Instability of high polygenic risk classification and mitigation by integrative scoring

Anika Misra, BS[1,2]; Buu Truong, MD[1,2]; Sarah M. Urbut, MD, PhD[1-3]; Yang Sui, PhD[1-3]; Akl C. Fahed, MD, MPH[1-3]; Jordan W. Smoller, MD, ScD[2-4]; Aniruddh P. Patel, MD[1-3]*; Pradeep Natarajan, MD, MMsc[1-3]*

[1]Cardiovascular Disease Initiative, Broad Institute of MIT and Harvard, Cambridge, Massachusetts
[2]Center for Genomic Medicine and Division of Cardiology, Department of Medicine, Massachusetts General Hospital, Boston, Massachusetts
[3]Harvard Medical School, Boston, Massachusetts
[4]Department of Psychiatry, Massachusetts General Hospital, Boston, Massachusetts

**These authors jointly supervised this work*

Please address correspondence to:
Aniruddh P. Patel, MD
185 Cambridge Street, CPZN 3.218
Boston, MA 02114
Phone: 617-643-3451
Fax: 617-726-2203
Email: apatel32@mgh.harvard.edu

Pradeep Natarajan, MD, MSc
185 Cambridge Street, CPZN 3.184
Boston, MA 02114
Phone: 617-643-3451
Fax: 617-726-2203
Email: pnatarajan@mgh.harvard.edu

Word Count: 1985

Polygenic risk scores (PRS) continue to improve with novel methods and expanding genome-wide association studies. Healthcare and third-party laboratories are increasingly deploying PRS reports to patients. Although new PRS show improving strengths of association with traits, it is unknown how the classification of high polygenic risk changes across individual PRS for the same trait. Here, we determined classification of high genetic risk from all cataloged PRS for three complex traits. While each PRS for each trait demonstrated generally consistent population-level strengths of associations, classification of individuals in the top 10% of each PRS distribution varied widely. Using the PRSMix framework, which incorporates information across several PRS to improve prediction, we generated sequential add-one-in (AOI) PRSMix_AOI scores based on order of publication. PRSMix_AOI$_n$ led to improved PRS performance and more consistent high-risk classification compared with the PRS$_n$. The PRSMix_AOI approach provides more stable and reliable classification of high-risk as new PRS continue to be generated toward PRS standardization.

Polygenic risk scores (PRS), which quantify the genetic risk for traits from common variants, have improved in their predictive performances over the past decade.[1] Building on classical approaches of pruning and thresholding, methods incorporating Bayesian approaches using prior knowledge about genetic architecture, relatedness of individuals, linkage disequilibrium patterns, and genetic effects across populations have improved assignment of variant weights and PRS performance.[2–5] In parallel, genome-wide association studies (GWAS) have continued to expand in size, with the most recent iteration of the GWASs for human height and coronary artery disease reaching 5.4 million and 1.3 million participants, respectively.[6,7] Incorporation of multi-ancestry and multi-trait GWAS data has further improved PRS prediction in diverse ancestral groups.[8] In attempting to quantify total inherited risk, each of these PRS iterations for a given trait captures unique information contingent on source GWAS, method, and training dataset.

PRS are increasingly being delivered to patients. Third-party genetic testing companies and healthcare system laboratories are already delivering polygenic scores for coronary artery disease, diabetes, cancers, and other diseases.[9,10] The eMERGE consortium has developed PRS reports for 10 diseases to return to participants within healthcare system as part of a larger effort to study genomic risk assessment and management.[11–13] Furthermore, researchers recently developed clinically valid assays, clinical workflows, and patient- and physician-oriented information materials to accompany PRS reports delivered within the Mass General Brigham Biobank and the Veterans Affairs Genomic Medicine at Veterans Affairs (GenoVA) Study.[14,15] New clinical trials are incorporating PRS into medical decision-making (NCT05819814, NCT05850091), and medical societies have begun to release initial statements on their utility.[16,17]

Despite ongoing progress toward clinical implementation, the variability of individual-level classification of 'high genetic risk' using different PRS for a given trait remains largely untested, and consensus PRS for any trait currently do not exist. Previous efforts have focused on population-level prediction metrics, rather than consistency of high-risk classification presented in individual clinical reports now as a part of clinical implementation workflows.[8] Prior limited availability of large holdout diverse datasets has limited individual-level benchmarking to assess agreement in classification between PRS. Furthermore, there is a need to aggregate and incorporate orthogonal data from available PRS while overcoming correlation between scores and maximize predictive performance.

