References:

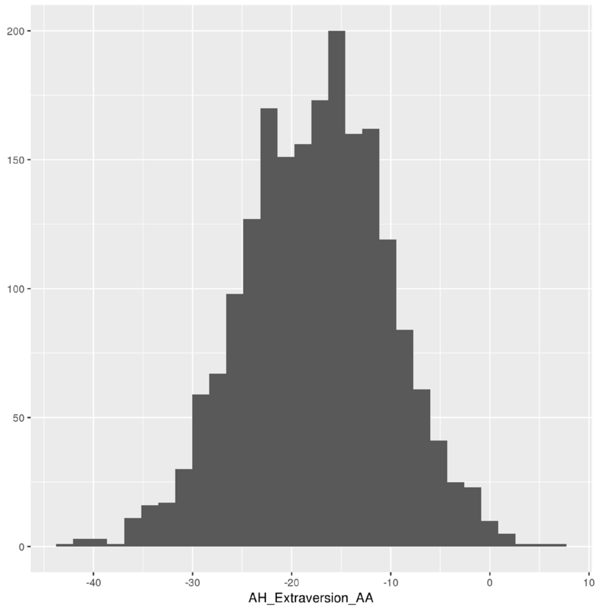
van den Berg, Stéphanie M. et al. 2014. "Harmonization of Neuroticism and Extraversion Phenotypes across Inventories and Cohorts in the Genetics of Personality Consortium: An Application of Item Response Theory." Behavior Genetics 44(4):295–313.

Extraversion

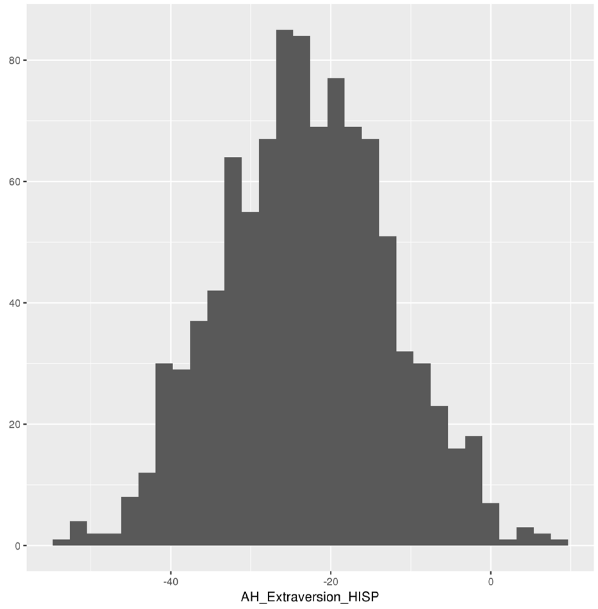
European ancestry raw PGS distribution



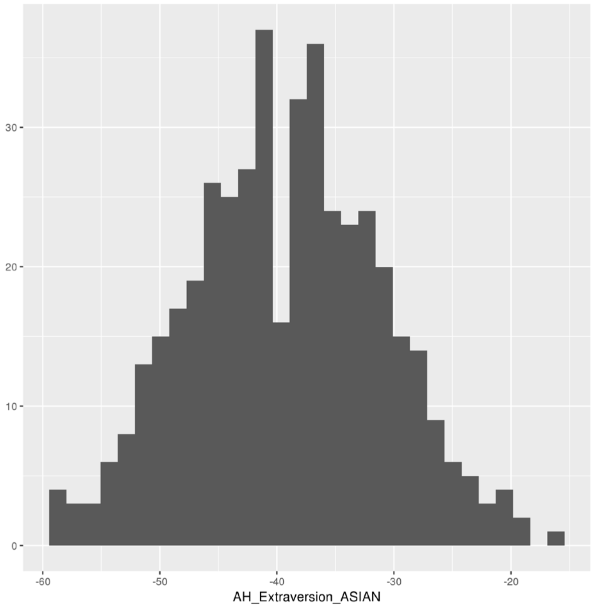
African ancestry raw PGS distribution



Hispanic ancestry raw PGS distribution



Asian ancestry raw PGS distribution



Add Health is a program project directed by Kathleen Mullan Harris and designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris at the University of North Carolina at Chapel Hill, and funded by grant P01-HD31921 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development, with cooperative funding from 23 other federal agencies and foundations. Information on how to obtain the Add Health data files is available on the Add Health website (http://www.cpc.unc.edu/addhealth).

Neuroticism (2018)

GWAS Summary Statistic Source: Nagel et al. 2018. "Meta-Analysis of Genome-Wide Association Studies for Neuroticism in 449,484 Individuals Identifies Novel Genetic Loci and Pathways." *Nature Genetics* 50(7):920–27.

GWAS Ancestry Group(s): European

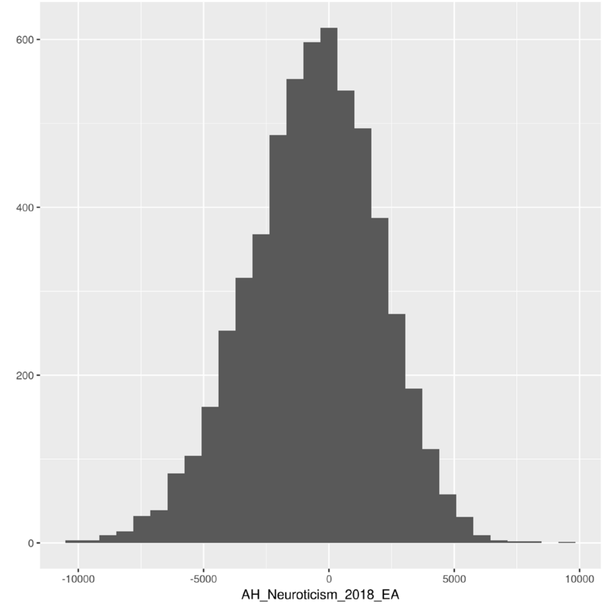
The PGSs for neuroticism were created using results from a 2018 study conducted by the Mats Nagel and colleagues. The GWAS meta-analysis files are publicly available and can be downloaded from: https://ctg.cncr.nl/software/summary_statistics.

The Neuroticism GWAS is based on the analysis of 449,484 individuals of European genetic ancestry from the UK Biobank, 23adnMe, and the Genetics of Personality Consortium (GPC). However, due to data use restrictions imposed by 23andMe, the summary statistics used to create the Add Health PGSs are based solely on a GWAS of the UK Biobank and GPC (N = 390,278). In the UK Biobank, neuroticism was measured as the average of 12 items from the Eysenck Personality Questionnaire while in the majority of GPC cohorts neuroticism was measured as the sum of 12 items from the NEO Five-Factor Inventory (see the accompanying Supplemental Materials from Nagel et al. 2018 for more details). GWAS analyses adjusted for different sets of covariates in the various cohorts, but generally included age, sex, and different numbers of principal components of the genetic data. More details are available in the Supplementary Information provided by Nagel et al. 2018.

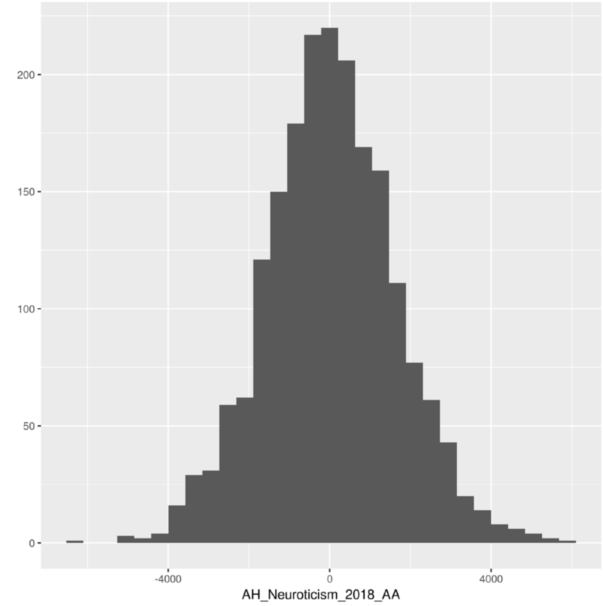
The Add Health PGSs for neuroticism are standardized (mean = 0, standard deviation = 1) within ancestry groups to control for population stratification. To give researchers an idea of the underlying raw PGS, we present density plots for the unstandardized PGSs below. **Please read the section entitled "Using Add Health PGSs" in the introductory portion of this document prior to conducting any analyses using the provided PGSs.**

Neuroticism (2018)

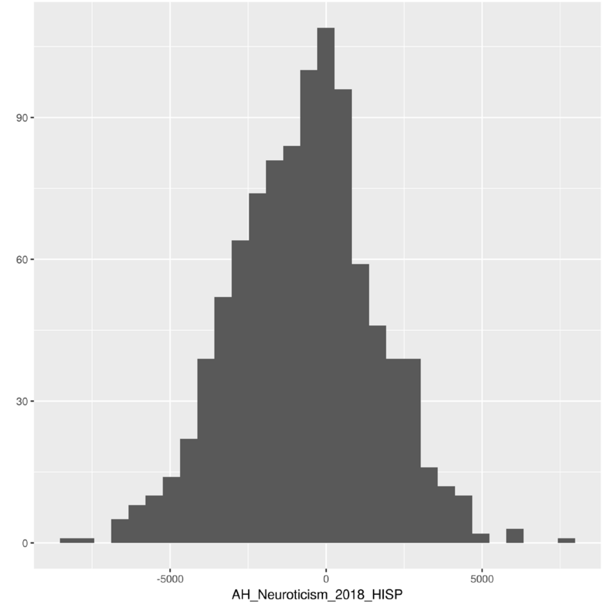
European ancestry raw PGS distribution



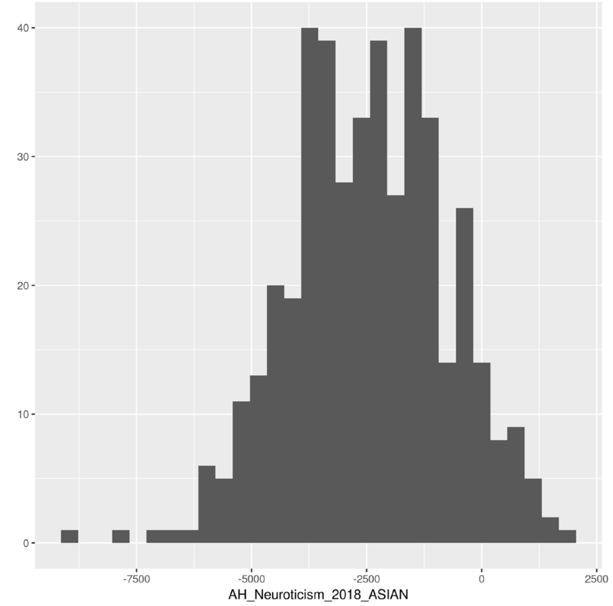
African ancestry raw PGS distribution



Hispanic ancestry raw PGS distribution



Asian ancestry raw PGS distribution



Add Health is a program project directed by Kathleen Mullan Harris and designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris at the University of North Carolina at Chapel Hill, and funded by grant P01-HD31921 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development, with cooperative funding from 23 other federal agencies and foundations. Information on how to obtain the Add Health data files is available on the Add Health website (http://www.cpc.unc.edu/addhealth).

Neuroticism (2019)

GWAS Summary Statistic Source: Baselmans et al. 2019. “Multivariate Genome-Wide Analyses of the Well-Being Spectrum.” *Nature Genetics* 51: 445-451.

GWAS Ancestry Group(s): European

The PGSs for neuroticism were created using results from a 2019 study conducted by the Bart Baselmans and colleagues including members of the BIOS Consortium and the Social Science Genetic Association Consortium. The GWAS meta-analysis files are publicly available and can be downloaded from: https://surfdrive.surf.nl/files/index.php/s/Ow1qCDpFT421ZOO.

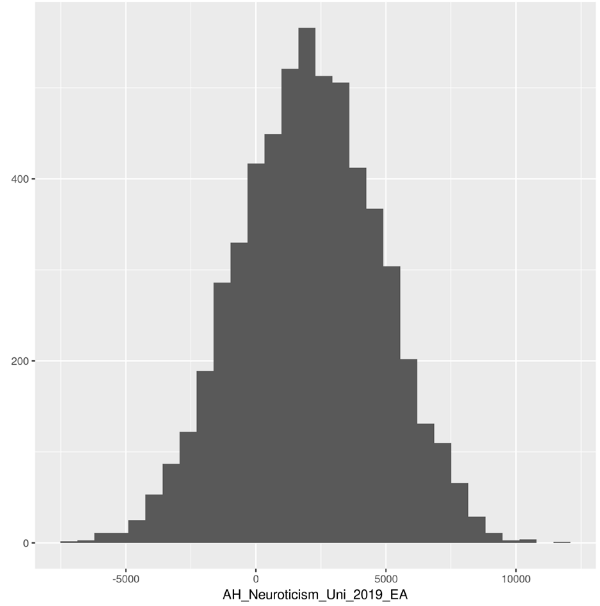
There are two PGSs for neuroticism from the Baselmans et al. (2019) study, one based on summary statistics from a traditional univariate GWAS meta-analyses and a second based on summary statistics from a novel statistical approach to genome-wide studies: model-averaging multivariate genome-wide association meta-analysis (MA-GWAMA). Details on the MA-GWAMA method can be found in Baselmans et al. (2019). Results from Baselmans and colleagues (2019) suggest that PGSs based on the MA-GWAMA summary statistics are up to approximately 50% more predictive than PGSs based on the traditional univariate GWAS meta-analysis methods.

The neuroticism GWAS meta-analyses and MA-GWAMA are based on the analysis of 582,989 individuals of European genetic ancestry. Details on the contributing cohorts can be found in Supplementary Figure 2 of the online supplement accompanying Baselmans et al. (2019).

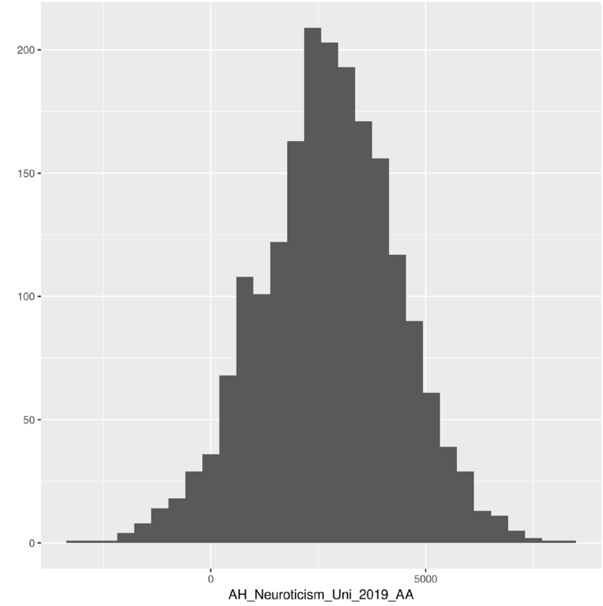
The Add Health PGSs for neuroticism are standardized (mean = 0, standard deviation = 1) within ancestry groups to control for population stratification. To give researchers an idea of the underlying raw PGS, we present density plots for the unstandardized PGSs below. **Please read the section entitled “Using Add Health PGSs” in the introductory portion of this document prior to conducting any analyses using the provided PGSs.**

Neuroticism (2019 – Univariate GWAS)

