NATIONAL HEALTH AND AGING TRENDS STUDY (NHATS)
Polygenic Scores User Guide

March 2024

Suggested Citation:

Ware EB, Skehan ME, Kasper JD, Freedman VA, Schrack JA. 2024. National Health and Aging Trends Study Polygenic Scores User Guide. Baltimore: Johns Hopkins Bloomberg School of Public Health. Available at www.NHATS.org. This work is supported by U01AG032947 from the National Institute on Aging.

Contents

I. Introduction

This guide describes the construction of polygenic scores (PGSs) for a variety of phenotypes for NHATS respondents who provided dried blood spot-based DNA. These scores serve as an attempt to harmonize research across studies and facilitate use among NHATS data users by making these scores available as sensitive data. PGSs for each phenotype are based on a single, replicated genome-wide association study (GWAS). This document describes the general method of construction with details on each phenotype included as appendices.

Rationale

Complex health outcomes or behaviors of interest to the research community are often highly polygenic or reflect the aggregate effect of many different genes so the use of single genetic variants or candidate genes may not capture the dynamic nature of more complex phenotypes. A PGS aggregates thousands to millions of individual loci across the human genome and weights them by effect sizes derived from a GWAS as an estimate of the strength of their association to produce a single quantitative measure of genetic risk and to increase power in genetic analyses.

II. Overview of NHATS Data Collection

Begun in 2011, the National Health and Aging Trends Study (NHATS) conducts annual in-person interviews with a nationally representative sample of Medicare beneficiaries ages 65 or older. Designed as a platform for scientific study of late-life disability trends and trajectories, NHATS fosters research to reduce disability, maximize independent functioning, and enhance quality of life at older ages [1].

Content areas include: the physical, social, technological and service environments; tests and self-reports of physical and cognitive capacity; use of assistive devices and rehabilitation services; help received with daily (self-care, mobility, household, and medical) activities; participation in valued activities; and wellbeing. Other topics focus on chronic conditions, symptoms, sensory impairments, transportation, subjective and economic wellbeing, and demographic factors. A last month of life interview focuses on quality of end-of-life care and a facility interview is conducted for those living in residential care settings. Links to Medicare records are also available. Caregivers of NHATS participants are interviewed occasionally in the companion National Study of Caregiving.

NHATS is led by the Johns Hopkins University Bloomberg School of Public Health and the University of Michigan's Institute for Social Research, with data collection by Westat. Support for NHATS is provided by the National Institute on Aging (U01AG032947).

NHATS is designed to provide a nationally representative cross-section of the Medicare population ages 65 and older at regular intervals. Oversamples by age and for Black non-Hispanic persons are embedded in the sample design [2]. In Round 5 (2015), a new sample was introduced to restore the sample to original size by age and race groups [3].

A dried blood spot (DBS) collection in Round 7 (2017) provided the biological material for genotyping. All participants in that round with a completed Sample Person interview were

considered eligible for the dried blood spot (DBS) collection. However, self-response was required for the DBS consent process so a small percentage who had proxy respondents were not invited to participate in the DBS collection. 4,903 (93.1%) provided consent to collect.

In all, 4,691 persons (95.7%) of those who consented had at least 1 card with DBS sample(s) available for assaying [4]. After several assays were conducted, 4,091 NHATS participants had genetic material available and had given permission for genotyping to be conducted.

III. Genotyping process

The 4,091 samples were genotyped at Erasmus Medical Center in Rotterdam, Netherlands. Samples were genotyped on the Illumina Infinium Global Screening Array v3.0. The array contains clinical and rare variants ideal for multiethnic populations. Additional information is provided through the Erasmus MC Human Genomics Facility HUGE-F website: http://glimdna.org/global-screening-array.html. In total, 725,831 SNPs are included in the dataset. Internal quality control (QC) methods for SNPs are described in the NHATS Quality Control and Imputation Report for Genotypic Data [5].

IV. Population structure and homogenous analytic groups

Please refer to the NHATS Quality Control and Imputation Report for Genotypic Data [5] for a full discussion on how analytic groups were identified. An overall summary of the quality control steps can be found in **Figure 1**. The European analytic group contains 2,827 individuals and the African ancestry group contains 664 individuals; see the decision flow chart in **Figure 2**.

Figure 1. Summary of the quality control process steps, National Health and Aging Trends Study genetic sample

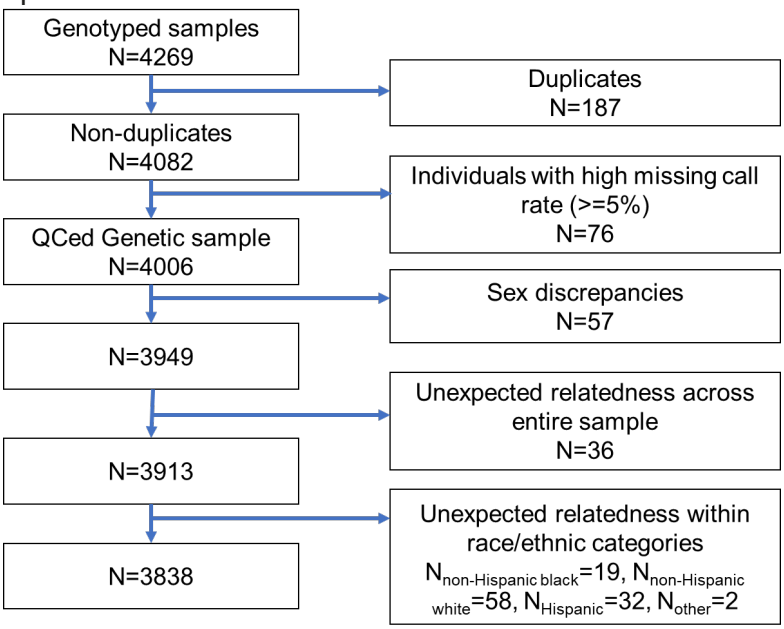
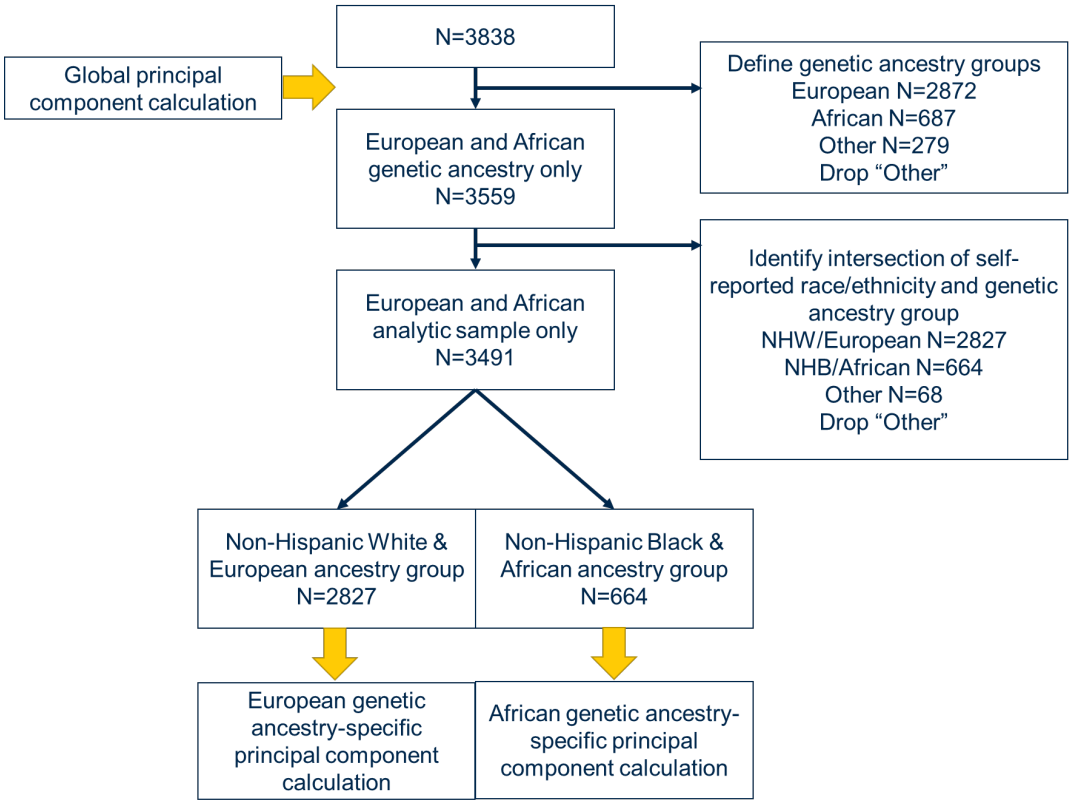


Figure 2. Decision flow chart to identify genetic analytic samples, National Health and Aging Trends Study



When performing analyses, we recommend doing so separately by analytic group (African, European) and controlling for local genetic principal components (those calculated within analytic group). The provided principal components should not be interpreted on their own and are included as precision variable.

In the African analytic group we recommend adjusting for at least two genetic principal components (**Figure 3**) in analyses and at least the first five European analytic group genetic principal components in the European analytic group (**Figure 4**).

Figure 3. Scatter plot matrix of the first five genetic principal components calculated within the analytic non-Hispanic Black/African genetic ancestry sample in the National Health and Aging Trends Study, n=664.

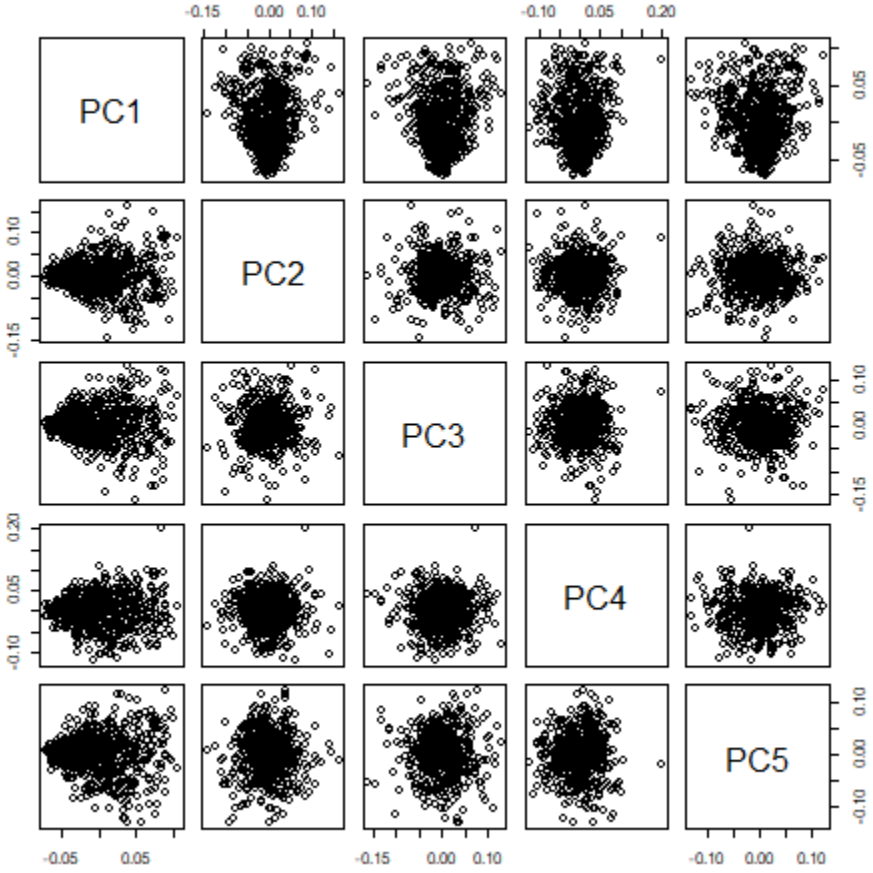
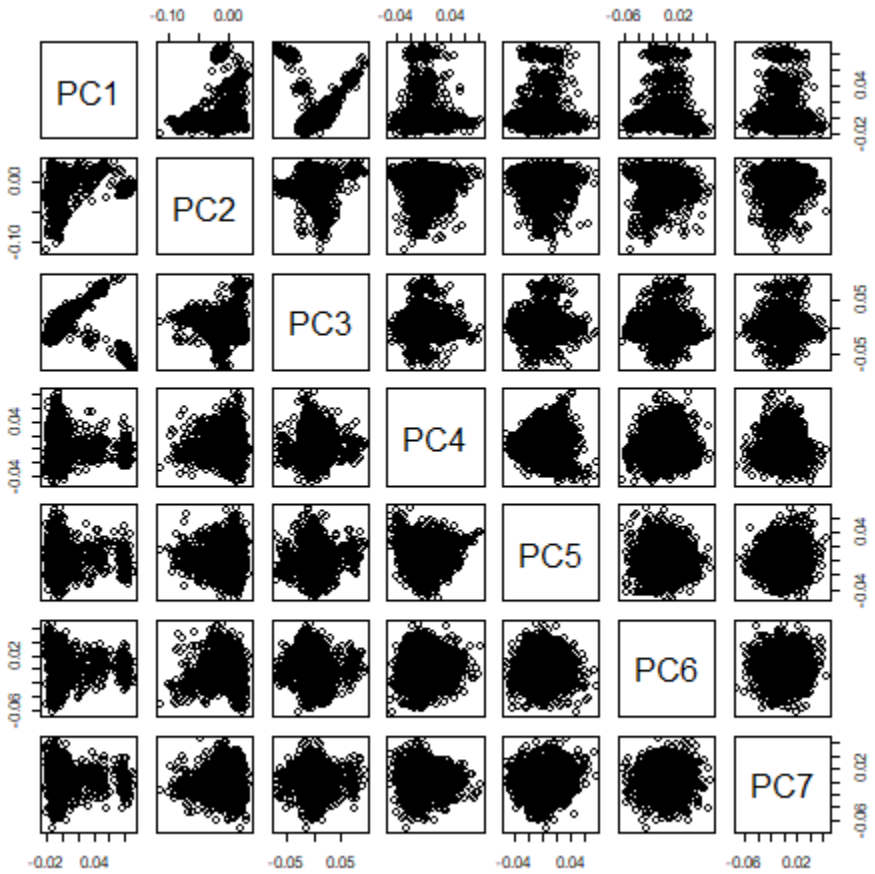


Figure 4. Scatter plot matrix of the first seven genetic principal components calculated within the analytic non-Hispanic White/European genetic ancestry sample in the National Health and Aging Trends Study, n=2827.



V. PGS Construction

While conceptually simple, there are numerous ways to estimate PGSs, not all achieving the same end goals. We followed recommendations by the Health and Retirement Study (HRS), whose researchers examined the predictive ability as well as the variability and co-variability in PGSs arising from different estimation approaches.[1]

Overall, results from these analyses concluded that including all available SNPs in a PGS (i.e. not accounting for any linkage disequilibrium or p-value thresholding) either demonstrated the largest predictive power (incremental R^2) of the score or produced a score that did not differ significantly from scores with similar predictive power that employed some degree of LD trimming or p-value thresholding. We have decided to use these methods to provide scores that include all available SNPs in the PGS that overlap between the GWAS meta-analysis and the NHATS genetic data.

---

[1] For additional information on this analysis, see:
Ware EB, Schmitz LL, Faul JD, Gard AM, Smith JA, Mitchell CM, Weir DR, Kardia SLR. (2017) *Method of Construction Affects Polygenic Score Prediction of Common Human Traits*. BiorXiv. 2017 doi: https://doi.org/10.1101/106062

Weighted sums were chosen to calculate the PGSs. Weights were defined by the odds ratio or beta estimate from the GWAS meta-analysis files corresponding to the phenotype of interest. PGSs are calculated using the following formula:

$$PGS_i = \sum_{j=1}^{J} W_j G_{ij}$$

where *i* is individual *i* (i=1 to N), *j* is SNP *j* (j=1 to J), *W* is the meta-analysis effect size for SNP *j* and *G* is the genotype, or the number of reference alleles (zero, one, or two), for individual *i* at SNP *j*. Due to the long-range linkage disequilibrium in this region, making linkage equilibrium difficult to obtain, the MHC region on chromosome 6 (26-33Mb) was omitted from all PGSs. PGSs were calculated using PRSIce2 (Choi et al. 2019).

Sources for SNP weights

To incorporate externally valid SNP weights from replicated GWAS, we performed a search of the literature to identify large GWAS meta-analysis studies related to the selected phenotype. SNP weights were downloaded from consortium webpages, requested from consortium authors, obtained from dbGap, or taken from published supplemental material. All base SNP files from GWAS meta-analyses were converted to NCBI build 37 annotation for compatibility with NHATS SNP data.

Notes about the use of PGSs

PGSs are released for both the European ancestry and African ancestry groups, separately. **However, it should be noted that the majority of GWAS used to inform the SNP weights come from GWAS on European ancestry groups and, as a result, PGSs for other ancestry groups may not have the same predictive capacity (Martin et al. 2017; Ware et al. 2017).**

To control for confounding from population stratification, or to account for any ancestry differences in genetic structures within populations that could bias estimates, **we highly recommend that users perform analyses separately by ancestral group and, at the very least, adjust for two ancestry-specific genetic principal components in the African ancestry group and five ancestry-specific genetic principal components in the European ancestry group**. The PCs control for any genetic aspects of common ancestry that could be spuriously correlated with the PGS and the outcome of interest (Price et al., 2006).

Polygenic scores were calculated at eight different P-value thresholds: 1, 0.5, 0.1, 0.05, 0.01, 0.001, $1x10^{-6}$, $5x10^{-8}$. If the GWAS used to inform the SNP weights did not find genome-wide suggestive (P-value < $1x10^{-6}$) or genome-wide significant (P-value < $5x10^{-8}$) associations, those scores were not calculable and therefore a set of polygenic scores may contain less than the eight thresholds. Which P-value threshold is used in analysis is at the user's discretion, though testing several thresholds is common.

References

Choi SW, and O'Reilly PF. "PRSice-2: Polygenic Risk Score Software for Biobank-Scale Data." GigaScience 8, no. 7 (July 1, 2019)

Martin, A. R., Gignoux, C. R., Walters, R. K., Wojcik, G. L., Neale, B. M., Gravel, S., ... & Kenny, E. E. (2017). Human demographic history impacts genetic risk prediction

across diverse populations. *The American Journal of Human Genetics*, *100*(4), 635-649.
Price, A. L., Patterson, N. J., Plenge, R. M., Weinblatt, M. E., Shadick, N. A., & Reich, D. (2006). Principal components analysis corrects for stratification in genome-wide association studies. *Nature genetics, 38*(8), 904-909.
Ware EB, Schmitz LL, Faul JD, Gard AM, Smith JA, Mitchell CM, Weir DR, Kardia SLR. (2017). Method of Construction Affects Polygenic Score Prediction of Common Human Traits. BiorXiv. 2017 doi: https://doi.org/10.1101/106062

VI. Polygenic score source information and distribution

a. Body Mass Index 2 (BMI2) – Genetic Investigation of ANthropometric Traits 2018

PGSs for body mass index (BMI) were created using results from a 2018 study conducted by the GIANT consortium. The GWAS meta-analysis files are publicly available on the Broad Institute data download page: (https://portals.broadinstitute.org/collaboration/giant/index.php/GIANT_consortium_data_files#2018_GIANT_and_UK_BioBank_Meta-analysis). The meta-analysis included 681,275 participants from a total of 15 cohorts of European ancestry. The 15 cohorts include the UK Biobank (UKB) and 14 cohorts from the previous GIANT GWAS of BMI (Locke et al., 2015; *Nature*). Authors performed a fixed effect inverse-variance weighted meta-analysis of the UKB results with GWAS summary statistics from Locke et al. (2015). 2,334,002 SNPs imputed from NCBI Build 37 HapMap phase 2 data were included in the meta-analysis. The GWAS of BMI in UKB was conducted in 456,426 participants of European Ancestry, using 16,653,239 SNPs imputed to the Haplotype Reference Consortium imputation reference panel. Associations were adjusted for 10 principal components to reduce confounding by population stratification, as well as for age, sex, recruitment center, and genotyping batch. The study identified 941 genome-wide significant SNPs ($P < 10^{-8}$) (**Figure 1** and **Table 1**).

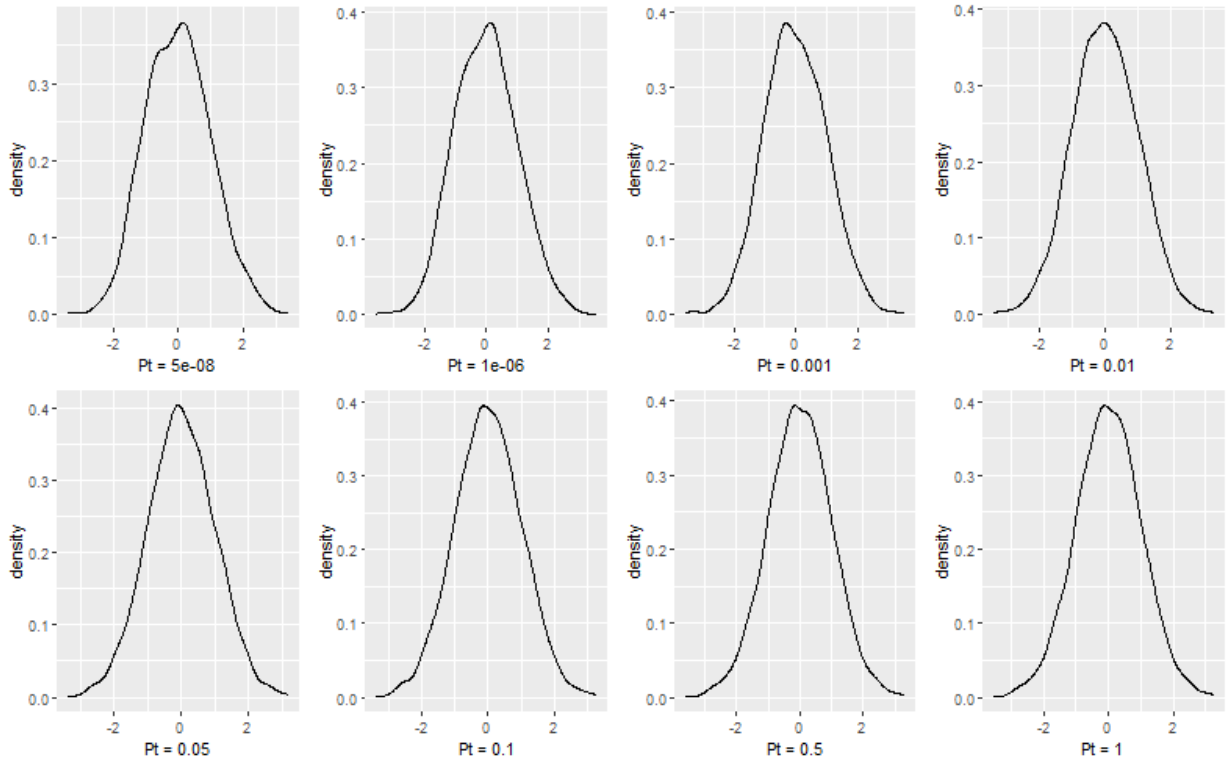
The PGSs contain 201,385 SNPs that overlapped between the NHATS genetic database for each analytic group and the GWAS meta-analysis. The posted PGSs have been standardized within genetic ancestry analysis group to a standard normal curve (mean=0, standard deviation = 1). Higher PGSs correspond to increased BMI.

N.B.: **Please note that the GIANT-BMI2 summary statistics are from a GWAS on individuals of European ancestry (see "Notes about the use of PGSs" for more information on the use of PGSs in other ancestry groups).**

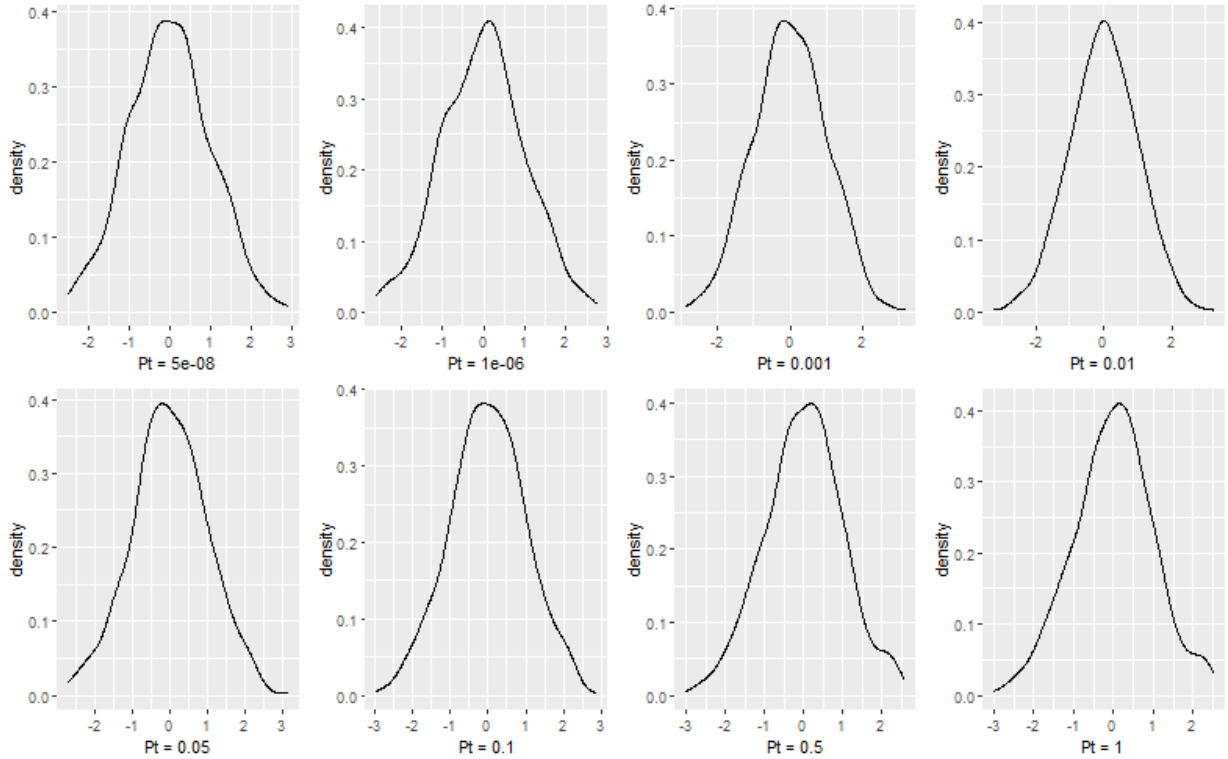
References

Yengo L, Sidorenko J, Kemper KE, et al. Meta-analysis of genome-wide association studies for height and body mass index in ~700000 individuals of European ancestry. Hum Mol Genet. 2018;27(20):3641–3649. doi:10.1093/hmg/ddy271

Distribution of polygenic scores by p-value threshold, NHATS European ancestry, n=2827



Distribution of polygenic scores by p-value threshold, NHATS African ancestry, n=664



b. Height 2 (Height2) – Genetic Investigation of ANthropometric Traits 2018

PGSs for Height were created using results from a 2018 study conducted by the GIANT consortium. The GWAS meta-analysis files are publicly available on the Broad Institute data download page: (https://portals.broadinstitute.org/collaboration/giant/index.php/GIANT_consortium_data_files#2018_GIANT_and_UK_BioBank_Meta-analysis). The GIANT meta-analysis included 693,529 participants from a total of 6 cohorts of European ancestry. The 6 cohorts include the UK Biobank (UKB) and 5 cohorts from the previous GIANT GWAS of Height (Wood et al., 2014; *Nature Genetics*). Authors performed a fixed effect inverse-variance weighted meta-analysis of the UKB results with GWAS summary statistics from Wood et al. (2014). NCBI Build 37 HapMap phase 2 data were used in the GWAS. The GWAS of Height in UKB was conducted in 456,426 participants of European Ancestry, using 16,653,239 SNPs imputed to the Haplotype Reference Consortium imputation reference panel. Associations were adjusted for 10 principal components to reduce confounding by population stratification, as well as for age, sex, recruitment center, and genotyping batch. The study identified 3,290 genome-wide significant SNPs ($P < 10^{-8}$) (**Figure 1** and **Table 1**).

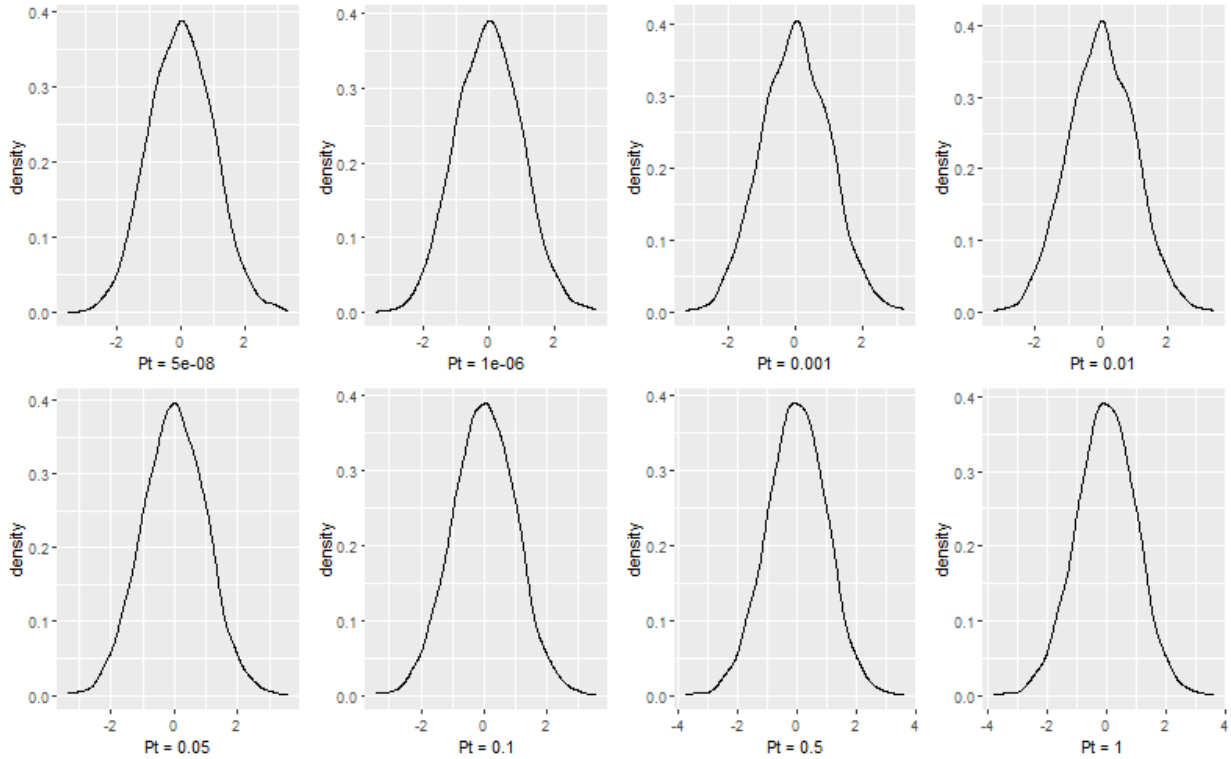
The PGSs contain 201,170 SNPs that overlapped between the NHATS genetic database for each analytic group and the GWAS meta-analysis. The posted PGSs have been standardized within genetic ancestry analysis group to a standard normal curve (mean=0, standard deviation = 1). Higher PGSs correspond to increased height.

N.B.: **Please note that the GIANT Height2 summary statistics are from a GWAS on individuals of European ancestry (see "Notes about the use of PGSs" for more information on the use of PGSs in other ancestry groups).**

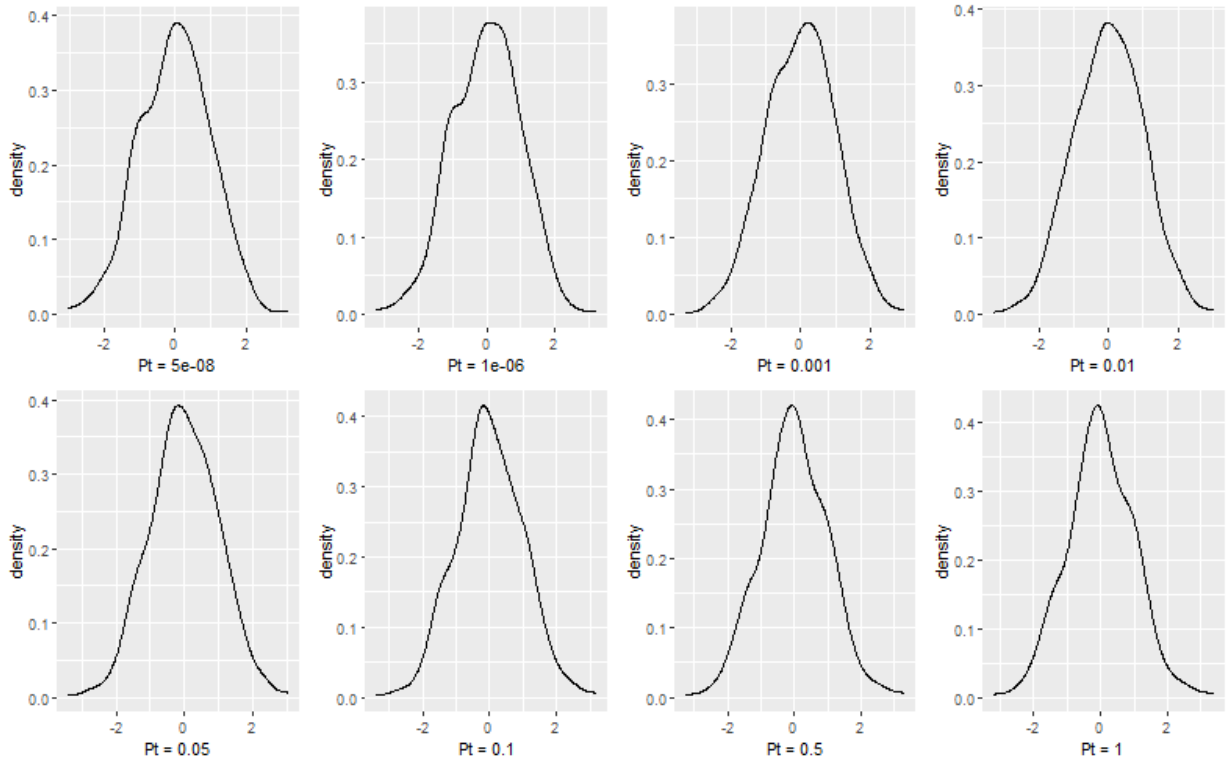
References

Yengo L, Sidorenko J, Kemper KE, et al. Meta-analysis of genome-wide association studies for height and body mass index in ~700000 individuals of European ancestry. Hum Mol Genet. 2018;27(20):3641–3649. doi:10.1093/hmg/ddy271

Distribution of polygenic scores by p-value threshold, NHATS European ancestry, n=2827



Distribution of polygenic scores by p-value threshold, NHATS African ancestry, n=664



c-g. Smoking behaviors – GWAS & Sequencing Consortium of Alcohol and Nicotine 2019

The substance behaviors PGSs were created using results from a 2019 study by the GWAS & Sequencing Consortium of Alcohol and Nicotine use (GSCAN) consortium. PGSs were created for smoking initiation (SI), cigarettes per day (CPD), smoking cessation (SC), age of initiation of regular smoking (AI), and drinks per day (DPW). SI was a binary phenotype indicating whether an individual had ever smoked regularly; the SI meta-analysis included 1,232,091 individuals of European ancestry. CPD captured heaviness of smoking, either as a current or former smoker; the CPD meta-analysis included 337,334 individuals of European ancestry. SC was a binary phenotype indicating whether an individual was a current smoker or former smoker, with never smokers coded as missing; the SC meta-analysis included 547,219 individuals of European ancestry. AI was measured as the age when an individual started smoking regularly; the AI meta-analysis included 341,427 individuals of European ancestry. DPW was defined as the average number of drinks a participant reported drinking each week, aggregated across all types of alcohol; the DPW meta-analysis 941,280 individuals of European ancestry. Genome-wide significant SNPs were identified in all meta-analyses: SI (N = 378), CPD (N = 55), SC (N = 24), AI (N = 10), and DPW (N = 99) (**Supplementary Tables 1 – 5; Supplemental Figures 2 – 12**). 23andMe results have been removed from the summary statistics (due to data use agreements). In the original analysis, 23andMe contributed 78,437 (SI), 73,380 (CPD), 403,931 (DPW), 234,398 (SC), 599,289 (SI).

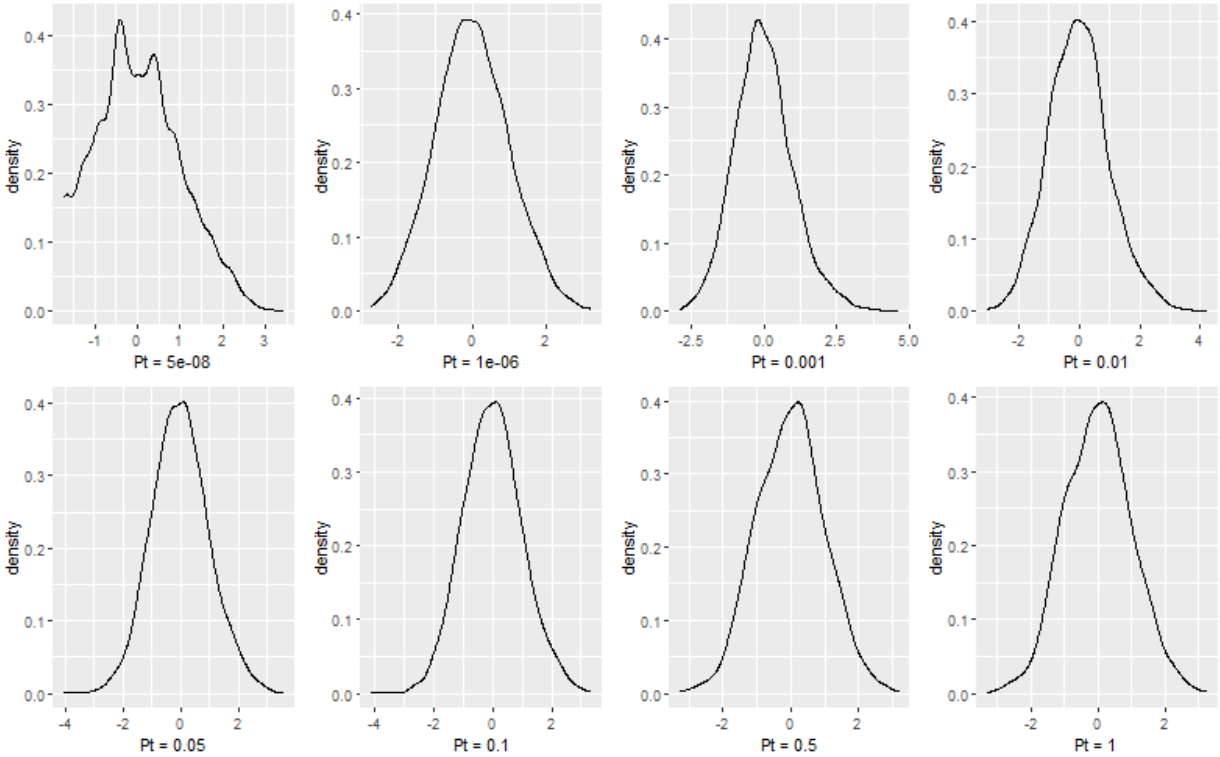
The PGSs (for both African and European ancestry participants) contains 505,549 (age at smoking initiation, AI), 505,552 (cigarettes per day, CPD), 506,501 (drinks per week, DPW), 502,367 (smoking cessation, SC), and 504,587 (smoking initiation, SI) SNPs that overlapped between the NHATS genetic database for each analytic group and each specific GWAS. Higher PGSs correspond to increased age at smoking initiation, cigarettes per day or drinks per week, or odds of smoking cessation or initiation. The posted PGSs have been standardized within genetic ancestry analysis group to a standard normal curve (mean=0, standard deviation = 1).

N.B.: **Please note that the GSCAN smoking behaviors summary statistics are from a GWAS on individuals of European ancestry (see "Notes about the use of PGSs" for more information on the use of PGSs in other ancestry groups).**

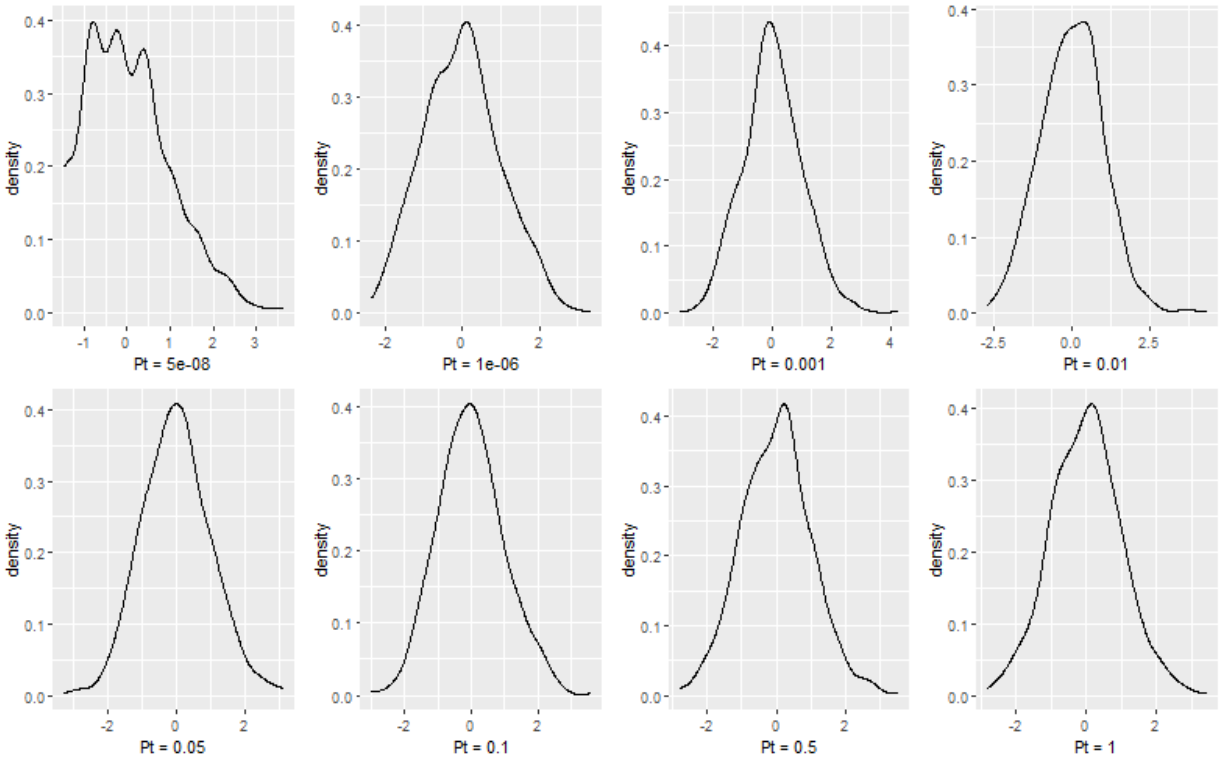
References

Liu M, Jiang Y, Wedow R, et al. Association studies of up to 1.2 million individuals yield new insights into the genetic etiology of tobacco and alcohol use. Nat Genet. 2019;51(2):237–244. doi:10.1038/s41588-018-0307-5

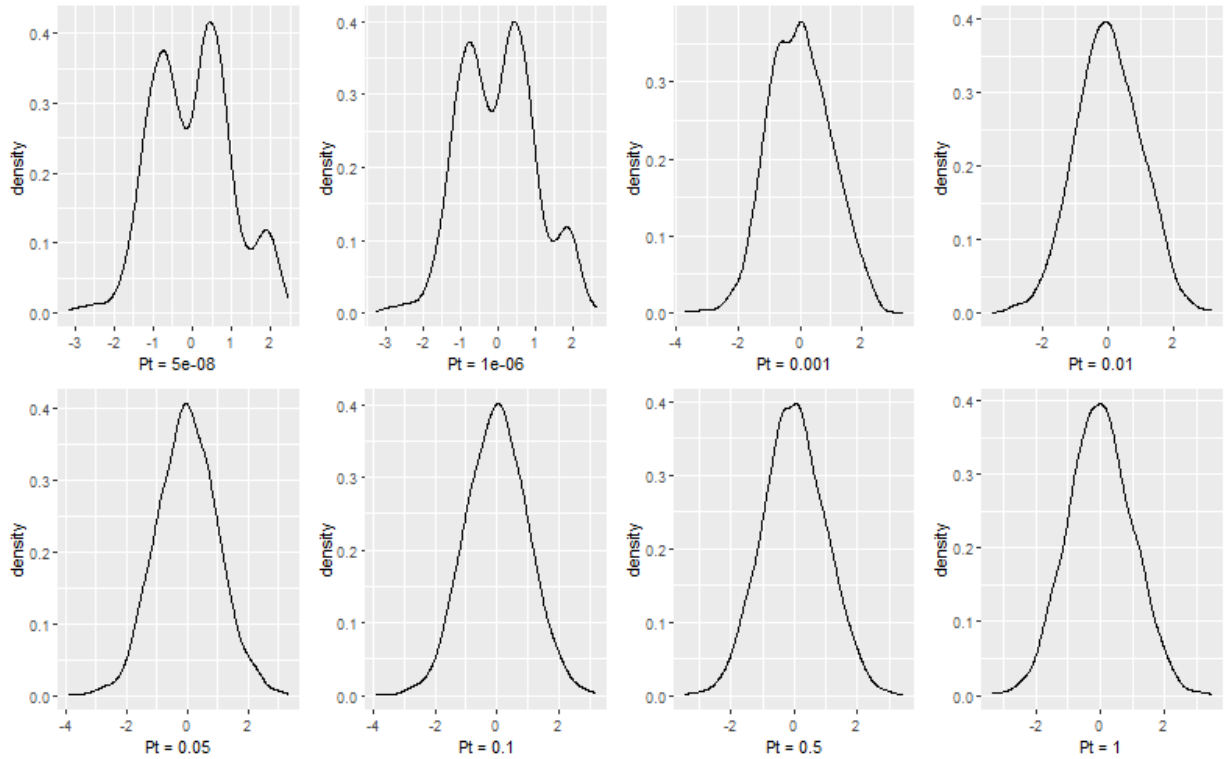
Distribution of polygenic scores by p-value threshold, NHATS European ancestry, n=2827, age at smoking initiation



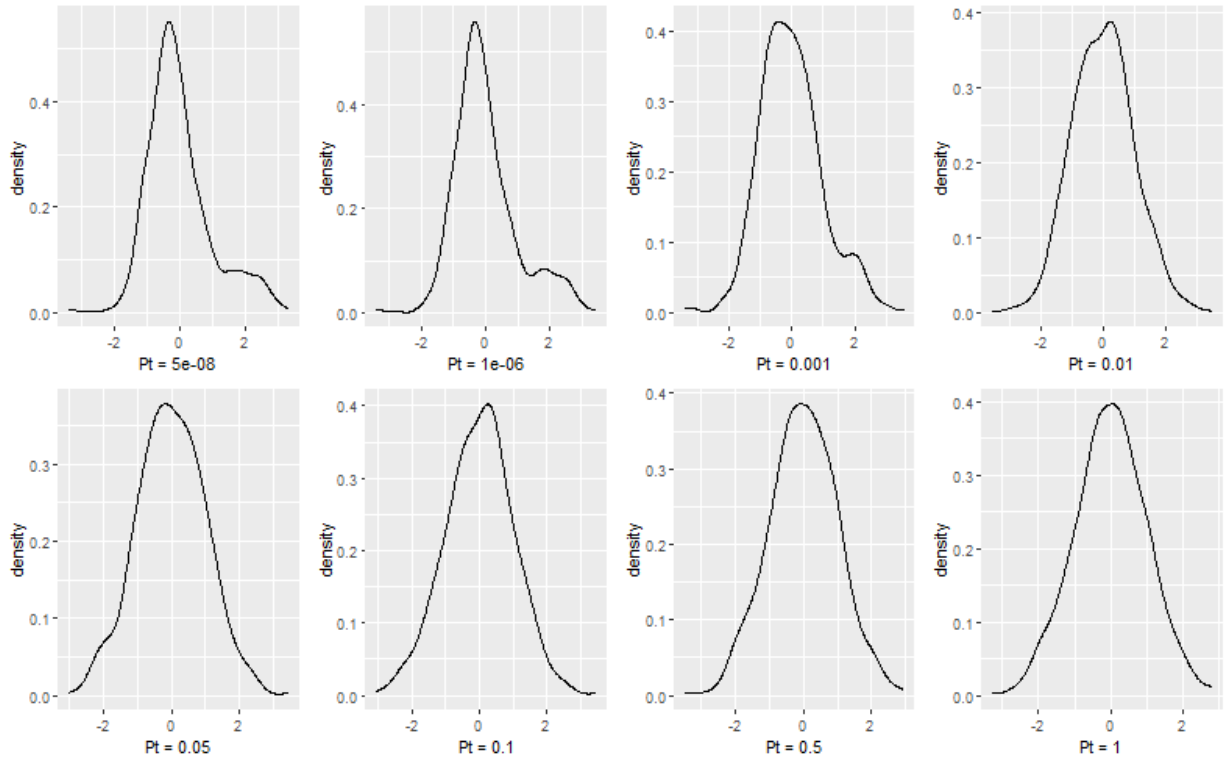
Distribution of polygenic scores by p-value threshold, NHATS African ancestry, n=664, age at smoking initiation



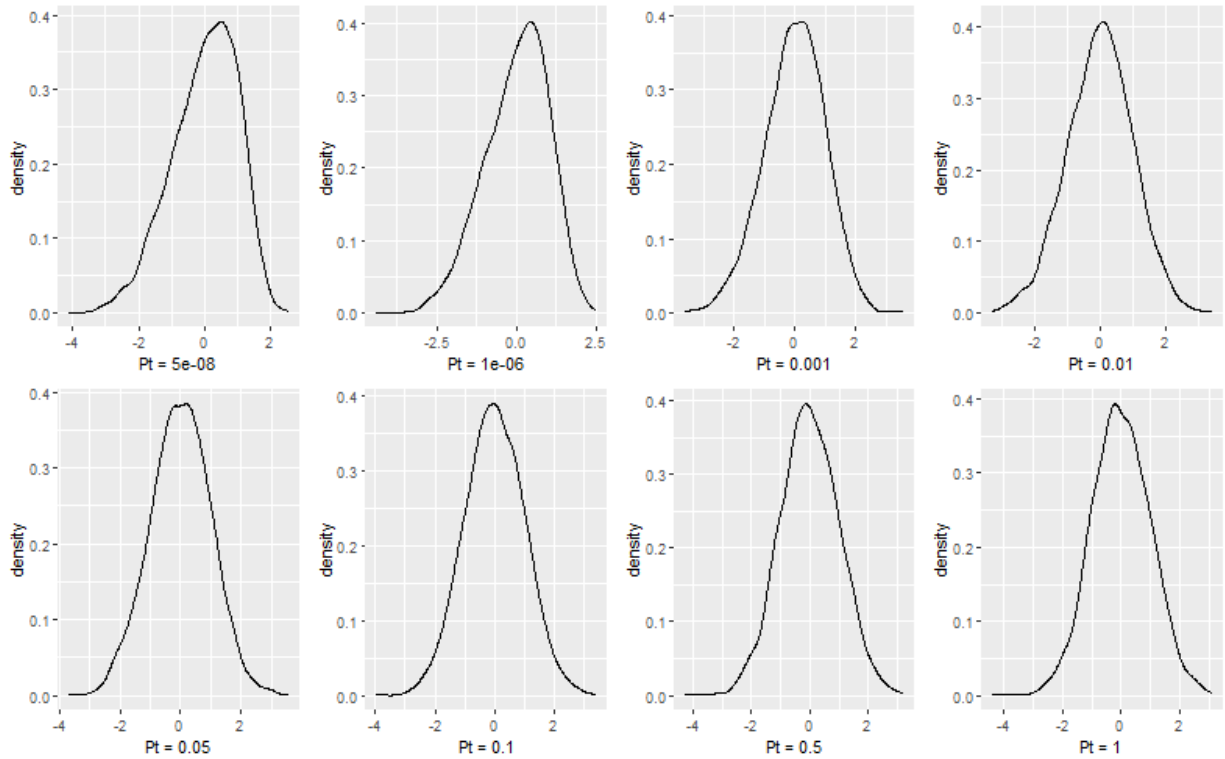
Distribution of polygenic scores by p-value threshold, NHATS European ancestry, n=2827, cigarettes per day



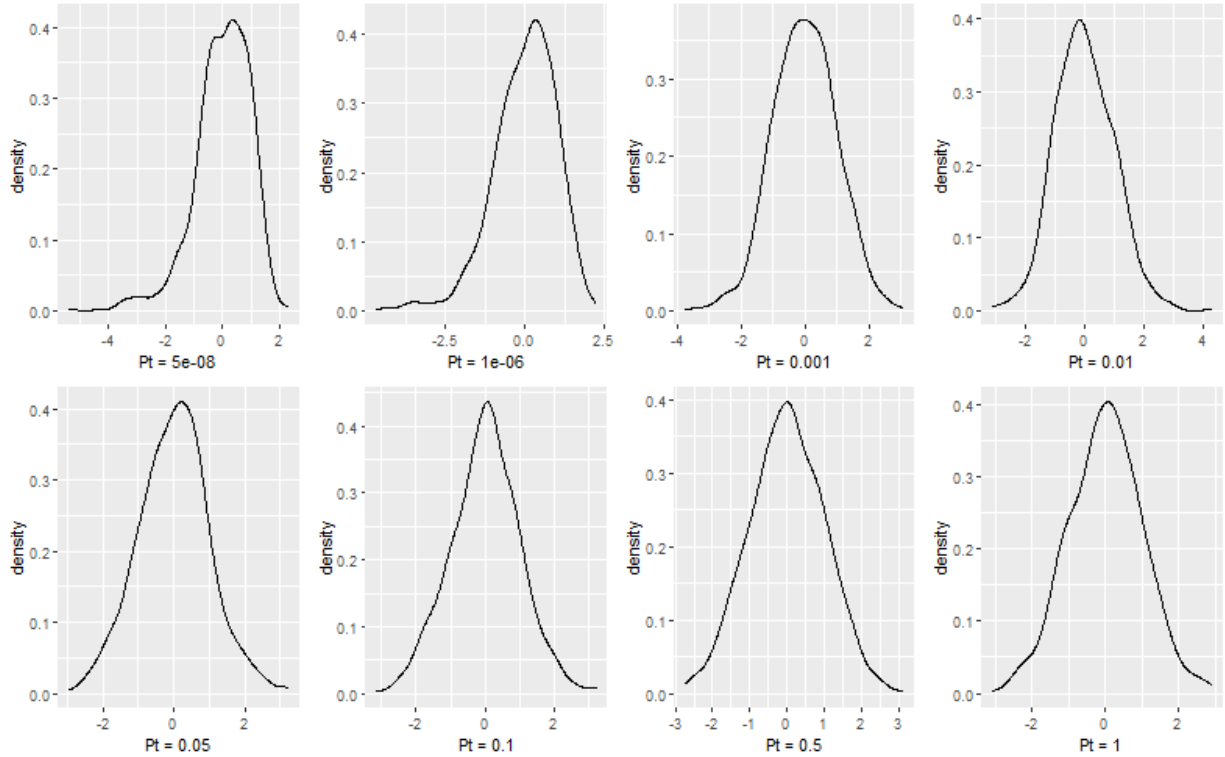
Distribution of polygenic scores by p-value threshold, NHATS African ancestry, n=664, cigarettes per day



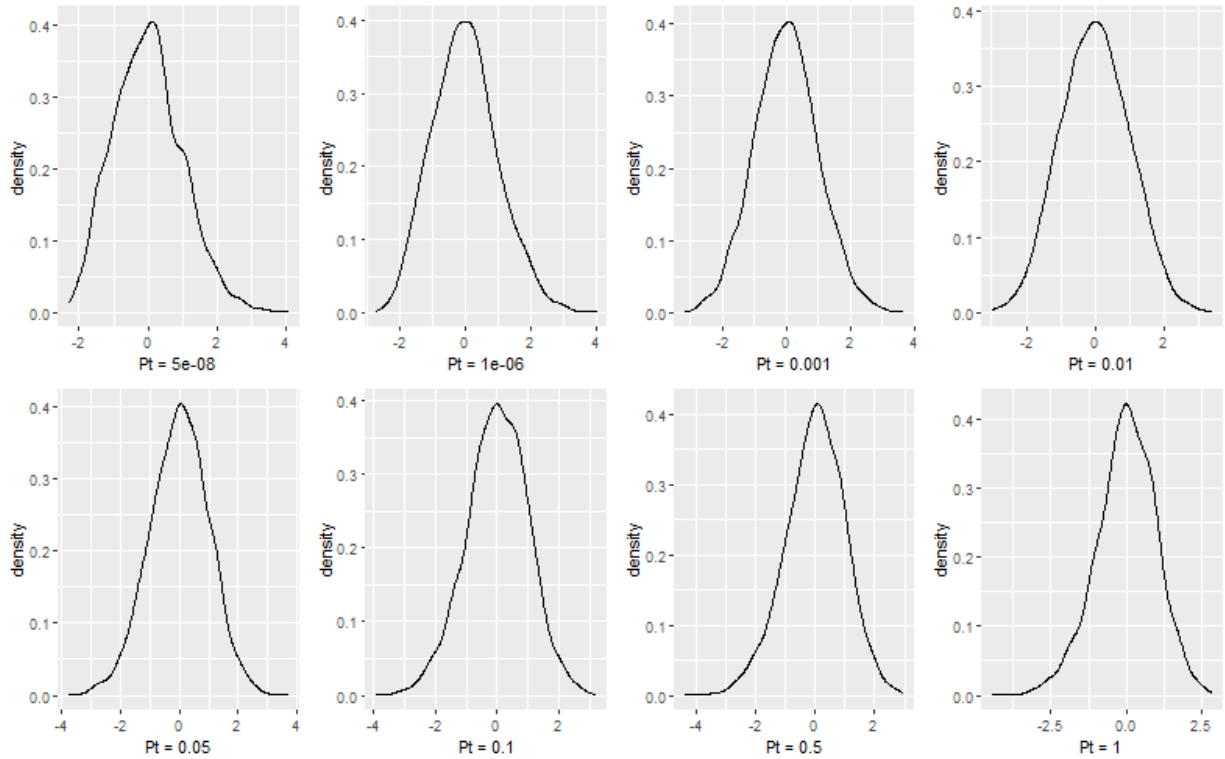
Distribution of polygenic scores by p-value threshold, NHATS European ancestry, n=2827, drinks per week



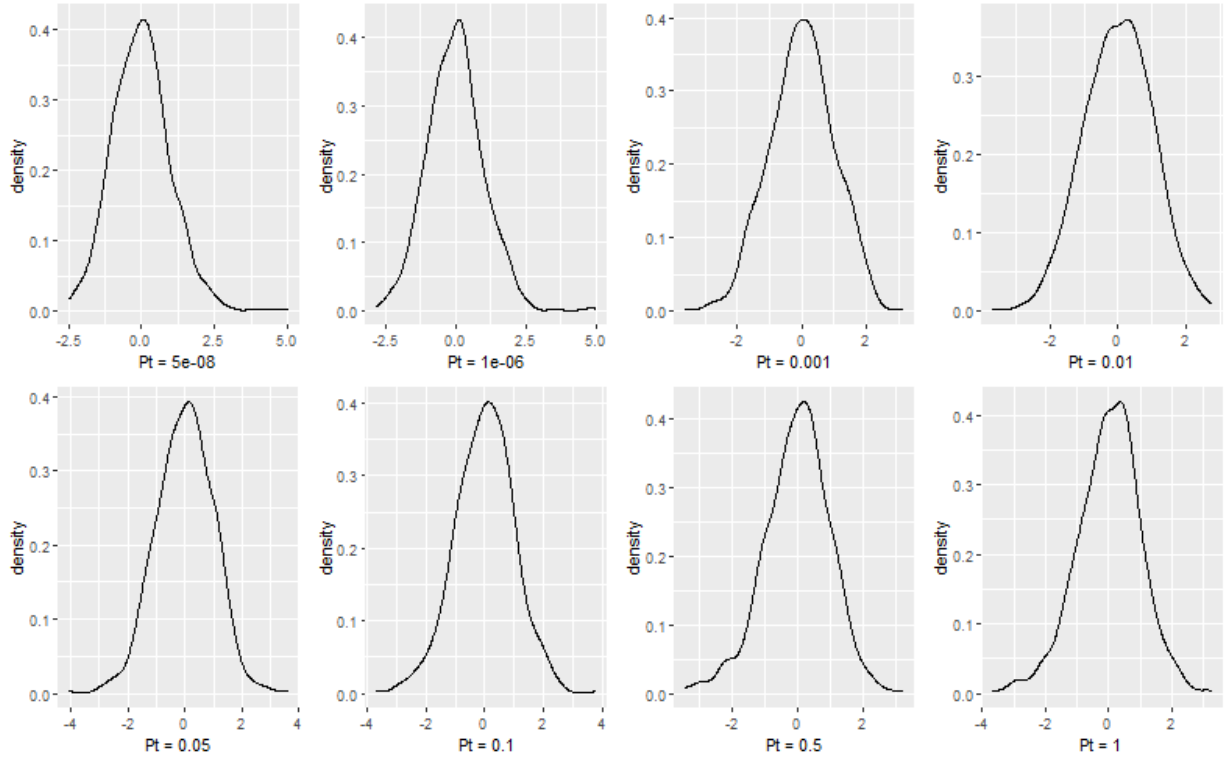
Distribution of polygenic scores by p-value threshold, NHATS African ancestry, n=664, drinks per week



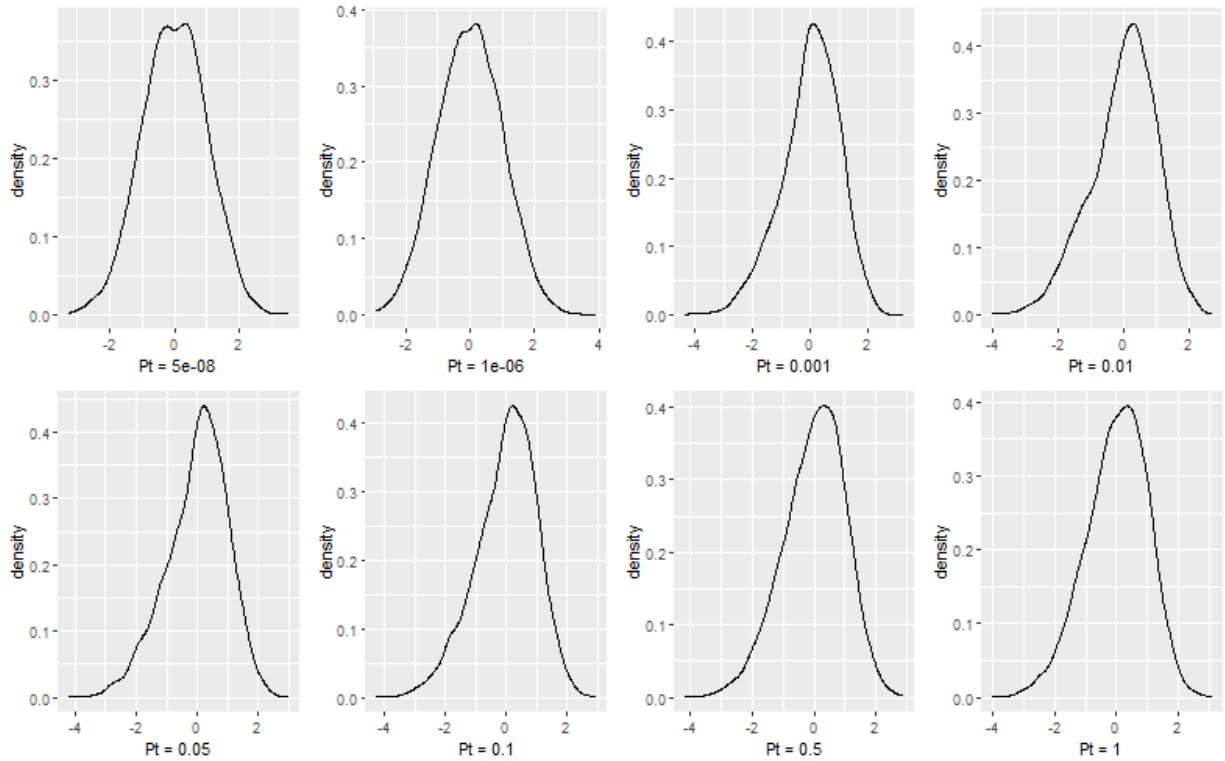
Distribution of polygenic scores by p-value threshold, NHATS European ancestry, n=2827, smoking cessation



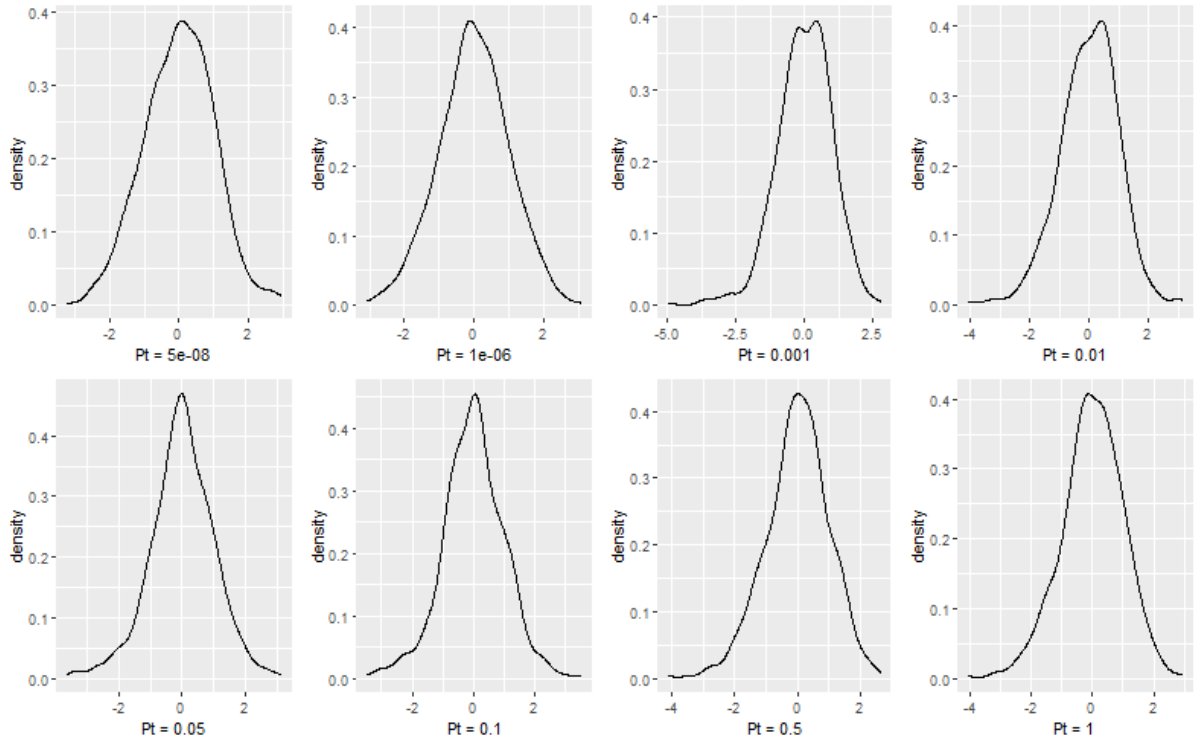
Distribution of polygenic scores by p-value threshold, NHATS African ancestry, n=664, smoking cessation



Distribution of polygenic scores by p-value threshold, NHATS European ancestry, n=2827, smoking initiation



Distribution of polygenic scores by p-value threshold, NHATS African ancestry, n=664, smoking initiation



h. Educational Attainment 3 – Social Science Genetic Association Consortium 2018

The educational attainment PGSs were created using results from a 2018 study by the Social Science Genetic Association Consortium (SSGAC). The meta-analysis included 405,073 individuals in the combined discovery and replication sample and 726,808 individuals that did not contribute to the analyses of the previous study and were used as replication in this study (total of 1,131,881 individuals). Genome-wide significant SNPs were identified in 1,271 loci (**Supplementary Information table 2** [1]). Approximately 10.2 million SNPs were included in the analyses, with all cohorts utilizing SNPs imputed to the 1000 genomes reference panel (1000G). The SSGAC provided SNP weights with 23andMe results removed (due to data use agreements). Study-specific GWASs controlled for the first ten principal components of the genotypic data, 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.

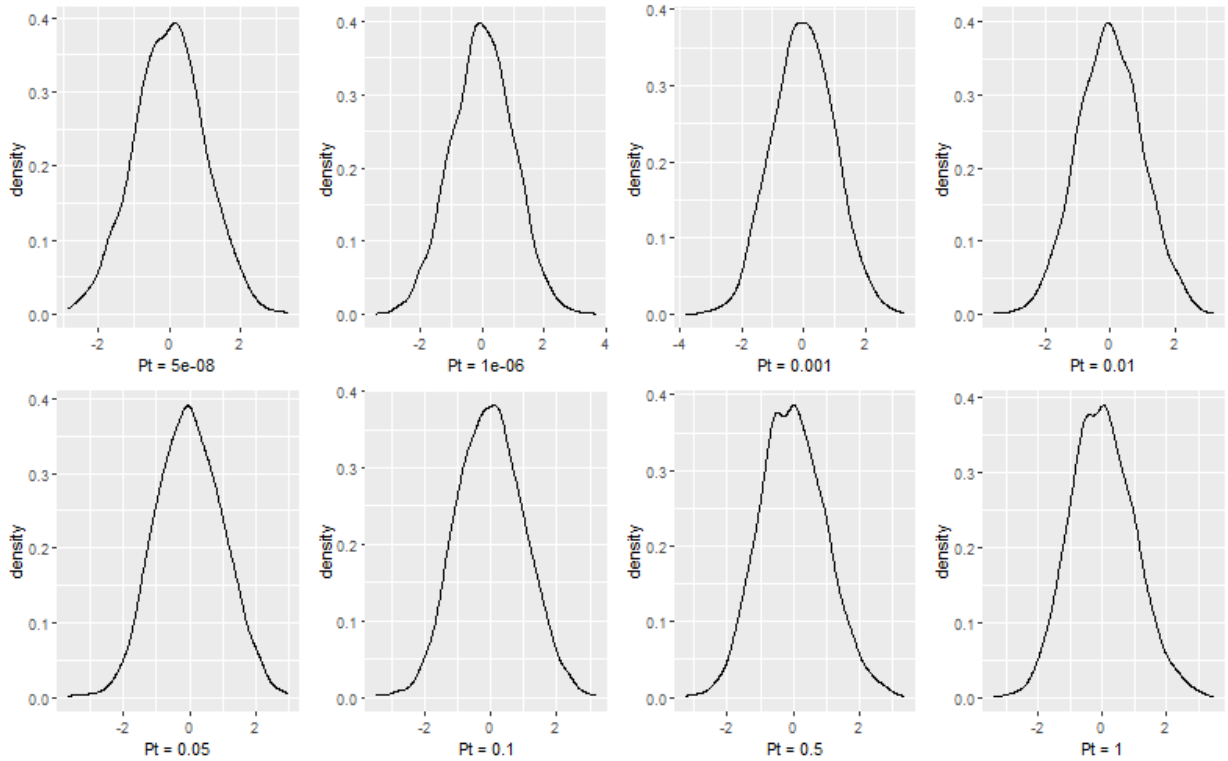
The PGSs contain 481,343 SNPs that overlapped between the NHATS genetic database for each analytic group and the GWAS meta-analysis. Higher PGSs correspond to increased education. The posted PGSs have been standardized within genetic ancestry analysis group to a standard normal curve (mean=0, standard deviation = 1).