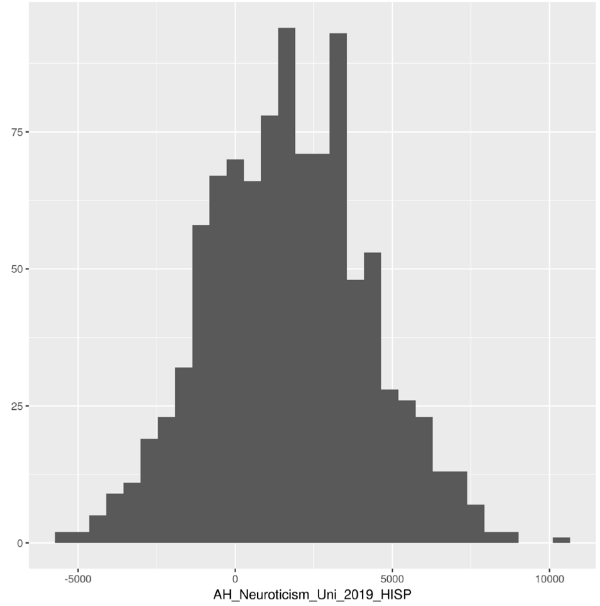
European ancestry raw PGS distribution



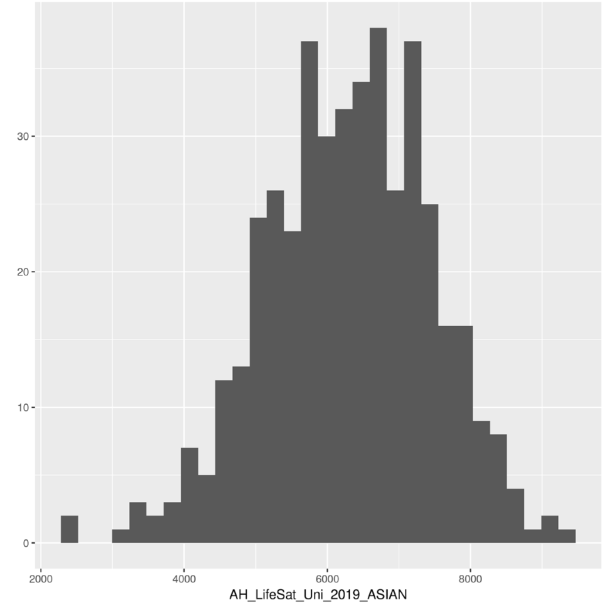
African ancestry raw PGS distribution



Hispanic ancestry raw PGS distribution

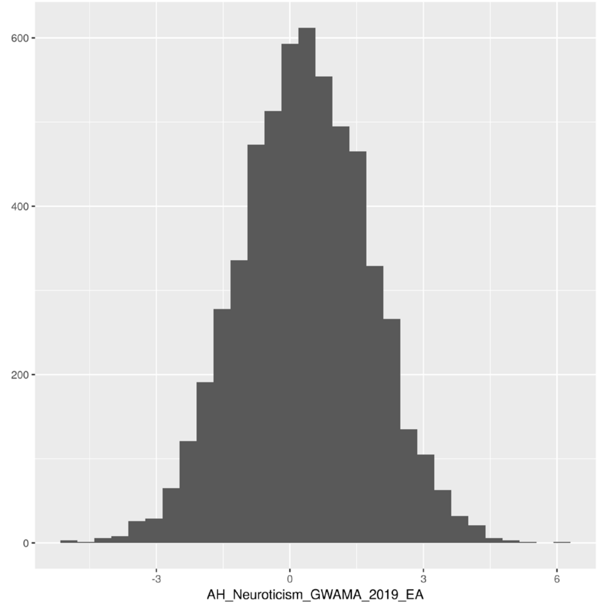


Asian ancestry raw PGS distribution

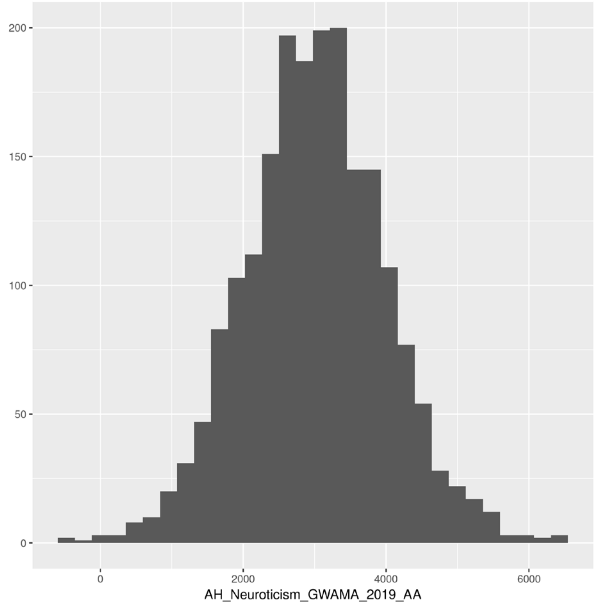


Neuroticism (2019 – MA-GWAMA)

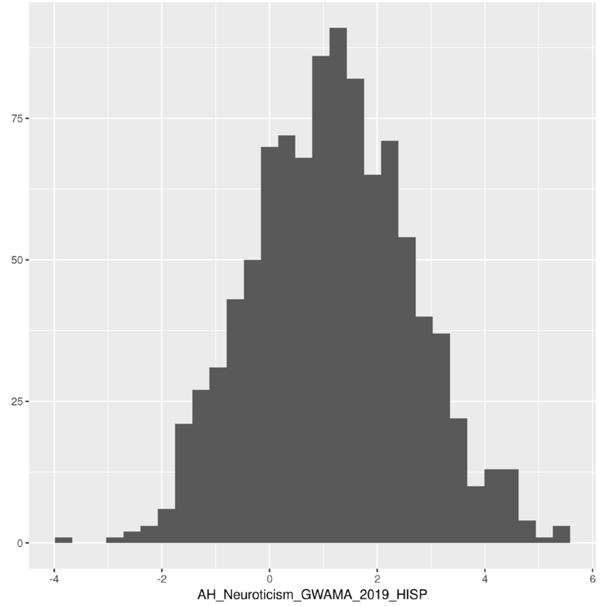
European ancestry raw PGS distribution



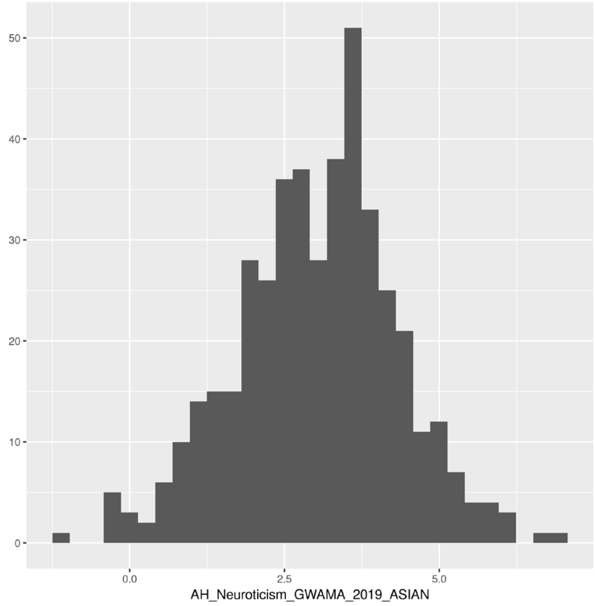
African ancestry raw PGS distribution



Hispanic ancestry raw PGS distribution



Asian ancestry raw PGS distribution



Anorexia Nervosa

GWAS Summary Statistic Source: Watson et al. 2019. “Genome-Wide Association Study Identifies Eight Risk Loci and Implicates Metabo-Psychiatric Origins for Anorexia Nervosa.” *Nature Genetics* 51(8):1207–14.

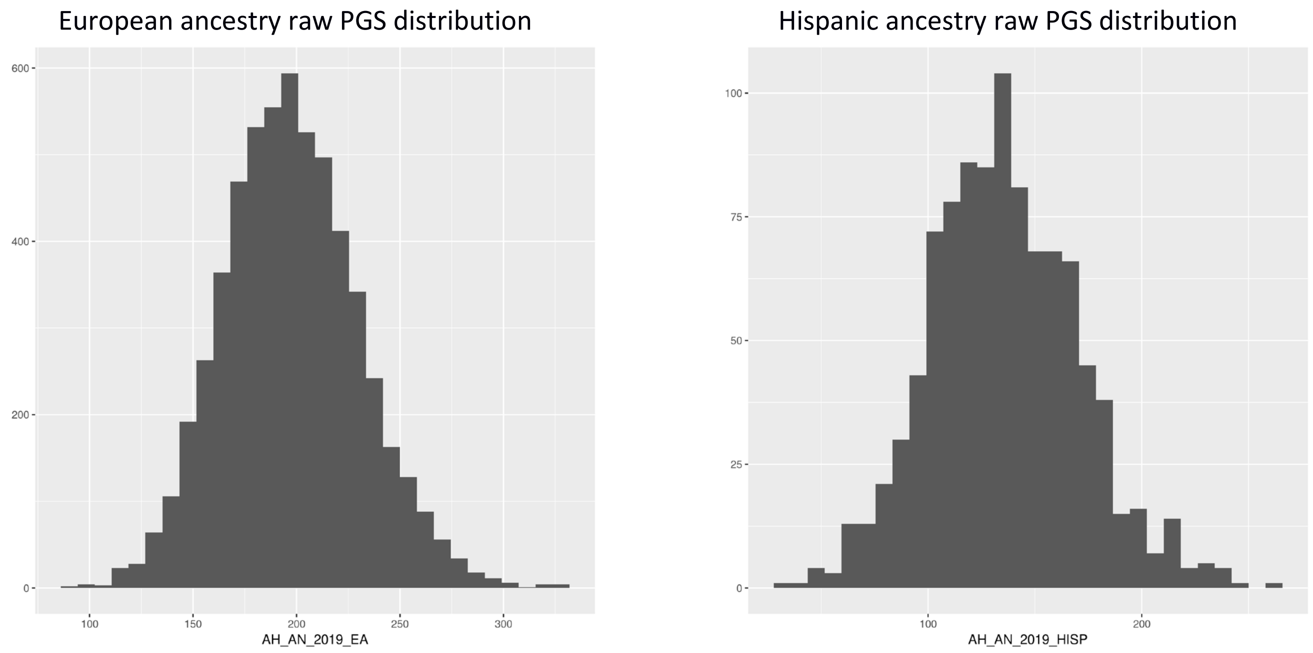
GWAS Ancestry Group(s): European

The PGSs for anorexia nervosa were created using results from a 2019 study conducted by Hunna Watson and colleagues. The GWAS meta-analysis files are publicly available and can be downloaded from: https://www.med.unc.edu/pgc/results-and-downloads/ed/.

The anorexia nervosa GWAS meta-analyses are based on the analysis of 72,517 (16,992 cases, 55,525 controls) individuals of European genetic ancestry. Cases are defined as ever being diagnosed, or meeting diagnostic criteria (e.g., DSM III-R and/or IV; ICD 8, 9, and 10) in your lifetime. Description of the GWAS analyses are vague, but generally, each analysis adjusted for five ancestry-specific PCs of the genetic data and an undisclosed number of other covariates associated with anorexia nervosa at the p<0.05 level based on univariate correlations (see Supplemental Information accompanying Watson et al. 2019 for more details).

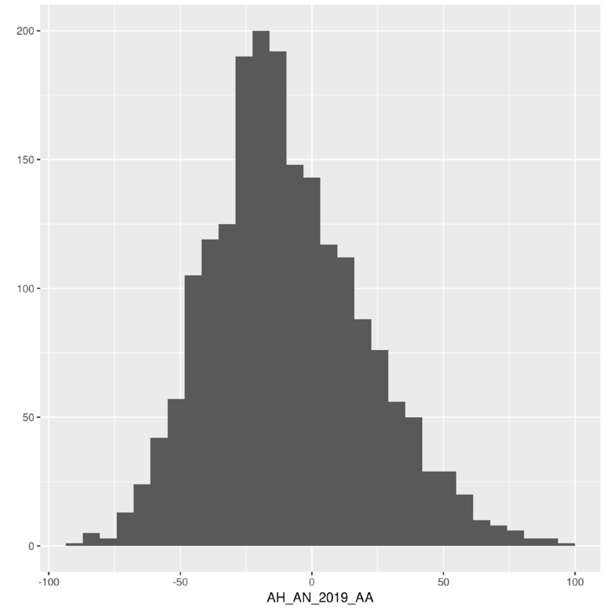
The Add Health PGSs for are standardized (mean = 0, standard deviation = 1) within ancestry groups to control for population stratification. To give researchers an idea of the underlying raw PGS, we present density plots for the unstandardized PGSs below. **Please read the section entitled “Using Add Health PGSs” in the introductory portion of this document prior to conducting any analyses using the provided PGSs.**

Anorexia Nervosa

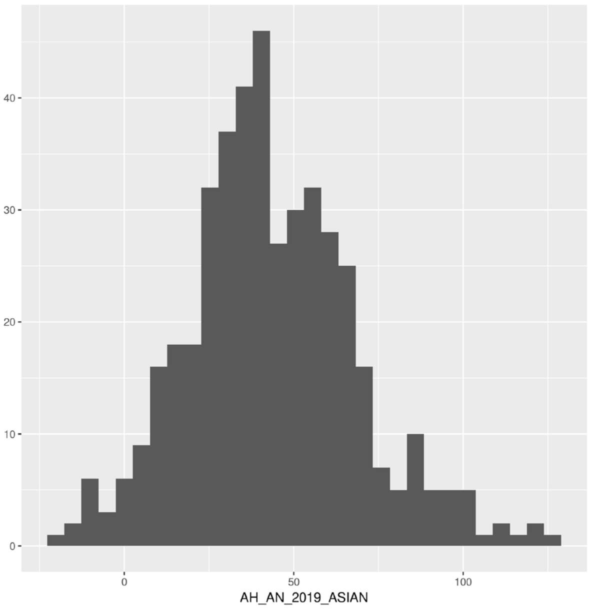


Add Health is a program project directed by Kathleen Mullan Harris and designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris at the University of North Carolina at Chapel Hill, and funded by grant P01-HD31921 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development, with cooperative funding from 23 other federal agencies and foundations. Information on how to obtain the Add Health data files is available on the Add Health website (http://www.cpc.unc.edu/addhealth).

African ancestry raw PGS distribution



Asian ancestry raw PGS distribution



Add Health is a program project directed by Kathleen Mullan Harris and designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris at the University of North Carolina at Chapel Hill, and funded by grant P01-HD31921 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development, with cooperative funding from 23 other federal agencies and foundations. Information on how to obtain the Add Health data files is available on the Add Health website (http://www.cpc.unc.edu/addhealth).

Mental Health Cross Disorder

GWAS Summary Statistic Source: Cross-Disorder Group of the Psychiatric Genomics Consortium. 2013. “Identification of Risk Loci with Shared Effects on Five Major Psychiatric Disorders: A Genome-Wide Analysis.” Lancet (London, England) 381(9875):1371–79.

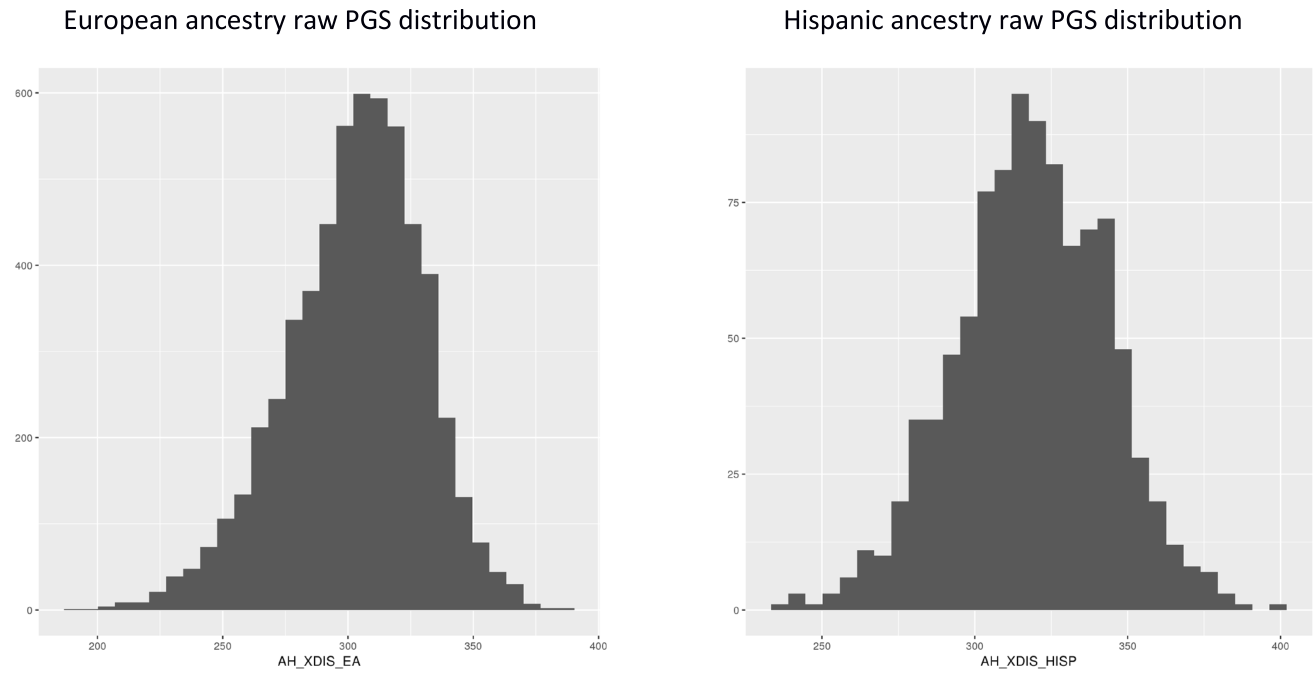
GWAS Ancestry Group(s): European

The PGSs for mental health cross-disorder were created using results from a 2013 GWAS conducted by the Cross Disorder working group of the Psychiatric GWAS Consortium. The GWAS meta-analysis files are publicly available from: http://www.med.unc.edu/pgc/results-and-downloads.

The discovery phase of the meta-analysis included 33,342 cases and 27,888 controls. Disorders that were counted as cases (DSM-III-R or DSM-IV criteria) included autism spectrum disorder, attention-deficit/hyperactivity disorder, bipolar disorder, major depressive disorder, and schizophrenia.