N.B.: **Please note that the SSGAC results are from a GWAS on individuals of European ancestry (see "Notes about the use of PGSs" for more information on the use of PGSs in other ancestry groups).**

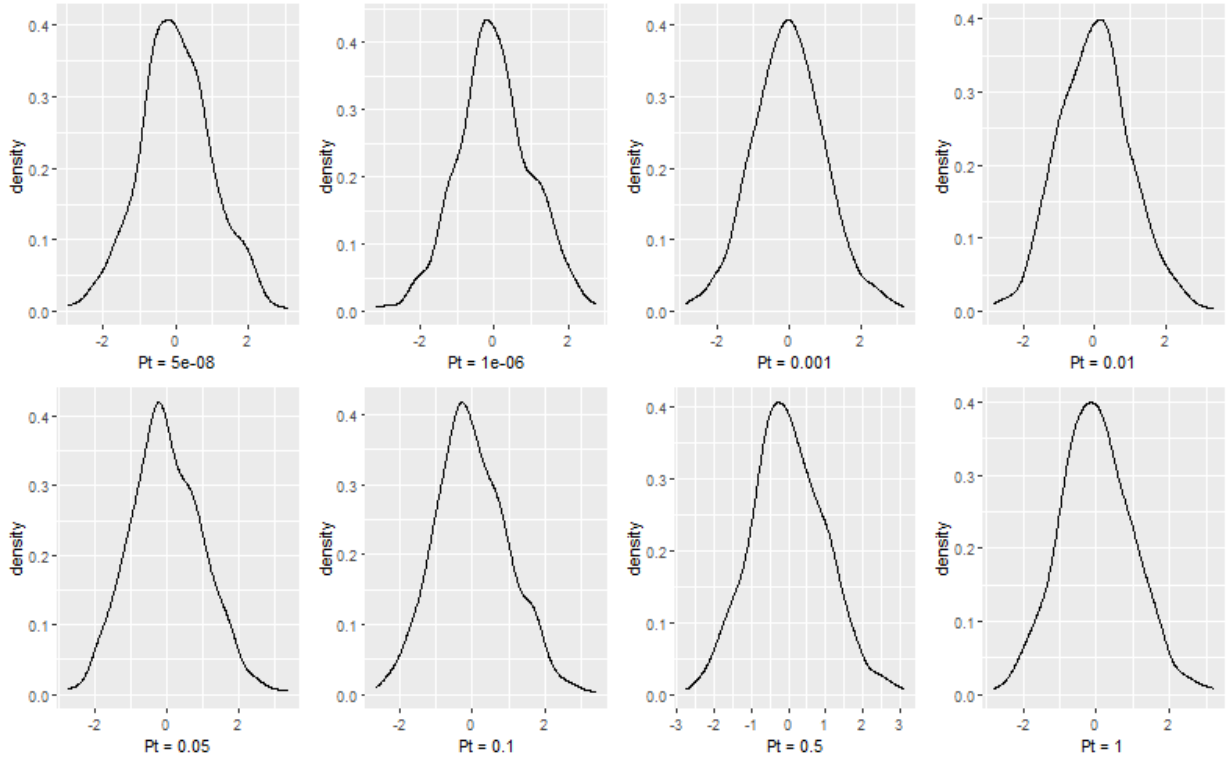
References

Lee JJ, Wedow R, Okbay A, Kong E, Maghzian O, Zacher M, … & Cesarini D. (2016). Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. *Nature genetics*, *doi: 10.1038/s41588-018-0147-3*

Distribution of polygenic scores by p-value threshold, NHATS European ancestry, n=2827



Distribution of polygenic scores by p-value threshold, NHATS African ancestry, n=664



i. Neuroticism – Social Science Genetic Association Consortium 2016

The PGSs for neuroticism were created using results from a 2016 auxiliary GWAS conducted by the Social Science Genetic Association Consortium (SSGAC) as part of their subjective wellbeing GWAS (see above). The GWAS meta-analysis files are publicly available on the SSGAC website: https://www.thessgac.org/data. The entire meta-analysis included 170,911 individuals. Meta-analysis was performed on publicly available results from the Genetics of Personality Consortium (GPC) (N=63,661) with results from UK Biobank data (N=107,245). The meta-analysis yielded 11 lead SNPs, 2 of which tag inversion polymorphisms (**Table 1**). A quasi-replication analysis tested whether these SNPs were associated with subjective wellbeing. A replication analysis was also performed using data from 23andMe (N=368,890). In UKB, the phenotype measure was the respondent's score on a 12-item version of the Eysenck Personality Inventory Neuroticism scale. The GPC harmonized different neuroticism batteries. In the UKB, analyses controlled for the first 15 PCs, indicator variables for genotyping array, sex, indicator variables for age ranges, and sex-by-age interactions. Model adjustments for the 29 cohorts contributing to the GPC meta-analysis varied (see de Moor et al., p. 644, 2015).

The PGSs contain 326,436 SNPs that overlapped between the NHATS genetic database for each analytic group and the GWAS meta-analysis. Higher PGSs correspond to increased neuroticism. The posted PGSs have been standardized within genetic ancestry analysis group to a standard normal curve (mean=0, standard deviation = 1).

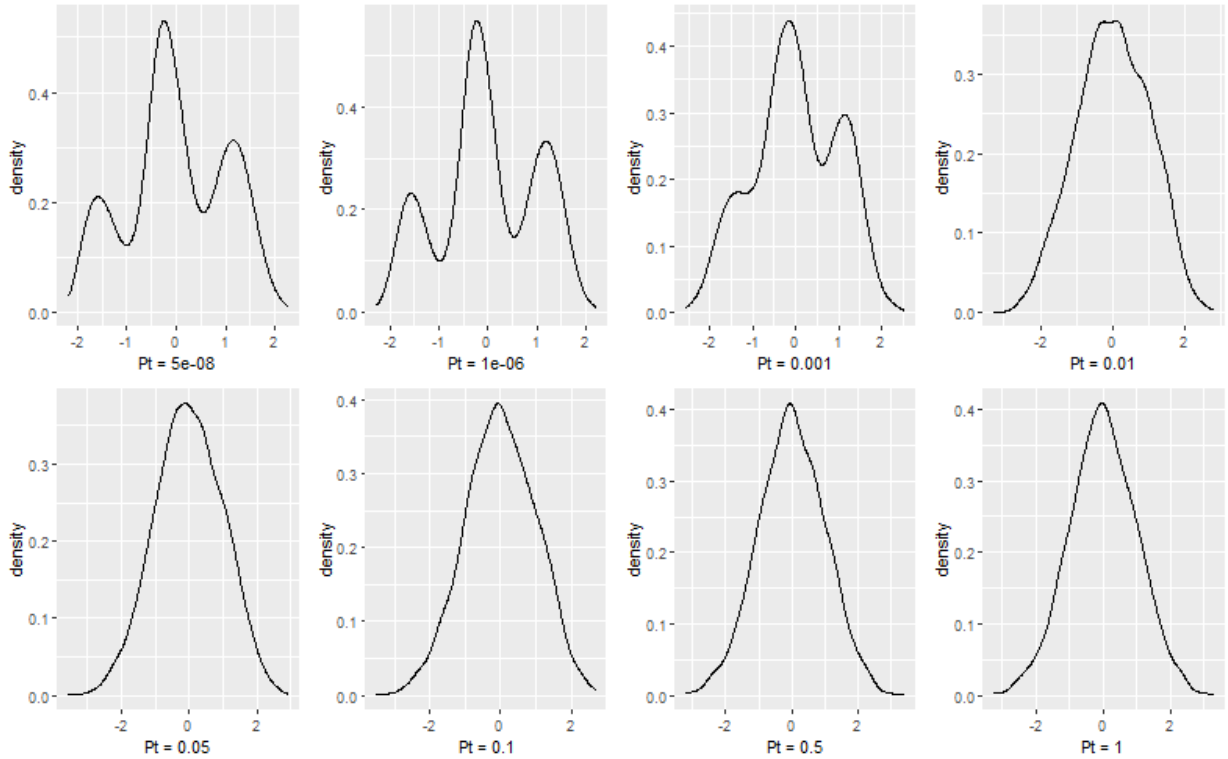
N.B.: **Please note that the SSGAC results are from a GWAS on individuals of European ancestry (see "Notes about the use of PGSs" for more information on the use of PGSs in other ancestry groups).**

References

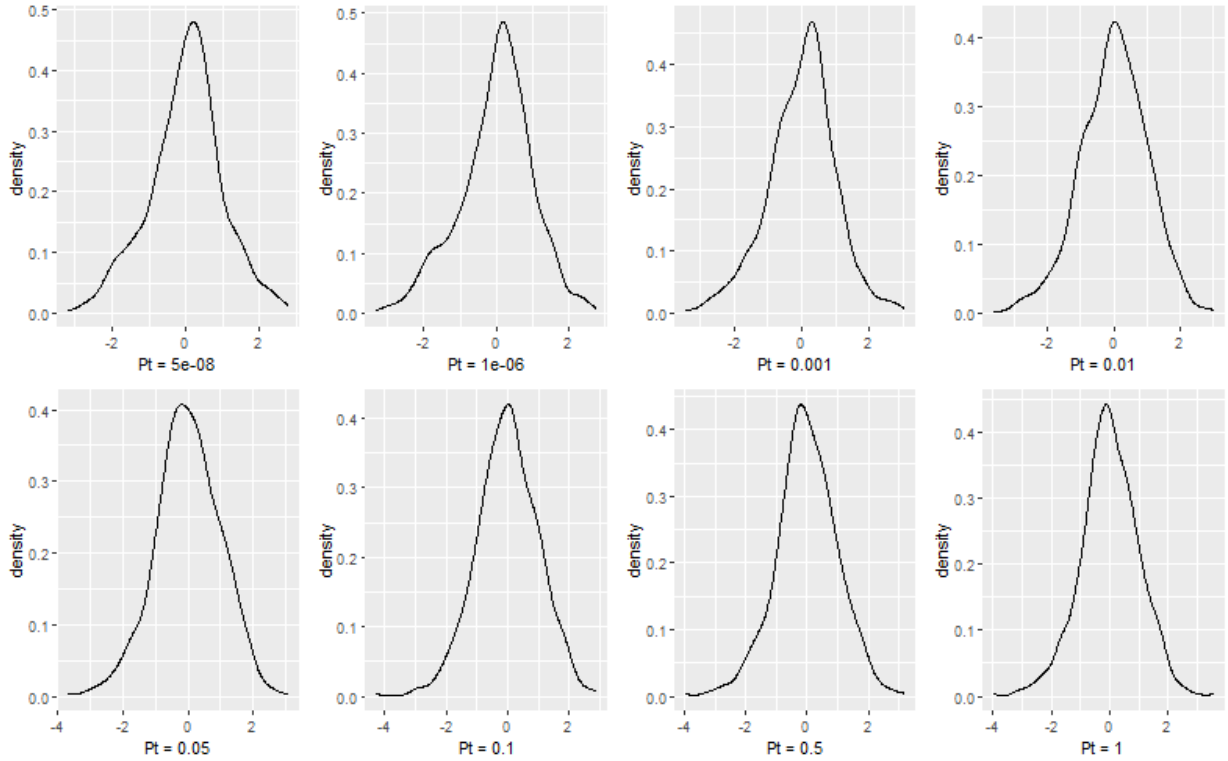
Okbay, A., Baselmans, B. M., De Neve, J. E., Turley, P., Nivard, M. G., Fontana, M. A., ... & Gratten, J. (2016). Genetic variants associated with subjective well-being, depressive symptoms, and neuroticism identified through genome-wide analyses. *Nature Genetics*, *48*(6), 624-633.

De Moor, M. H., Van Den Berg, S. M., Verweij, K. J., Krueger, R. F., Luciano, M., Vasquez, A. A., ... & Gordon, S. D. (2015). Meta-analysis of genome-wide association studies for neuroticism, and the polygenic association with major depressive disorder. *JAMA Psychiatry*, *72*(7), 642-650.

Distribution of polygenic scores by p-value threshold, NHATS European ancestry, n=2827



Distribution of polygenic scores by p-value threshold, NHATS African ancestry, n=664



j. Type II Diabetes – DIAbetes Genetics Replication And Meta-analysis, Genetic Epidemiology Research on Aging and the full cohort release of the UK Biobank 2018

The PGSs for Type II Diabetes (T2D) were created using GWAS meta-analysis results from a 2018 study conducted by Xue et al 2018. We conducted a meta-analysis of genome-wide association studies (GWAS) with ~16 million genetic variants in 62,892 T2D cases and 596,424 controls of European ancestry, by combining three GWAS data sets: DIAbetes Genetics Replication And Meta-analysis (DIAGRAM), Genetic Epidemiology Research on Aging (GERA) and the full cohort release of the UK Biobank (UKB). Full details of genotyping, QC, association analysis for each study are provided in the Methods section of Xue et al. Nature Communications, 2018. Meta-analysis results were available for for ~5.1 million common SNPs.

The GWAS meta-analysis files can be downloaded from the program in complex trait genomics website: https://cnsgenomics.com/data.html (Xue_et_al_T2D_META_Nat_Commun_2018.gz). We identify 139 common and 4 rare variants associated with T2D, 42 of which (39 common and 3 rare variants) are independent of the known variants.

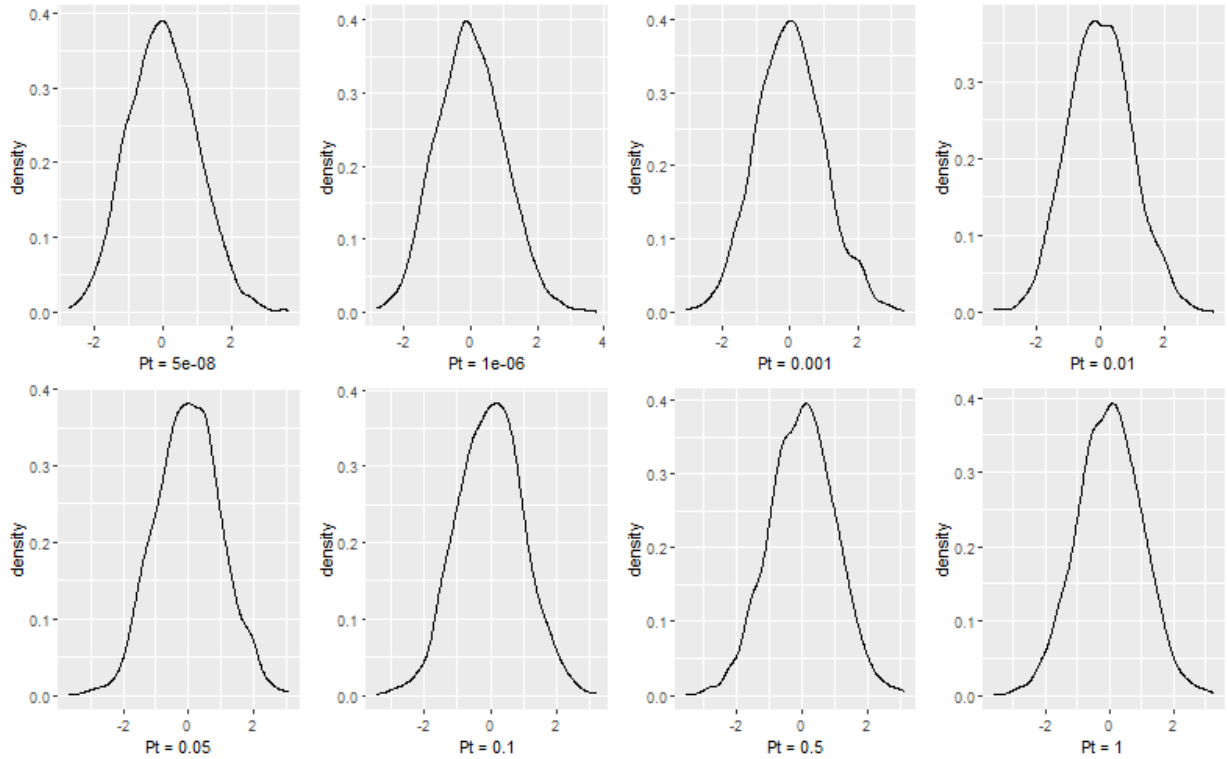
The PGSs contain 221,048 SNPs that overlapped between the NHATS genetic database for each analytic group and the GWAS meta-analysis. The posted PGSs have been standardized within ethnicity, to a standard normal curve (mean=0, standard deviation = 1).

N.B.: The effect estimates for SNPs come from the discovery stage I meta-analysis of European descent individuals **(see "Notes about the use of PGSs" for more information on the use of PGSs in other ancestry groups).**

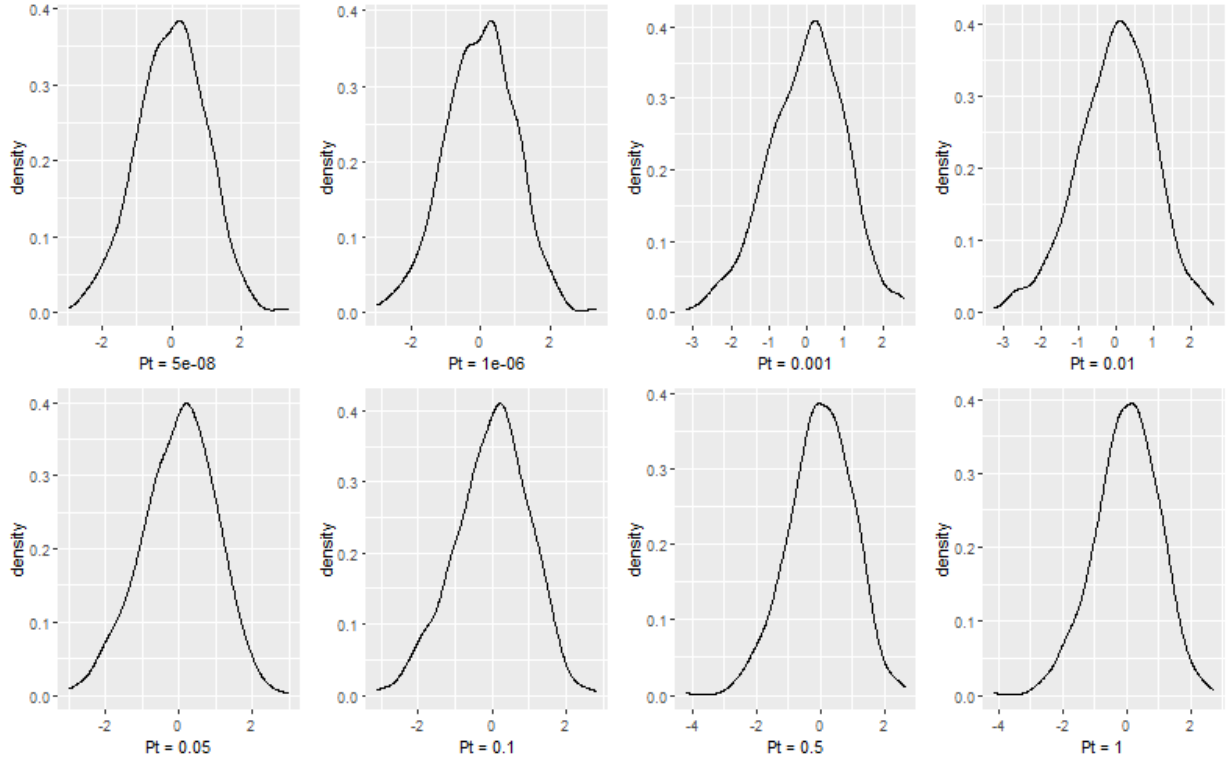
References

Xue A, Wu Y, Zhu Z, Zhang F, Kemper KE, Zheng Z, Yengo L, Lloyd-Jones LR, Sidorenko J, Wu Y; eQTLGen Consortium; McRae AF, Visscher PM, Zeng J, Yang J. Genome-wide association analyses identify 143 risk variants and putative regulatory mechanisms for type 2 diabetes. Nat Commun. 2018 Jul 27;9(1):2941. doi: 10.1038/s41467-018-04951-w. PMID: 30054458; PMCID: PMC6063971.

Distribution of polygenic scores by p-value threshold, NHATS European ancestry, n=2827



Distribution of polygenic scores by p-value threshold, NHATS African ancestry, n=664



k. Coronary Artery Disease - Coronary ARtery DIsease Genome wide Replication and Meta-analysis 2011

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 www.cardiogramplusc4d.org (cad.add.160614.website.txt). The 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). Analysis identified 13 new genome-wide significant loci and confirmed 10 previously reported CAD loci **(Tables 1 and 2)**. Study-specific GWAS adjusted for age of onset (cases) or age of recruitment (controls), gender, and genetic principal components.

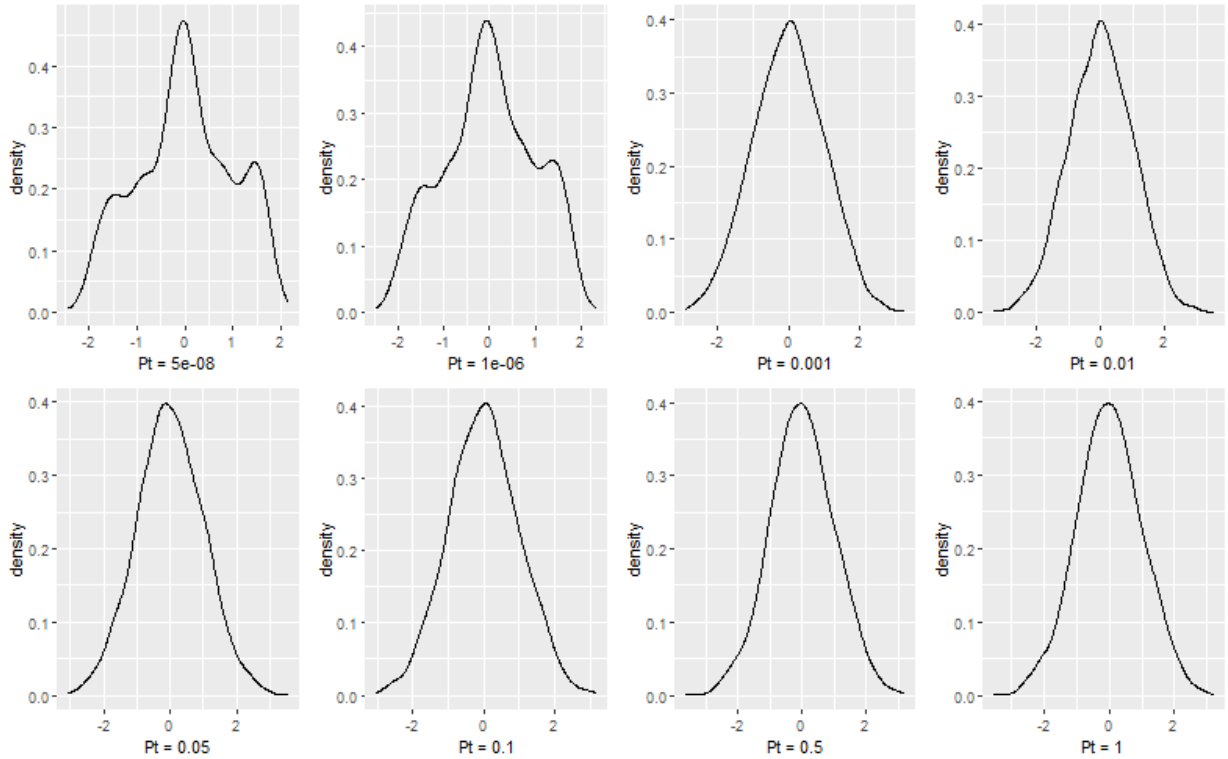
The PGSs contain 211,880 SNPs that overlapped between the NHATS genetic database for each analytic group and the GWAS meta-analysis. The posted PGSs have been standardized within genetic ancestry analysis group to a standard normal curve (mean=0, standard deviation = 1). Higher PGSs correspond to increased odds of CAD. Weights are represented as log(OR).

N.B.: **Please note that the CARDIoGRAM results are from a GWAS on individuals of European ancestry (see "Notes about the use of PGSs" for more information on the use of PGSs in other ancestry groups).**

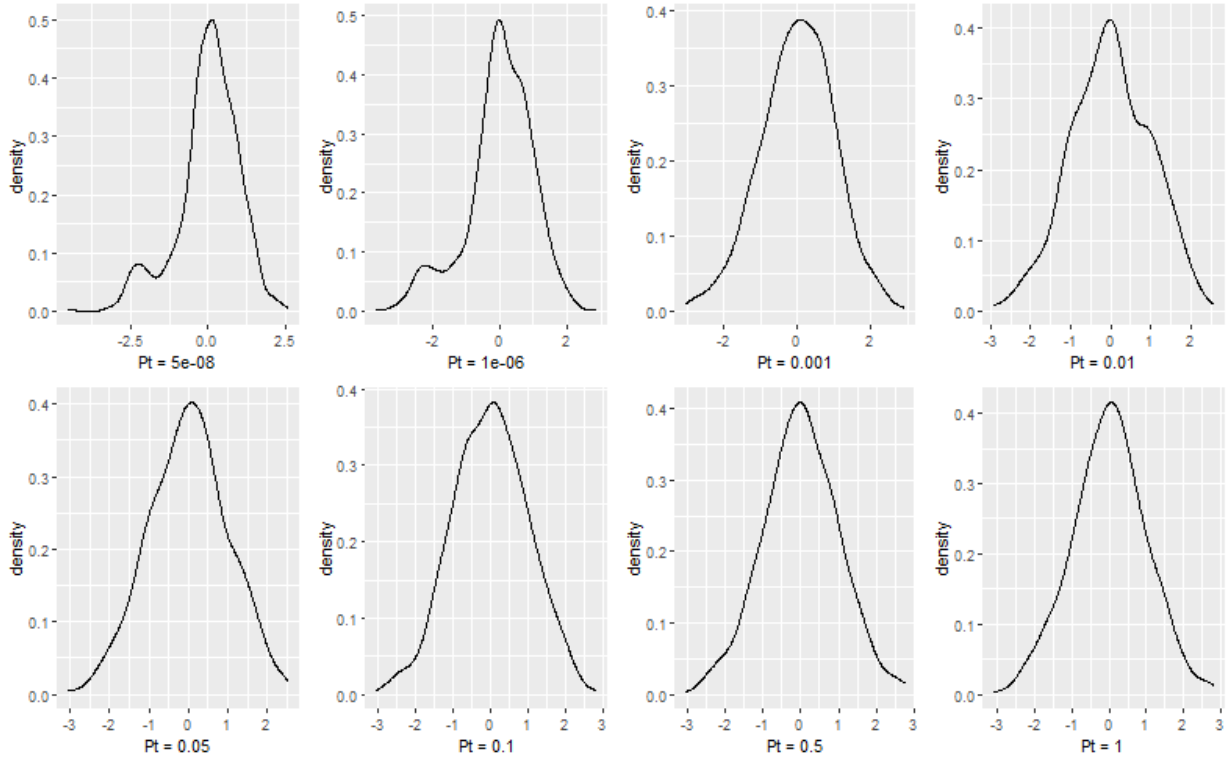
References

Schunkert, H., König, I. R., Kathiresan, S., Reilly, M. P., Assimes, T. L., Holm, H., ... & Absher, D. (2011). Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease. *Nature Genetics*, *43*(4), 333-338.

Distribution of polygenic scores by p-value threshold, NHATS European ancestry, n=2827



Distribution of polygenic scores by p-value threshold, NHATS African ancestry, n=664



I. Myocardial Infarction - Coronary ARtery DIsease Genome wide Replication and Meta-analysis 2015

The PGSs for myocardial infarction (MI) were created using 2015 results from a subgroup analysis of coronary artery disease (CAD) 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 www.cardiogramplusc4d.org (mi.add.030315.website.txt). The GWAS is a meta-analysis of 48 studies of mainly European, South Asian, and East Asian, descent imputed using the 1000 Genomes phase 1 v3 training set with 38 million variants. The study interrogated 9.4 million variants and involved 60,801 CAD cases and 123,504 controls. Case status was defined by an inclusive CAD diagnosis (for example, myocardial infarction, acute coronary syndrome, chronic stable angina or coronary stenosis of >50%). Thirty-seven previous loci and ten new loci achieved genome-wide significance (**Supplementary Table 2**). MI subgroup analysis was performed in cases with a reported history of MI (~70% of the total number of cases). No additional loci reached genome-wide significance in the MI analysis.

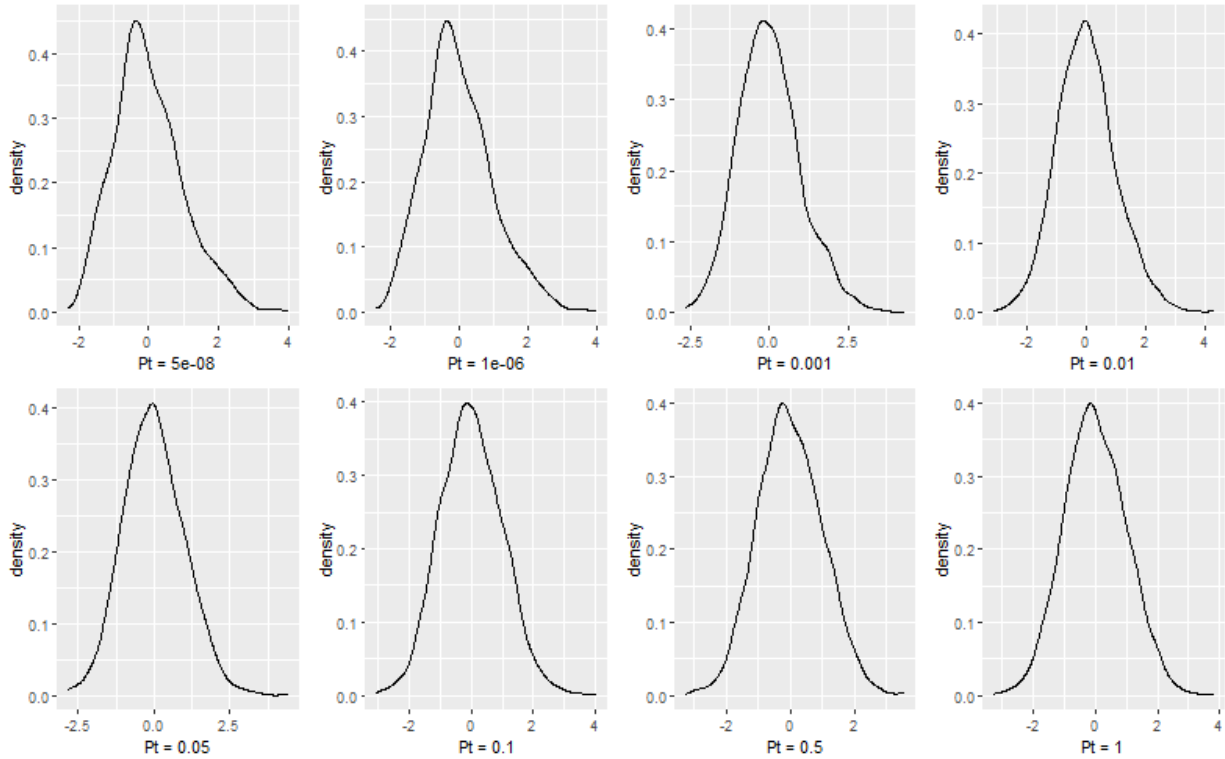
The PGSs contain 448,973 SNPs that overlapped between the NHATS genetic database for each analytic group and the GWAS meta-analysis. Higher PGSs correspond to increased odds of MI. The posted PGSs have been standardized within genetic ancestry analysis group to a standard normal curve (mean=0, standard deviation = 1). Weights are represented as log(OR).

N.B.: **Please note that the CARDIoGRAM results are from a GWAS on individuals of mostly European ancestry (see "Notes about the use of PGSs" for more information on the use of PGSs in other ancestry groups).**

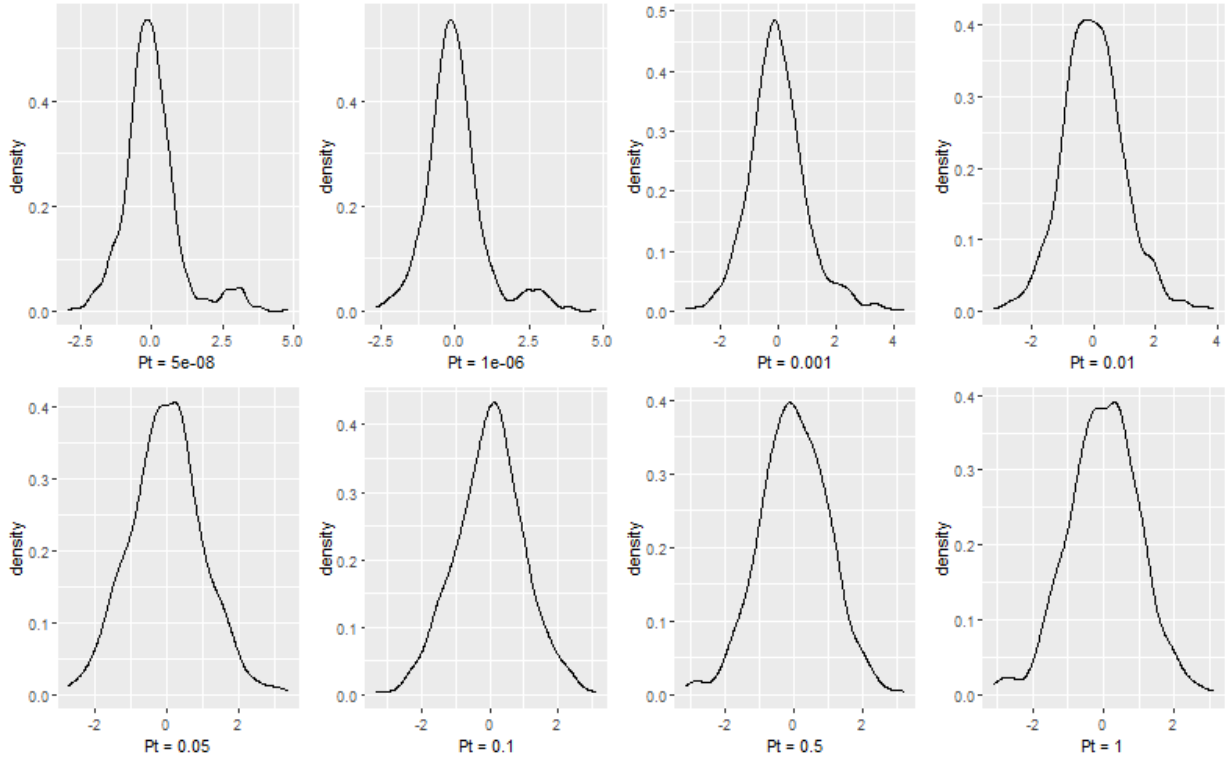
References

CARDIoGRAMplusC4D Consortium. (2015). A comprehensive 1000 Genomes-based genome-wide association meta-analysis of coronary artery disease. *Nature Genetics*, *47*(10), 1121-1130.

Distribution of polygenic scores by p-value threshold, NHATS European ancestry, n=2827



Distribution of polygenic scores by p-value threshold, NHATS African ancestry, n=664



m. General Cognition 2 (Gencog2) – Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE), the Cognitive Genomics Consortium (COGENT) consortia, and UK Biobank 2018

PGSs for general cognitive function were created using results from a 2018 study conducted by the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE), the Cognitive Genomics Consortium (COGENT) consortia, and UK Biobank. Access to the full GWAS summary results for general cognitive function can be requested by application to the chairs of the CHARGE and COGENT consortia. The general cognition GWAS meta-analysis included 300,486 individuals aged 16-102 of European ancestry. To compute the PGSs, we used SNPs weights from a GWAS with several NHLBI cohorts removed from the meta-analysis. There are 12,871,898 SNPs in the summary statistics file imputed to NCBI Build 37/UCSC hg 19. Inclusion/exclusion criteria for cases and controls for each cohort are available in **Supplementary Data 18**. The study identified 148 loci with 434 independent genome-wide significant SNPs (**Supplementary Data 1 and 2**).

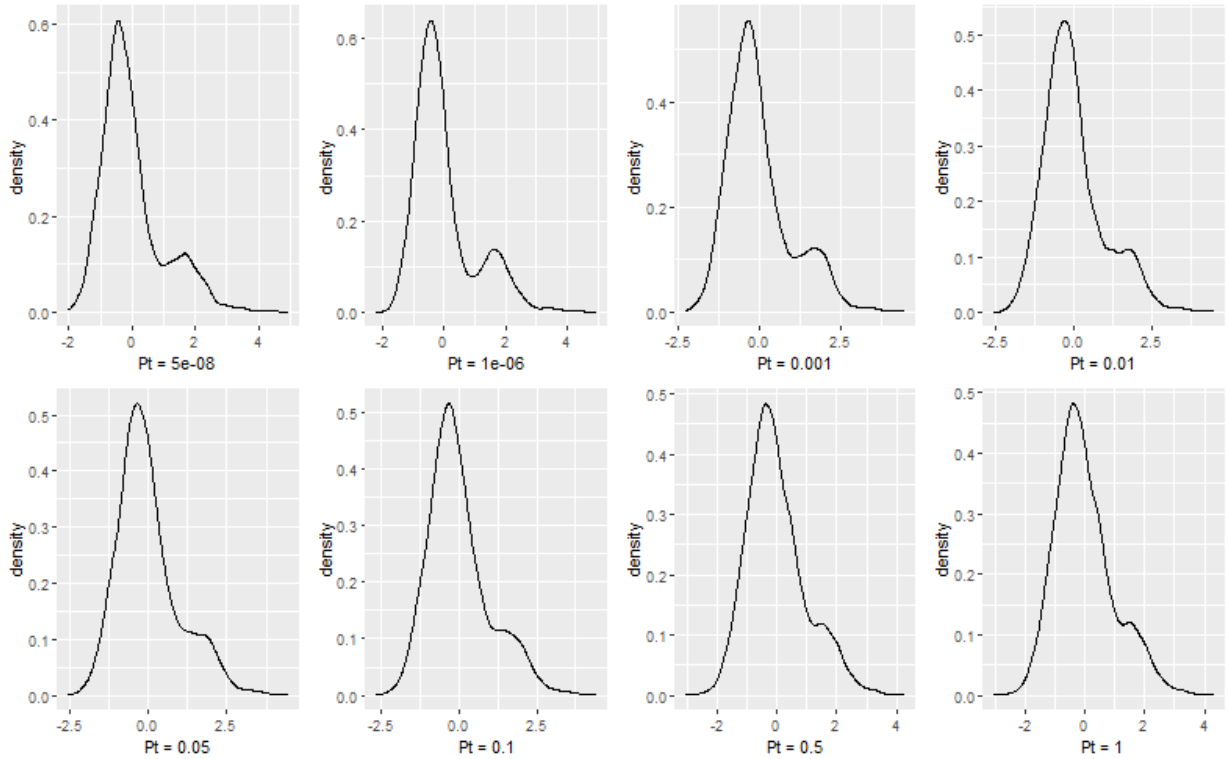
The PGSs contain 494,909 SNPs that overlapped between the NHATS genetic database and the GWAS meta-analysis. Higher PGSs correspond to increased cognition. The posted PGSs have been standardized within genetic ancestry analysis group to a standard normal curve (mean=0, standard deviation = 1).

N.B.: **Please note that the CHARGE-Gencog2 weights are from a GWAS on individuals of European ancestry (see "Notes about the use of PGSs" for more information on the use of PGSs in other ancestry groups).** The following CHARGE cohorts requested that their data be removed due to restrictions on their use: The Aging Gene-Environment Susceptibility – Reykjavik Study (AGES), The Atherosclerosis Risk in Communities Study (ARIC), The Cardiovascular Health Study (CHS), The Framingham Heart Study (FHS), and The Genetic Epidemiology Network of Arteriopathy (GENOA).

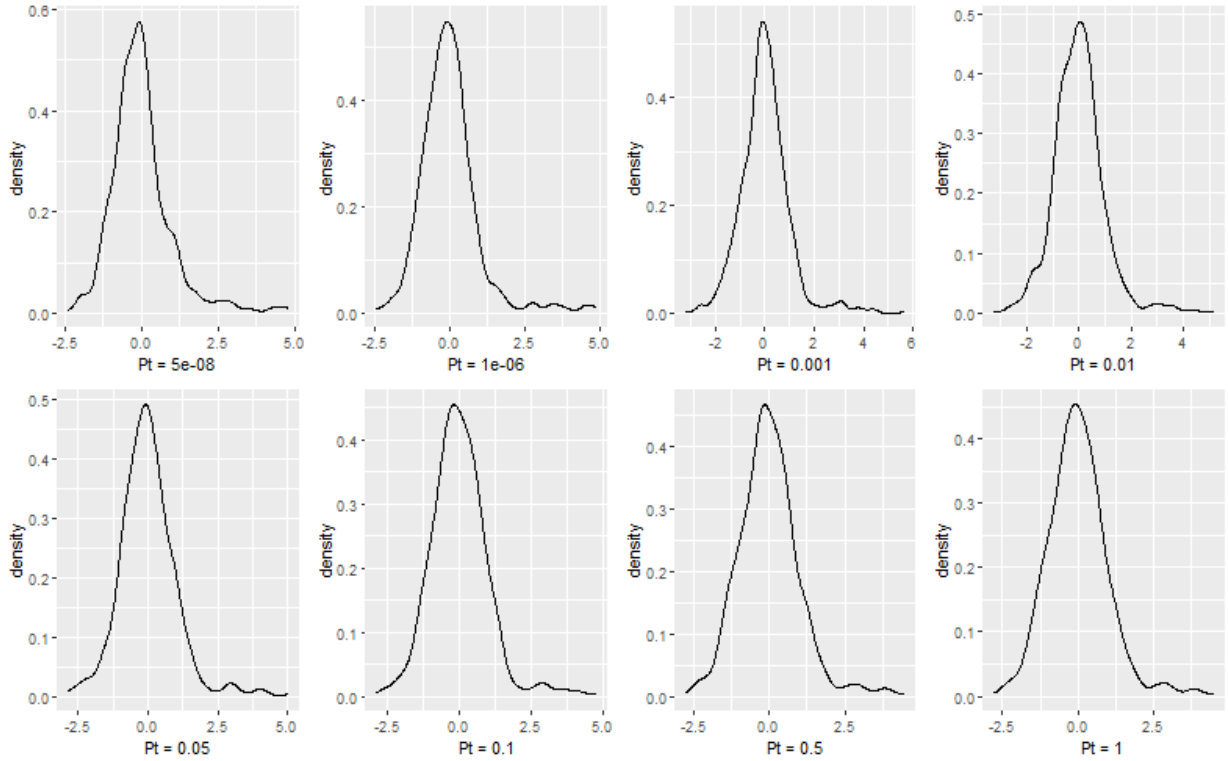
References

Davies, G., Lam, M., Harris, S. E., Trampush, J. W., Luciano, M., Hill, W. D., … Deary, I. J. (2018). Study of 300,486 individuals identifies 148 independent genetic loci influencing general cognitive function. Nature Communications, 9, 2098. http://doi.org/10.1038/s41467-018-04362-x

Distribution of polygenic scores by p-value threshold, NHATS European ancestry, n=2827



Distribution of polygenic scores by p-value threshold, NHATS African ancestry, n=664



n-s. Kidney function – Chronic Kidney Disease Genetics consortium 2019

PGSs for kidney function phenotypes were created using results from a 2019 study conducted by the Chronic Kidney Disease Genetics (CKDGen) Consortium. The GWAS meta-analysis files are publicly available on the CKDGen data download page: http://ckdgen.imbi.uni-freiburg.de/ (20171016_MW_eGFR_overall_ALL_nstud61.dbgap.txt.gz, 20171017_MW_eGFR_overall_EA_nstud42.dbgap.txt.gz, BUN_overall_ALL_YL_20171017_METAL1_nstud_33.dbgap.txt.gz, BUN_overall_EA_YL_20171108_METAL1_nstud24.dbgap.txt.gz, CKD_overall_ALL_JW_20180223_nstud30.dbgap.txt.gz, CKD_overall_EA_JW_20180223_nstud23.dbgap.txt.gz). The CKDGen meta-analysis included GWAS on estimated glomerular filtration rate (eGFR), blood urea nitrogen (BUN) and chronic kidney disease (CKD) using a European ancestry only sample and also a trans-ethnic sample encompassing individuals of European, East Asian, African, South Asian, and Hispanic ancestry. The trans-ethnic GWAS of eGFR included 121 studies with an n of 765,348 and found 308 loci associated with eGFR. The European ancestry eGFR GWAS included 85 studies and an n of 567,460 with 256 discovered loci. The trans-ethnic BUN discovery analysis included an n of 416,178. The European ancestry BUN GWAS included an n of 243,029. The CKD trans-ethnic analysis included 625,219 individuals. The European ancestry CKD analysis included 480,698 individuals (41,395 cases and 439,303 controls). The GWAS meta-analyses included ~9 million imputed SNPs on NCBI Build 37/UCSC hg 19.

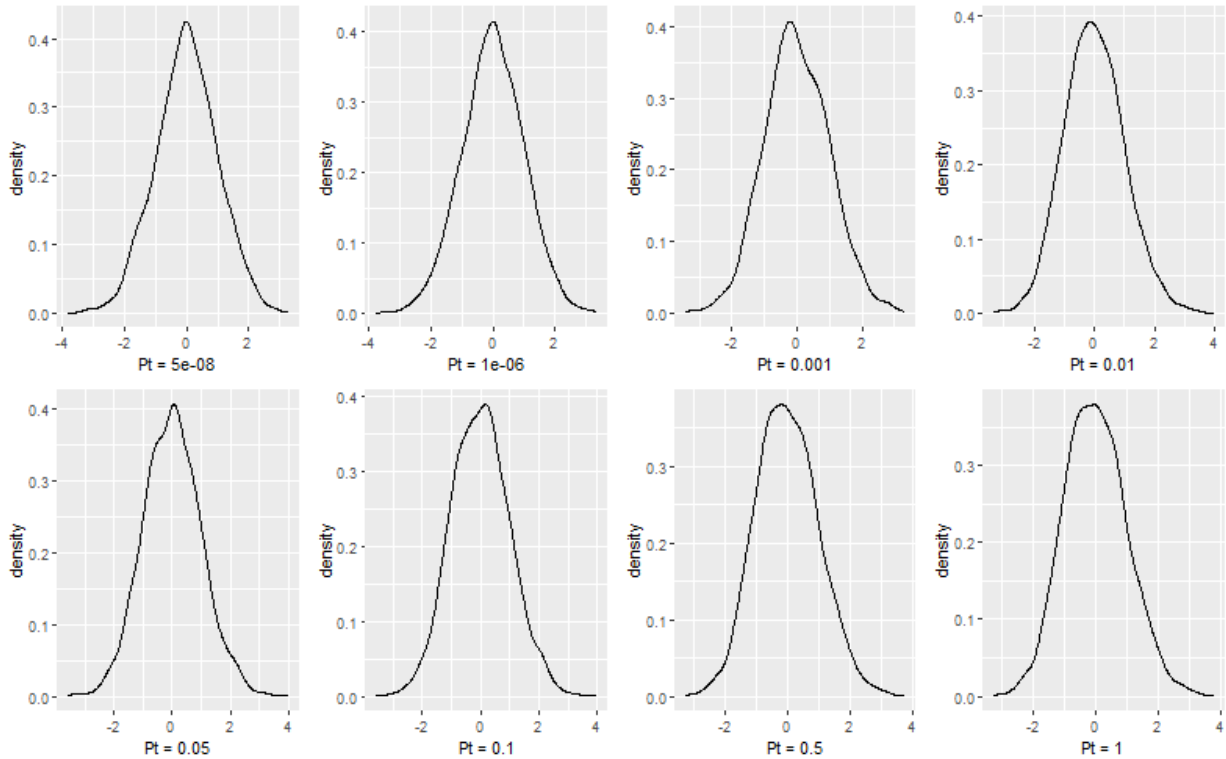
The PGSs (for both African and European ancestry participants) contains 469,290 (BUN), 477,385 (CKD), and 475,581 (eGFR) SNPs that overlapped between the NHATS genetic database in the European ancestry-based GWAS meta-analysis; and 454,908 (BUN), 492,575 (CKD), and 453,972 (eGFR) SNPs that overlapped between the NHATS genetic database and the trans-ancestry based GWAS meta-analysis. Higher PGSs correspond to increased BUN, odds of CKD, or eGFR. The posted PGSs have been standardized within genetic ancestry analysis group to a standard normal curve (mean=0, standard deviation = 1).

N.B.: **Please note that the European ancestry-based summary statistics are from a GWAS on individuals of European ancestry (see "Notes about the use of PGSs" for more information on the use of PGSs in other ancestry groups).**

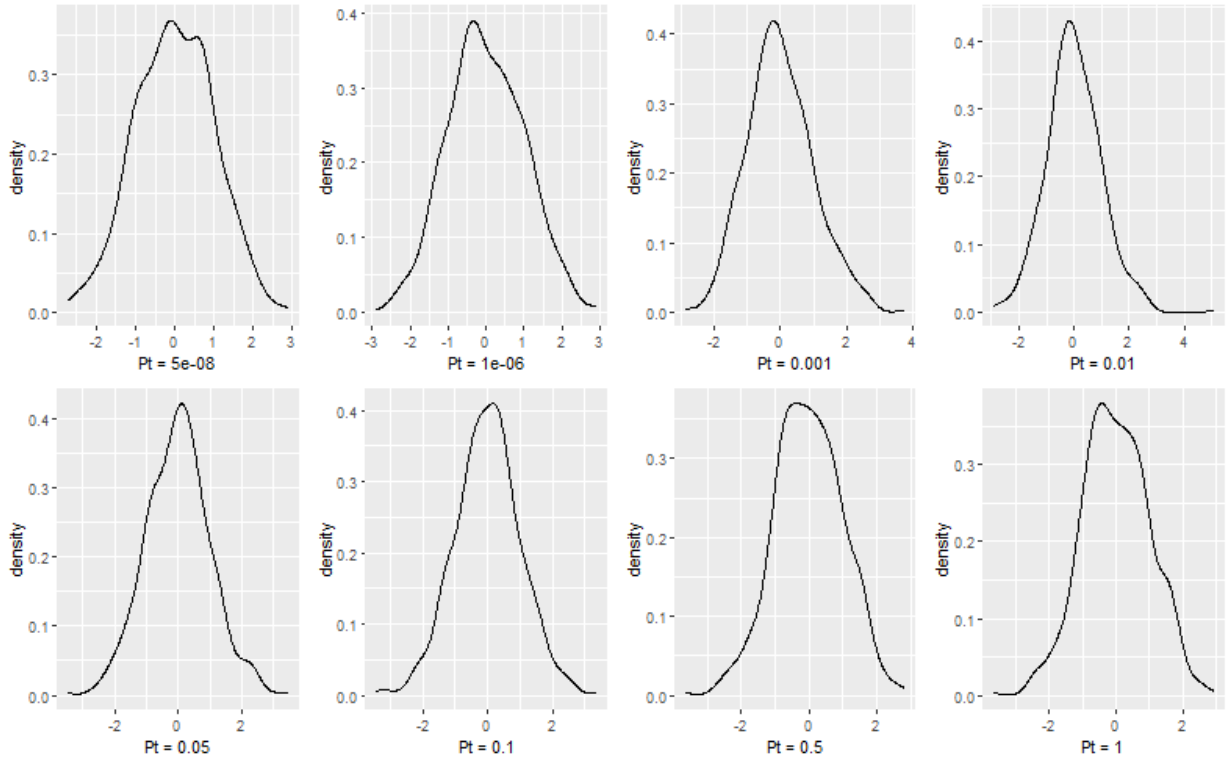
References

Wuttke M, Li Y, Li M, et al. A catalog of genetic loci associated with kidney function from analyses of a million individuals. Nat Genet. 2019;51(6):957–972. doi:10.1038/s41588-019-0407-x

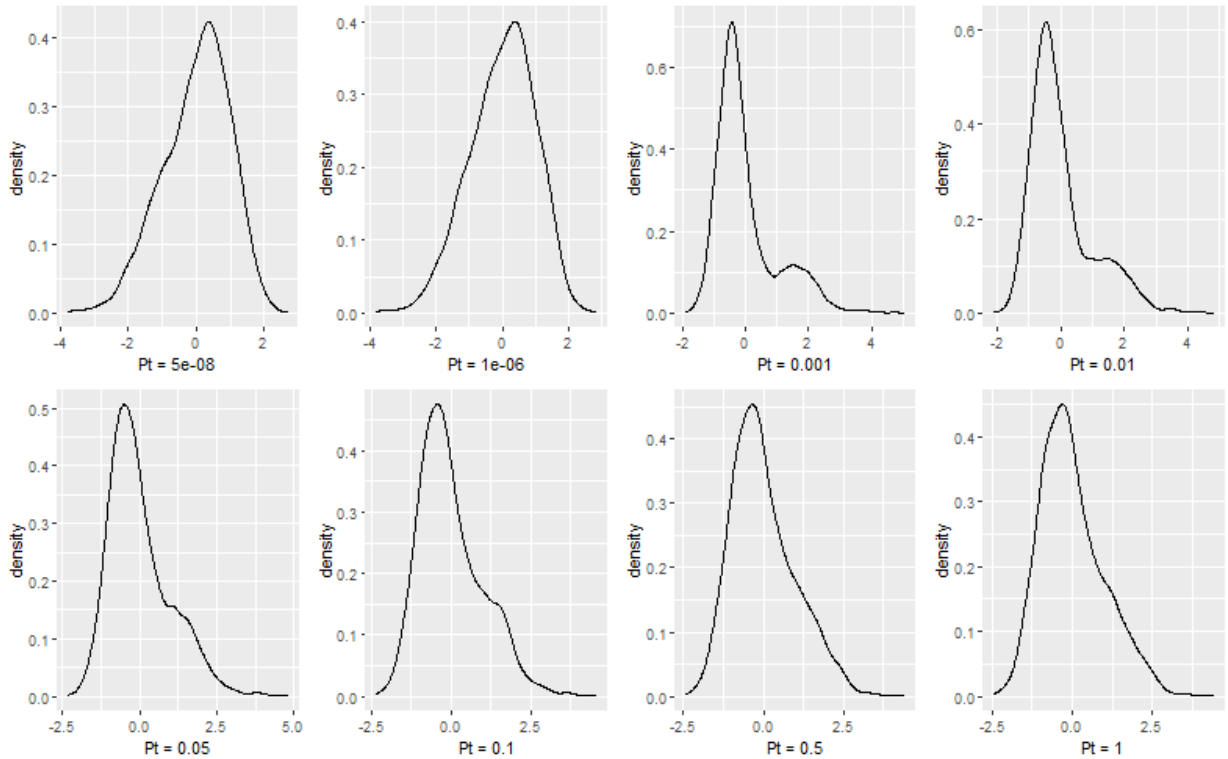
Distribution of polygenic scores by p-value threshold, NHATS European ancestry, n=2827, blood urea nitrogen from the European-ancestry based GWAS



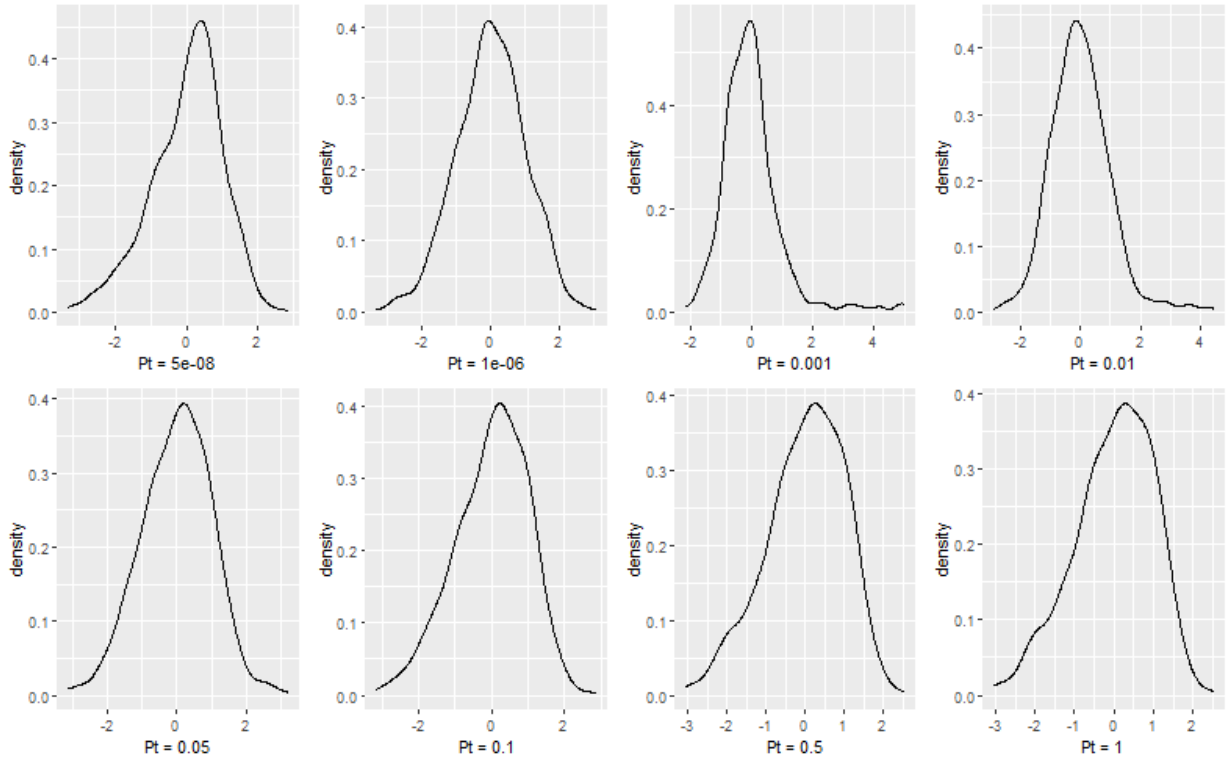
Distribution of polygenic scores by p-value threshold, NHATS African ancestry, n=664, blood urea nitrogen from the European-ancestry based GWAS



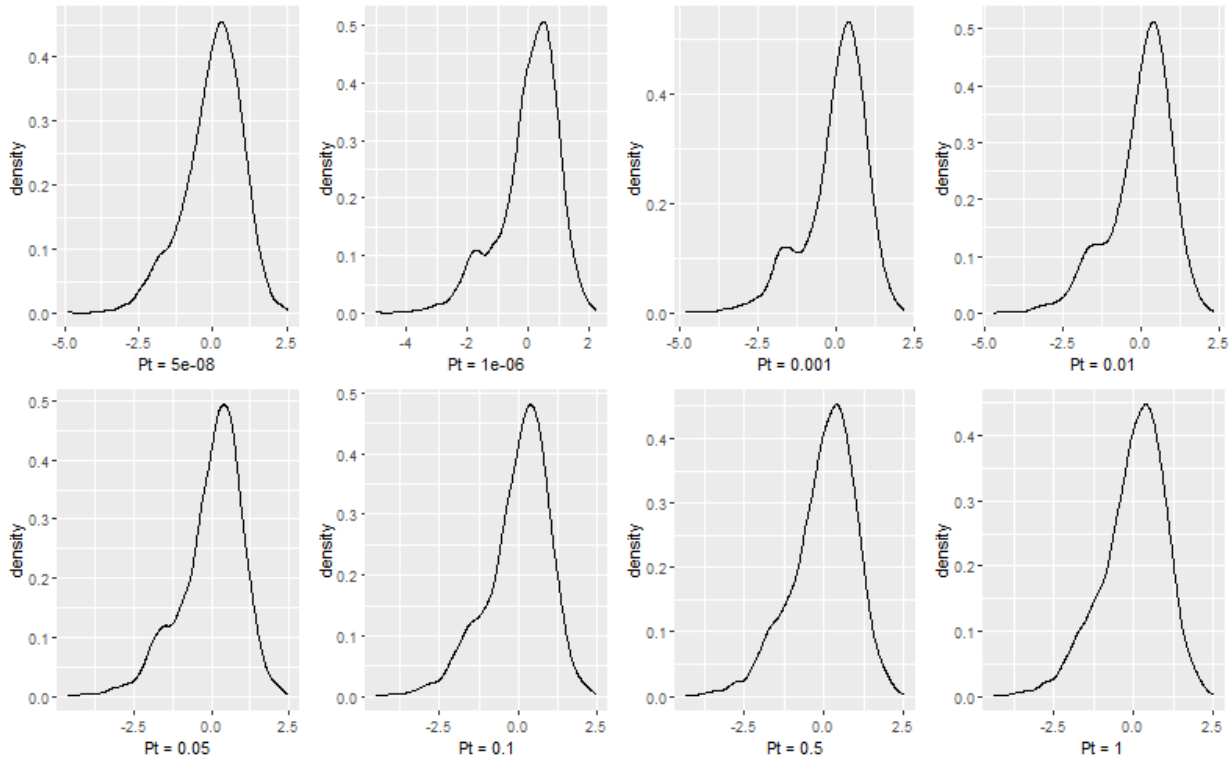
Distribution of polygenic scores by p-value threshold, NHATS European ancestry, n=2827, chronic kidney disease from the European-ancestry based GWAS



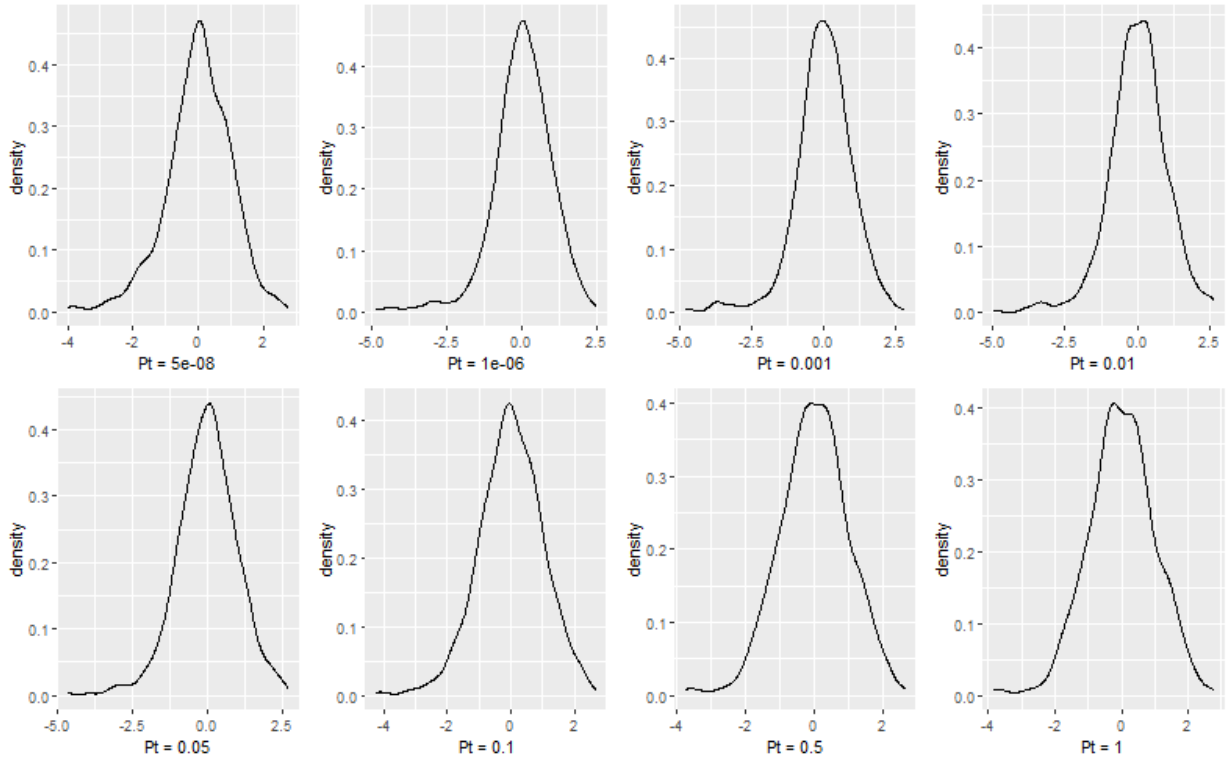
Distribution of polygenic scores by p-value threshold, NHATS African ancestry, n=664, chronic kidney disease from the European-ancestry based GWAS



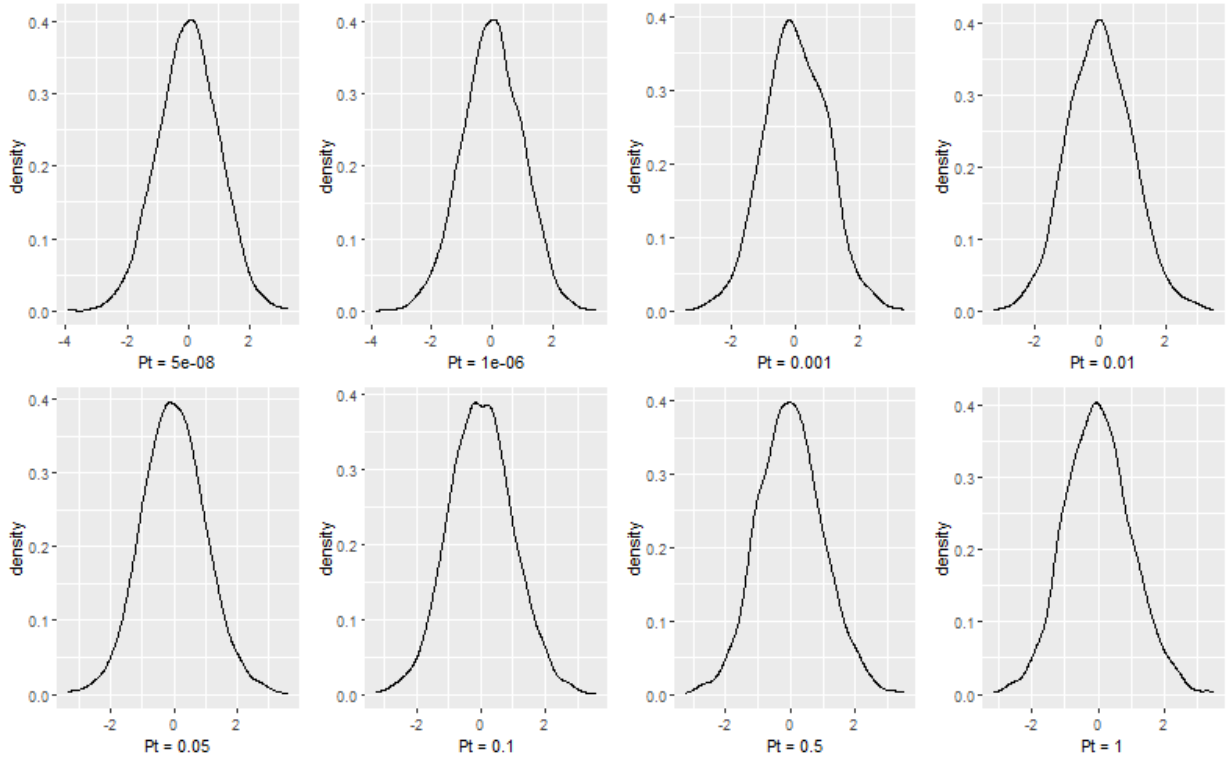
Distribution of polygenic scores by p-value threshold, NHATS European ancestry, n=2827, estimated glomerular filtration rate from the European-ancestry based GWAS



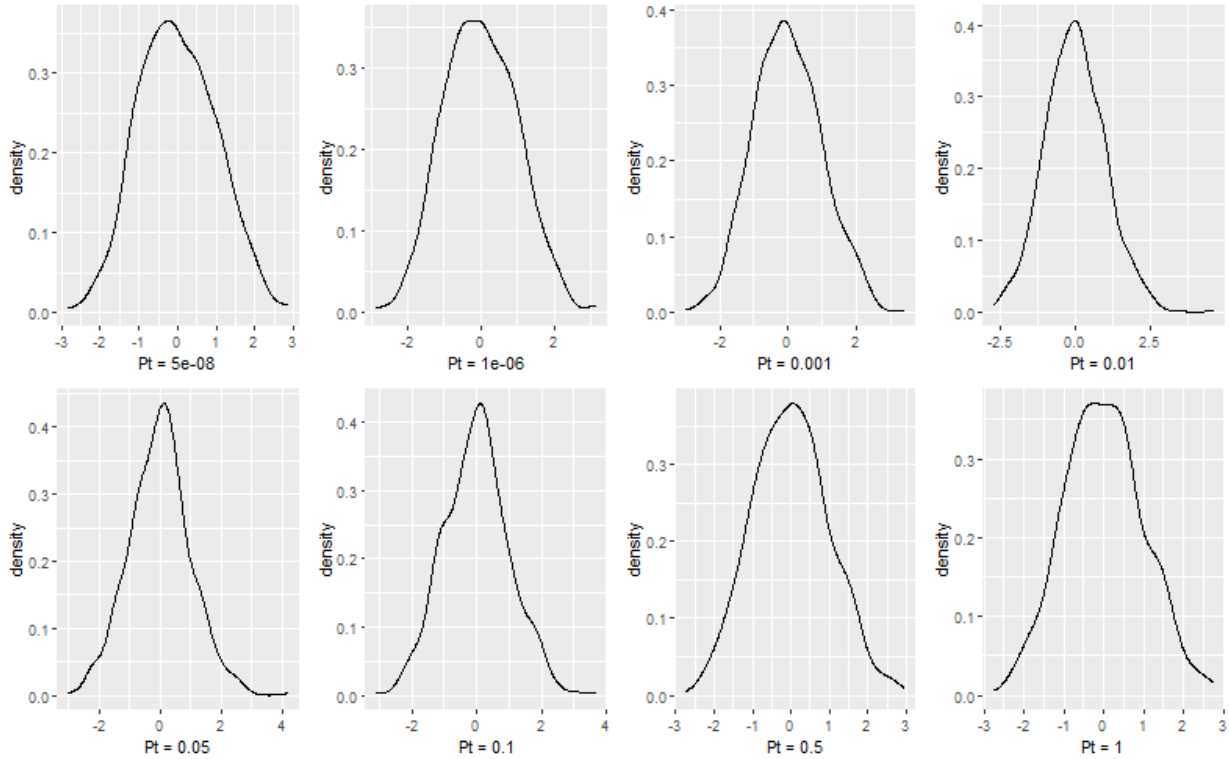
Distribution of polygenic scores by p-value threshold, NHATS African ancestry, n=664, estimated glomerular filtration rate from the European-ancestry based GWAS