The Add Health PGSs for mental health cross-disorder are standardized (mean = 0, standard deviation = 1) within ancestry groups to control for population stratification. To give researchers an idea of the underlying raw PGS, we present density plots for the unstandardized PGSs below. **Please read the section entitled “Using Add Health PGSs” in the introductory portion of this document prior to conducting any analyses using the provided PGSs.**

Cross Disorder

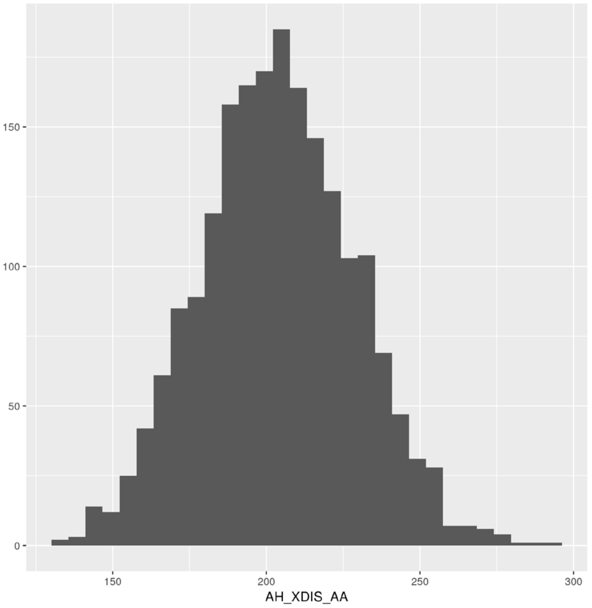


European ancestry raw PGS distribution

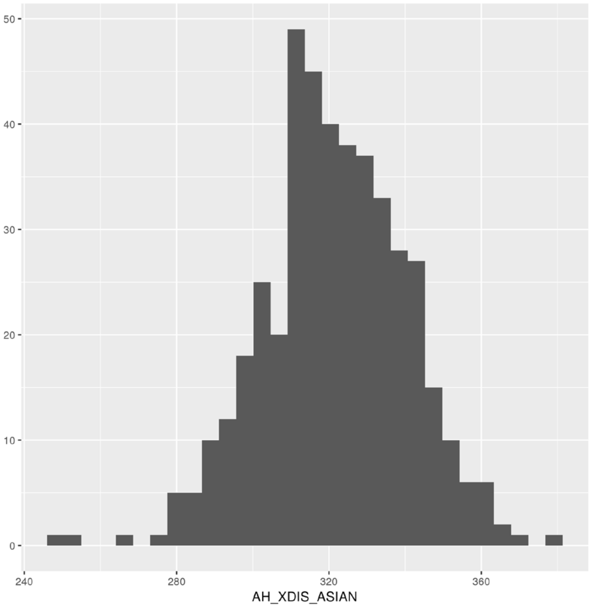
Hispanic ancestry raw PGS distribution

Add Health is a program project directed by Kathleen Mullan Harris and designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris at the University of North Carolina at Chapel Hill, and funded by grant P01-HD31921 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development, with cooperative funding from 23 other federal agencies and foundations. Information on how to obtain the Add Health data files is available on the Add Health website (http://www.cpc.unc.edu/addhealth).

African ancestry raw PGS distribution



Asian ancestry raw PGS distribution



Add Health is a program project directed by Kathleen Mullan Harris and designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris at the University of North Carolina at Chapel Hill, and funded by grant P01-HD31921 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development, with cooperative funding from 23 other federal agencies and foundations. Information on how to obtain the Add Health data files is available on the Add Health website (http://www.cpc.unc.edu/addhealth).

Social Isolation

GWAS Summary Statistic Source: Day, Felix R., Ken K. Ong, and John R. B. Perry. 2018. "Elucidating the Genetic Basis of Social Interaction and Isolation." *Nature Communications* 9(1):2457.

GWAS Ancestry Group(s): European

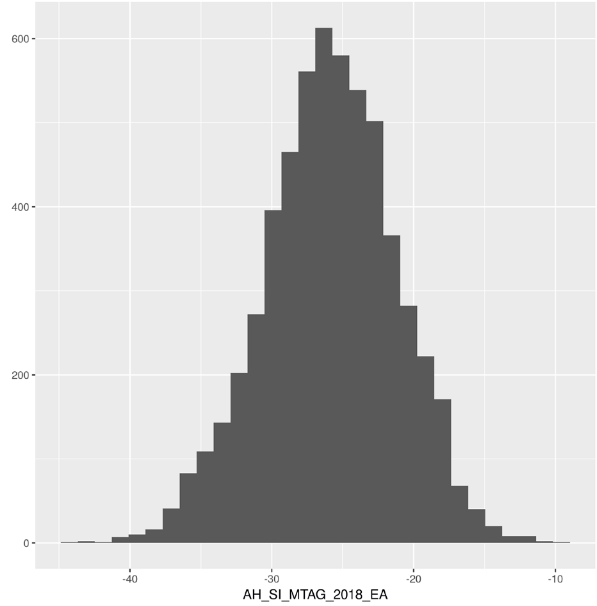
The PGSs for social isolation were created using results from a 2018 study conducted by Felix Day, Ken Ong, and John Perry. The GWAS meta-analysis files are publicly available and can be downloaded from: https://www.med.unc.edu/pgc/results-and-downloads.

The social isolation GWAS meta-analyses are based on the analysis of 487,647 individuals of European genetic ancestry from the UK Biobank and are conducted using the multi-trait GWAS (MTAG) method developed by Turley et al. (2018) and summary statistics from GWASs of perceived loneness (responses of yes to the question, "'Do you often feel lonely?" coded as cases), a composite measure of living alone (cases are defined based on responses to the following two questions: 1) "Including yourself, how many people are living together in your household?" and 2) 'How often do you visit friends or family or have them visit you?" with cases identified as individuals who lived alone and who indicated that they either never visited or had no friends or family outside their household; controls were defined as those who either did not live alone, or had friends who visited at least once a week), and the quality of social interactions based on responses to the question, "How often are you able to confide in someone close to you?" with cases defined as responses of "Never or almost never" and controls defined as responses of "Almost daily." See the methods section in Day et al. (2018) for more details.

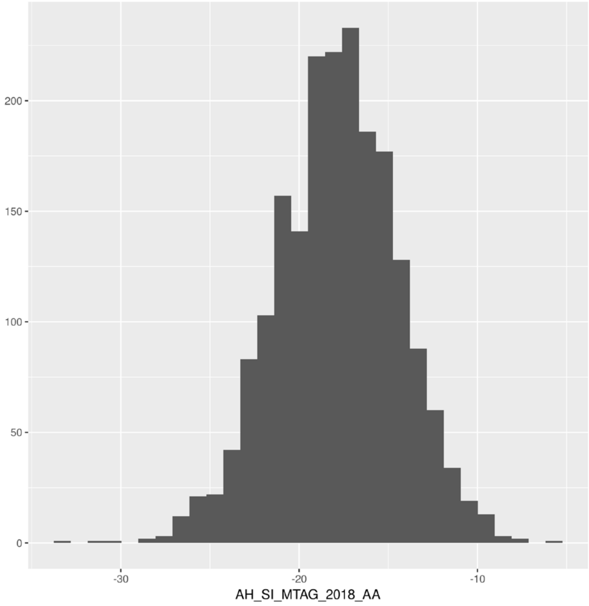
The Add Health PGSs for social isolation are standardized (mean = 0, standard deviation = 1) within ancestry groups to control for population stratification. To give researchers an idea of the underlying raw PGS, we present density plots for the unstandardized PGSs below. **Please read the section entitled "Using Add Health PGSs" in the introductory portion of this document prior to conducting any analyses using the provided PGSs.**

Social Isolation

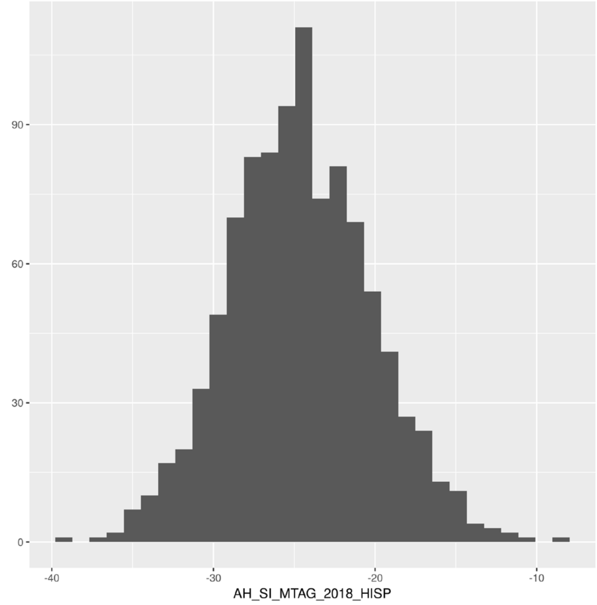
European ancestry raw PGS distribution



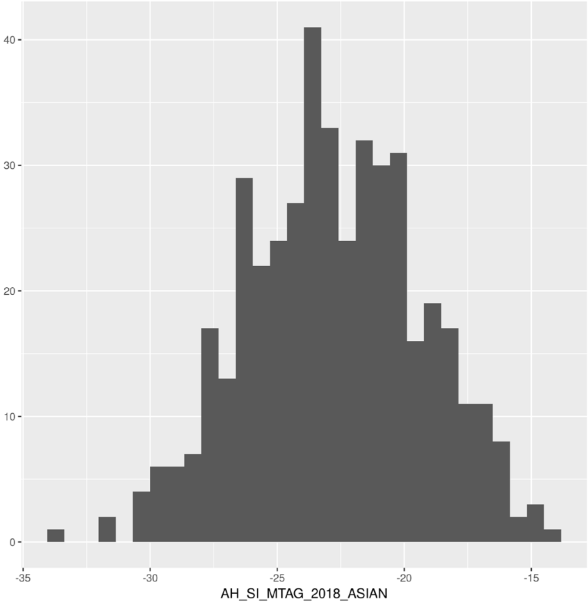
African ancestry raw PGS distribution



Hispanic ancestry raw PGS distribution



Asian ancestry raw PGS distribution



Lonely

GWAS Summary Statistic Source: Day, Felix R., Ken K. Ong, and John R. B. Perry. 2018. “Elucidating the Genetic Basis of Social Interaction and Isolation.” *Nature Communications* 9(1):2457.

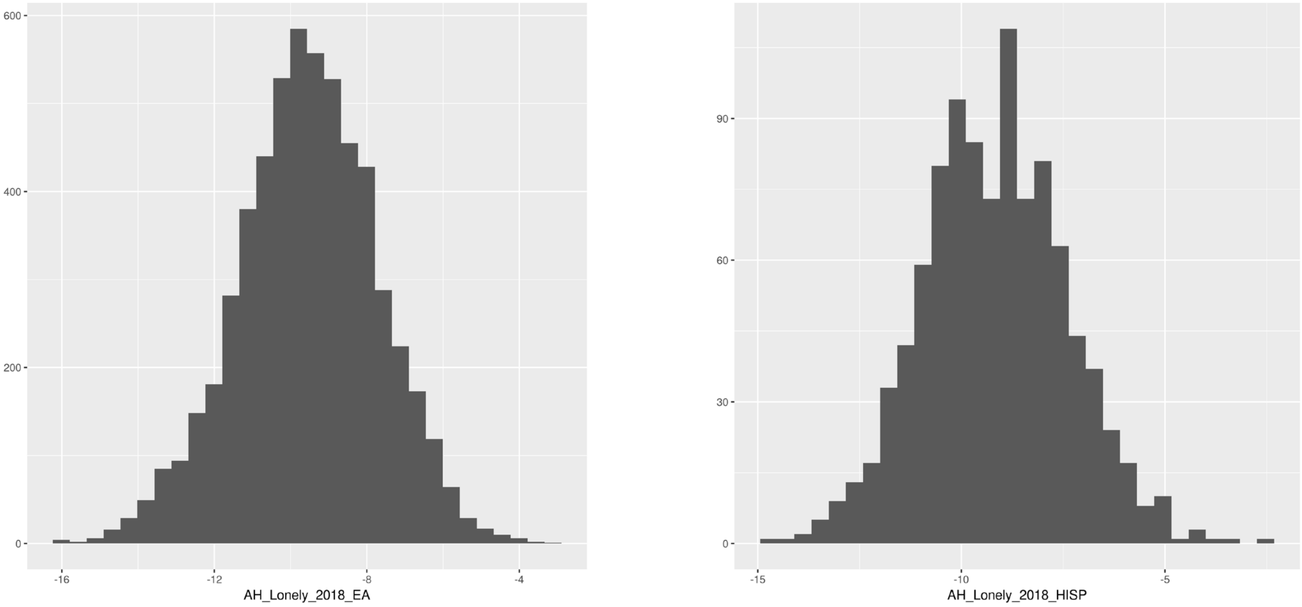
GWAS Ancestry Group(s): European

The PGSs for social isolation were created using results from a 2018 study conducted by Felix Day, Ken Ong, and John Perry. The GWAS meta-analysis files are publicly available and can be downloaded from: https://www.med.unc.edu/pgc/results-and-downloads.

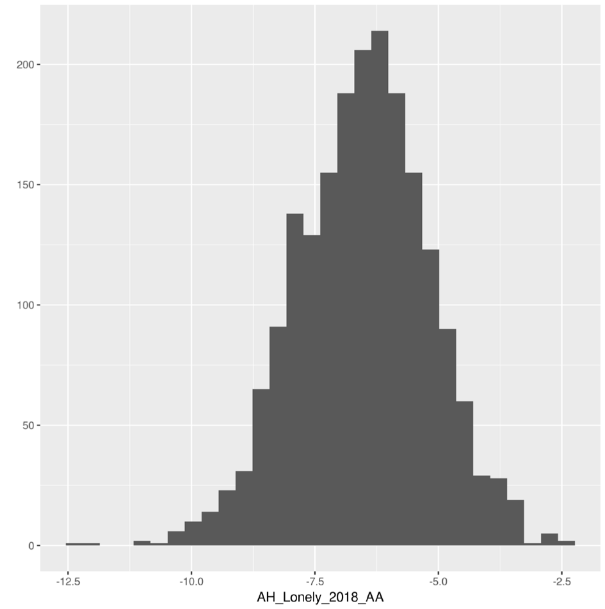
The being lonely GWAS meta-analyses are based on the analysis of 445,024 (80,134 cases and 364,890 controls) individuals of European genetic ancestry from the UK Biobank. The measure of being lonely is a dichotomous indicator based on respondents’ answers to the question: “Are you lonely?” with answers of “yes” coded as cases and answers of “no” coded as controls. See the methods section for in Day et al. (2018) for more details.

The Add Health PGSs for being lonely are standardized (mean = 0, standard deviation = 1) within ancestry groups to control for population stratification. To give researchers an idea of the underlying raw PGS, we present density plots for the unstandardized PGSs below. **Please read the section entitled “Using Add Health PGSs” in the introductory portion of this document prior to conducting any analyses using the provided PGSs.**

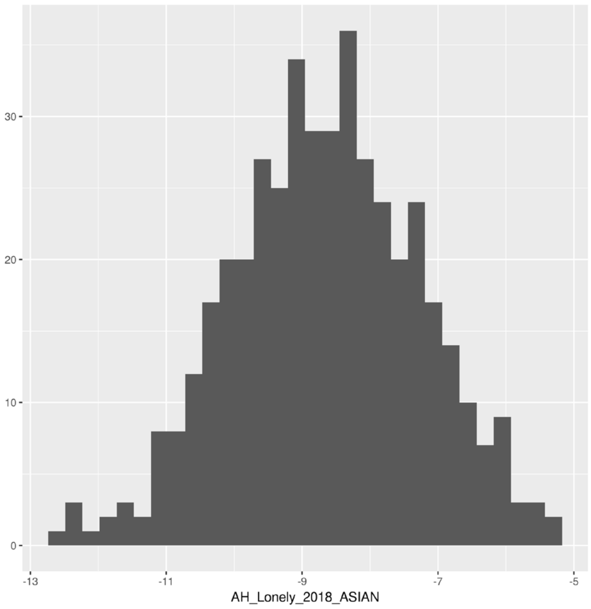
Lonely


European ancestry raw PGS distribution

Hispanic ancestry raw PGS distribution



African ancestry raw PGS distribution



Asian ancestry raw PGS distribution

Going to Pubs/Social Clubs Weekly

GWAS Summary Statistic Source: Day, Felix R., Ken K. Ong, and John R. B. Perry. 2018. “Elucidating the Genetic Basis of Social Interaction and Isolation.” *Nature Communications* 9(1):2457.

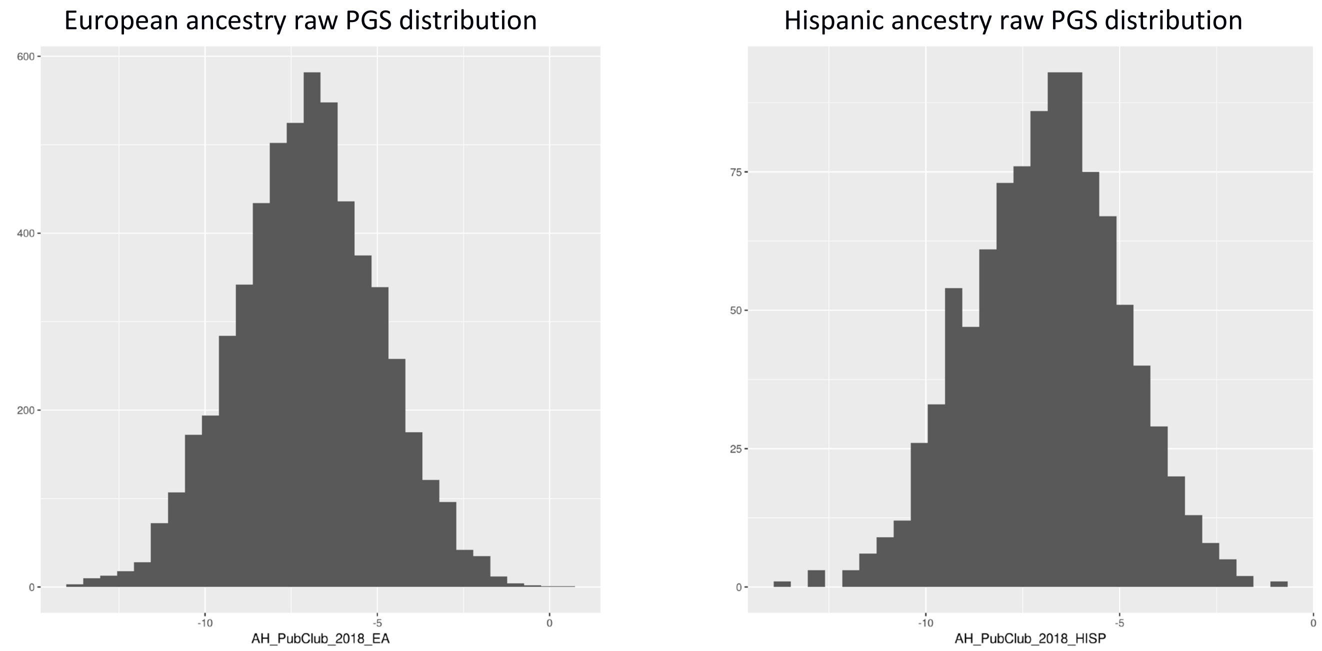
GWAS Ancestry Group(s): European

The PGSs for visiting a pub and/or social club at least once a week were created using results from a 2018 study conducted by Felix Day, Ken Ong, and John Perry. The GWAS meta-analysis files are publicly available and can be downloaded from: https://www.med.unc.edu/pgc/results-and-downloads.

The going to a pub and/or social club weekly GWAS meta-analyses are based on the analysis of 452,302 (124,047 cases, 328,255 controls) individuals of European genetic ancestry from the UK Biobank. The measure of pub/social club attendance is a dichotomous indicator based on respondents’ answers to the question: “Do you attend a Pub or Social Club on at least a weekly basis?” with answers of “yes” coded as cases and answers of “no” coded as controls. See the methods section for in Day et al. (2018) for more details.

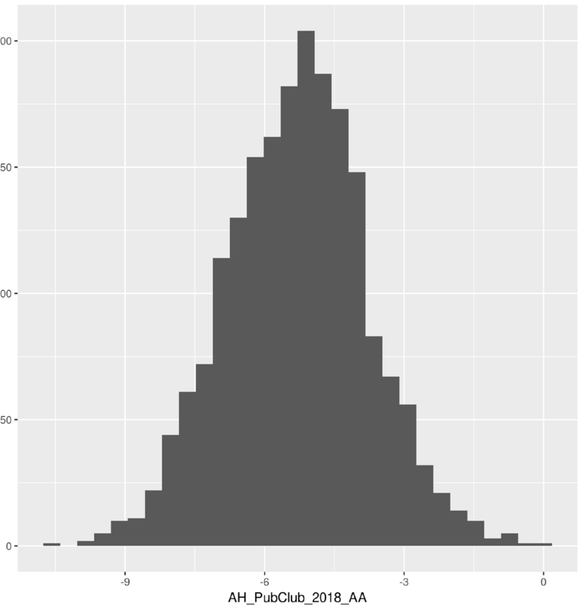
The Add Health PGSs for weekly attendance at a pub and/or social club are standardized (mean = 0, standard deviation = 1) within ancestry groups to control for population stratification. To give researchers an idea of the underlying raw PGS, we present density plots for the unstandardized PGSs below. **Please read the section entitled “Using Add Health PGSs” in the introductory portion of this document prior to conducting any analyses using the provided PGSs.**

Going to Pubs and/or Social Clubs

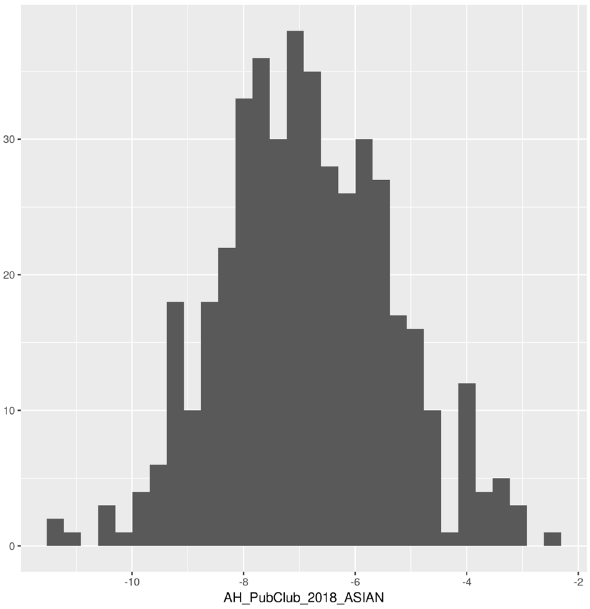


European ancestry raw PGS distribution

Hispanic ancestry raw PGS distribution



African ancestry raw PGS distribution



Asian ancestry raw PGS distribution

Going to a Gym/Sports Club Weekly

GWAS Summary Statistic Source: Day, Felix R., Ken K. Ong, and John R. B. Perry. 2018. “Elucidating the Genetic Basis of Social Interaction and Isolation.” *Nature Communications* 9(1):2457.

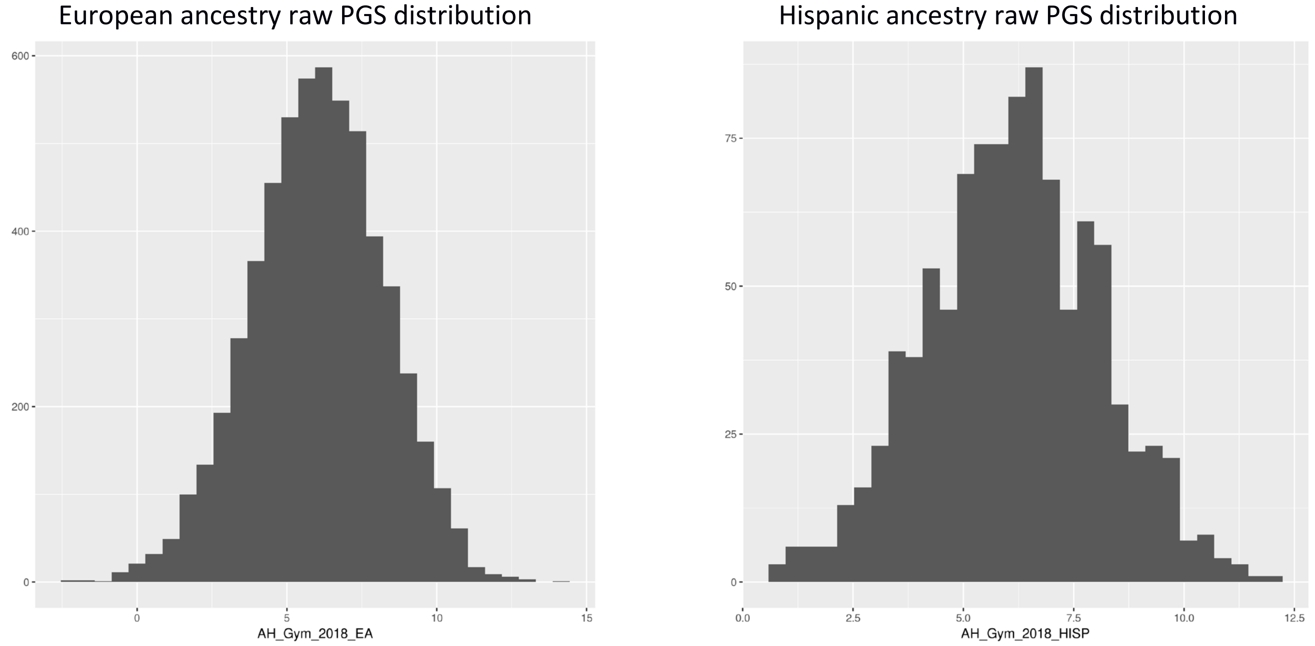
GWAS Ancestry Group(s): European

The PGSs for visiting a gym and/or sports club at least once a week were created using results from a 2018 study conducted by Felix Day, Ken Ong, and John Perry. The GWAS meta-analysis files are publicly available and can be downloaded from: https://www.med.unc.edu/pgc/results-and-downloads.

The weekly attendance at a gym and/or sports club GWAS meta-analyses are based on the analysis of 452,302 (135,060 cases, 317,242 controls) individuals of European genetic ancestry from the UK Biobank. The measure of pub/social club attendance is a dichotomous indicator based on respondents’ answers to the question: “Do you attend a Gym or Sports Club on at least a weekly basis?” with answers of “yes” coded as cases and answers of “no” coded as controls. See the methods section for in Day et al. (2018) for more details.

The Add Health PGSs for weekly attendance at a gym and/or sports club are standardized (mean = 0, standard deviation = 1) within ancestry groups to control for population stratification. To give researchers an idea of the underlying raw PGS, we present density plots for the unstandardized PGSs below. **Please read the section entitled “Using Add Health PGSs” in the introductory portion of this document prior to conducting any analyses using the provided PGSs.**

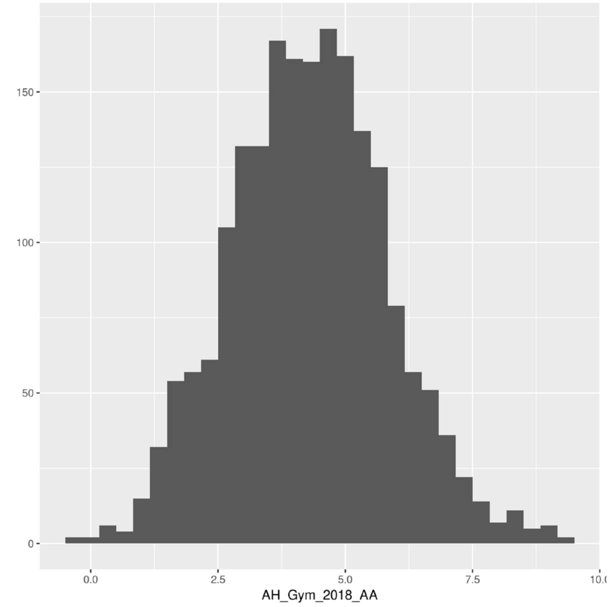
Going to Gyms/Sports Clubs


European ancestry raw PGS distribution

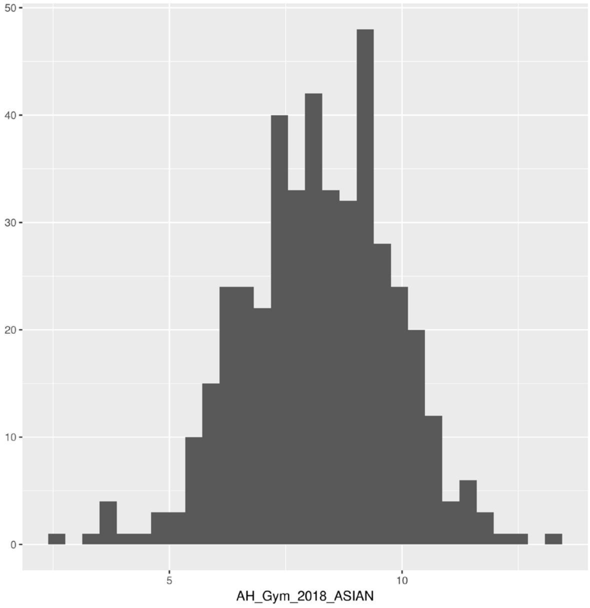
Hispanic ancestry raw PGS distribution

Add Health is a program project directed by Kathleen Mullan Harris and designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris at the University of North Carolina at Chapel Hill, and funded by grant P01-HD31921 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development, with cooperative funding from 23 other federal agencies and foundations. Information on how to obtain the Add Health data files is available on the Add Health website (http://www.cpc.unc.edu/addhealth).

African ancestry raw PGS distribution



Asian ancestry raw PGS distribution



Attending Religious Meetings/Groups

GWAS Summary Statistic Source: Day, Felix R., Ken K. Ong, and John R. B. Perry. 2018. “Elucidating the Genetic Basis of Social Interaction and Isolation.” *Nature Communications* 9(1):2457.

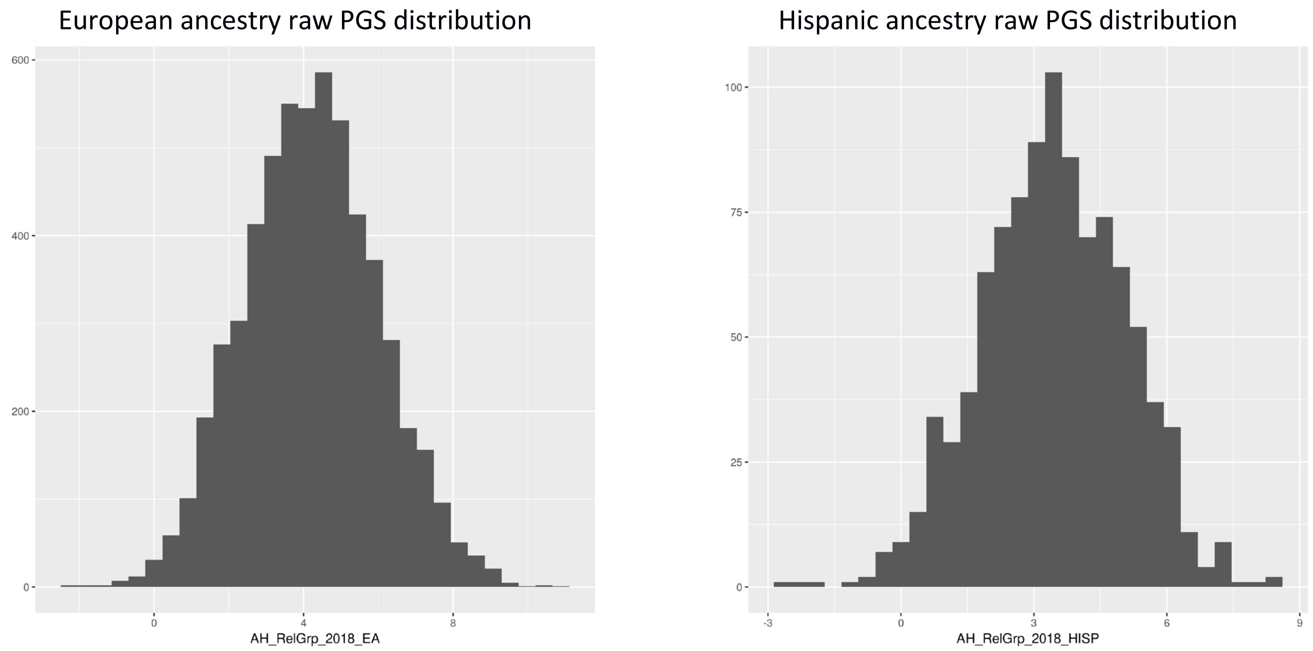
GWAS Ancestry Group(s): European

The PGSs for weekly attendance at religious meetings were created using results from a 2018 study conducted by Felix Day, Ken Ong, and John Perry. The GWAS meta-analysis files are publicly available and can be downloaded from: https://www.med.unc.edu/pgc/results-and-downloads.

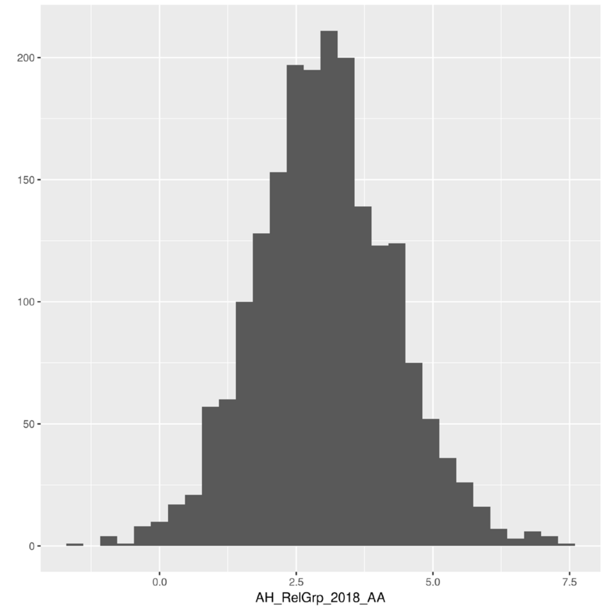
The weekly attendance at religious meetings GWAS meta-analyses are based on the analysis of 452,302 (66,259 cases, 386,043 controls) individuals of European genetic ancestry from the UK Biobank. The measure of pub/social club attendance is a dichotomous indicator based on respondents’ answers to the question: “Do you attend a Religious group on at least a weekly basis??” with answers of “yes” coded as cases and answers of “no” coded as controls. See the methods section for in Day et al. (2018) for more details.