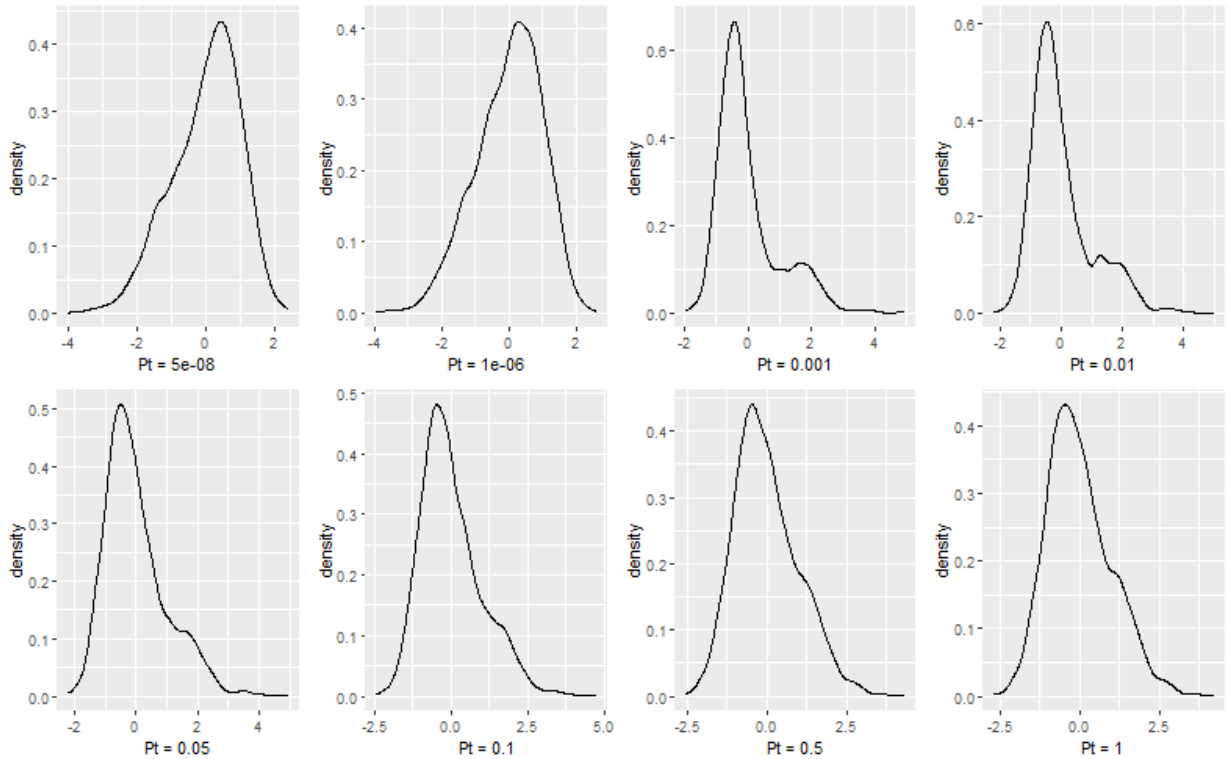
Distribution of polygenic scores by p-value threshold, NHATS European ancestry, n=2827, blood urea nitrogen from the Trans-ancestry based GWAS



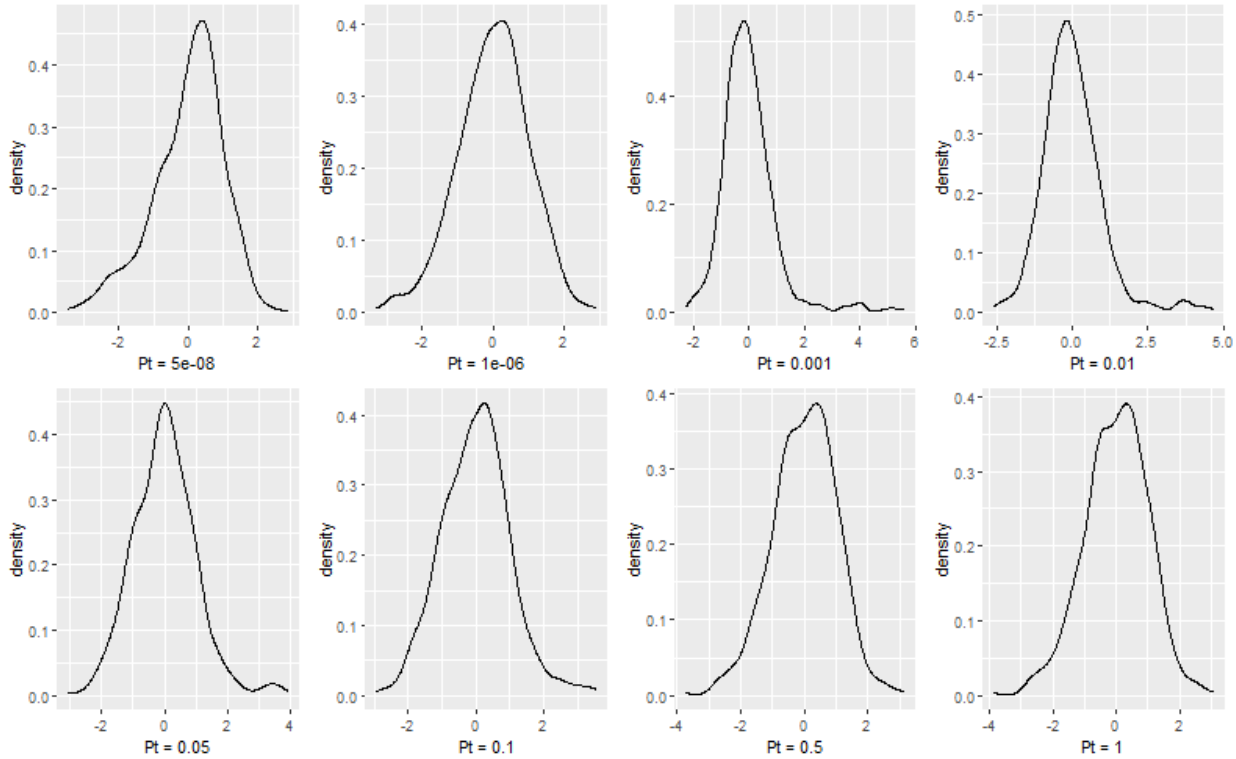
Distribution of polygenic scores by p-value threshold, NHATS African ancestry, n=664, blood urea nitrogen from the Trans -ancestry based GWAS



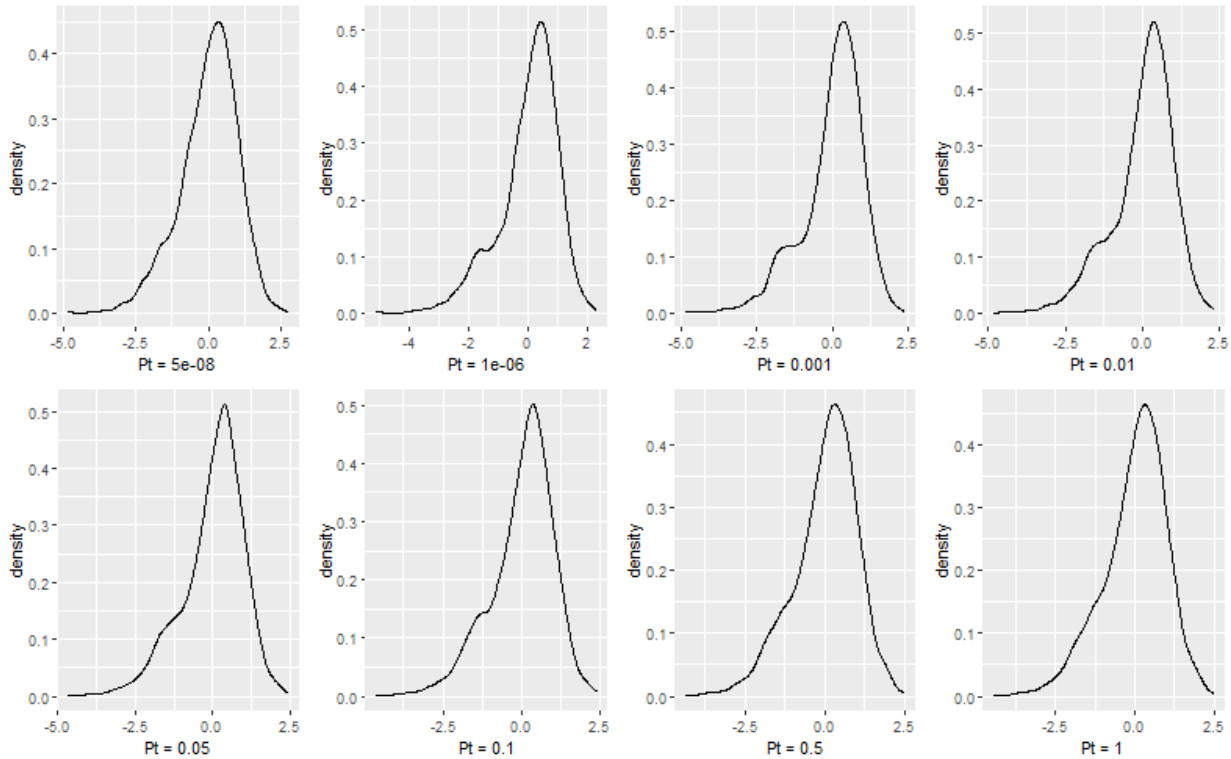
Distribution of polygenic scores by p-value threshold, NHATS European ancestry, n=2827, chronic kidney disease from the Trans -ancestry based GWAS



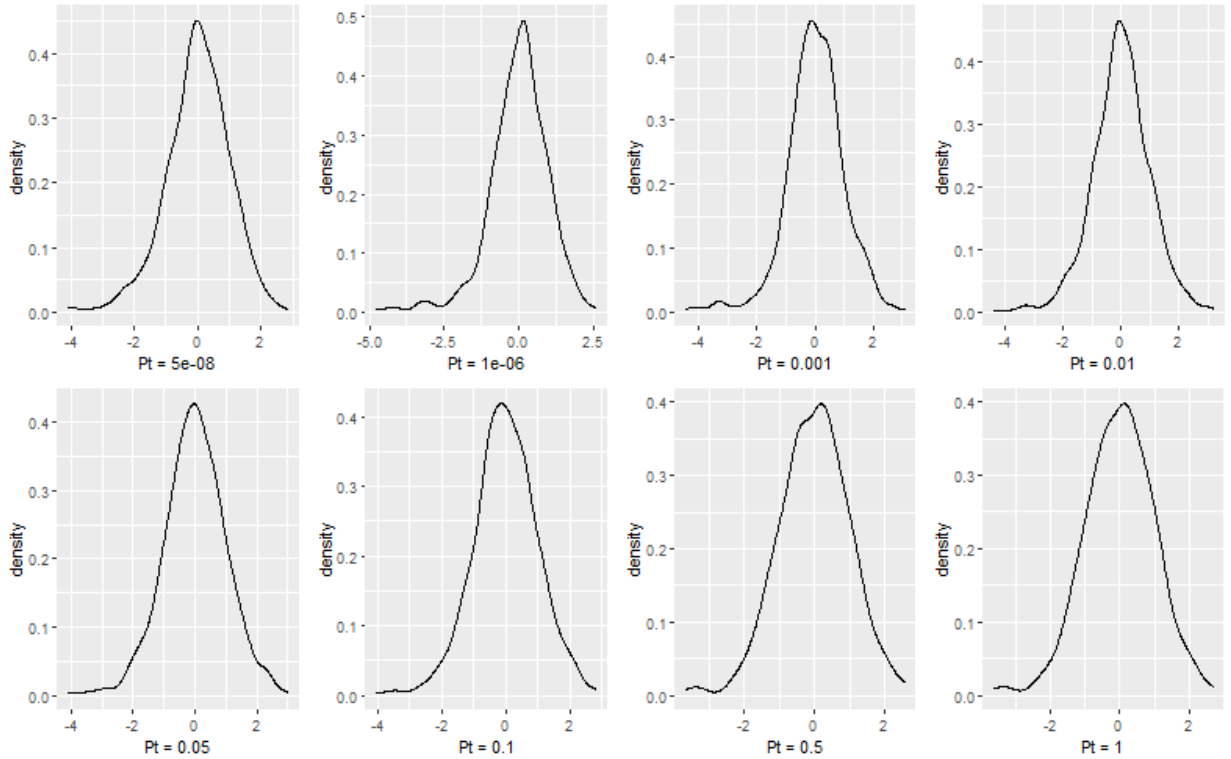
Distribution of polygenic scores by p-value threshold, NHATS African ancestry, n=664, chronic kidney disease from the Trans -ancestry based GWAS



Distribution of polygenic scores by p-value threshold, NHATS European ancestry, n=2827, estimated glomerular filtration rate from the Trans -ancestry based GWAS



Distribution of polygenic scores by p-value threshold, NHATS African ancestry, n=664, estimated glomerular filtration rate from the Trans -ancestry based GWAS



t, u. Waist Circumference and Waist-to-Hip Ratio – Genetic Investigation of ANthropometric Traits 2015

PGSs for waist circumference (WC) and waist-to-hip ratio (WHR) 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 on their data download page:
https://www.broadinstitute.org/collaboration/giant/index.php/GIANT_consortium_data_files
(WC: GIANT 2015 WC COMBINED EUR.txt.gz, WHR: GIANT 2015 WHR COMBINED EUR.txt.gz).
GWAS meta-analysis was performed on a sample of 142,762 individuals from 57 studies across 2,507,022 SNPs, and separately in a Metabochip (MC) meta-analysis on a sample of 67,326 individuals from 44 studies across 124,196 SNPs. A joint GWAS and MC meta-analysis was then conducted on 210,088 individuals across 93,057 SNPs. The GWAS identified 49 loci associated with WHR and an additional 19 loci associated with WC at the genome-wide significance level (**Table 1**). Association analyses adjusted for age, age^2, study-specific covariates if necessary, and BMI.

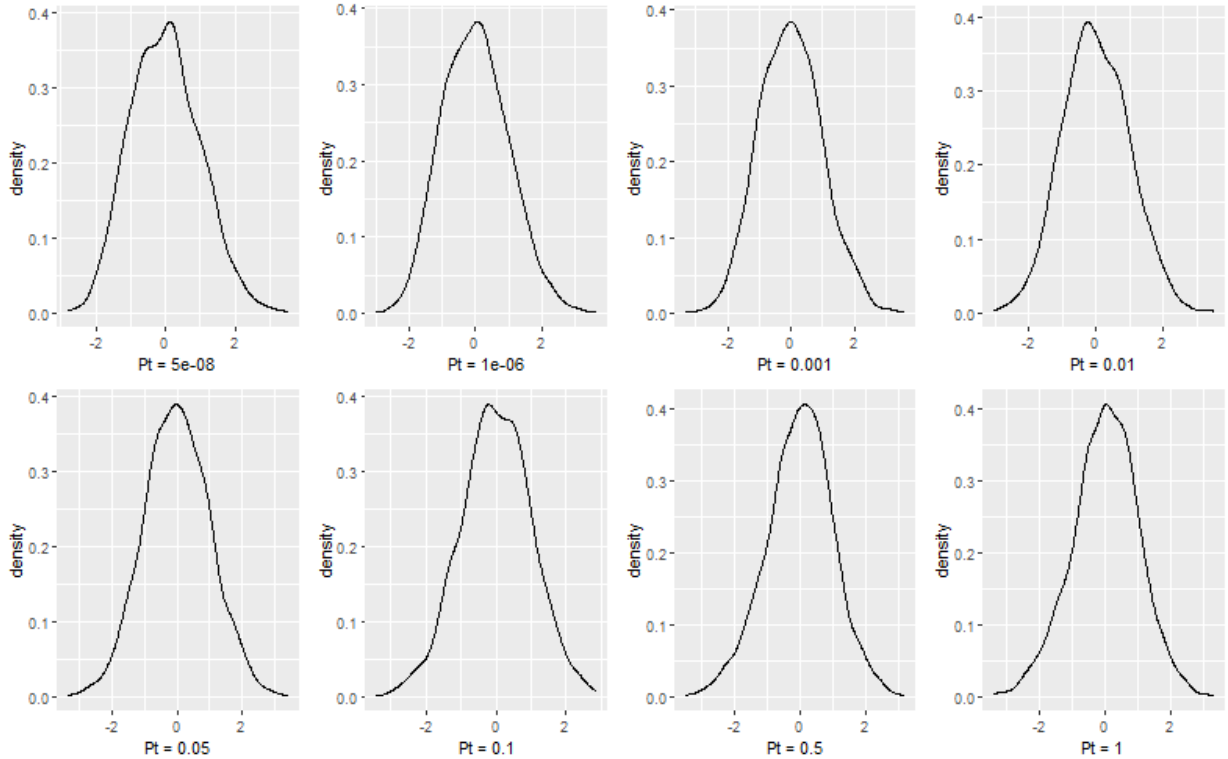
The PGSs (for both African and European ancestry participants) contains 222,093 (WC), 221,993 (WHR), SNPs that overlapped between the NHATS genetic database. Higher PGSs correspond to increased waist circumference or waist-to-hip ratio. The posted PGSs have been standardized within genetic ancestry analysis group to a standard normal curve (mean=0, standard deviation = 1).

N.B.: These weights are from the joint analysis of GWAS and MC meta-analysis conducted on 210,088 individuals. **Please note that the GIANT results are from a GWAS on individuals of European ancestry (see Section C. "Notes about the use of PGSs" for more information on the use of PGSs in other ancestry groups).**

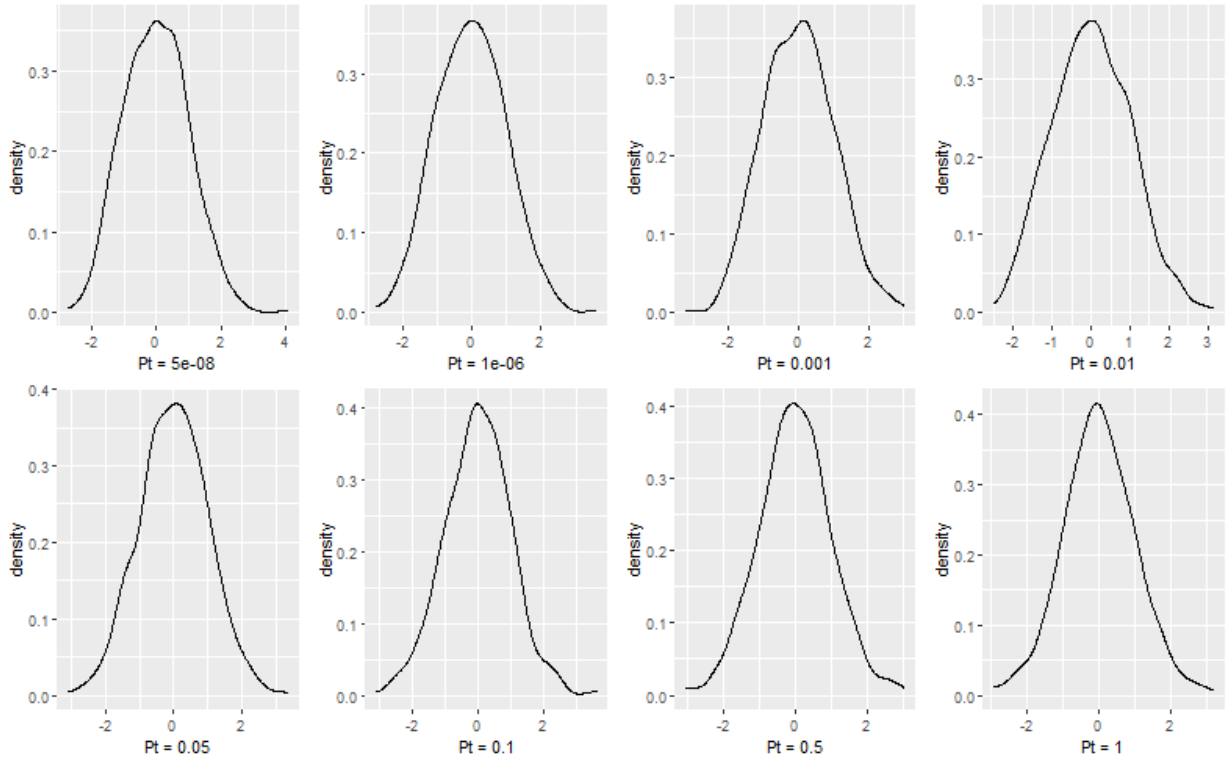
References

Shungin, D., Winkler, T. W., Croteau-Chonka, D. C., Ferreira, T., Locke, A. E., Mägi, R., ... & Workalemahu, T. (2015). New genetic loci link adipose and insulin biology to body fat distribution. *Nature*, *518*(7538), 187.

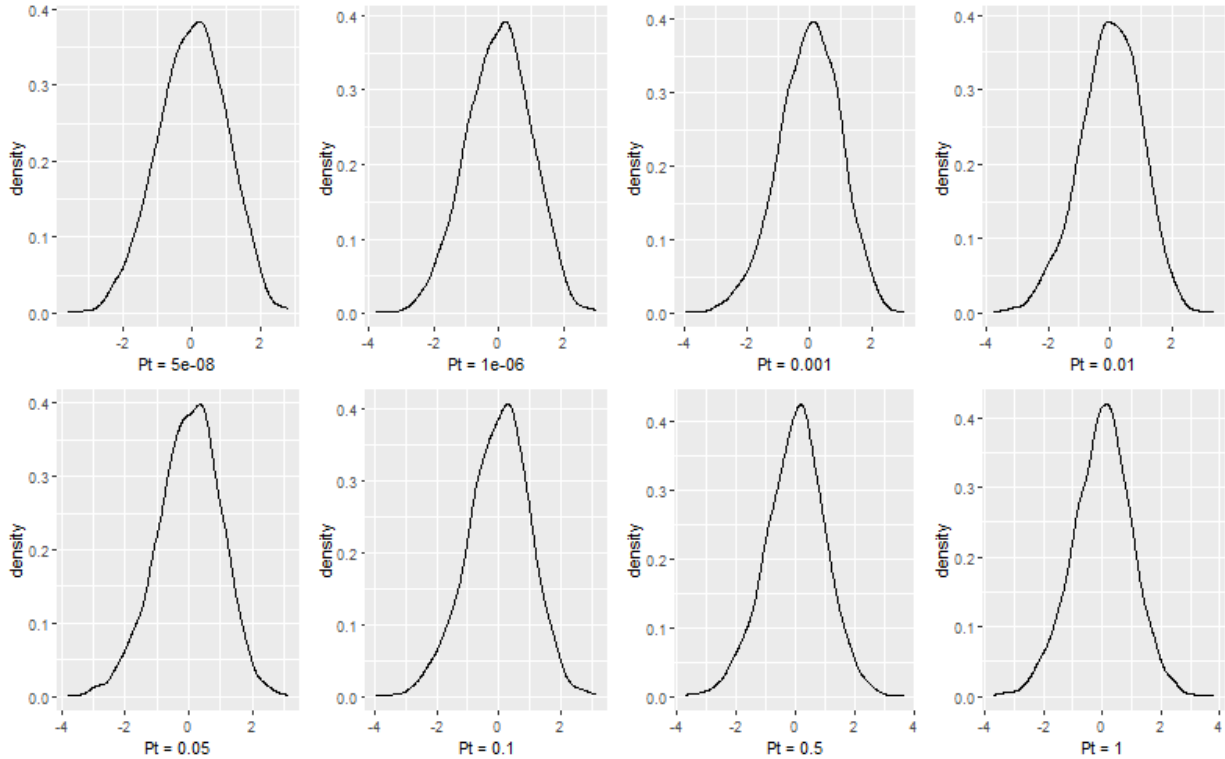
Distribution of polygenic scores by p-value threshold, NHATS European ancestry, n=2827, Waist Circumference (WC)



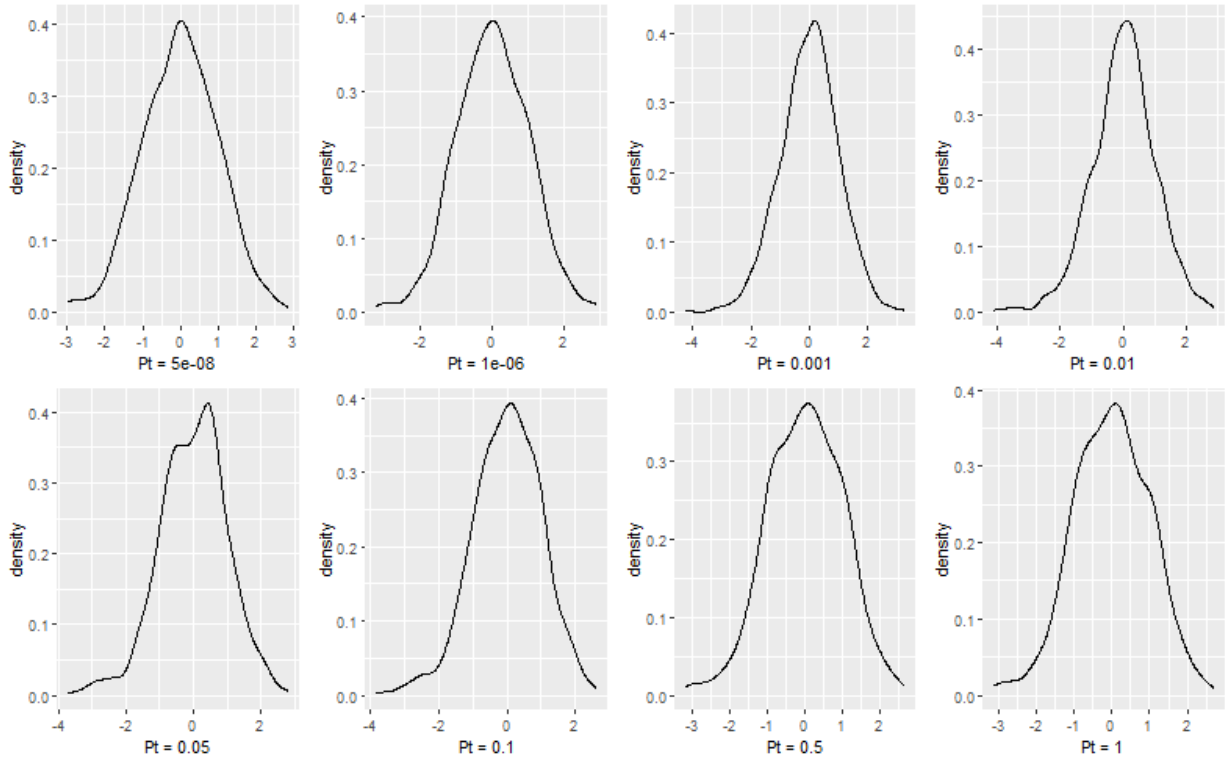
Distribution of polygenic scores by p-value threshold, NHATS African ancestry, n=664, Waist Circumference (WC)



Distribution of polygenic scores by p-value threshold, NHATS European ancestry, n=2827, Waist-to-Hip ratio (WHR)



Distribution of polygenic scores by p-value threshold, NHATS African ancestry, n=664, Waist-to-Hip ratio (WHR)



v-y. Lipid traits (HDL, LDL, total cholesterol, triglycerides) – Global Lipid Genetics Consortium 2013

The HDL, LDL, and TC PGS were created using results from a 2013 study by the Global Lipid Genetics Consortium (Willer et al. 2013). Authors conducted separate GWAS for European (n=188,578) and non-European (n=7,898) ancestries followed by a meta-analysis of 7,168 individuals in a single ancestry group. Only European samples were used for discovery of novel genome-wide significant loci; non-European samples were meta-analyzed and examined only for fine-mapping analyses. Results are available for download directly from the Center for Statistical Genetic's website (http://csg.sph.umich.edu/willer/public/lipids2013/) and results from the joint analysis of metabochip and GWAS data were used to create the PGSs. Results files were slightly modified on 11/26/2013. Sites with N<50,000 were removed from the joint meta-analysis results, sites with N<20,000 were removed from the Metabochip-only results and an rsid column was added to each dataset. Data was sourced by collecting summary statistics from 23 studies of European ancestry genotyped with GWAS arrays and 46 studies genotyped with Metabochip arrays, of which 37 studies consisted primarily of individuals of European ancestry. Nine studies using Metabochip arrays were of non-European ancestry: two studies were South Asian, two studies were East Asian, and five studies were African. Blood lipid levels were typically measured after > 8 hours of fasting and individuals known to be on lipid-lowering medication were excluded when possible. Hapmap release 22 CEU reference was used. In cases where Metabochip and GWAS array data were available for the same individuals, Metabochip data was used to ensure key variants were directly genotyped, rather than imputed. The study identified 157 loci associated with lipid levels at P <5×01-8, including 62 loci not previously associated with lipid levels in humans. Adjustments for population structure using principal component analysis or mixed model approaches were carried out in 24 studies (35% of individuals).

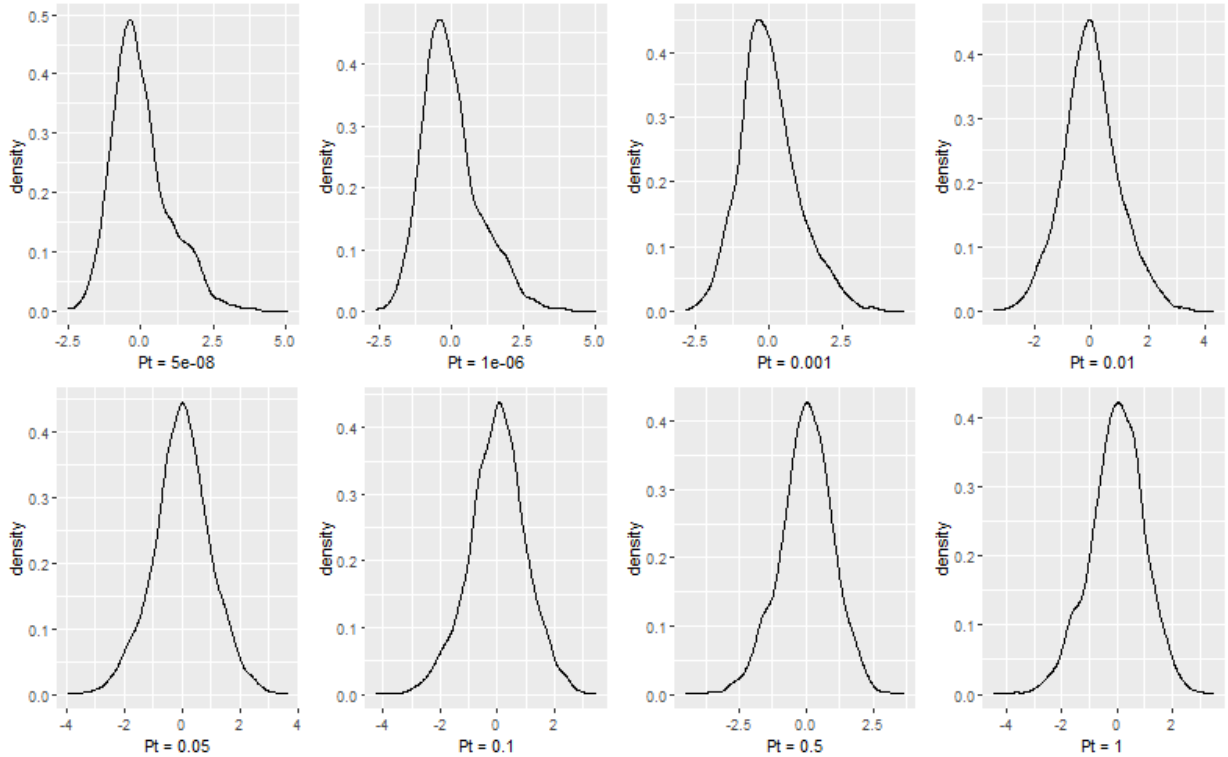
The PGSs (for both African and European ancestry participants) contains 212,447 (HDL), 211,316 (LDL), 212,409 (total cholesterol), and 211,437 (triglycerides) SNPs that overlapped between the NHATS genetic database. Higher PGSs correspond to increased lipid trait. The posted PGSs have been standardized within genetic ancestry analysis group to a standard normal curve (mean=0, standard deviation = 1).

N.B.: **Please note that the GLGC-lipid results contain PGSs from European ancestry backgrounds (see "Notes about the use of PGSs" for more information on the use of PGSs in other ancestry groups).**

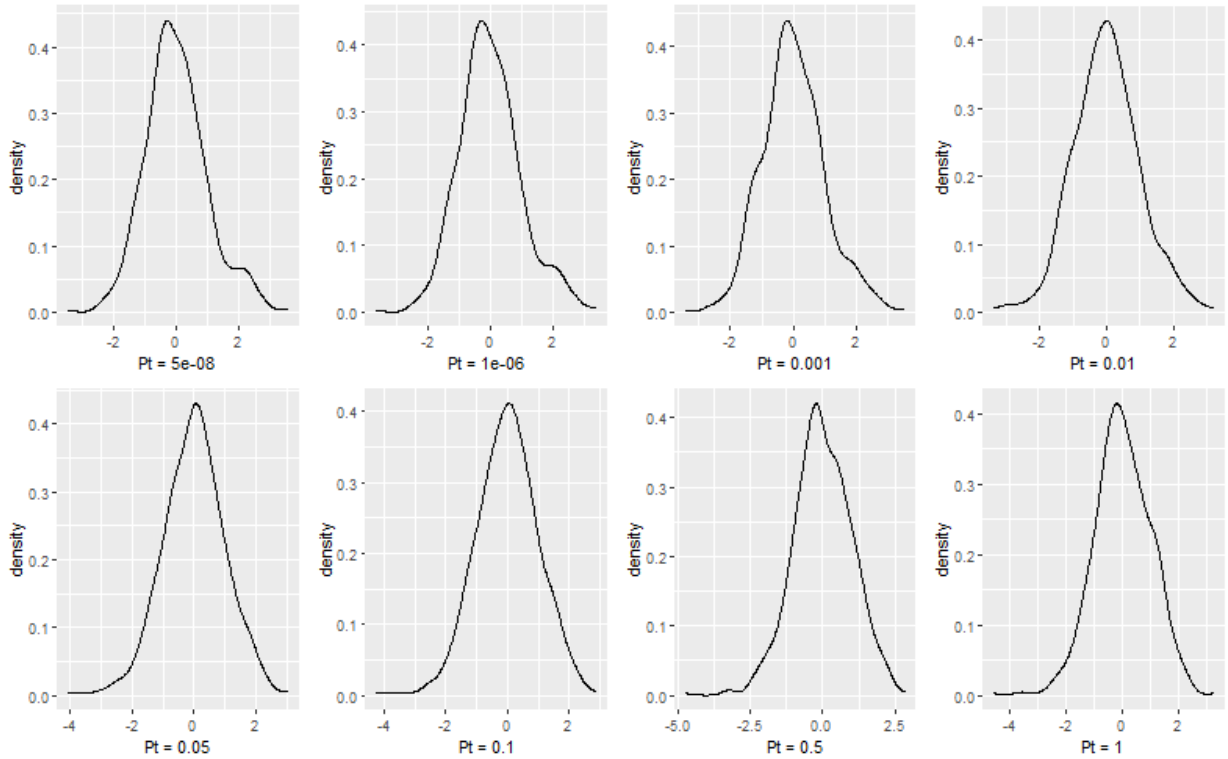
References

Willer, C.J., Schmidt, E.M., Sengupta, S., Peloso, G.M., Gustafsson, S., … & Global Lipids Genetic Consortium. (2013) Discovery and Refinement of Loci Associated with Lipid Levels. Nat Genet. 45(11), 1274-1283. doi:10.1038/ng.2797.

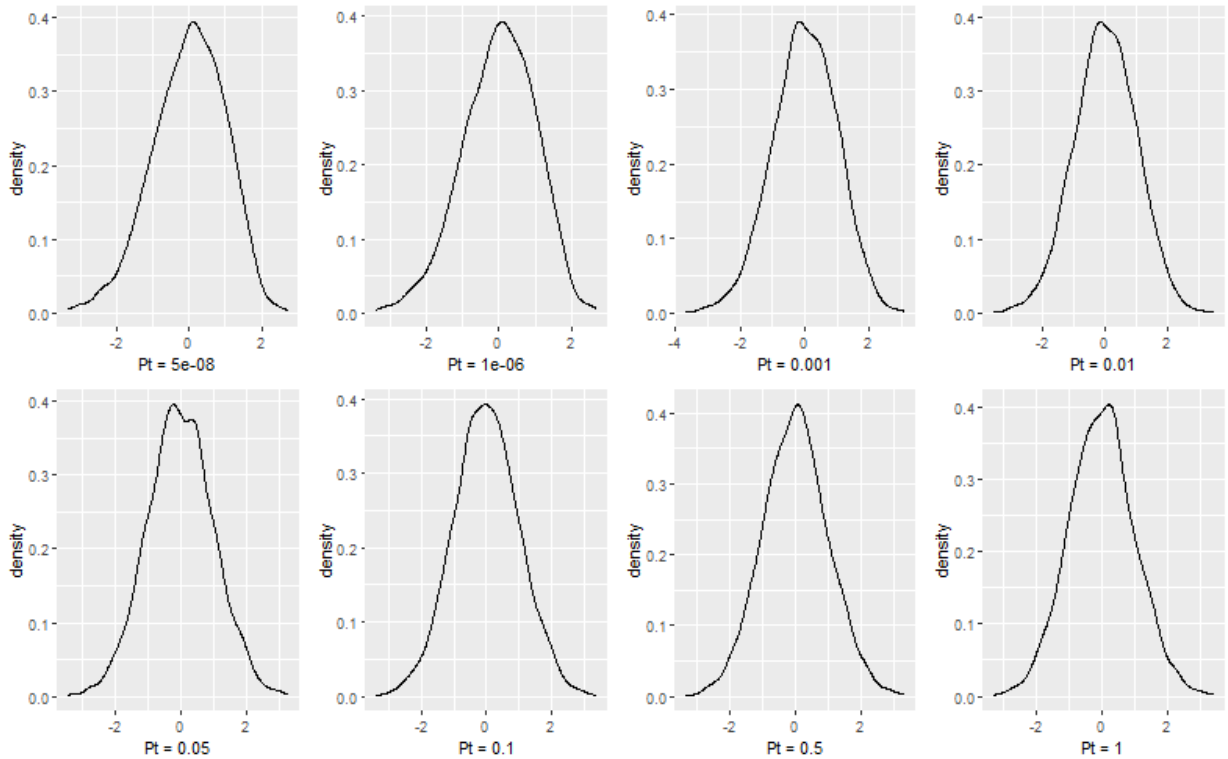
Distribution of polygenic scores by p-value threshold, NHATS European ancestry, n=2827, high density lipoprotein (HDL)



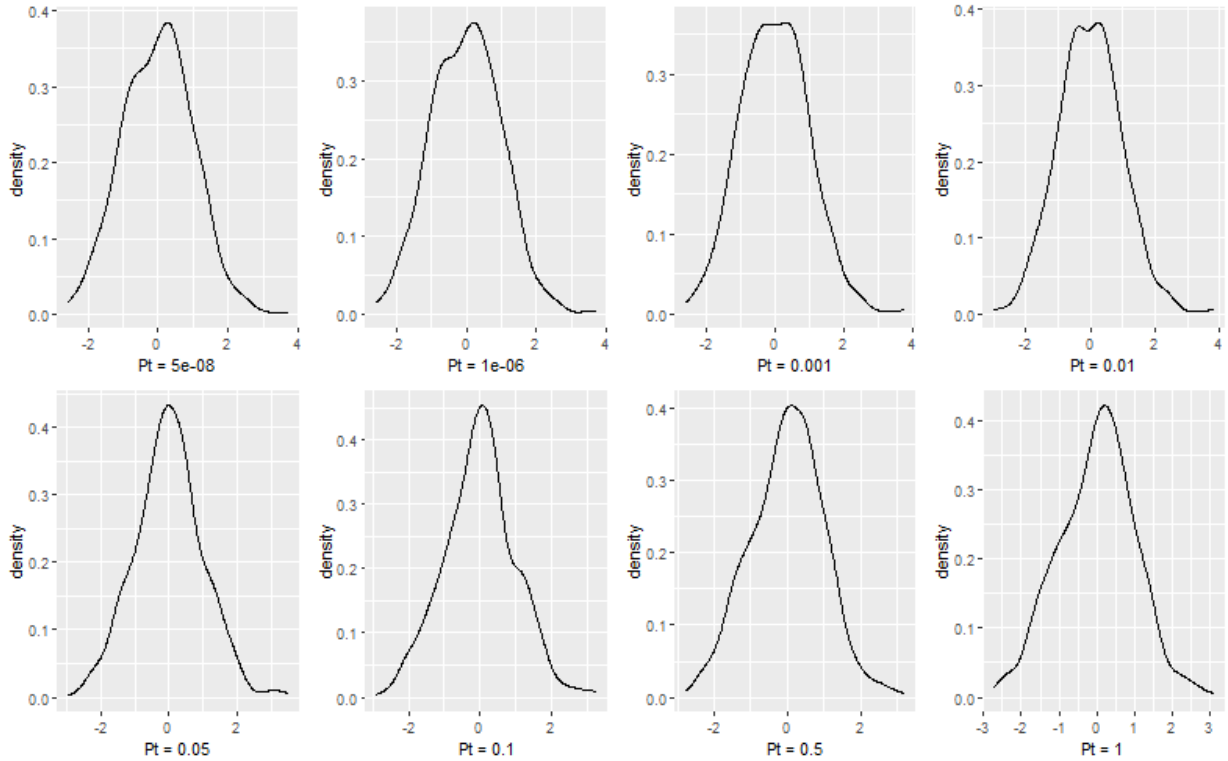
Distribution of polygenic scores by p-value threshold, NHATS African ancestry, n=664, high density lipoprotein (HDL)



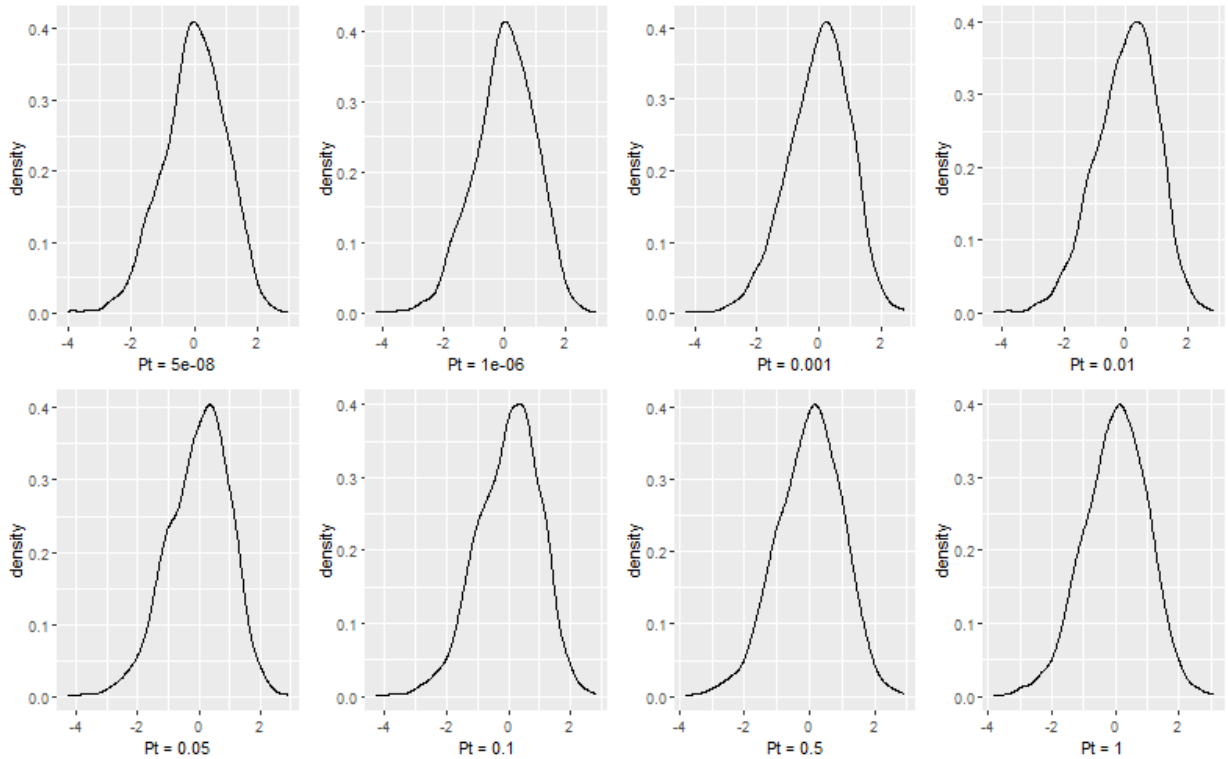
Distribution of polygenic scores by p-value threshold, NHATS European ancestry, n=2827, low density lipoprotein (LDL)



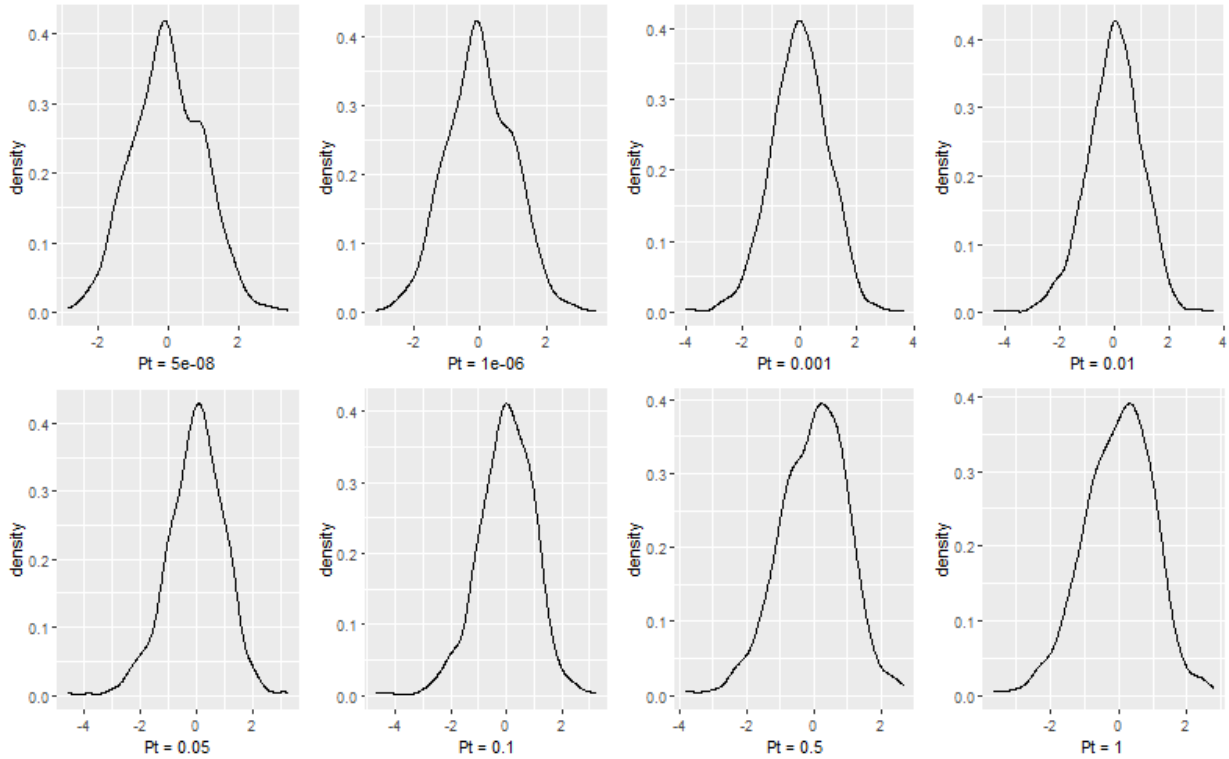
Distribution of polygenic scores by p-value threshold, NHATS African ancestry, n=664, low density lipoprotein (LDL)



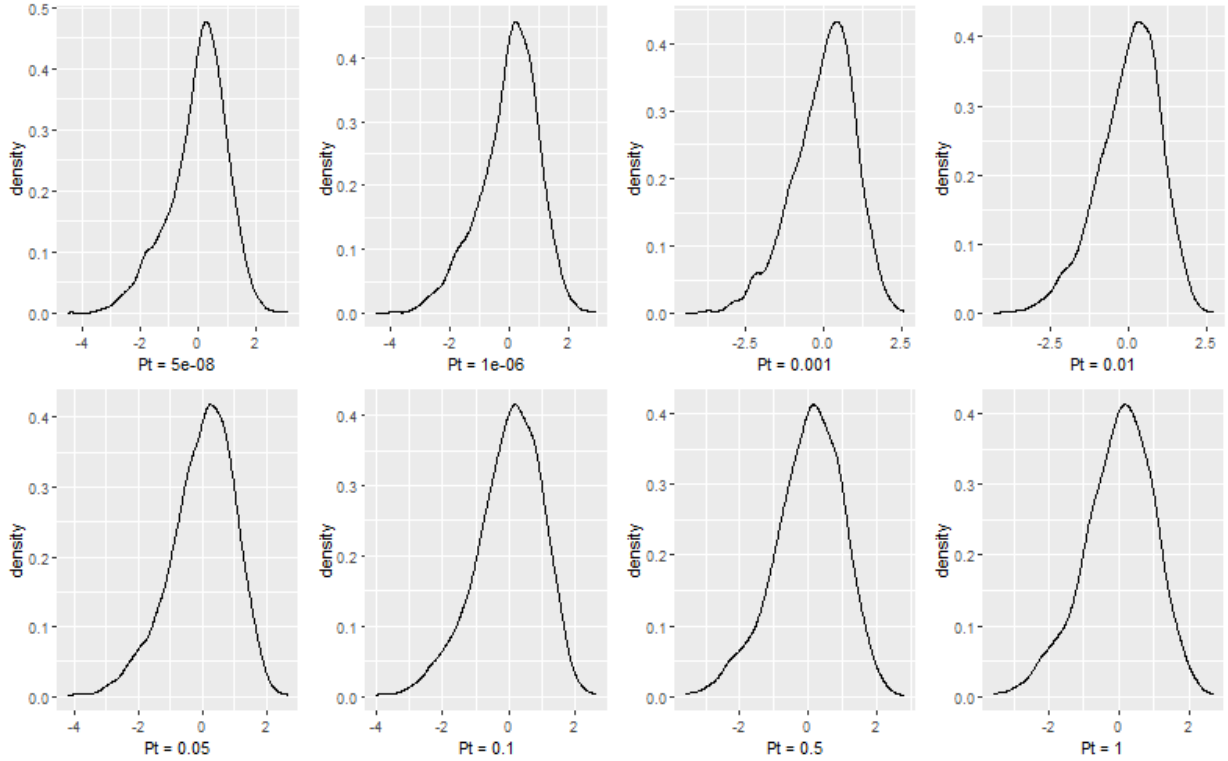
Distribution of polygenic scores by p-value threshold, NHATS European ancestry, n=2827, total cholesterol (TC)



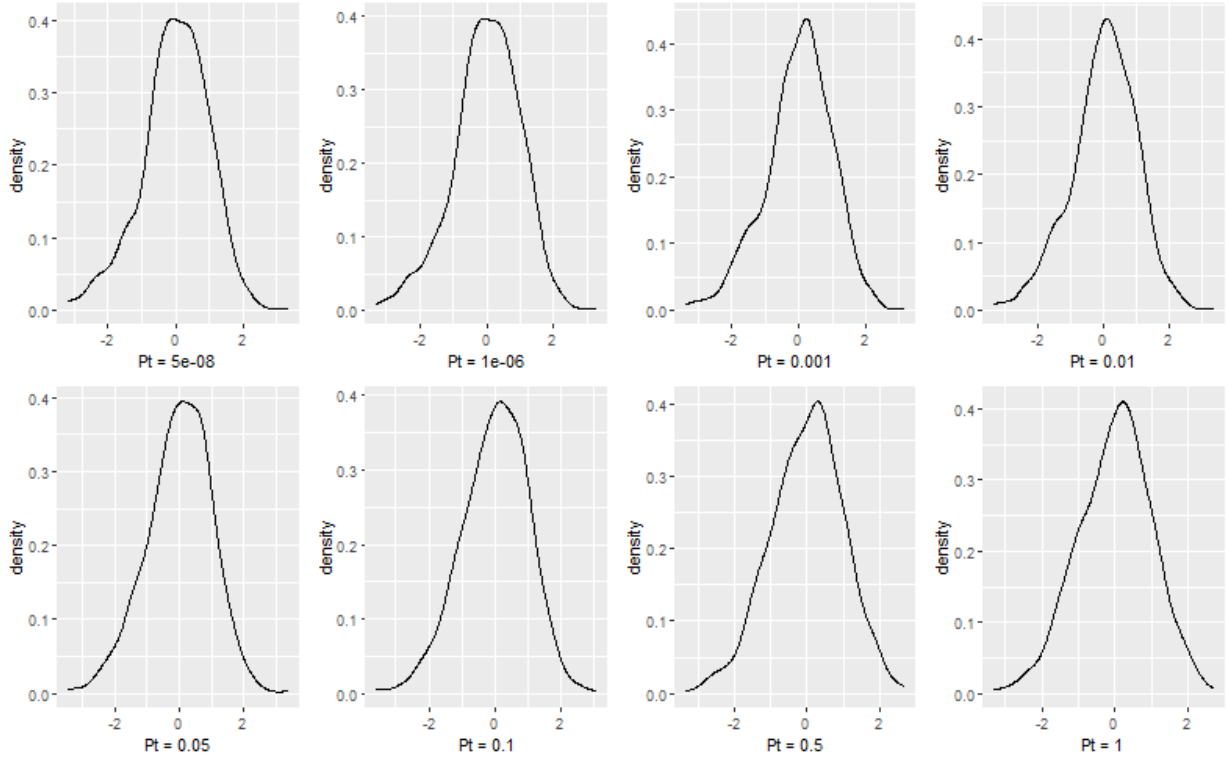
Distribution of polygenic scores by p-value threshold, NHATS African ancestry, n=664, total cholesterol (TC)



Distribution of polygenic scores by p-value threshold, NHATS European ancestry, n=2827, triglycerides (TG)



Distribution of polygenic scores by p-value threshold, NHATS African ancestry, n=664, low triglycerides (TG)



z. Age at Menarche – REPROductive GENetics consortium 2014

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 on the ReproGen data download page:
http://www.reprogen.org/data_download.html
(Menarche_Nature2014_GWASMetaResults_17122014.txt). The ReproGen meta-analysis included 182,416 women of European descent from 57 studies imputed to HapMap Phase 2 CEU build 35 or 36 with at total of 2,441,815 autosomal SNPs. Birth year was the only covariate included to allow for the secular trends in menarche timing. The study reported 3,915 genome-wide significant SNPs (**Figure 1**). Of these, the authors identified 123 independent signals for age at menarche, which they assessed further in an independent sample of 8,689 women from the EPIC-InterAct study.

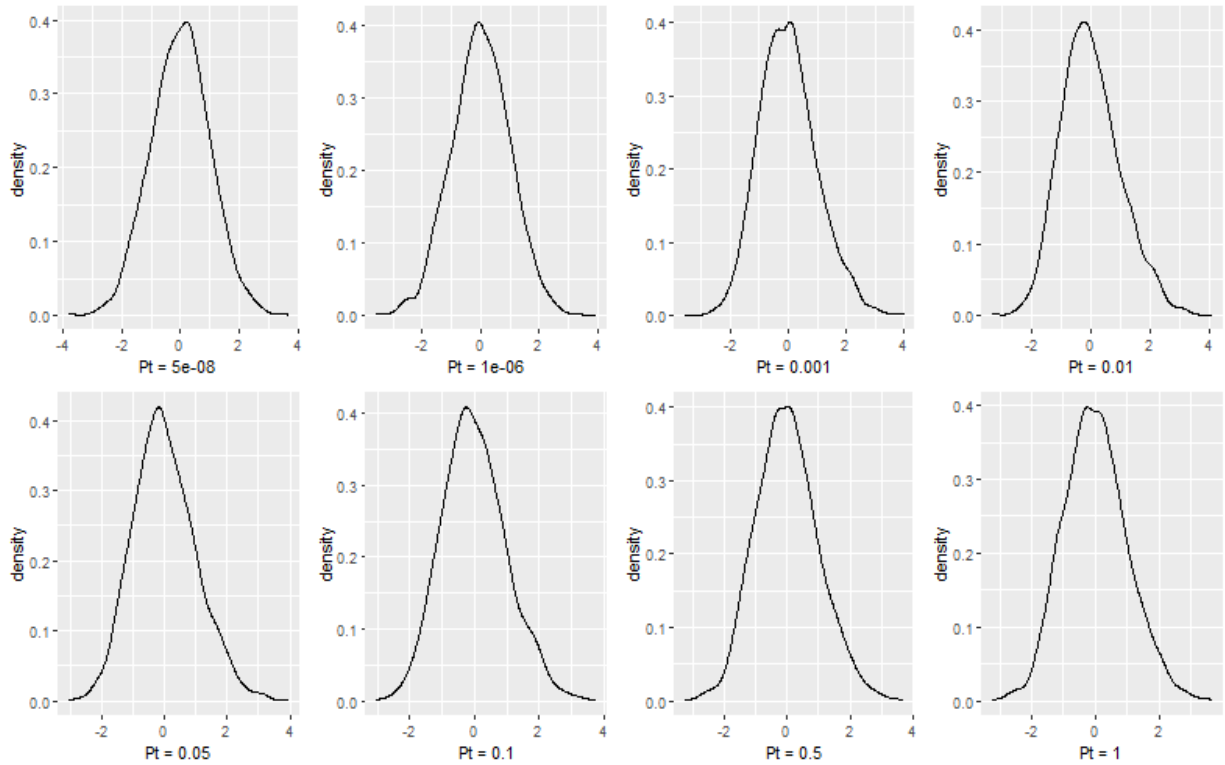
The PGSs contain 215,707 SNPs that overlapped between the NHATS genetic database and the GWAS meta-analysis. Higher PGSs correspond to increased age at menarche. The posted PGSs have been standardized within genetic ancestry analysis group to a standard normal curve (mean=0, standard deviation = 1).

N.B.: **Please note that the ReproGen results are from a GWAS on individuals of European ancestry (see "Notes about the use of PGSs" for more information on the use of PGSs in other ancestry groups).**

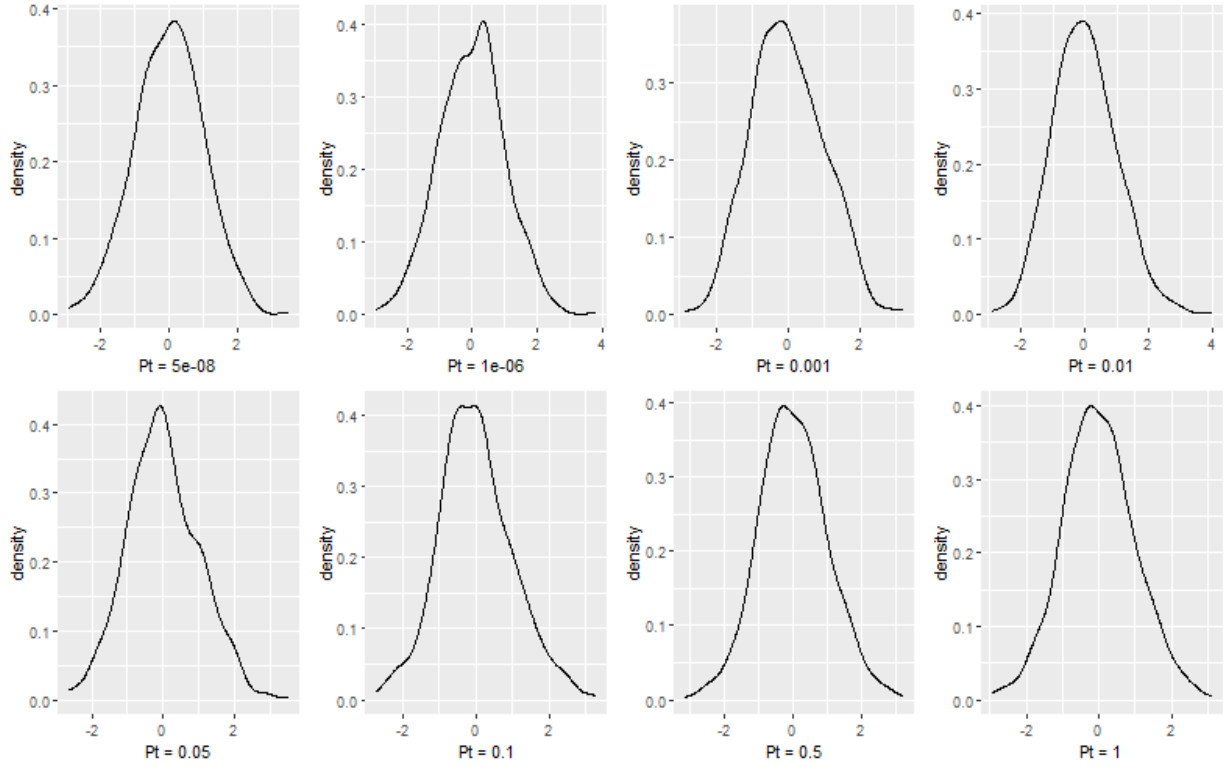
References:

Perry, J. R., Day, F., Elks, C. E., Sulem, P., Thompson, D. J., Ferreira, T., ... & Albrecht, E. (2014). Parent-of-origin-specific allelic associations among 106 genomic loci for age at menarche. *Nature*, *514*(7520), 92-97.

Distribution of polygenic scores by p-value threshold, NHATS European ancestry, n=2827



Distribution of polygenic scores by p-value threshold, NHATS African ancestry, n=664



aa. Age at Menopause – REPROductive GENetics consortium 2015

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 on the ReproGen data download page: http://www.reprogen.org/data_download.html. The ReproGen meta-analysis included 69,360 women of European descent from 33 studies imputed to HapMap Phase 2 CEU build 35 or 36. The study reported 44 genome-wide significant regions (**Table 1**).

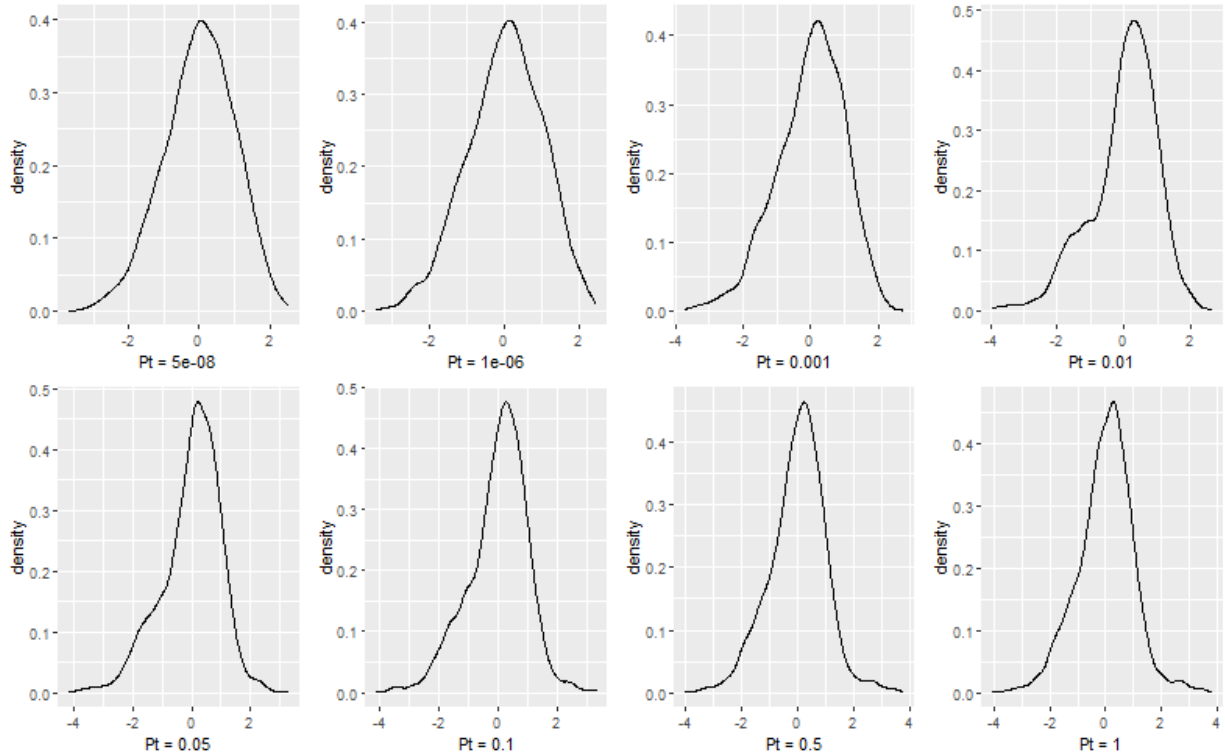
The PGSs contain 212,384 SNPs that overlapped between the NHATS genetic database and the GWAS meta-analysis. Higher PGSs correspond to increased age at menopause. The posted PGSs have been standardized within genetic ancestry analysis group to a standard normal curve (mean=0, standard deviation = 1).

N.B.: **Please note that the ReproGen results are from a GWAS on individuals of European ancestry (see "Notes about the use of PGSs" for more information on the use of PGSs in other ancestry groups).**

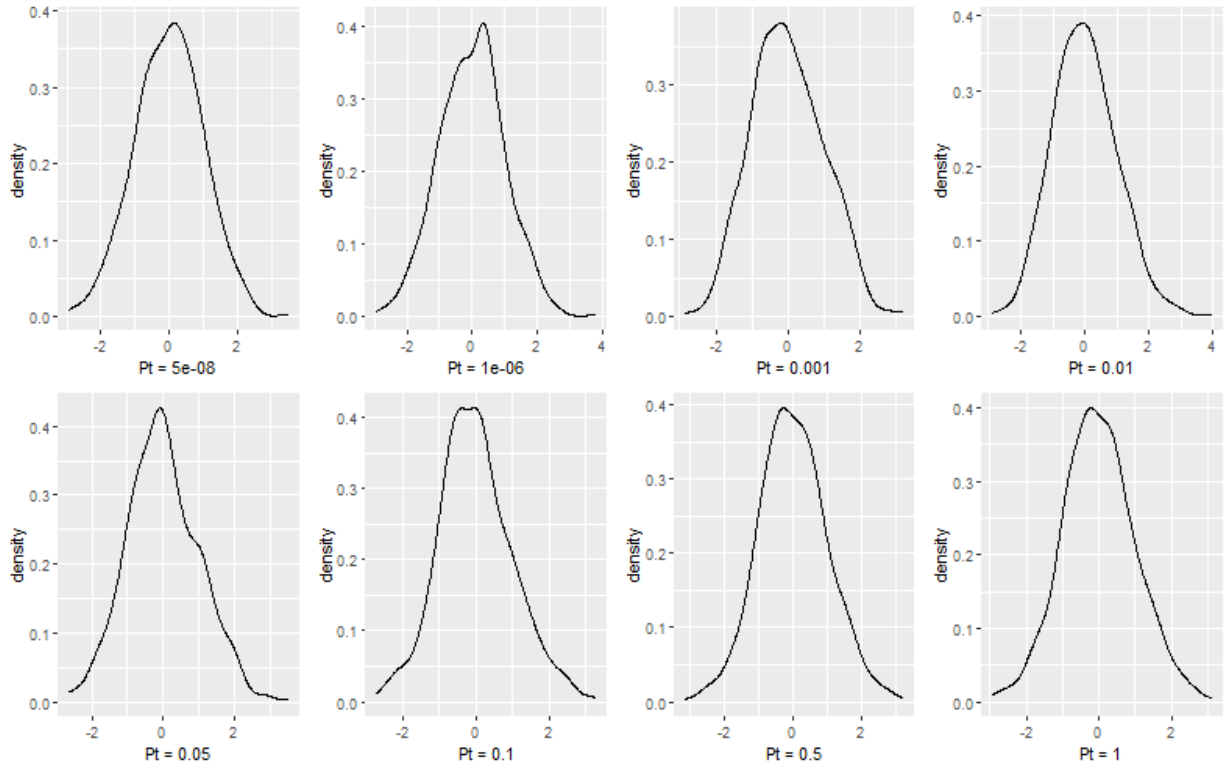
References:

Day, F. R., Ruth, K. S., Thompson, D. J., Lunetta, K. L., Pervjakova, N., Chasman, D. I., … Murray, A. (2015). Large-scale genomic analyses link reproductive ageing to hypothalamic signaling, breast cancer susceptibility and BRCA1-mediated DNA repair. Nature Genetics, 47(11), 1294–1303. http://doi.org/10.1038/ng.3412

Distribution of polygenic scores by p-value threshold, NHATS European ancestry, n=2827



Distribution of polygenic scores by p-value threshold, NHATS African ancestry, n=664



ab, ac. Alzheimer's Disease - International Genomics of Alzheimer's Project 2013

The PGSs for Alzheimer's disease (AD) were created using results from a 2013 GWAS conducted by the International Genomics of Alzheimer's Project (IGAP): http://web.pasteur-lille.fr/en/recherche/u744/igap/igap_download.php. A GWAS meta-analysis of AD was conducted across 20 independent studies using data from four international consortia: Alzheimer 's Disease Genetic Consortium (ADGC), the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium, the European Alzheimer's Disease Initiative (EADI), and the Genetic and Environmental Risk in Alzheimer's Disease (GERAD) Consortium. The stage 1 meta-analysis included 54,162 participants (N_{cases}=17,008 and $N_{controls}$=37,154) of European decent with a total of 7,055,881 SNPs imputed to 1000 Genomes (2010 release). The stage 2 replication sample included 19,884 participants of European ancestry (N_{cases}=8,572 and $N_{controls}$=11,312) with a total of 11,632 genotyped SNPs. In addition to the *APOE* locus (encoding apolipoprotein E), the two-stage combined discovery and replication GWAS revealed 19 SNPs that reached genome-wide significant associations with AD (**Table 2**). Adjustment covariates within each contributing cohort included age, sex, and genetic principal components.