The Add Health PGSs for weekly attendance at religious meetings are standardized (mean = 0, standard deviation = 1) within ancestry groups to control for population stratification. To give researchers an idea of the underlying raw PGS, we present density plots for the unstandardized PGSs below. **Please read the section entitled “Using Add Health PGSs” in the introductory portion of this document prior to conducting any analyses using the provided PGSs.**

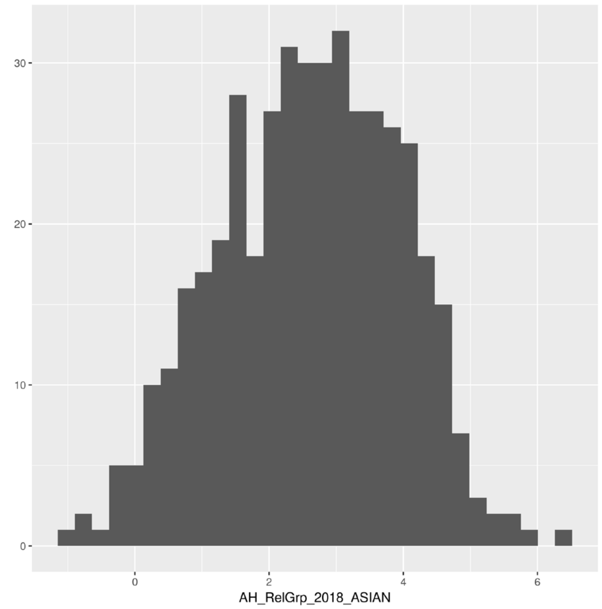
Attending Religious Meetings/Group



African ancestry raw PGS distribution



Asian ancestry raw PGS distribution



Risk Tolerance

GWAS Summary Statistic Source: Karlsson Linnér et al. 2019. "Genome-wide association analysis of risk tolerance and risky behaviors in over 1 million individuals identify hundreds of loci and shared genetic influences." *Nature Genetics* 51: 245-257.

GWAS Ancestry Group(s): European

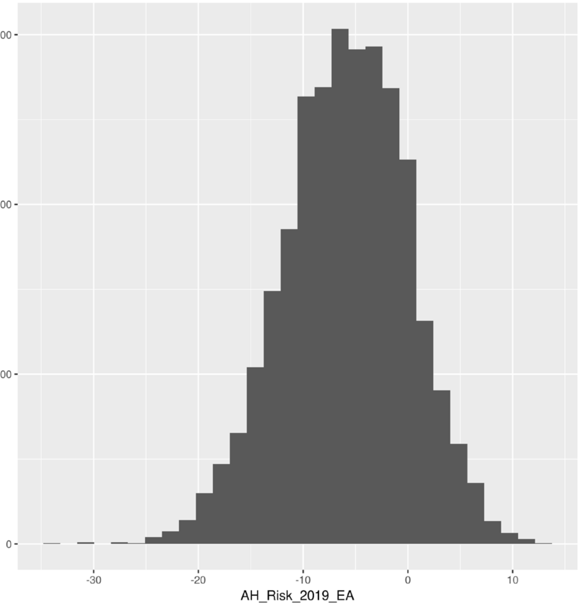
The PGSs for risk tolerance were created using results from a 2018 study conducted by Richard Karlsson Linnér and colleagues. The GWAS meta-analysis files are publicly available and can be downloaded from: https://www.thessgac.org/data.

The risk tolerance GWAS meta-analyses are based on the analysis of 939,908 individuals of European genetic ancestry from the UK Biobank (431,126) and 23andMe (508,782). However, due to data use restrictions imposed by 23andMe, the summary statistics used to create the Add Health PGSs exclude the 23andMe sample. As a benefit to the larger scientific community, Karlsson Linnér and colleagues re-estimated the original GWAS including the replication sample (N = 35,445), thus the summary statistics used to create the Add Health PGSs are based on the combined UK Biobank and replication sample (N = 466, 571). In the UK Biobank, risk tolerance is measured as responses to the question, “Would you describe yourself as someone who takes risks?” with answers of “yes” coded as cases. See the methods section in Karlsson Linnér et al. (2019) for more details.

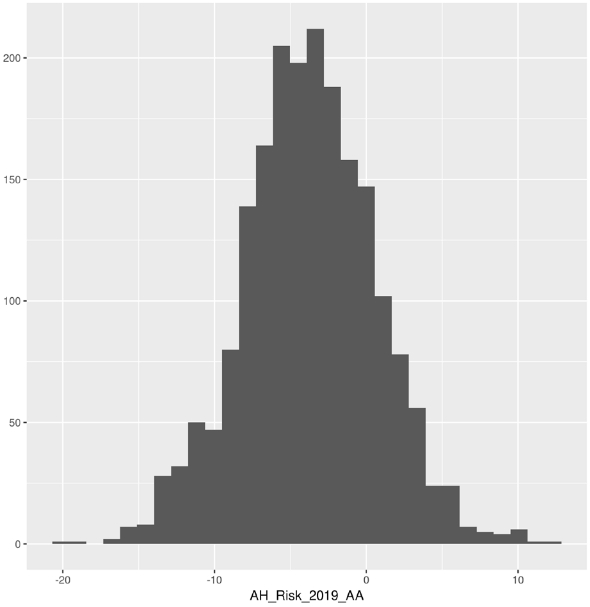
The Add Health PGSs for risk tolerance are standardized (mean = 0, standard deviation = 1) within ancestry groups to control for population stratification. To give researchers an idea of the underlying raw PGS, we present density plots for the unstandardized PGSs below. **Please read the section entitled “Using Add Health PGSs” in the introductory portion of this document prior to conducting any analyses using the provided PGSs.**

Risk Tolerance

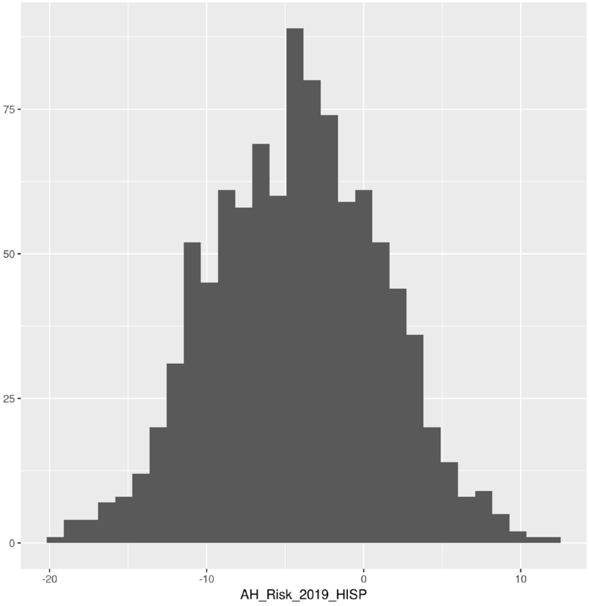
European ancestry raw PGS distribution



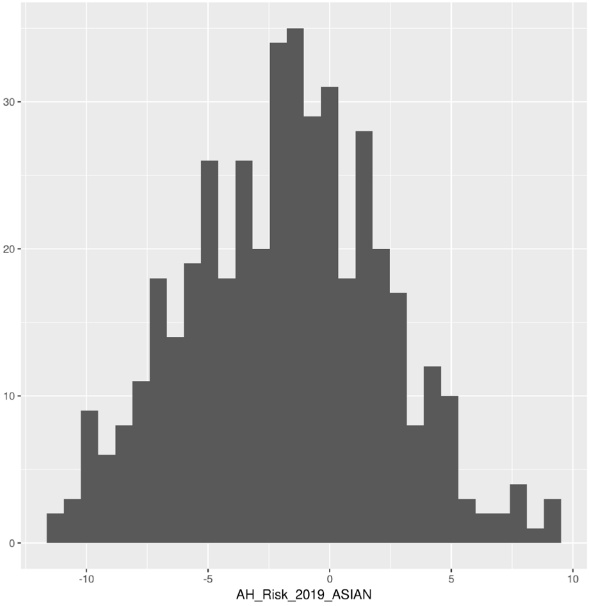
African ancestry raw PGS distribution



Hispanic ancestry raw PGS distribution



Asian ancestry raw PGS distribution



Propensity for Speeding while Driving

GWAS Summary Statistic Source: Karlsson Linnér et al. 2019. "Genome-wide association analysis of risk tolerance and risky behaviors in over 1 million individuals identify hundreds of loci and shared genetic influences." *Nature Genetics* 51: 245-257.

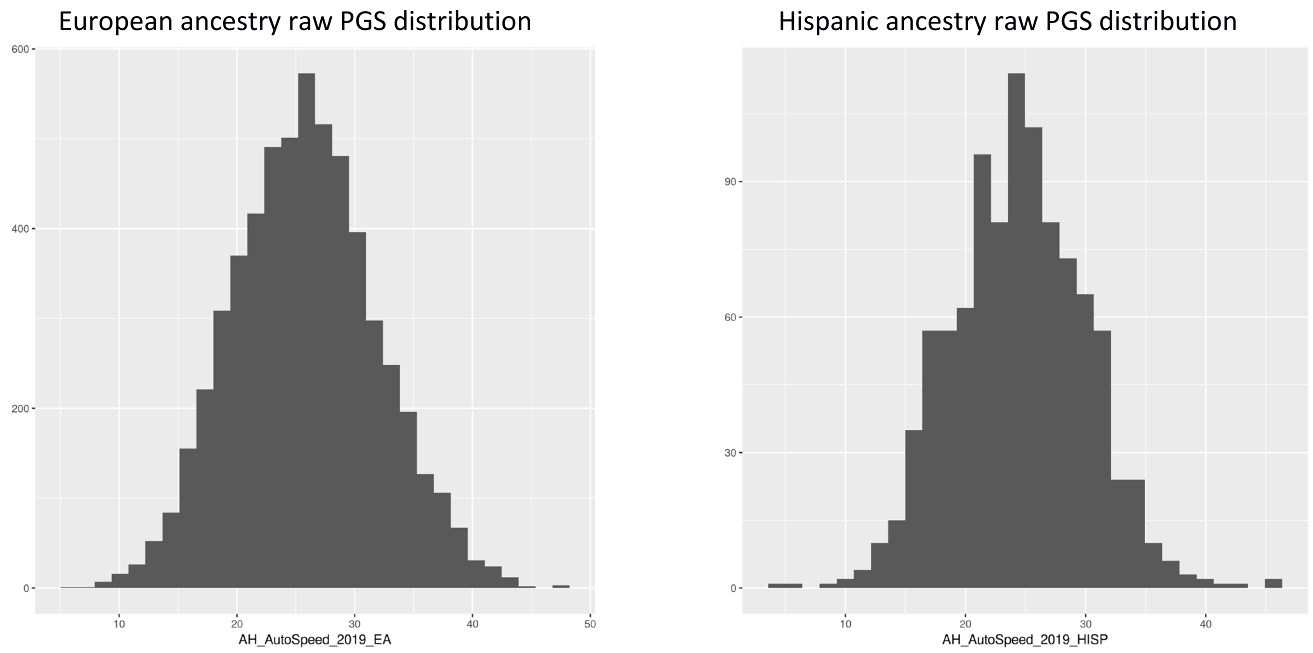
GWAS Ancestry Group(s): European

The PGSs for propensity for speeding while driving were created using results from a 2018 study conducted by Richard Karlsson Linnér and colleagues. The GWAS meta-analysis files are publicly available and can be downloaded from: https://www.thessgac.org/data.

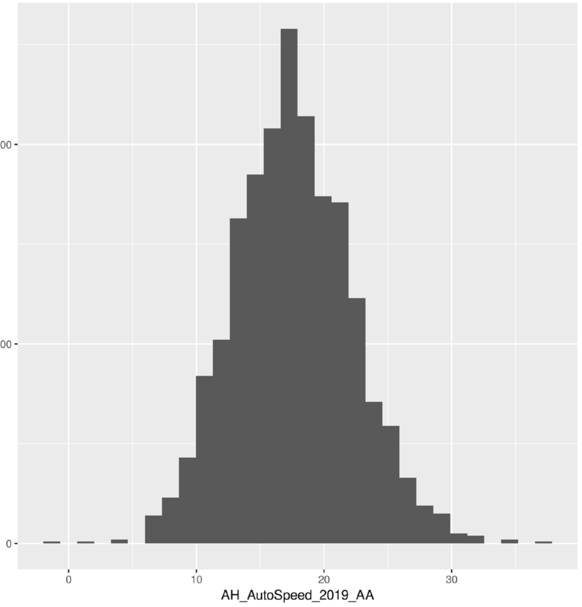
The propensity for speeding while driving GWAS meta-analyses are based on the analysis of 404,291 individuals of European genetic ancestry from the UK Biobank. The measure of a propensity for automobile speeding is a sex-specific normalized score based on answers to the question, “How often do you drive faster than the speed limit on the motorway?” with Likert scale options ranging from 1 ("never/rarely”) to 4 (“most of the time”). Respondents who answered that they do not drive on the motorway, were excluded from all analyses. See the methods section in Karlsson Linnér et al. (2019) for more details.

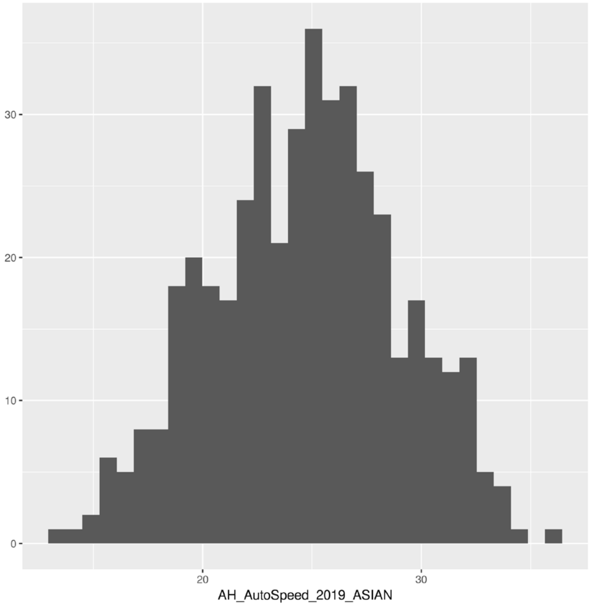
The Add Health PGSs for propensity for speeding while driving are standardized (mean = 0, standard deviation = 1) within ancestry groups to control for population stratification. To give researchers an idea of the underlying raw PGS, we present density plots for the unstandardized PGSs below. **Please read the section entitled “Using Add Health PGSs” in the introductory portion of this document prior to conducting any analyses using the provided PGSs.**

Propensity for Speeding while Driving



Add Health is a program project directed by Kathleen Mullan Harris and designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris at the University of North Carolina at Chapel Hill, and funded by grant P01-HD31921 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development, with cooperative funding from 23 other federal agencies and foundations. Information on how to obtain the Add Health data files is available on the Add Health website (http://www.cpc.unc.edu/addhealth).


African ancestry raw PGS distribution


Asian ancestry raw PGS distribution

Add Health is a program project directed by Kathleen Mullan Harris and designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris at the University of North Carolina at Chapel Hill, and funded by grant P01-HD31921 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development, with cooperative funding from 23 other federal agencies and foundations. Information on how to obtain the Add Health data files is available on the Add Health website (http://www.cpc.unc.edu/addhealth).