The Alzheimer's disease "no *APOE* region" PGS contains 375,856 SNPs that overlapped between the NHATS genetic database and the GWAS meta-analysis. The Alzheimer's disease PGSs that retain the *APOE* gene region contain 375,880 SNPs that overlapped between the NHATS genetic database and the GWAS meta-analysis. The posted PGSs have been standardized within genetic ancestry analysis group to a standard normal curve

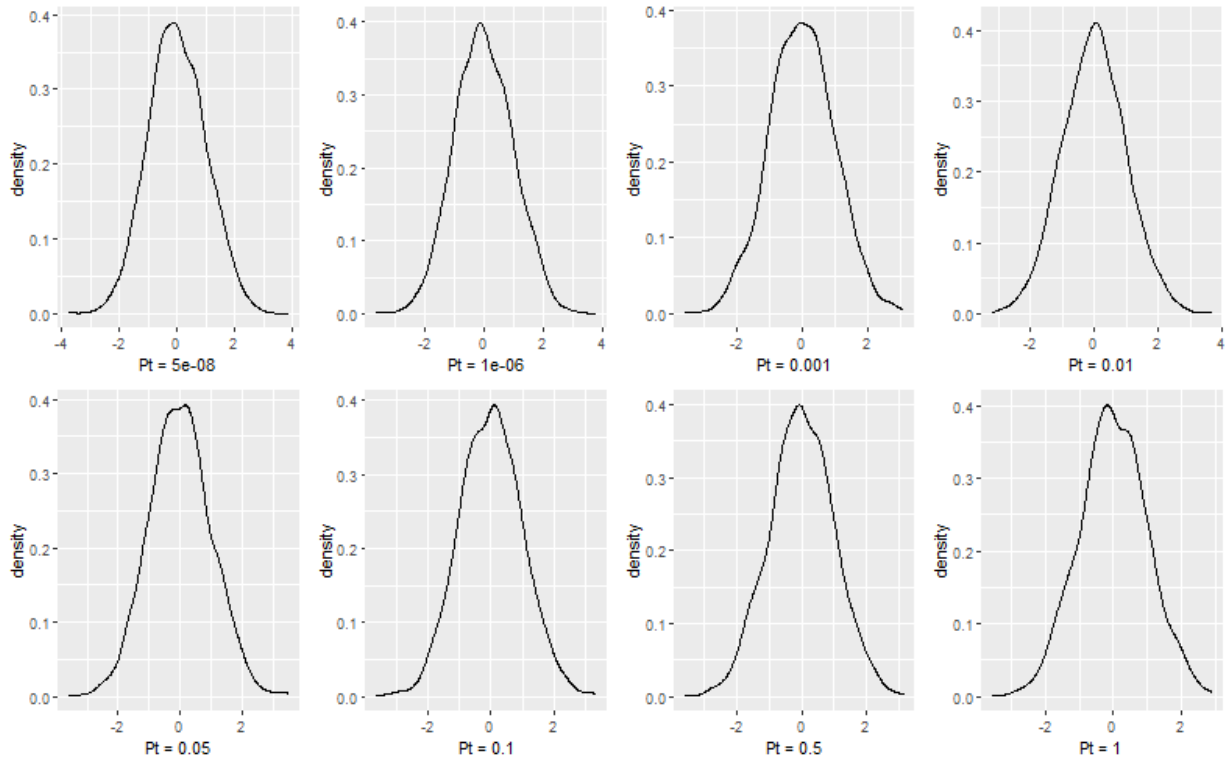
(mean=0, standard deviation = 1). Higher PGSs correspond to increased odds of Alzheimer's disease. The *APOE* gene region was considered to be chr19: 45,384,477 to 45,432,606, build 37/hg 19. This is *not* the same as the *APOE-epsilon* haplotype.

N.B.: There are TWO types of Alzheimer's disease scores published here representing PGS with and without the *APOE* gene region. The correlation between the two scores at a genome-wide significance level is 0.711 for the European ancestry sample and 0.575 for the African ancestry sample. At a p-value threshold of 1, the correlation between the two scores is 0.977 for the European ancestry sample and 0.954 for the African ancestry sample.
Please note that the IGAP results are from a GWAS on individuals of European ancestry (see "Notes about the use of PGSs" for more information on the use of PGSs in other ancestry groups).

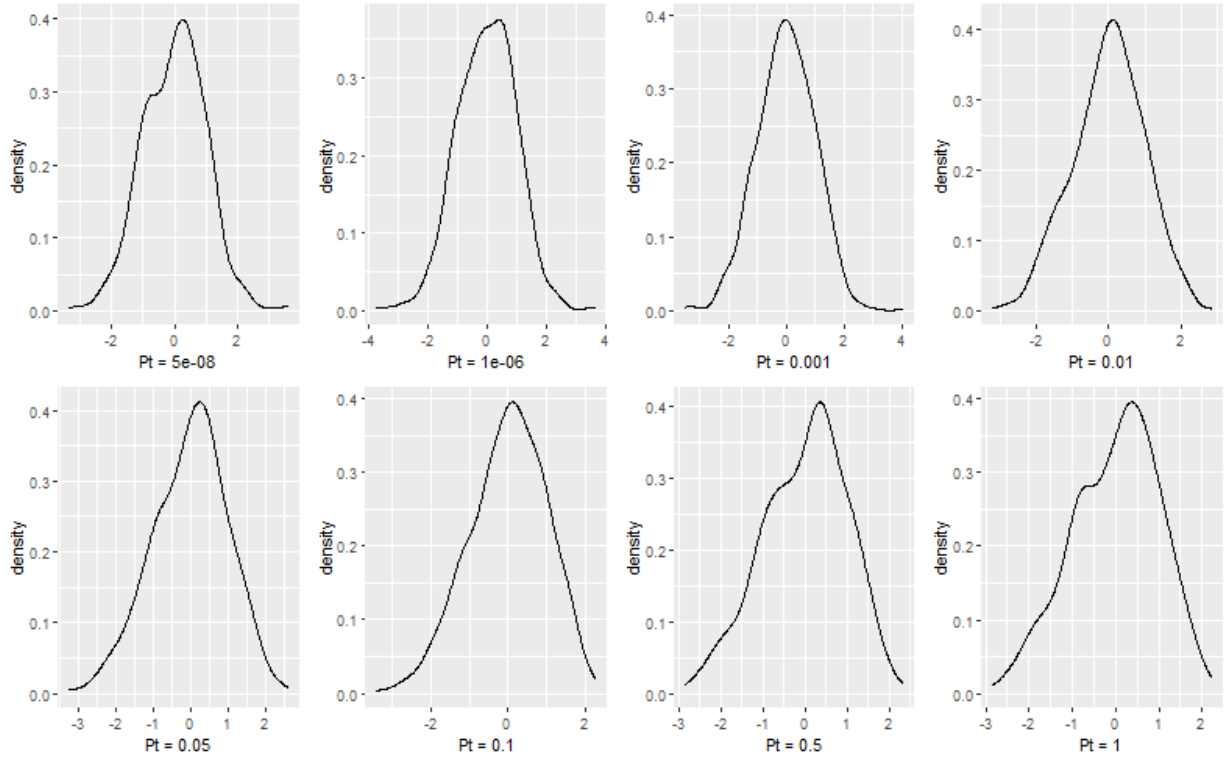
References

Lambert JC, Ibrahim-Verbaas CA, Harold D, et al. Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nature Genetics.* 2013;45(12):1452-1458.

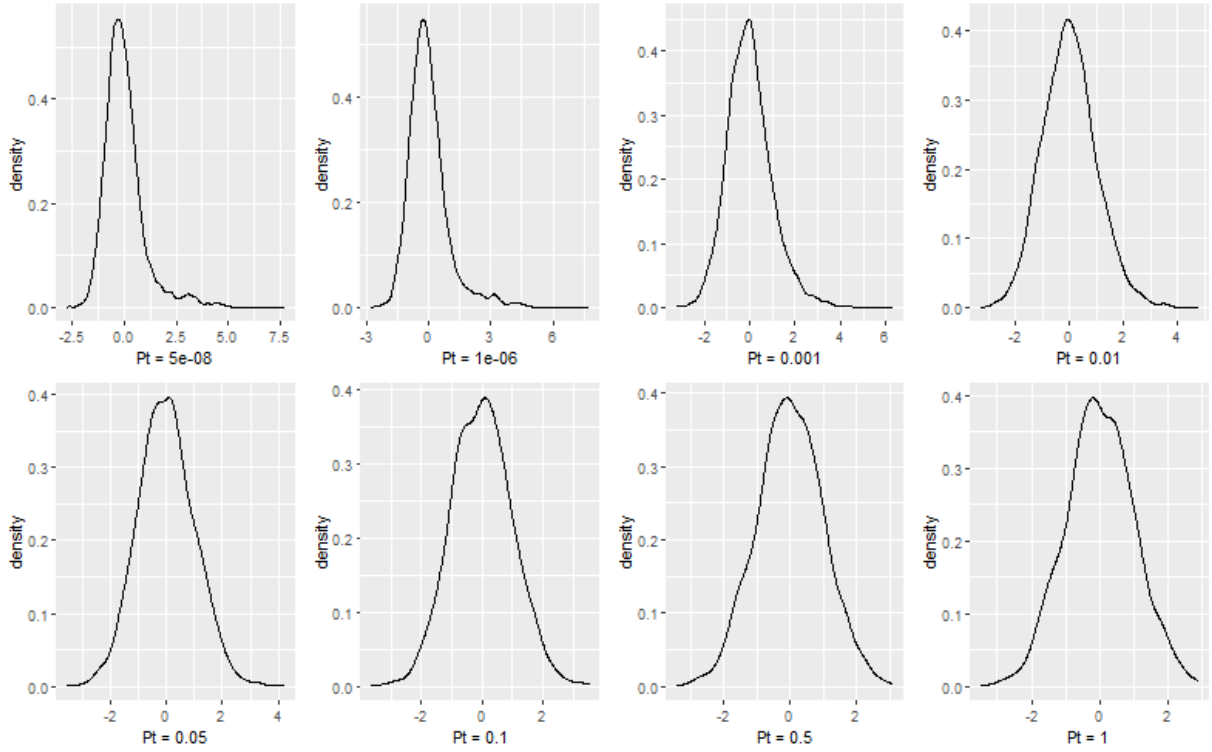
Distribution of polygenic scores by p-value threshold, NHATS European ancestry, n=2827, Alzheimer's disease without the *APOE* gene region (chr19: 45,384,477 to 45,432,606, build 37/hg 19)



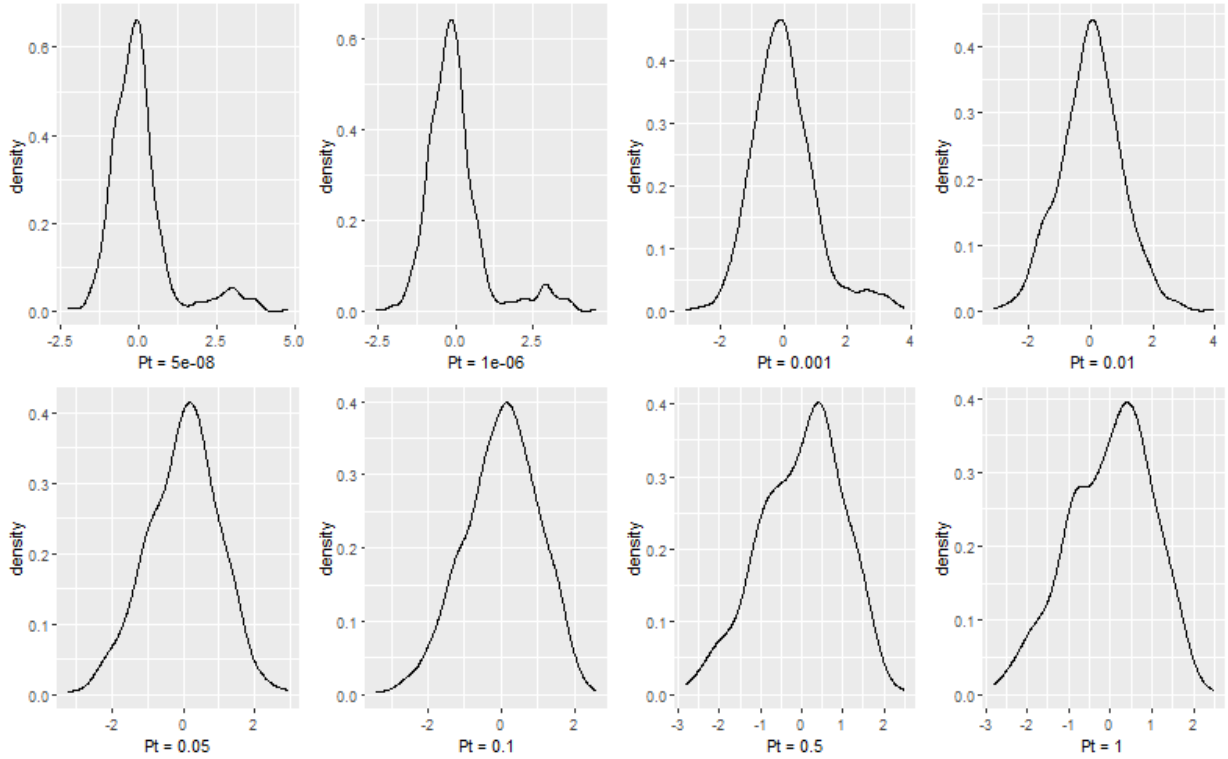
Distribution of polygenic scores by p-value threshold, NHATS African ancestry, n=664, Alzheimer's disease without the *APOE* gene region (chr19: 45,384,477 to 45,432,606, build 37/hg 19)



Distribution of polygenic scores by p-value threshold, NHATS European ancestry, n=2827, Alzheimer's disease with the *APOE* gene region



Distribution of polygenic scores by p-value threshold, NHATS African ancestry, n=664, Alzheimer's disease with the *APOE* gene region



ad, ah. HbA1c – Meta-Analyses of Glucose and Insulin-related traits Consortium 2017

PGS for glycated hemoglobin A1c (HbA1c) were created using results from a 2017 study conducted by the Meta-Analyses of Glucose and Insulin-related traits Consortium (MAGIC). The GWAS meta-analysis files are publicly available on the MAGIC investigators download page (https://www.magicinvestigators.org/downloads/; ftp://ftp.sanger.ac.uk/pub/magic/HbA1c_METAL_AfricanAmerican_updatedSept2018.txt.gz, and ftp://ftp.sanger.ac.uk/pub/magic/HbA1c_METAL_European.txt.gz). Ancestry-specific and transethnic genome-wide meta-analysis summary statistics for association with HbA1c in up to 159,940 individuals from 82 cohorts of European (N=123,665), African (N=7,564), East Asian (N=20,838) and South Asian (N=8,874) ancestry. HbA1c trait values are untransformed and adjusted for age, sex and study-specific covariates. SNP weights were constructed from publicly available data for the European and African-based discovery analysis. The GWAS meta-analysis included 3,009,839 million SNPs (African-based discovery results) and 2,586,698 million SNPs (European-based discovery results) imputed to NCBI Build 37/UCSC hg 19. The GWAS meta-analysis identified 60 common genetic variants associated with HbA1c. Variants were classified as implicated in glycemic, erythrocytic, or unclassified biology. Where possible, studies reported HbA1c as a National Glycohemoglobin Standardization Program (NGSP) percent (**S1 Table**).

The PGSs (for both African and European ancestry participants) contains 222,225 SNPs that overlapped with the NHATS genetic database and the European-based GWAS summary statistics. The PGSs (for both African and European ancestry participants)

contains 219,725 SNPs that overlapped with the NHATS genetic database and the African-based GWAS summary statistics. Higher PGSs correspond to increased HbA1c. The posted PGSs have been standardized within genetic ancestry analysis group to a standard normal curve (mean=0, standard deviation = 1).

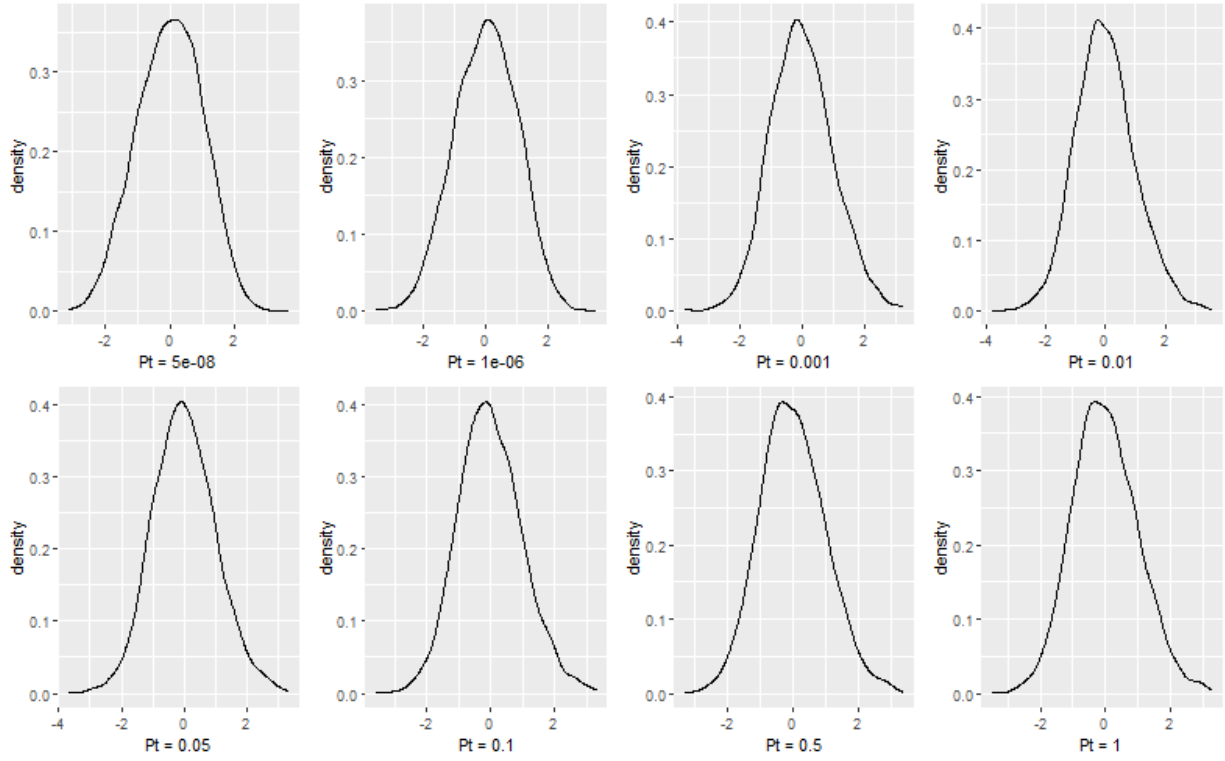
N.B.: **Please note that the MAGIC-HbA1c weights are from ancestry-specific discovery GWAS (see "Notes about the use of PGSs" for more information on the use of PGSs in other ancestry groups).**

N.B.: In the African ancestry-based Genome-wide association study, there were no variants that reached genome-wide suggestive levels (our second threshold cutoff). Therefore, there are no polygenic scores created for the genome-wide significant or genome-wide suggestive p-value threshold cutoffs.

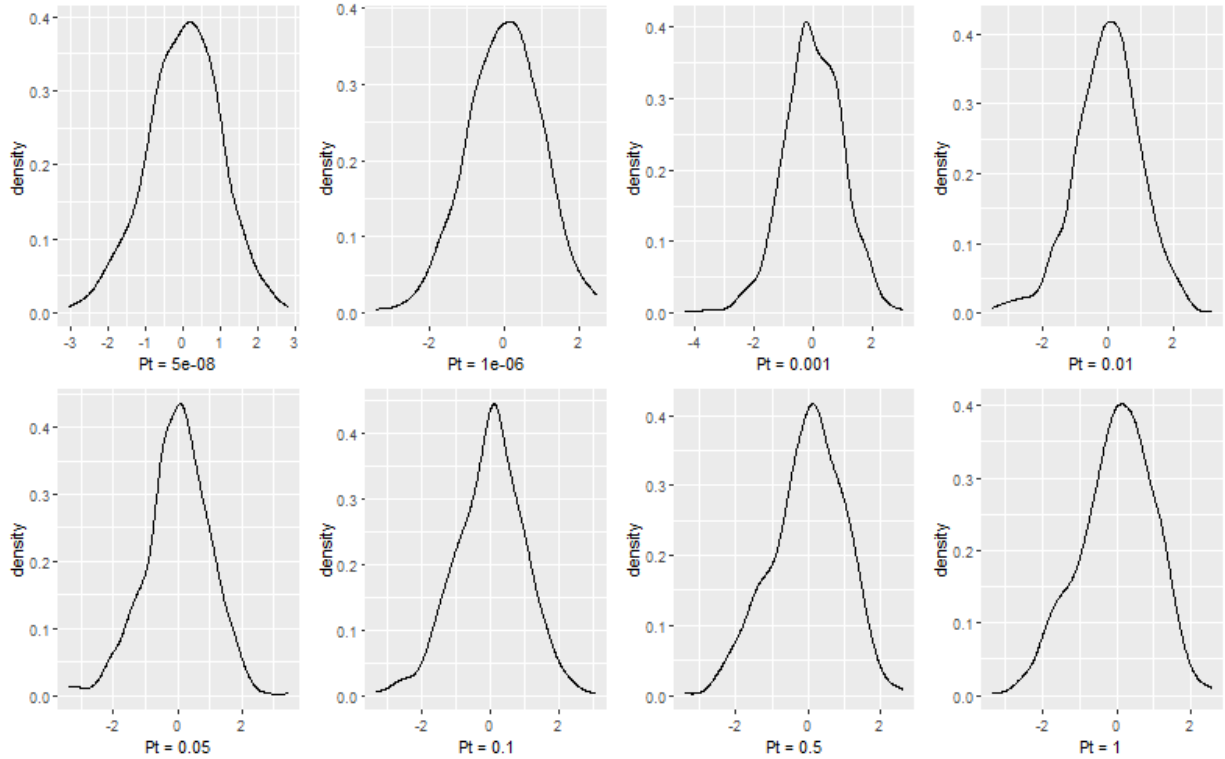
References

Wheeler E, Leong A, Liu CT, et al. Impact of common genetic determinants of Hemoglobin A1c on type 2 diabetes risk and diagnosis in ancestrally diverse populations: A transethnic genome-wide meta-analysis. PLoS Med. 2017;14(9):e1002383. Published 2017 Sep 12. doi:10.1371/journal.pmed.1002383

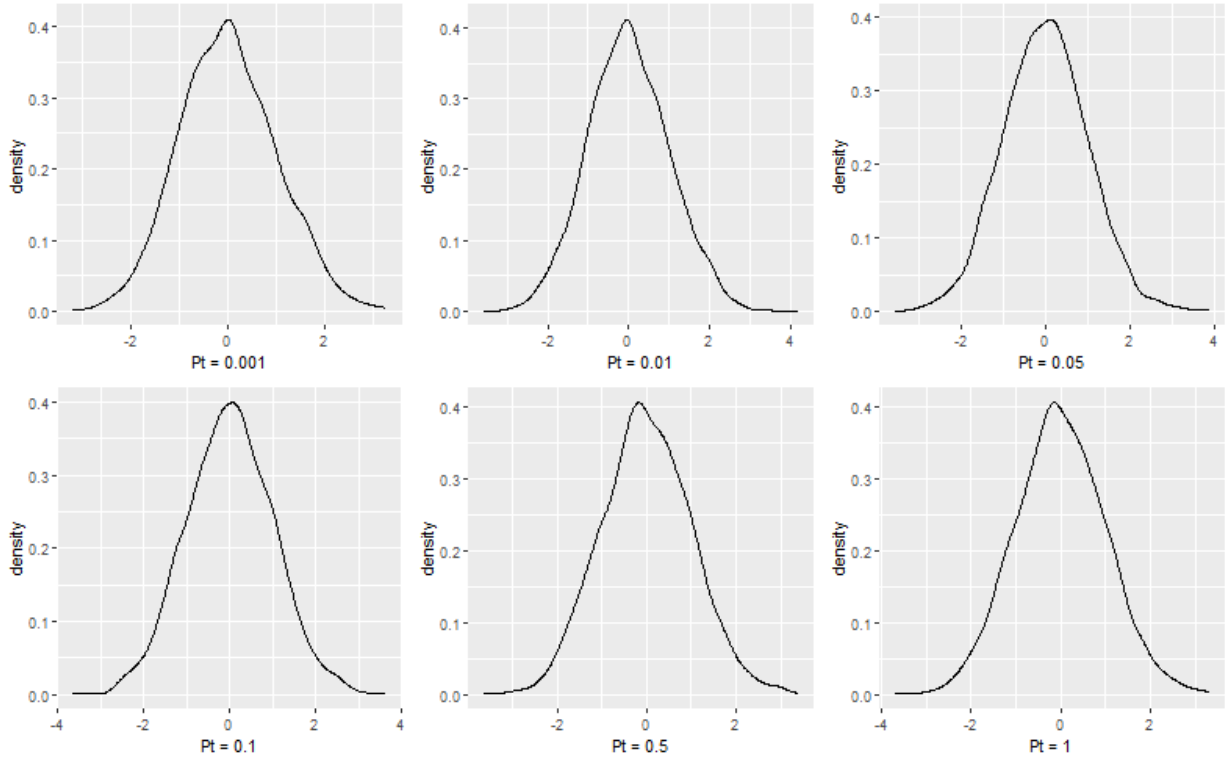
Distribution of polygenic scores by p-value threshold, NHATS European ancestry, n=2827, European ancestry based HbA1c GWAS



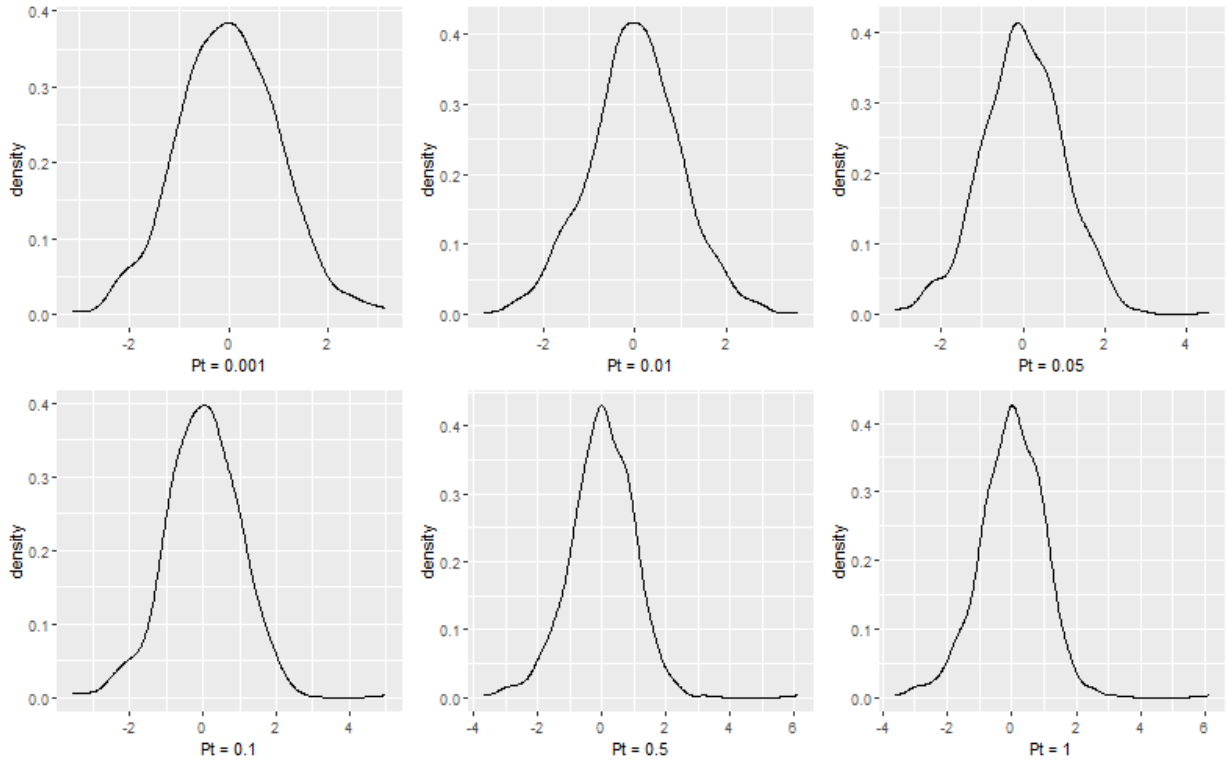
Distribution of polygenic scores by p-value threshold, NHATS African ancestry, n=664, European ancestry based HbA1c GWAS



Distribution of polygenic scores by p-value threshold, NHATS European ancestry, n=2827, African ancestry based HbA1c GWAS



Distribution of polygenic scores by p-value threshold, NHATS African ancestry, n=664
African ancestry based HbA1c GWAS



ae. Depressive Symptoms – Social Science Genetic Association Consortium 2016

The PGSs for depressive symptoms were created using results from a 2016 auxiliary GWAS conducted by the Social Science Genetic Association Consortium (SSGAC) as part of their subjective wellbeing GWAS (see above). The GWAS meta-analysis files are publicly available on the SSGAC website: https://www.thessgac.org/data. The GWAS included 180,866 individuals and meta-analyzed publicly available results from a study performed by the Psychiatric Genomics Consortium (PGC) (Ncases = 9,240, Ncontrols = 9,519) with results from analyses of UK Biobank (UKB) data (N = 105,739), and the Resource for Genetic Epidemiology Research on Aging (GERA) Cohort (Ncases = 7,231, Ncontrols = 49,316). The meta-analysis identified two genome-wide significant SNPs (**Table 1**). A quasi-replication analysis tested whether these SNPs were associated with subjective wellbeing. A replication analysis was also performed using data from 23andMe (N=368,890). In UKB, a continuous phenotype measure was used that combined responses to two questions, which ask about the frequency in the past two weeks with which the respondent experienced feelings of unenthusiasm/disinterest and depression/hopelessness. The PGC and GERA cohorts utilized case-control data on major depressive disorder. In the UKB, analyses controlled for the first 15 PCs, indicator variables for genotyping array, sex, indicator variables for age ranges, and sex-by-age interactions. In GERA, analyses controlled for the first four PCs of the genotypic data, sex, and 14 indicator variables for age

ranges. The PGC included controls for five PCs, sex, age, and cohort fixed effects (for details see Ripke et al., 2013).

The PGSs contain 326,450 SNPs that overlapped between the NHATS genetic database and the GWAS meta-analysis. Higher PGSs correspond to increased depressive symptoms. The posted PGSs have been standardized within genetic ancestry analysis group to a standard normal curve (mean=0, standard deviation = 1).

N.B.: **Please note that the SSGAC results are from a GWAS on individuals of European ancestry (see "Notes about the use of PGSs" for more information on the use of PGSs in other ancestry groups).**

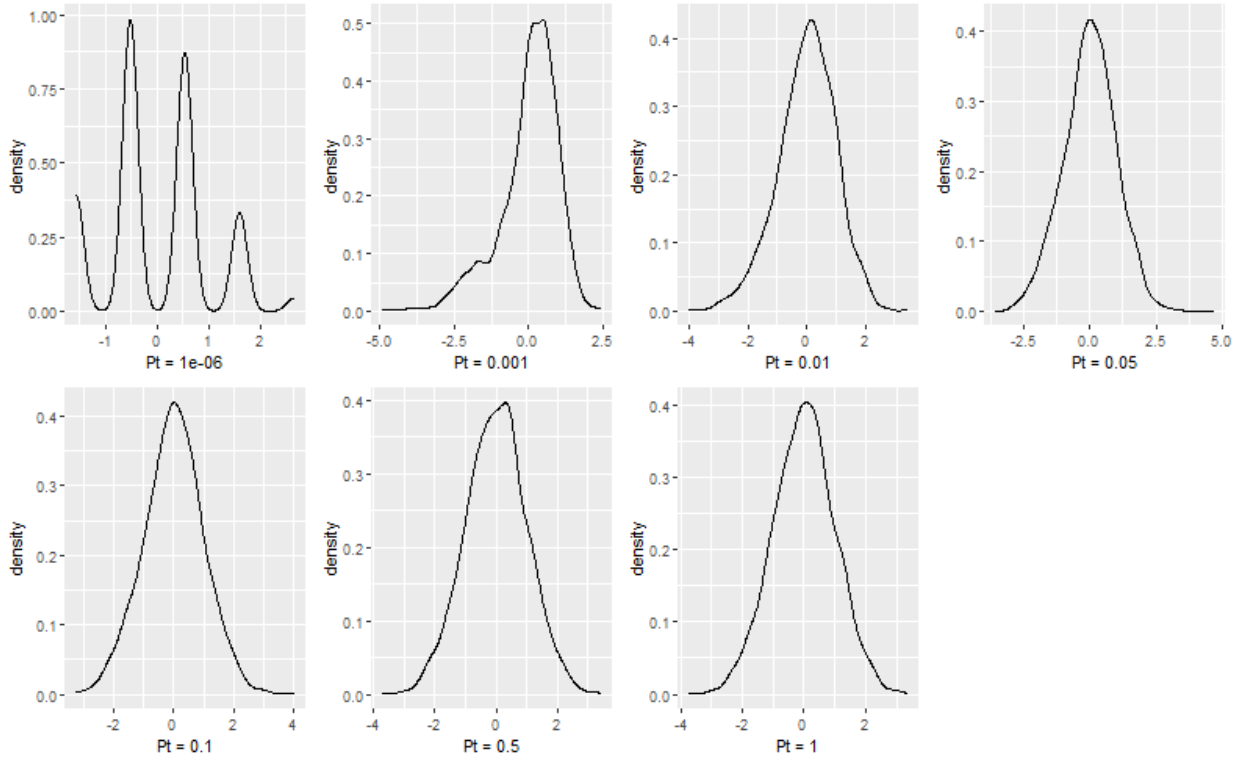
N.B.: No variants that reached genome-wide significance in this GWAS were present in the NHATS data; therefore, we do not include a genome-wide significant PGS.