Lifetime Number of Sexual Partners

GWAS Summary Statistic Source: Karlsson Linnér et al. 2019. "Genome-wide association analysis of risk tolerance and risky behaviors in over 1 million individuals identify hundreds of loci and shared genetic influences." *Nature Genetics* 51: 245-257

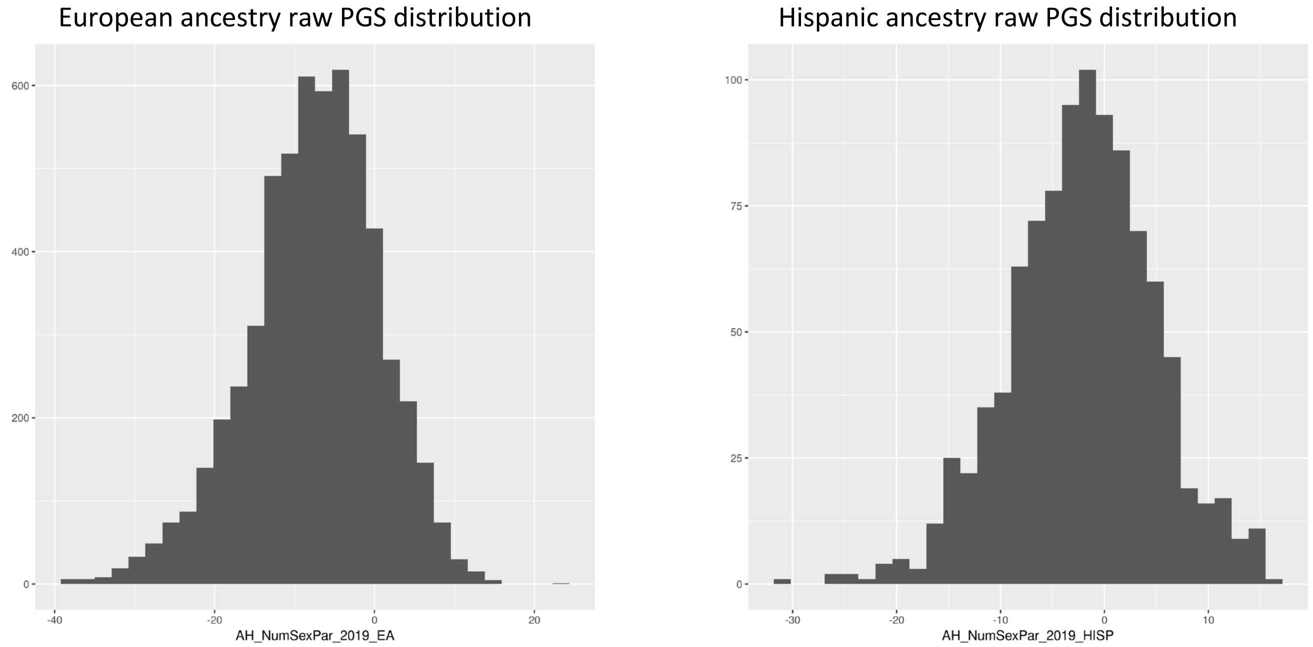
GWAS Ancestry Group(s): European

The PGSs for lifetime number of sexual partners were created using results from a 2018 study conducted by Richard Karlsson Linnér and colleagues. The GWAS meta-analysis files are publicly available and can be downloaded from: https://www.thessgac.org/data.

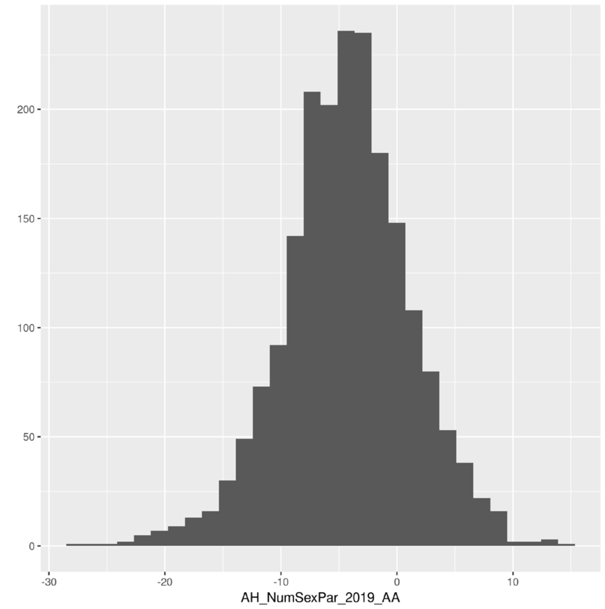
The lifetime number of sexual partners GWAS meta-analyses are based on the analysis of 404,291 individuals of European genetic ancestry from the UK Biobank. The measure of lifetime number of sexual partners is a sex-specific normalized score based on answers to the question, “About how many sexual partners have you had in your lifetime?” See the methods section for in Karlsson Linnér et al. (2019) for more details.

The Add Health PGSs for lifetime number of sexual partners are standardized (mean = 0, standard deviation = 1) within ancestry groups to control for population stratification. To give researchers an idea of the underlying raw PGS, we present density plots for the unstandardized PGSs below. **Please read the section entitled “Using Add Health PGSs” in the introductory portion of this document prior to conducting any analyses using the provided PGSs.**

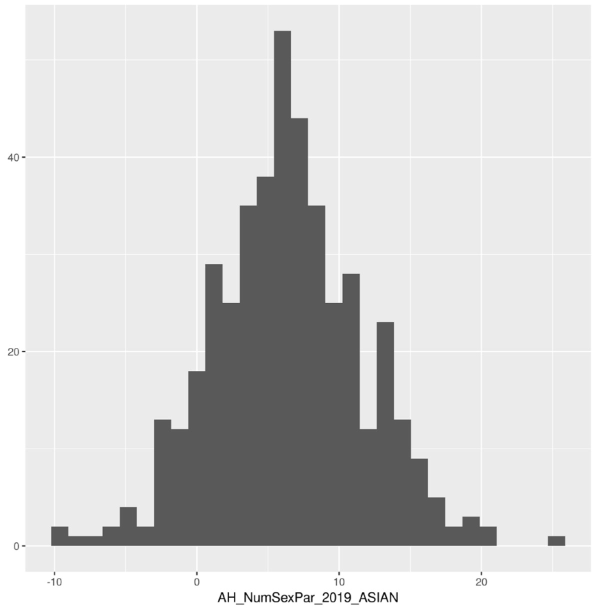
Number of Sexual Partners



African ancestry raw PGS distribution



Asian ancestry raw PGS distribution



Insomnia

GWAS Summary Statistic Source: Jansen et al. 2019. “Genome-Wide Analysis of Insomnia in 1,331,010 Individuals Identifies New Risk Loci and Functional Pathways.” *Nature Genetics* 51: 394-403.

GWAS Ancestry Group(s): European

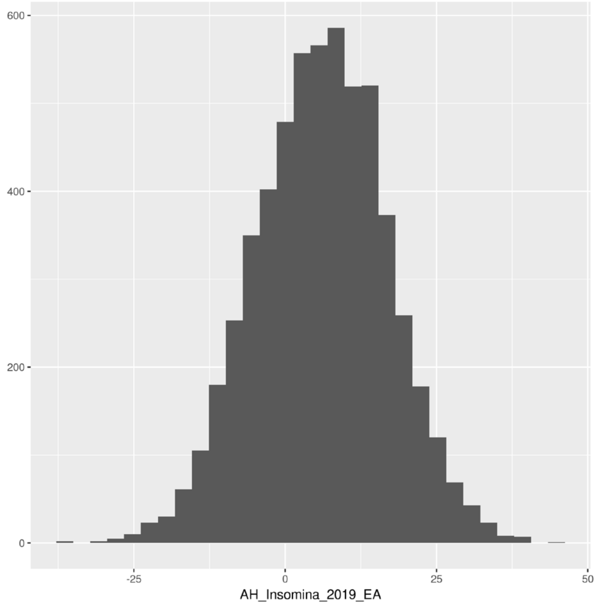
The PGSs for Insomnia were created using results from a 2019 study conducted by the Philip Jansen and colleagues. The GWAS meta-analysis files are publicly available and can be downloaded from: https://ctg.cncr.nl/software/summary_statistics.

The Insomnia GWAS meta-analysis consisted of 2 cohorts (the UK Biobank and 23andMe) resulting in a total of 1,331,010 individuals of European genetic ancestry. However, due to data use restrictions imposed by 23andMe, the summary statistics used to create the Add Health PGSs are based solely on a GWAS of the 386,533 individuals from the UK Biobank. Insomnia was measured using self-reports from responses to the question, “Do you have trouble falling asleep at night or do you wake up in the middle of the night?” from which researchers coded respondents who responded as “usually” as insomnia cases while participants who responded “sometimes” or “never/rarely” were coded as controls. The authors report that this measurement of insomnia has been validated with similar survey items from the Netherlands Sleep Register (see Supplemental Note in Jansen et al. 2019 for more details). To the best of our knowledge Jansen and colleagues do not provide any information concerning if or how GWAS analyses adjusted for other covariates.

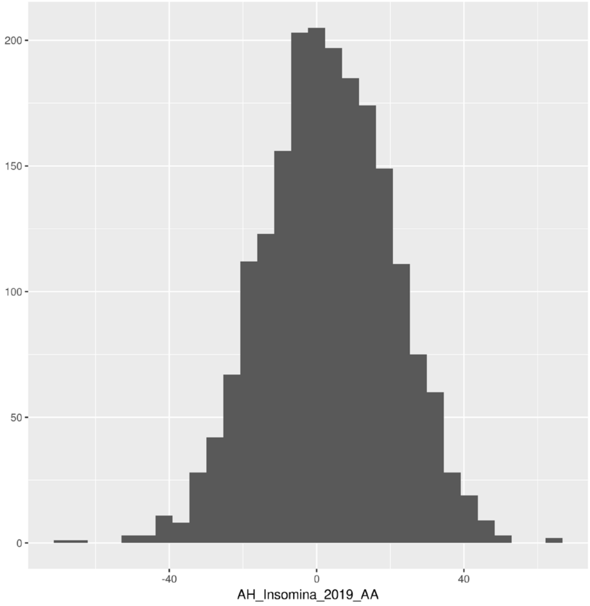
The Add Health PGSs for insomnia are standardized (mean = 0, standard deviation = 1) within ancestry groups to control for population stratification. To give researchers an idea of the underlying raw PGS, we present density plots for the unstandardized PGSs below. **Please read the section entitled “Using Add Health PGSs” in the introductory portion of this document prior to conducting any analyses using the provided PGSs.**

Insomnia

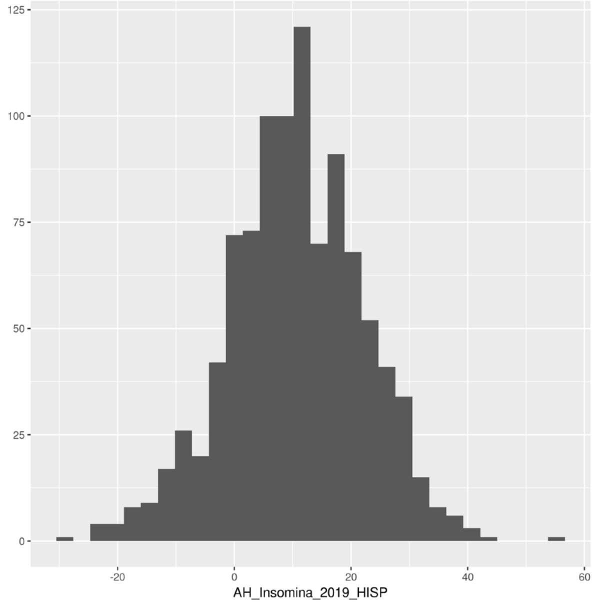
European ancestry raw PGS distribution



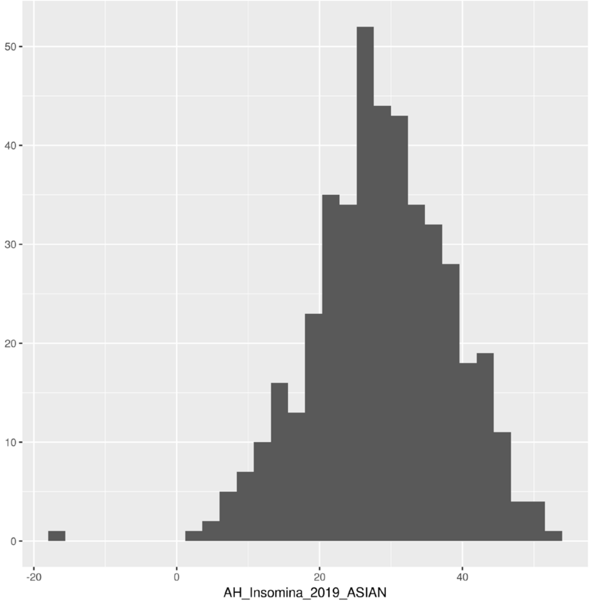
African ancestry raw PGS distribution



Hispanic ancestry raw PGS distribution



Asian ancestry raw PGS distribution



Daytime Sleepiness (Dozing)

GWAS Summary Statistic Source: Jansen et al. 2019. “Genome-Wide Analysis of Insomnia in 1,331,010 Individuals Identifies New Risk Loci and Functional Pathways.” *Nature Genetics* 51: 394-403.

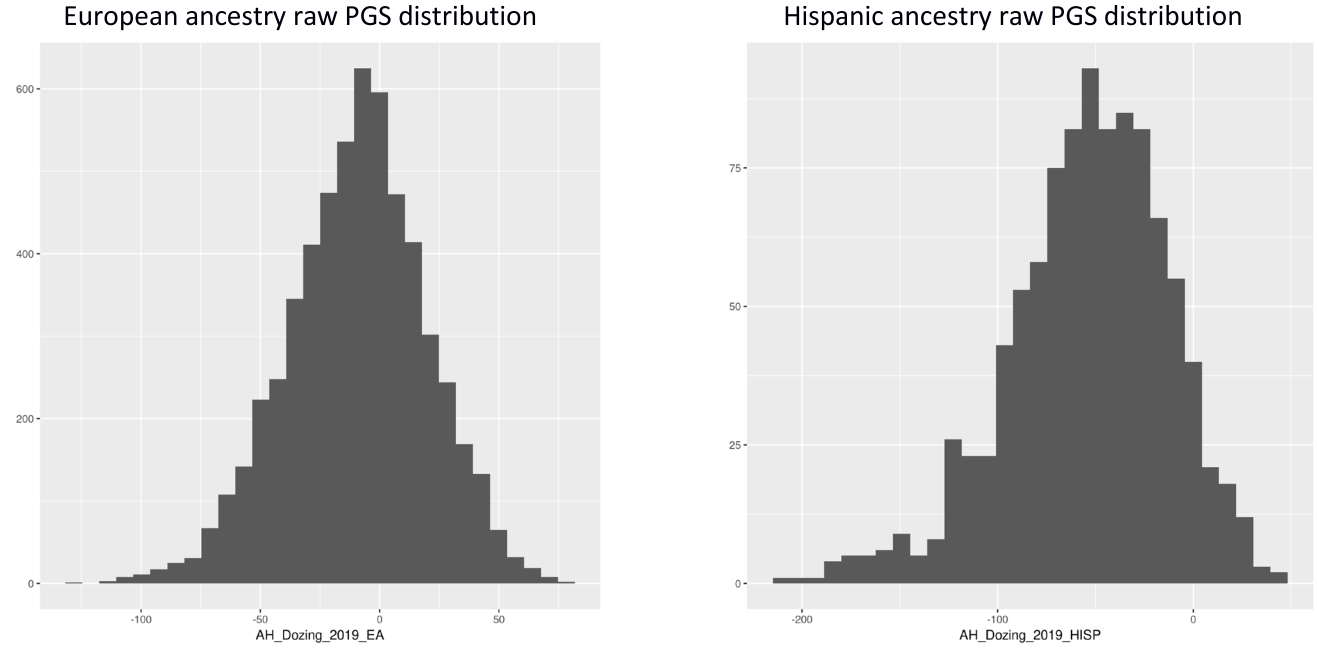
GWAS Ancestry Group(s): European

The PGSs for daytime sleepiness were created using results from a 2019 study conducted by the Philip Jansen and colleagues. The GWAS meta-analysis files are publicly available and can be downloaded from: https://ctg.cncr.nl/software/summary_statistics.

The daytime sleepiness (Dozing) GWAS analysis is based on 386,548 individuals of European genetic ancestry from the UK Biobank. Daytime sleepiness was measured by dichotomizing responses to the question “How likely are you to doze off or fall asleep during the daytime when you don’t mean to?” with responses of “often” and “all the time” as cases of daytime dozing and responses of “never/rarely” and “sometimes” as controls (see Supplemental Note in Jansen et al. 2019 for more details). To the best of our knowledge Jansen and colleagues do not provide any information concerning if or how GWAS analyses adjusted for other covariates.

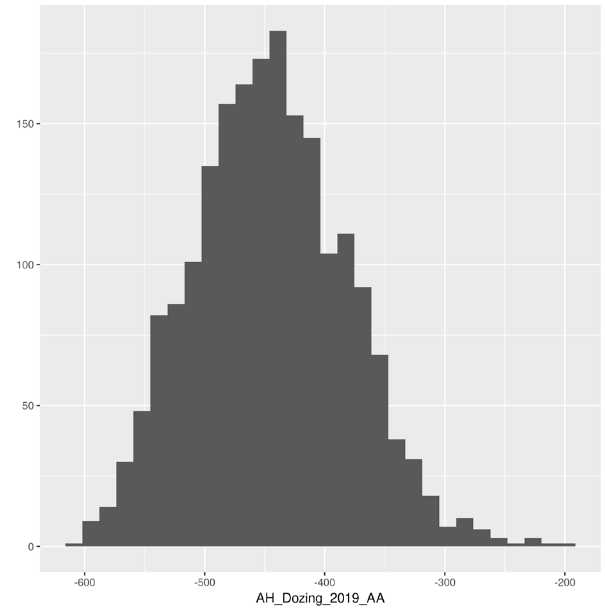
The Add Health PGSs for Dozing are standardized (mean = 0, standard deviation = 1) within ancestry groups to control for population stratification. To give researchers an idea of the underlying raw PGS, we present density plots for the unstandardized PGSs below. **Please read the section entitled “Using Add Health PGSs” in the introductory portion of this document prior to conducting any analyses using the provided PGSs.**

Dozing

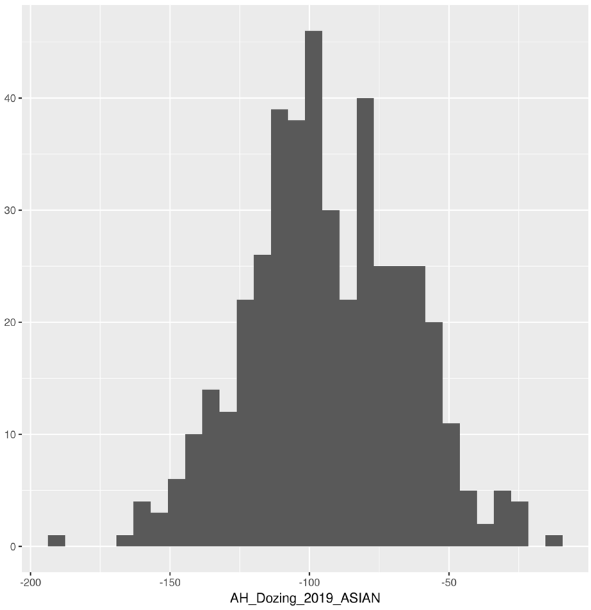

European ancestry raw PGS distribution

Hispanic ancestry raw PGS distribution

African ancestry raw PGS distribution



Asian ancestry raw PGS distribution



Add Health is a program project directed by Kathleen Mullan Harris and designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris at the University of North Carolina at Chapel Hill, and funded by grant P01-HD31921 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development, with cooperative funding from 23 other federal agencies and foundations. Information on how to obtain the Add Health data files is available on the Add Health website (http://www.cpc.unc.edu/addhealth).