References

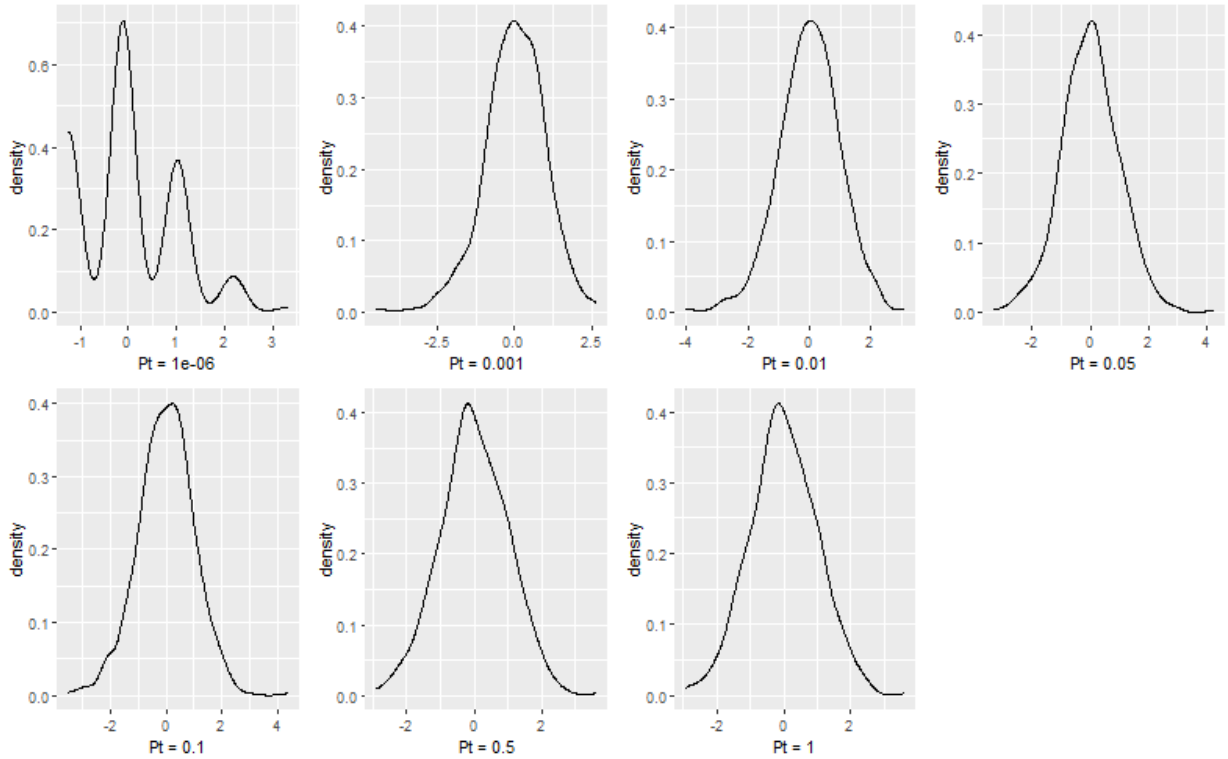
Okbay, A., Baselmans, B. M., De Neve, J. E., Turley, P., Nivard, M. G., Fontana, M. A., ... & Gratten, J. (2016). Genetic variants associated with subjective well-being, depressive symptoms, and neuroticism identified through genome-wide analyses. *Nature Genetics*, *48*(6), 624-633.

Ripke, S., Wray, N. R., Lewis, C. M., Hamilton, S. P., Weissman, M. M., Breen, G., ... & Heath, A.C. (2013). A mega-analysis of genome-wide association studies for major depressive disorder. *Molecular Psychiatry*, *18*(4), 497.

Distribution of polygenic scores by p-value threshold, NHATS European ancestry, n=2827



Distribution of polygenic scores by p-value threshold, NHATS African ancestry, n=664



af. Subjective Wellbeing – Social Science Genetic Association Consortium 2016

The PGSs for subjective wellbeing were created using results from a 2016 GWAS conducted by the Social Science Genetic Association Consortium (SSGAC). The subjective wellbeing analyses included 298,420 European ancestry individuals in the discovery sample. Genome-wide significant SNPs were identified in 3 loci (**Table 1**). A quasi-replication analysis tested whether these three SNPs were associated with depressive symptoms and neuroticism. The phenotype measure was life satisfaction, positive affect, or in some cohorts a measure combining both. Approximately 9.3 million SNPs were included in the analyses, with cohorts utilizing SNPs imputed to the 1000 genomes reference panel (1000G) or the HapMap 2 Panel. Adjustments for age, age^2, sex, and population stratification (first four PCs from the genotypic data) were included in study-specific GWAS association analyses. Cohorts were also asked to include any study-specific covariates such as study site or batch effectsThe SSGAC provided SNP weights with 23andMe results (due to data use agreements) removed.

The PGSs contain 192,787 SNPs that overlapped between the NHATS genetic database and the GWAS meta-analysis. Higher PGSs correspond to increased wellbeing. The posted PGSs have been standardized within genetic ancestry analysis group to a standard normal curve (mean=0, standard deviation = 1).

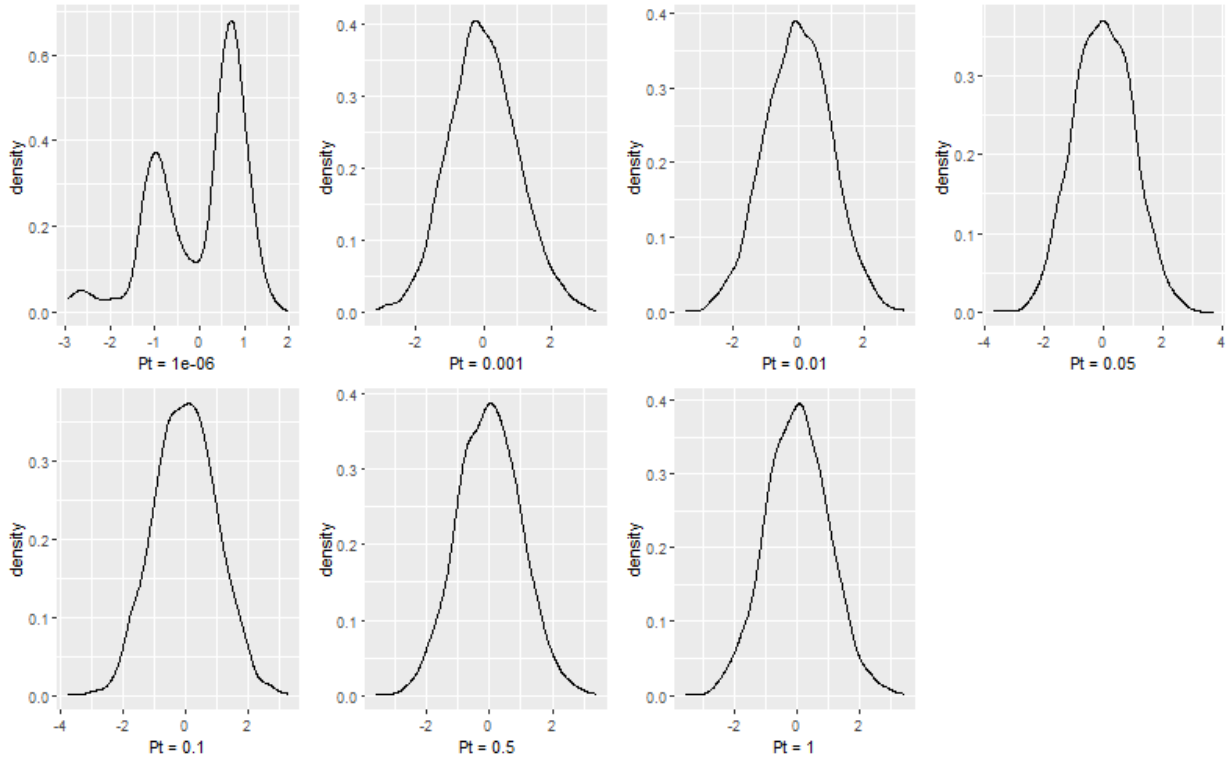
N.B.: **Please note that the SSGAC results are from a GWAS on individuals of European ancestry (see "Notes about the use of PGSs" for more information on the**

use of PGSs in other ancestry groups). No variants that reached genome-wide significance in this GWAS were present in the NHATS data; therefore, we do not include a genome-wide significant PGS.

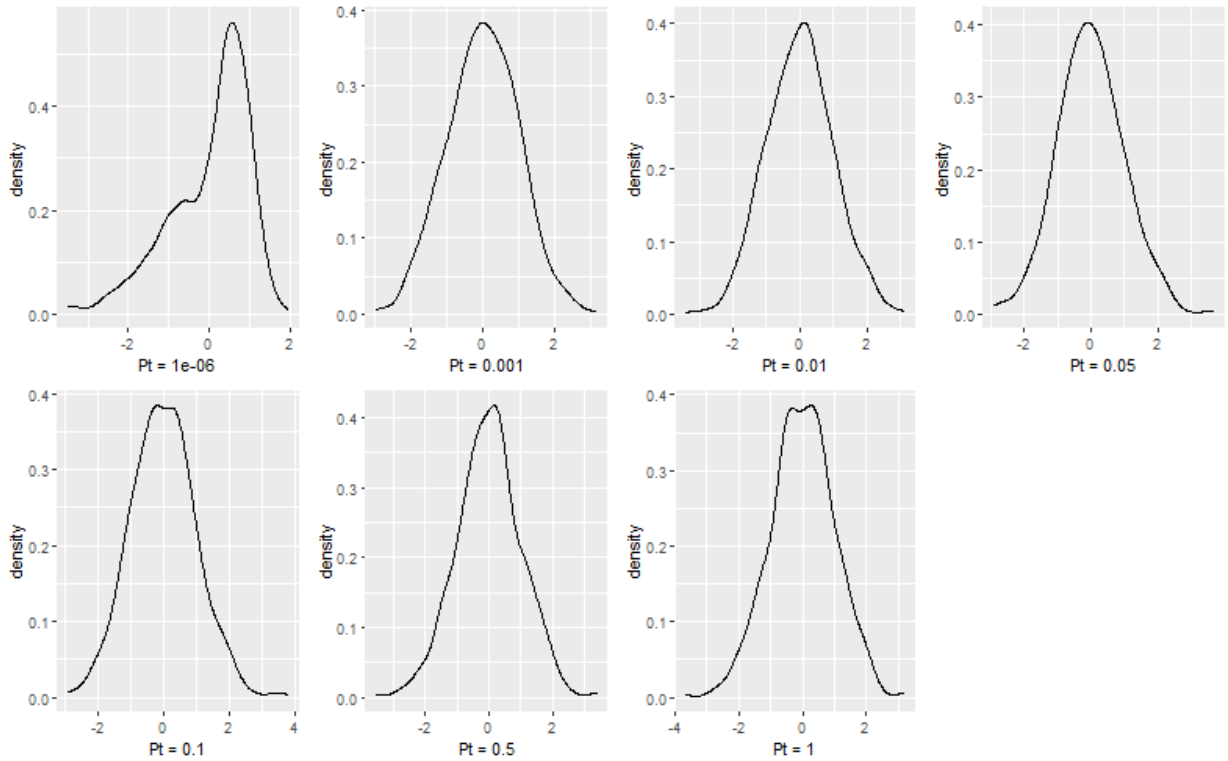
References

Okbay, A., Baselmans, B. M., De Neve, J. E., Turley, P., Nivard, M. G., Fontana, M. A., ... & Gratten, J. (2016). Genetic variants associated with subjective well-being, depressive symptoms, and neuroticism identified through genome-wide analyses. *Nature Genetics*, *48*(6), 624-633.

Distribution of polygenic scores by p-value threshold, NHATS European ancestry, n=2827



Distribution of polygenic scores by p-value threshold, NHATS African ancestry, n=664



ag. Longevity – Cohorts for Heart and Aging Research in Genomic Epidemiology 2014

The longevity PGSs were created using summary statistics from a 2015 GWAS conducted by the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) consortia. The GWAS meta-analysis summary statistics were requested from the CHARGE consortia. The meta-analysis includes 6,036 longevity cases (age ≥90 years) and 3,757 controls that died between ages 55 and 80 and were of European descent. Genetic measures were imputed to ~2.5 million SNPs using the HapMap 22 CEU (Build 36) genotyped samples as a reference. Logistic regression analyses were used to test each SNP for association with longevity using an additive model adjusting for sex and genetic principal components to adjust for population stratification. Meta-analysis raw results with were filtered for HetPval>0.2 and HetDF>5.

The PGSs contain 226,274 SNPs that overlapped between the NHATS genetic database and the GWAS meta-analysis. Higher PGSs correspond to increased longevity. The posted PGSs have been standardized within genetic ancestry analysis group to a standard normal curve (mean=0, standard deviation = 1).

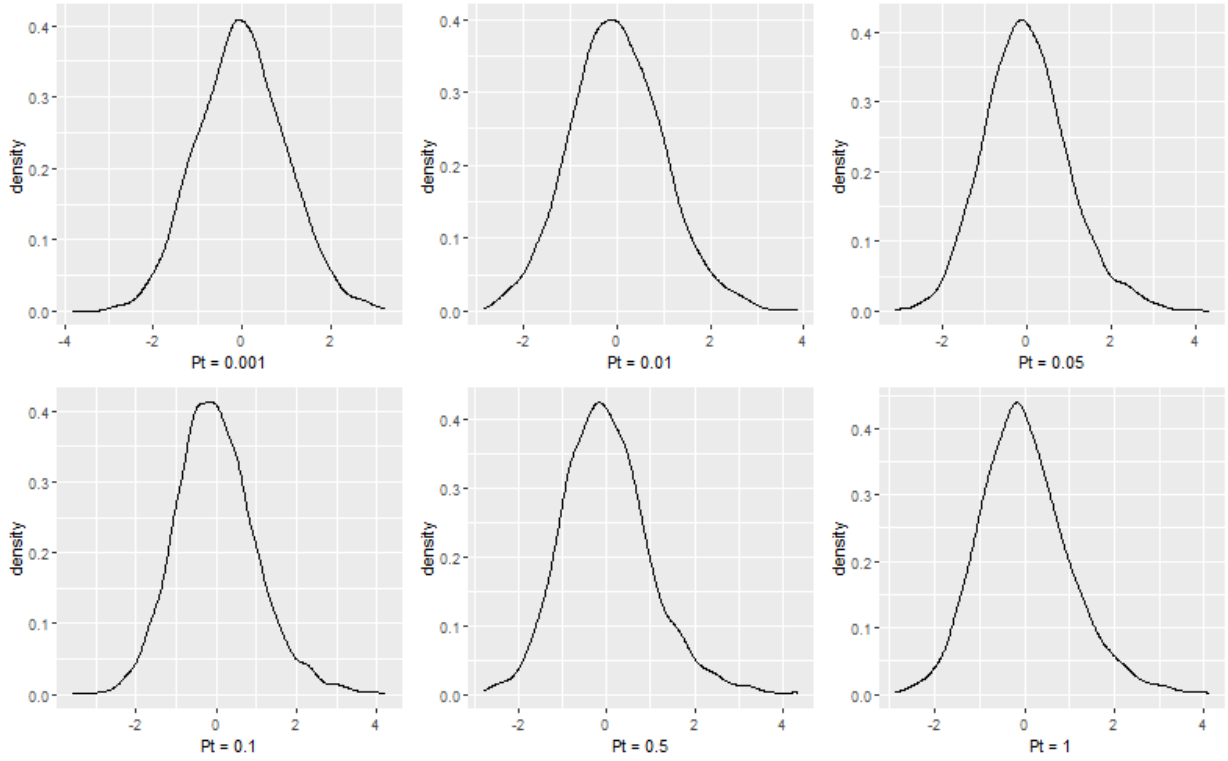
N.B.: **Please note that the CHARGE results are from a GWAS on individuals of European ancestry (see "Notes about the use of PGSs" for more information on the use of PGSs in other ancestry groups).**

N.B.: There were no variants that reached genome-wide suggestive levels (our second threshold cutoff). Therefore, there are no polygenic scores created for the genome-wide significant or genome-wide suggestive p-value threshold cutoffs.

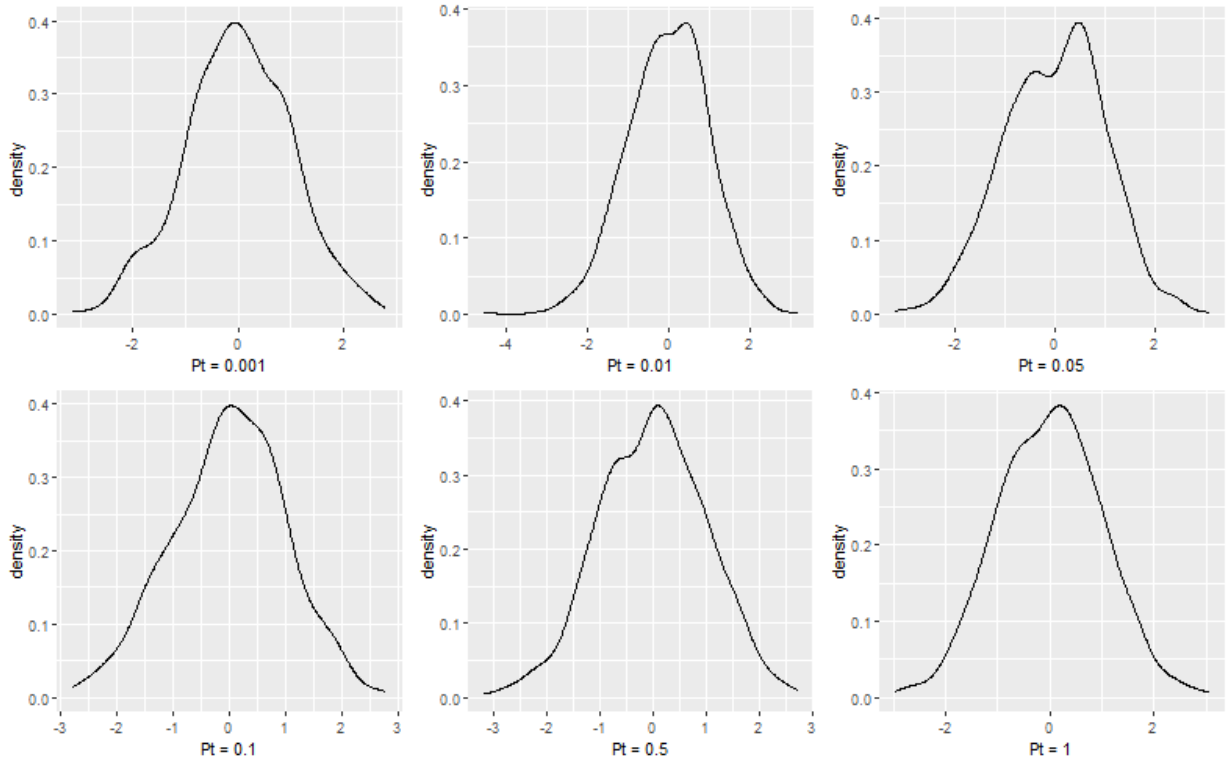
References:

Broer L, Buchman, A. S., Deelen, J., Evans, D. S., Faul, J. D., Lunetta, K. L., ... & Yu, L. (2014). GWAS of longevity in CHARGE consortium confirms APOE and FOXO3 candidacy. *Journals of Gerontology Series A: Biomedical Sciences and Medical Sciences*, *70*(1), 110-118.

Distribution of polygenic scores by p-value threshold, NHATS European ancestry, n=2827



Distribution of polygenic scores by p-value threshold, NHATS African ancestry, n=664



ai,aj. Longevity – Cohorts for Heart and Aging Research in Genomic Epidemiology 2019

The longevity PGSs were created using summary statistics from a 2019 GWAS conducted by the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) consortia. The GWAS meta-analysis summary statistics for the age >90th and >99th survival percentile versus controls was downloaded from the https://www.longevitygenomics.org/downloads website (Results_90th_percentile.txt.gz, Results_99th_percentile.txt.gz). The meta-analysis for age >90 includes 11,262 European ancestry cases, 25,483 European ancestry controls. The meta-analysis for age >99th survival percentile includes 3,484 European ancestry cases, 25,483 European ancestry controls. Logistic regression analyses were used to test each SNP for association with longevity using an additive model adjusting for sex and genetic principal components to adjust for population stratification. SNP effect weights are provided as log-odds betas and oriented to GRCh Build 37 and are truncated to the 10000th place.

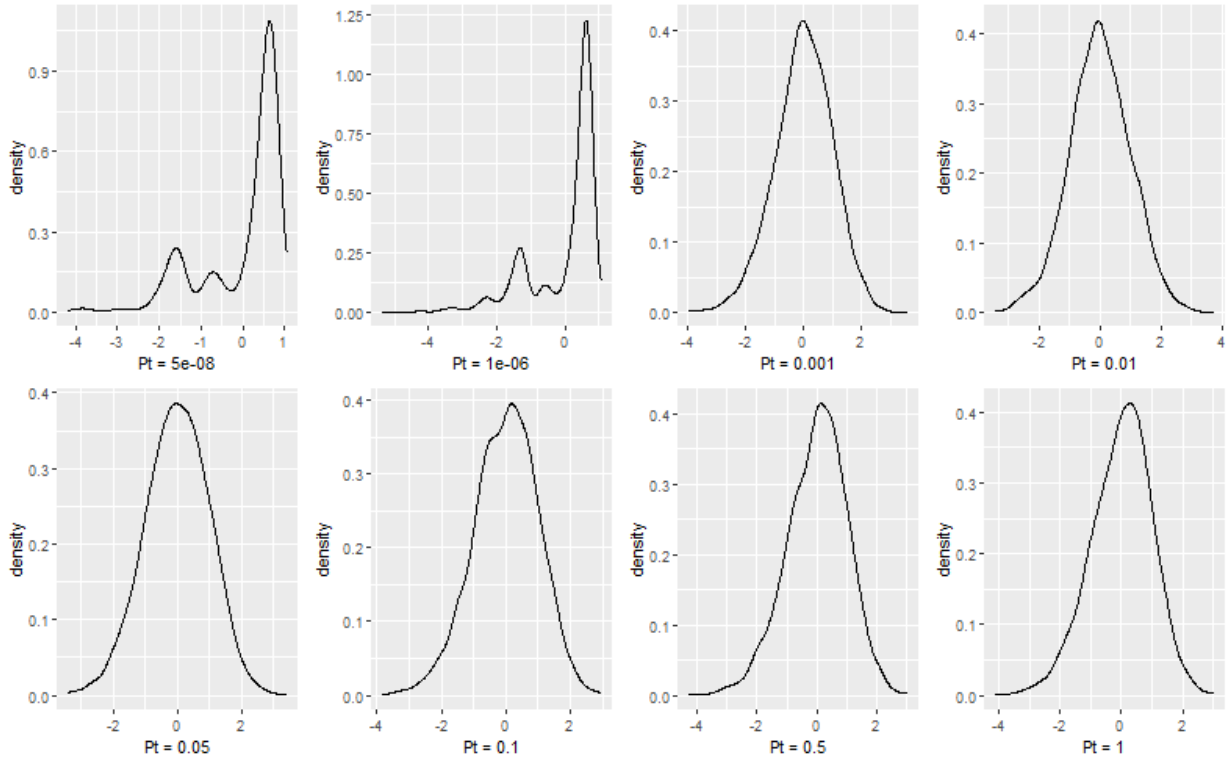
The PGSs for >90th survival percentile versus controls contains 448,076 SNPs that overlapped between the NHATS genetic database and the GWAS meta-analysis. The PGSs for >99th survival percentile versus controls contains 435,802 SNPs that overlapped between the NHATS genetic database and the GWAS meta-analysis. Higher PGSs correspond to increased longevity. The posted PGSs have been standardized within genetic ancestry analysis group to a standard normal curve (mean=0, standard deviation = 1).

N.B.: **Please note that the CHARGE results are from a GWAS on individuals of European ancestry (see "Notes about the use of PGSs" for more information on the use of PGSs in other ancestry groups).**

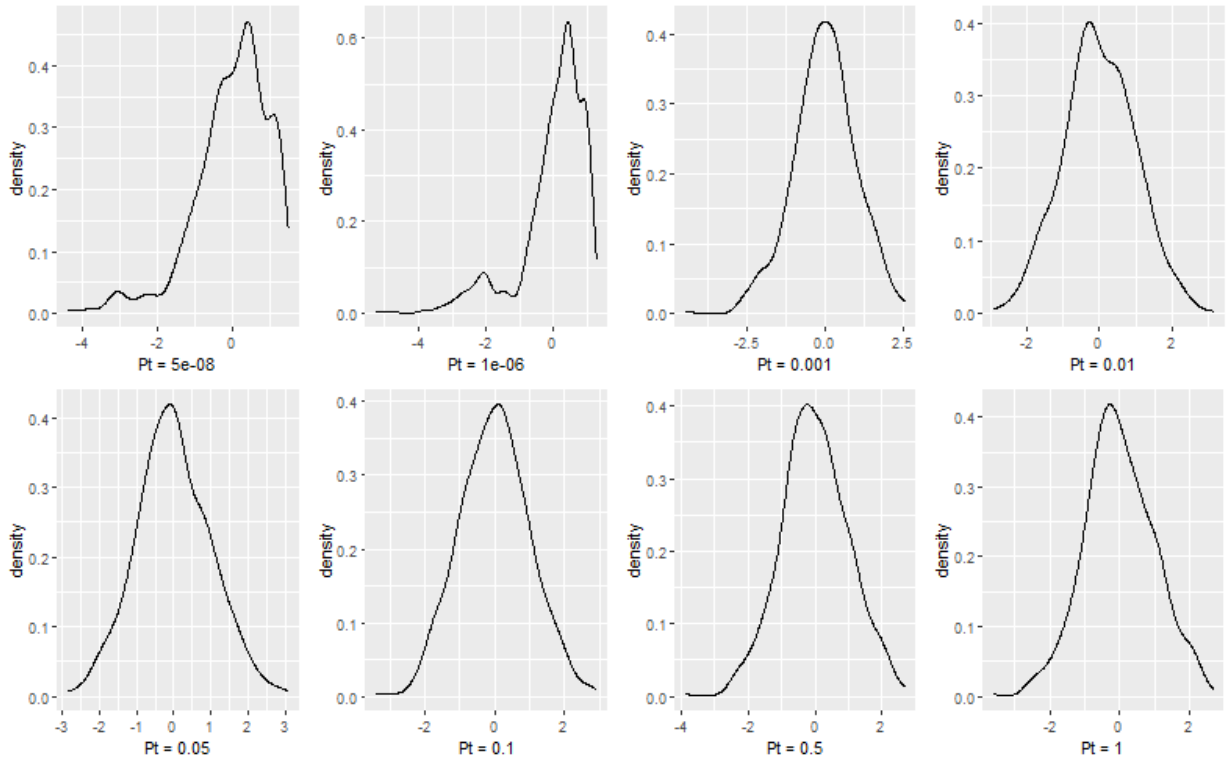
References:

Deelen, J., Evans, D.S., Arking, D.E. et al. A meta-analysis of genome-wide association studies identifies multiple longevity genes. Nat Commun 10, 3669 (2019). https://doi.org/10.1038/s41467-019-11558-2

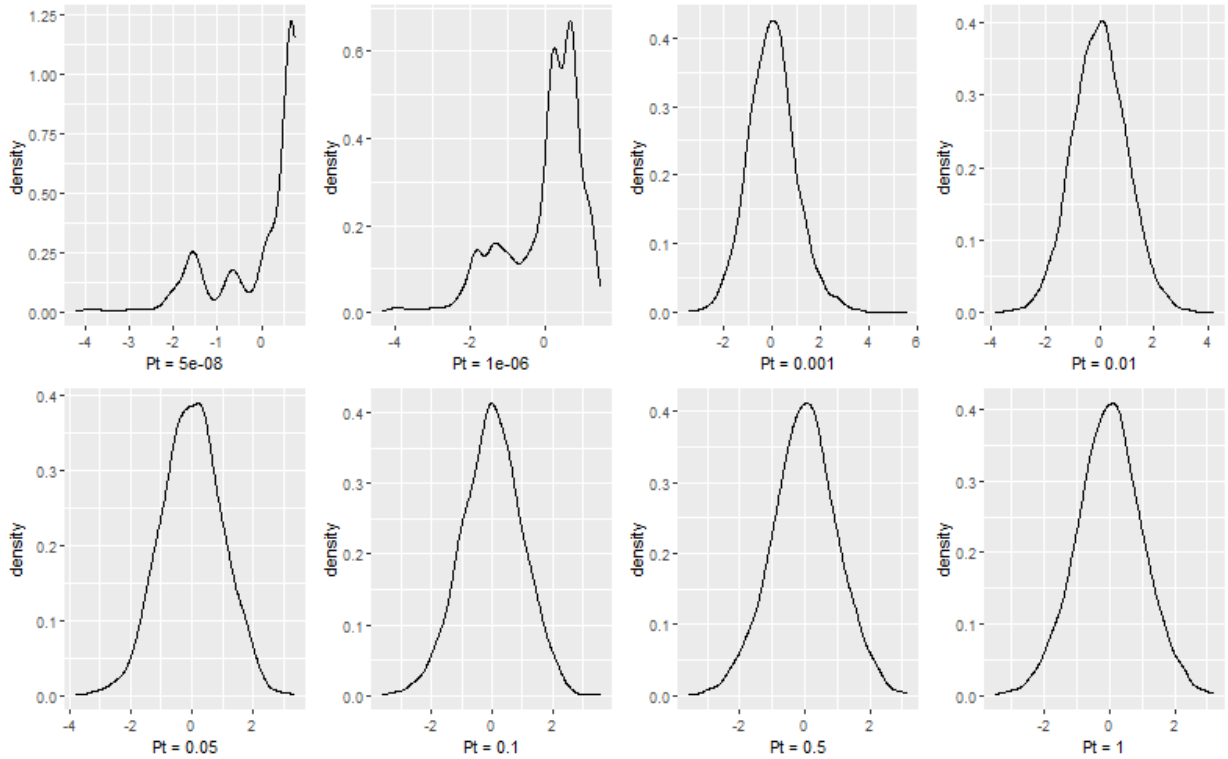
Distribution of polygenic scores by p-value threshold, NHATS European ancestry, n=2827, >90th survival percentile versus controls



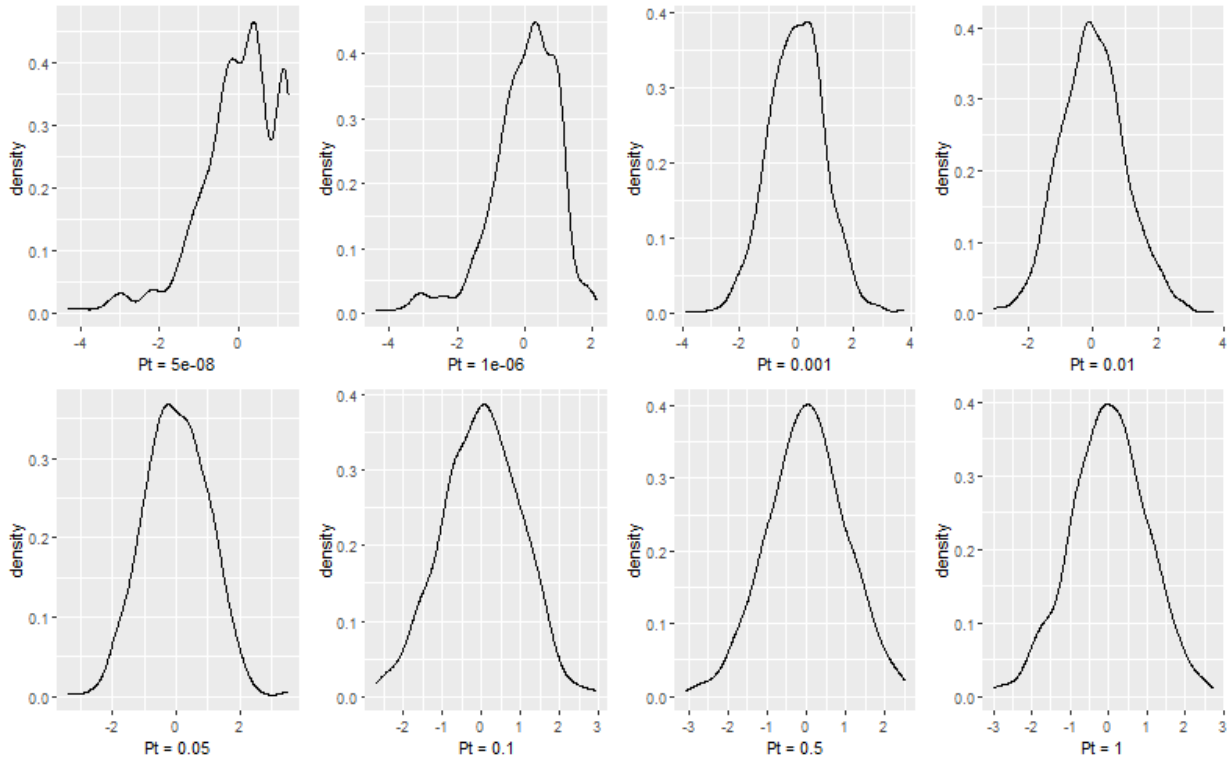
Distribution of polygenic scores by p-value threshold, NHATS African ancestry, n=664, >90th survival percentile versus controls



Distribution of polygenic scores by p-value threshold, NHATS European ancestry, n=2827, >99th survival percentile versus controls



Distribution of polygenic scores by p-value threshold, NHATS African ancestry, n=664, >99th survival percentile versus controls



VII. Data Documentation

Files

The PGS Tracker file contains an indicator of the ancestry analytic group. The PGS AA file contains polygenic scores and local genetic principal components for the African ancestry group. The PGS EA file contains polygenic scores and local genetic principal components for the European ancestry group.

Data Documentation

NHATS documentation is available at www.NHATS.org as sensitive data. The Round 7 Data Collection Instrument, including the Blood Collection section, and PGS crosswalk can be found at https://nhats.org/researcher/nhats.

Variable Names and Labels

Variable names for NHATS polygenic scores follow a standard convention. Polygenic scores variables start with “gn” (for Genetics), followed by the NHATS round number (7), an indicator of the ancestry group (“a” for African ancestry and “e” for European ancestry), a letter corresponding to the phenotypes described in this guide, and a number corresponding to the p-value threshold of the variable.

Local genetic principal components variables start with “gn” (for Genetics), followed by the NHATS round number (7), an indicator of the ancestry group (“a” for African ancestry and “e” for European ancestry), “pc” for principal components, and a number from 1-20.

Variable labels provide more detailed information about the polygenic scores and p-values.

VIII. Obtaining PGS Data

PGS files are designated as Sensitive for purposes of data release. To obtain the data files and codebook, go to https://nhats.org/researcher/data-access/sensitive-data-files?id=nsoc_other_sensitive_files. Download the document titled “Obtaining Sensitive Data from the National Health and Aging Trends Study” and follow the instructions.

References

[1] Freedman VA, Kasper JD. Cohort Profile: The National Health and Aging Trends Study (NHATS). Int J Epidemiol. 2019;48(4):1044-1045g. doi:10.1093/ije/dyz109

[2] Montaquila J, Freedman VA, Edwards B, Kasper JD. *National Health and Aging Trends Study Round 1 Sample Design and Selection. NHATS Technical Paper #1*. Baltimore, MD: Johns Hopkins University School of Public Health, 2012

[3] DeMatteis J, Freedman VA, Kasper JD. National Health and Aging Trends Study Round 5 Sample Design and Selection. NHATS Technical Paper #16. Baltimore, MD: Johns Hopkins University School of Public Health, 2016

[4] Kasper JD, Skehan ME, Seeman T, Freedman VA. Dried Blood Spot (DBS) Based Biomarkers in the National Health and Aging Trends Study User Guide: Final Release. Baltimore, MD: Johns Hopkins University Bloomberg School of Public Health, 2019

[5] Ware EB, Battle SL, Skehan M, Arking DE, Freedman VA, Schrack JA, Kasper JD. National Health and Aging Trends Study Quality Control and Imputation Report for Genotypic Data. Baltimore, MD: Johns Hopkins Bloomberg School of Public Health, 2023