Ease of getting up in the morning

GWAS Summary Statistic Source: Jansen et al. 2019. “Genome-Wide Analysis of Insomnia in 1,331,010 Individuals Identifies New Risk Loci and Functional Pathways.” *Nature Genetics* 51: 394-403.

GWAS Ancestry Group(s): European

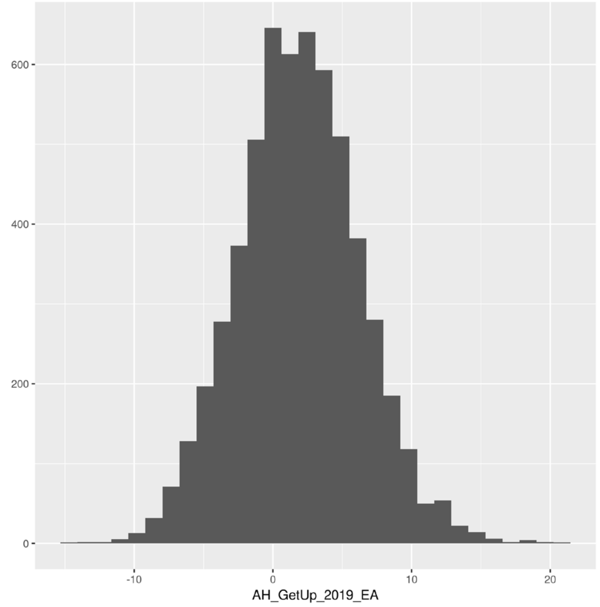
The PGSs for ease of getting up in the morning were created using results from a 2019 study conducted by the Philip Jansen and colleagues. The GWAS meta-analysis files are publicly available and can be downloaded from: https://ctg.cncr.nl/software/summary_statistics.

The GWAS of ease of getting up in the morning is based on the analysis of 385,949 individuals of European genetic ancestry from the UK Biobank. Ease of getting up was measured as a continuous variable based on responses to the question “On an average day, how easy do you find getting up in the morning?” with possible responses of “not at all easy”, “not very easy”, “fairly easy” and “very easy” (see Supplemental Note in Jansen et al. 2019 for more details). To the best of our knowledge Jansen and colleagues do not provide any information concerning if or how GWAS analyses adjusted for other covariates.

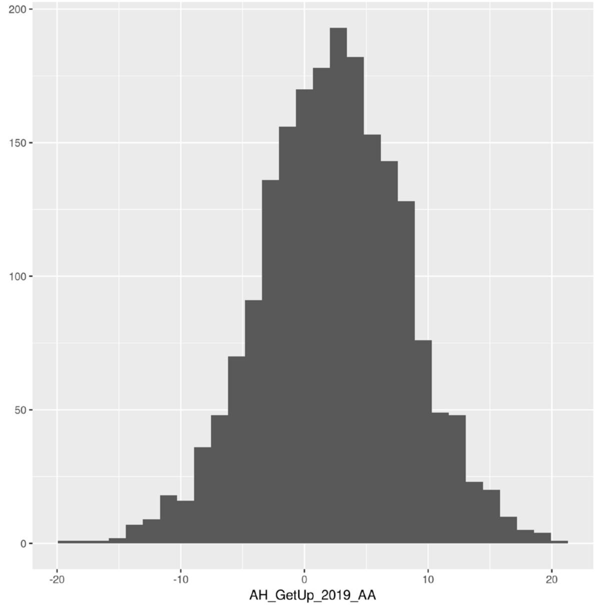
The Add Health PGSs for ease of getting up in the morning are standardized (mean = 0, standard deviation = 1) within ancestry groups to control for population stratification. To give researchers an idea of the underlying raw PGS, we present density plots for the unstandardized PGSs below. **Please read the section entitled “Using Add Health PGSs” in the introductory portion of this document prior to conducting any analyses using the provided PGSs.**

Ease of Getting Up

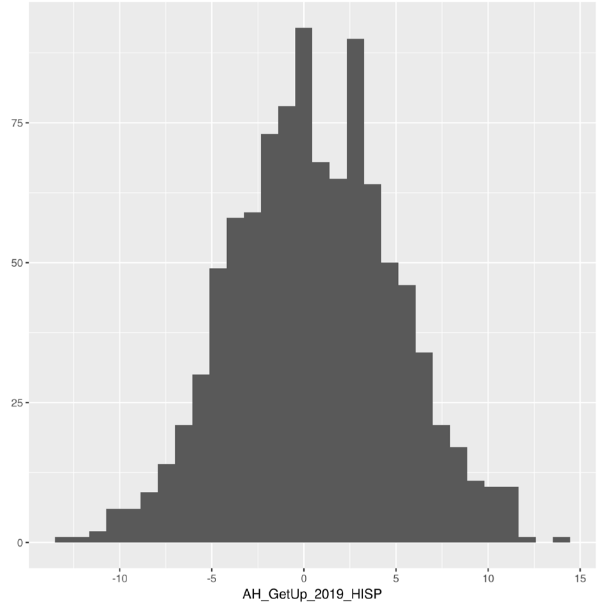
European ancestry raw PGS distribution



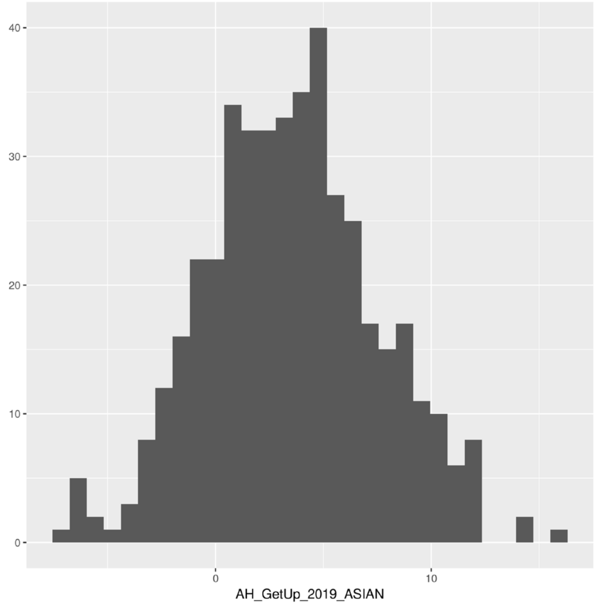
African ancestry raw PGS distribution



Hispanic ancestry raw PGS distribution



Asian ancestry raw PGS distribution



Add Health is a program project directed by Kathleen Mullan Harris and designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris at the University of North Carolina at Chapel Hill, and funded by grant P01-HD31921 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development, with cooperative funding from 23 other federal agencies and foundations. Information on how to obtain the Add Health data files is available on the Add Health website (http://www.cpc.unc.edu/addhealth).

Morning person

GWAS Summary Statistic Source: Jansen et al. 2019. “Genome-Wide Analysis of Insomnia in 1,331,010 Individuals Identifies New Risk Loci and Functional Pathways.” *Nature Genetics* 51: 394-403.

GWAS Ancestry Group(s): European

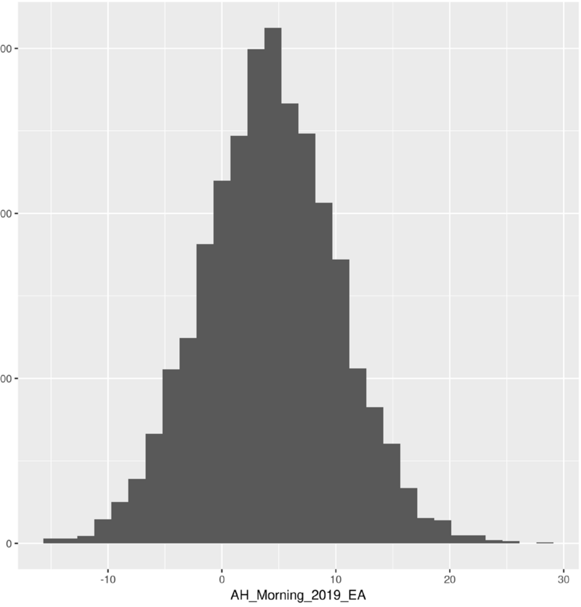
The PGSs for whether an individual is a morning person were created using results from a 2019 study conducted by the Philip Jansen and colleagues. The GWAS meta-analysis files are publicly available and can be downloaded from: https://ctg.cncr.nl/software/summary_statistics.

The GWAS of whether an individual is a morning person is based on the analysis of 345,552 individuals of European genetic ancestry from the UK Biobank. Being a morning person (as opposed to being an evening person) was measured as a continuous variable based on responses to the question “Do you consider yourself to be?”, with response options including “Definitely a ‘morning’ person”, “More a ‘morning’ than ‘evening’ person”, “More an ‘evening’ than a 145 ‘morning’ person”, and “Definitely an ‘evening’ person” (see Supplemental Note in Jansen et al. 2019 for more details). To the best of our knowledge Jansen and colleagues do not provide any information concerning if or how GWAS analyses adjusted for other covariates.

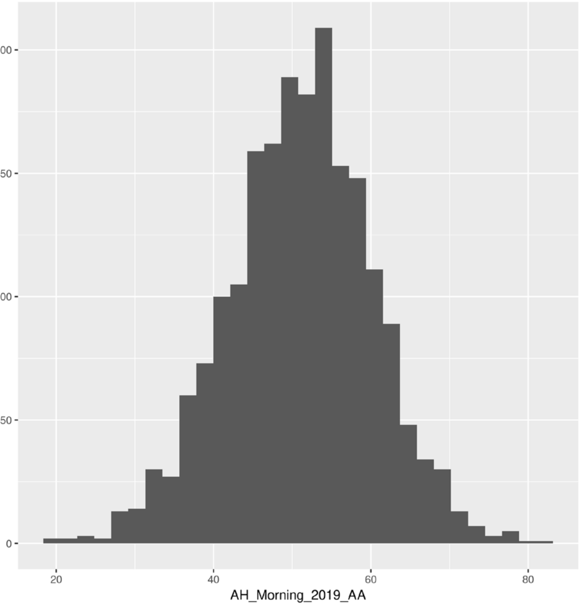
The Add Health PGSs for being a morning person are standardized (mean = 0, standard deviation = 1) within ancestry groups to control for population stratification. To give researchers an idea of the underlying raw PGS, we present density plots for the unstandardized PGSs below. **Please read the section entitled “Using Add Health PGSs” in the introductory portion of this document prior to conducting any analyses using the provided PGSs.**

Morning Person

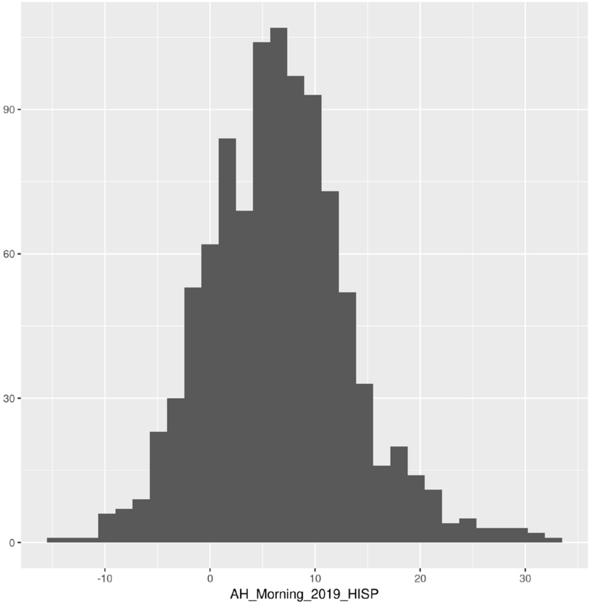
European ancestry raw PGS distribution



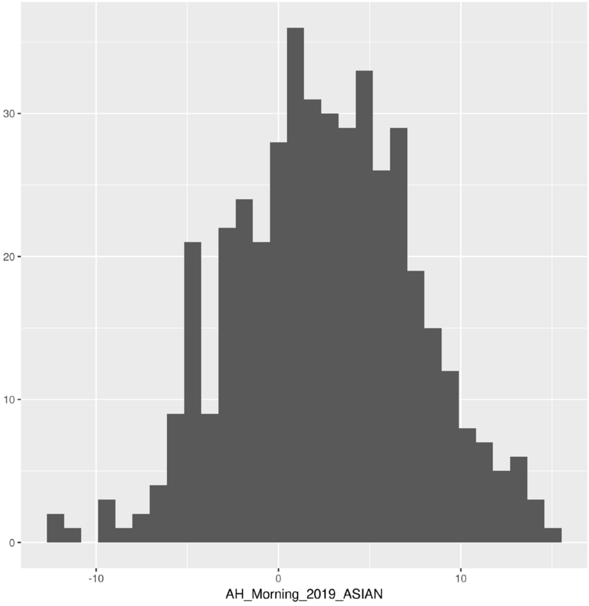
African ancestry raw PGS distribution



Hispanic ancestry raw PGS distribution



Asian ancestry raw PGS distribution



Daytime Napping

GWAS Summary Statistic Source: Jansen et al. 2019. “Genome-Wide Analysis of Insomnia in 1,331,010 Individuals Identifies New Risk Loci and Functional Pathways.” *Nature Genetics* 51: 394-403.

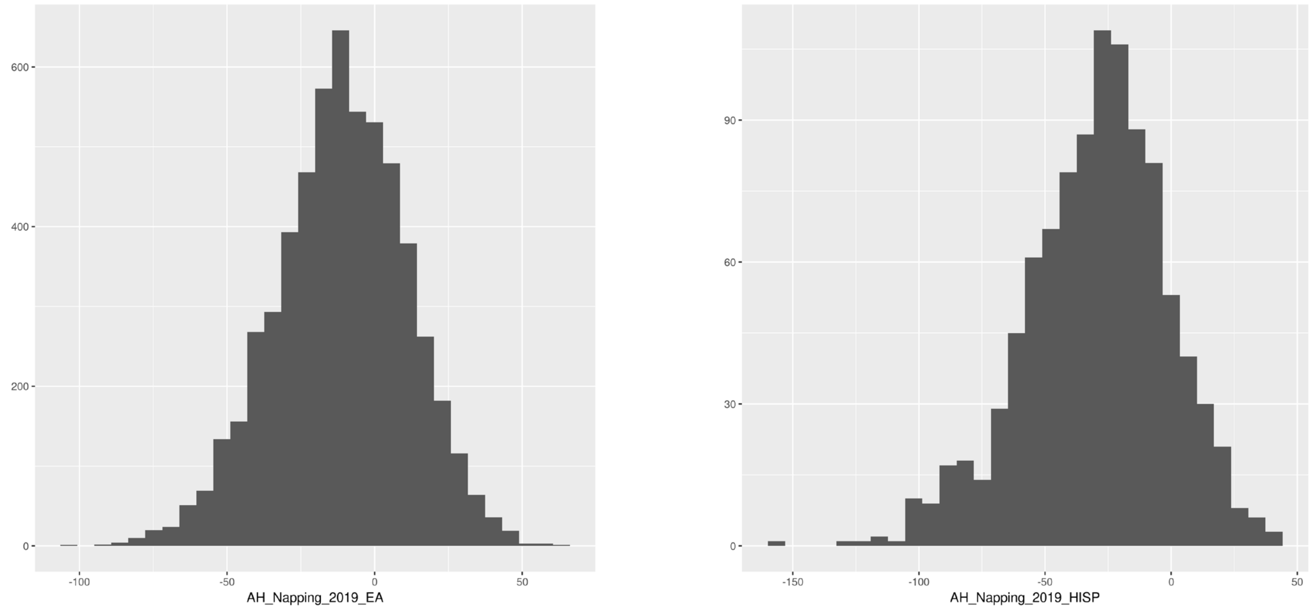
GWAS Ancestry Group(s): European

The PGSs for daytime napping were created using results from a 2019 study conducted by the Philip Jansen and colleagues. The GWAS meta-analysis files are publicly available and can be downloaded from: https://ctg.cncr.nl/software/summary_statistics.

The daytime napping GWAS is based on the analysis of 386,577 individuals of European genetic ancestry from the UK Biobank. Daytime napping was measured as a dichotomous variable based on responses to the question “Do you have a nap during the day?”, with responses of “usually” coded as cases and responses of “never/rarely” or “sometimes” coded as controls (see Supplemental Note in Jansen et al. 2019 for more details). To the best of our knowledge Jansen and colleagues do not provide any information concerning if or how GWAS analyses adjusted for other covariates.

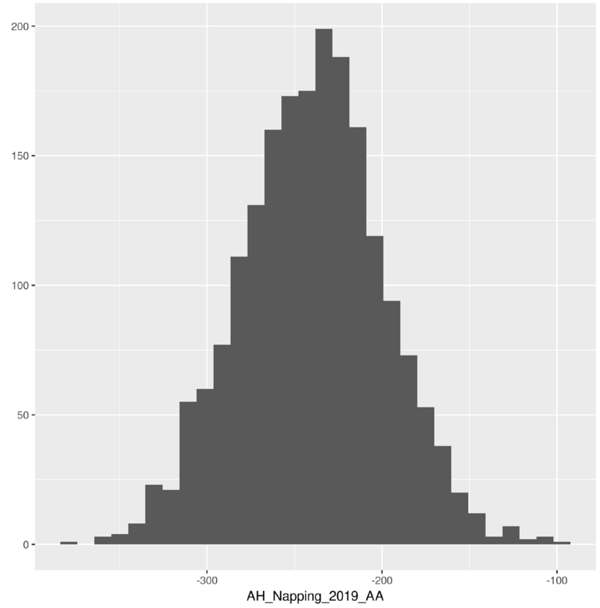
The Add Health PGSs for daytime napping are standardized (mean = 0, standard deviation = 1) within ancestry groups to control for population stratification. To give researchers an idea of the underlying raw PGS, we present density plots for the unstandardized PGSs below. **Please read the section entitled “Using Add Health PGSs” in the introductory portion of this document prior to conducting any analyses using the provided PGSs.**

Daytime Napping

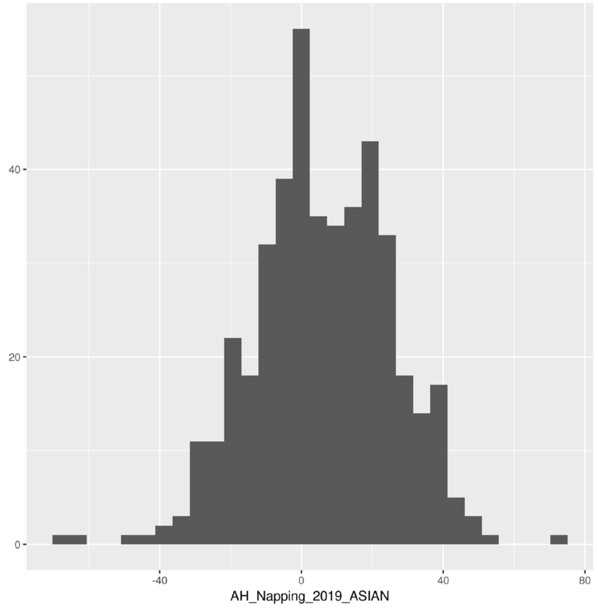


European ancestry raw PGS distribution

Hispanic ancestry raw PGS distribution



African ancestry raw PGS distribution



Asian ancestry raw PGS distribution

Sleep Duration

GWAS Summary Statistic Source: Jansen et al. 2019. “Genome-Wide Analysis of Insomnia in 1,331,010 Individuals Identifies New Risk Loci and Functional Pathways.” *Nature Genetics* 51: 394-403.

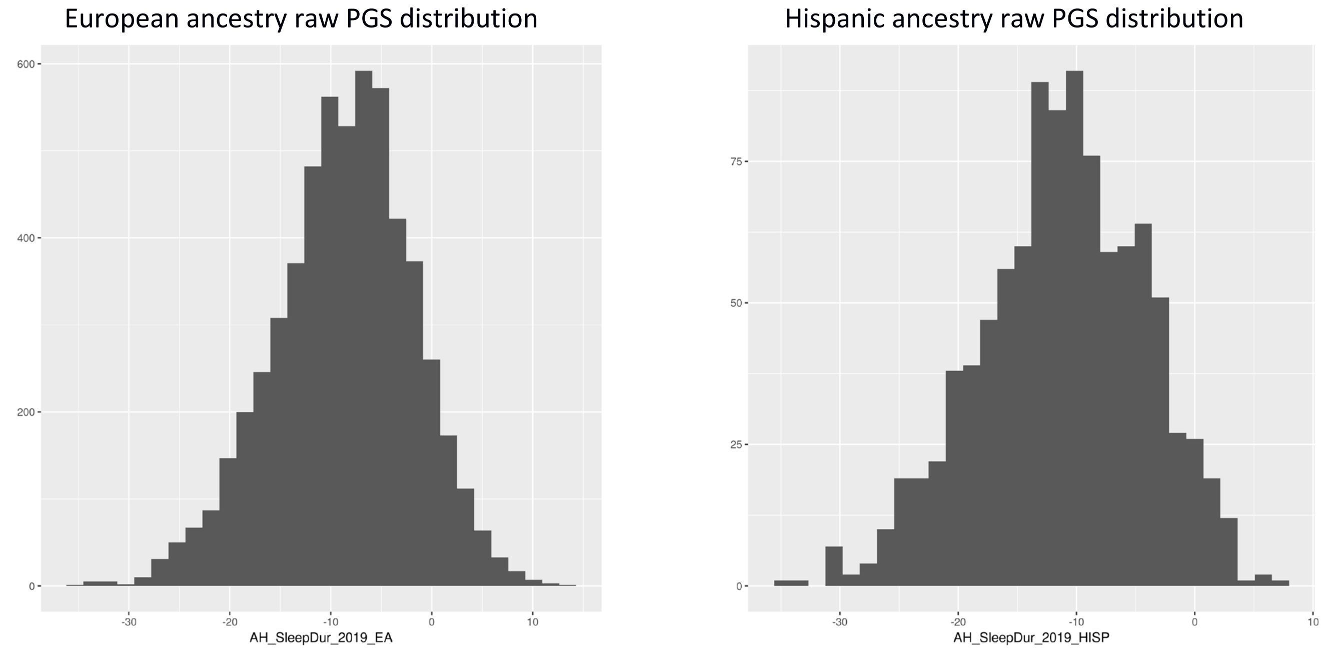
GWAS Ancestry Group(s): European

The PGSs for sleep duration were created using results from a 2019 study conducted by the Philip Jansen and colleagues. The GWAS meta-analysis files are publicly available and can be downloaded from: https://ctg.cncr.nl/software/summary_statistics.

The sleep duration GWAS is based on the analysis of 384,317 individuals of European genetic ancestry from the UK Biobank. Sleep duration was measured as a continuous variable based on responses to the question, “About how many hours sleep do you get in every 24 hours? (please include naps)" with answers rounded to the nearest integer (see Supplemental Note in Jansen et al. 2019 for more details). To the best of our knowledge Jansen and colleagues do not provide any information concerning if or how GWAS analyses adjusted for other covariates.

The Add Health PGSs for sleep duration are standardized (mean = 0, standard deviation = 1) within ancestry groups to control for population stratification. To give researchers an idea of the underlying raw PGS, we present density plots for the unstandardized PGSs below. **Please read the section entitled “Using Add Health PGSs” in the introductory portion of this document prior to conducting any analyses using the provided PGSs.**

Sleep Duration

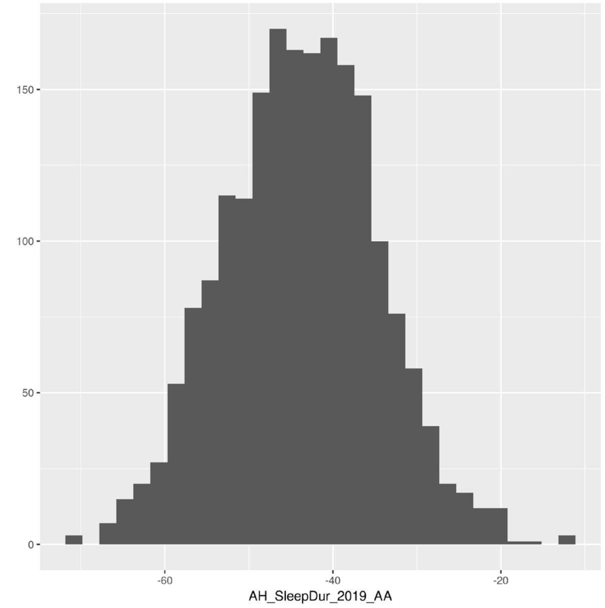


European ancestry raw PGS distribution

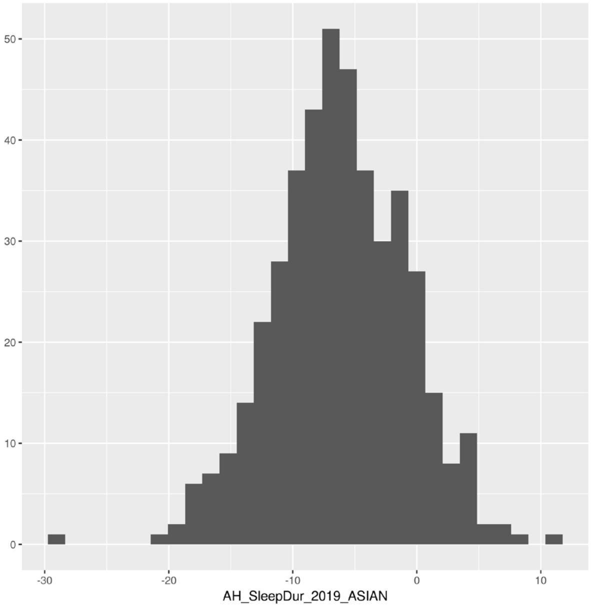
Hispanic ancestry raw PGS distribution

Add Health is a program project directed by Kathleen Mullan Harris and designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris at the University of North Carolina at Chapel Hill, and funded by grant P01-HD31921 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development, with cooperative funding from 23 other federal agencies and foundations. Information on how to obtain the Add Health data files is available on the Add Health website (http://www.cpc.unc.edu/addhealth).

African ancestry raw PGS distribution



Asian ancestry raw PGS distribution



Snoring

GWAS Summary Statistic Source: Jansen et al. 2019. “Genome-Wide Analysis of Insomnia in 1,331,010 Individuals Identifies New Risk Loci and Functional Pathways.” *Nature Genetics* 51: 394-403.

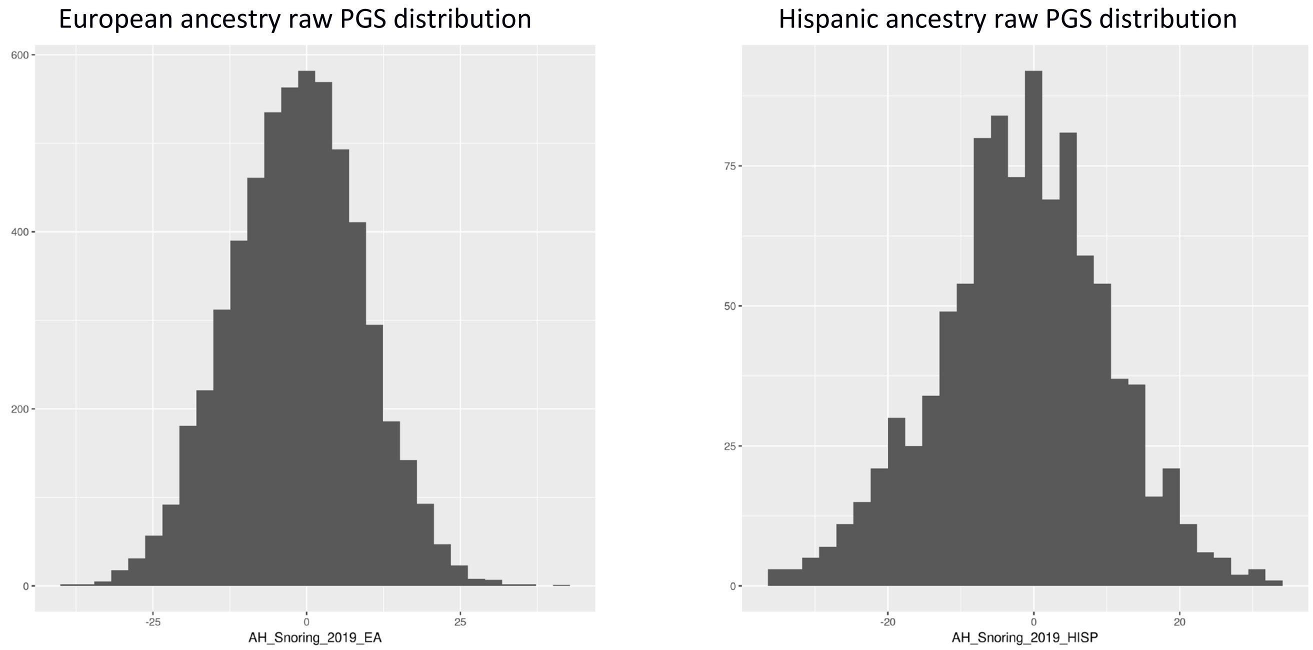
GWAS Ancestry Group(s): European

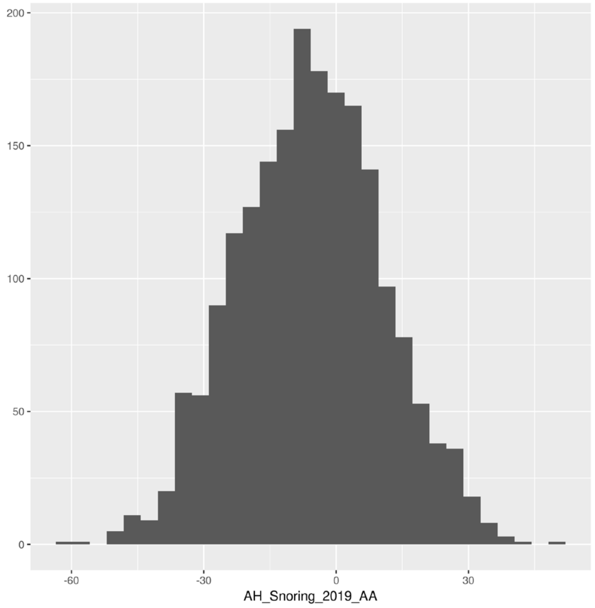
The PGSs for snoring were created using results from a 2019 study conducted by the Philip Jansen and colleagues. The GWAS meta-analysis files are publicly available and can be downloaded from: https://ctg.cncr.nl/software/summary_statistics.

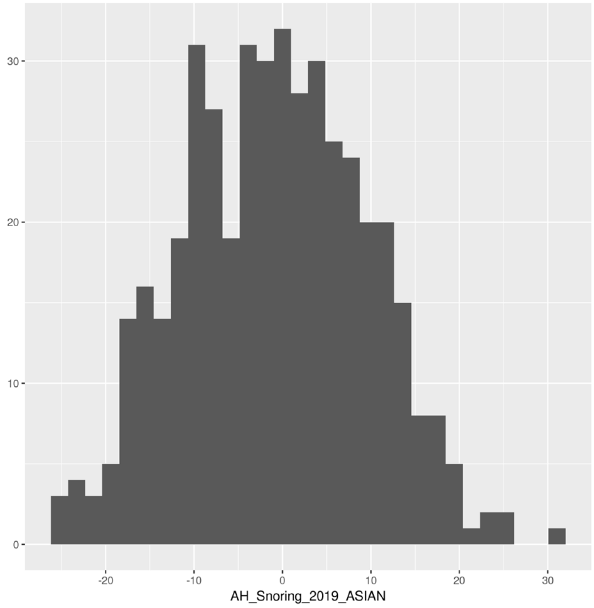
The snoring GWAS is based on the analysis of 359,916 individuals of European genetic ancestry from the UK Biobank. Snoring was measured as a dichotomous outcome based on responses to the question, "Does your partner or a close relative or friend complain about your snoring?" (see Supplemental Note in Jansen et al. 2019 for more details). To the best of our knowledge Jansen and colleagues do not provide any information concerning if or how GWAS analyses adjusted for other covariates.

The Add Health PGSs for snoring are standardized (mean = 0, standard deviation = 1) within ancestry groups to control for population stratification. To give researchers an idea of the underlying raw PGS, we present density plots for the unstandardized PGSs below. **Please read the section entitled “Using Add Health PGSs” in the introductory portion of this document prior to conducting any analyses using the provided PGSs.**

Snoring




African ancestry raw PGS distribution


Asian ancestry raw PGS distribution

Add Health is a program project directed by Kathleen Mullan Harris and designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris at the University of North Carolina at Chapel Hill, and funded by grant P01-HD31921 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development, with cooperative funding from 23 other federal agencies and foundations. Information on how to obtain the Add Health data files is available on the Add Health website (http://www.cpc.unc.edu/addhealth).