Using the large and ancestrally-diverse *All of Us* (AOU) cohort[18], we set out to compare the classification of individuals with high genetic risk based on published polygenic scores for three common, complex diseases: coronary artery disease (CAD), type 2 diabetes (T2DM), and major depressive disorder (MDD). We also test the effect of using PRSMix[19] — a tool that agnostically incorporates information across several PRS for a given trait to improve prediction accuracy for a target population — in influencing high genetic risk classification over iterations of polygenic scores.

We determined the associations of published PRS for complex traits from the Polygenic Score Catalog[20] in AOU, which has aggregated genotype data and extensive phenotypic information on 236,393 participants (average enrollment age: 51.8 years; 60.6% female; genetically inferred ancestry of 54.7% EUR, 22.9% AFR, 18.9% AMR, 2.3% EAS and 1.1% SAS).[21] Specifically, we calculated 57 scores for CAD, 129 scores for T2DM, and 18 scores for MDD. We tested the associations of these scores with corresponding outcomes. In each group of trait-specific scores, we chose the strongest target disease-associated score deposited by each unique publication in the group to form the subset of scores called PRS_n, with n referring to the chronological order of the PRS_n publication among the group of all trait-specific PRS_n chosen – this resulted in 40 CAD PRS_n, 39 T2DM PRS_n, and 7 MDD PRS_n. Then, for each of the three trait-specific groups of chronologically ordered PRS_n, we generated a parallel group of sequential add-one-in (AOI) $PRSMix_AOI_n$ scores wherein, using PRSMix, each PRS_n was combined with all other PRS_n in its group published before it to produce its corresponding $PRSMix_AOI_n$ score. Finally, we generated three trait-specific $PRSMix_{all}$ scores, each describing the PRSMix of all published PRS for the given trait. $PRSMix_{all}$ was used to establish a benchmark definition and classification of high genetic risk, allowing consistent comparison between individual PRS_n and their paired $PRSMix_AOI_n$. (Figure 1)

First, we calculated the associations and classification stability of high polygenic risk, defined as the top 10% of the score distribution, with the corresponding trait. Using contemporary PRS for common diseases, defining a threshold of top 10% as high genetic risk is associated with approximately two-fold greater risk of disease in prior studies.[14] The ORs associated with top 10% risk classification ranged from 1.25-2.50 for all CAD PRS_n with CAD, 1.30-2.58 for all T2DM PRS_n with T2DM, and 0.99-1.47 for all MDD PRS_n with MDD.(Extended Figure 1) While each PRS_n demonstrated consistent strengths of association, classification of individuals in the top 10% of each PRS distribution varied widely. The Jaccard index, a metric of similarity, is the proportion of observations in agreement between two sets of data relative to the total number of observations. The median [interquartile range] Jaccard index for top 10% classification by CAD PRS_n was 0.17 [0.13-0.22], indicating poor agreement across scores. Similarly, the median Jaccard index for the top 10% classification was 0.18 [0.15-0.22] for T2DM PRS_n and 0.11 [0.08-0.17] for MDD PRS_n (Figure 2). This trend was invariant of threshold choice. (Supplementary Table 1) When restricting to polygenic scores from the five most recent publications, we continued to observe poor agreement of high-risk classifications, with median [IQR] Jaccard index estimates of 0.15 [0.13-0.22] for CAD PRS_n, 0.26 [0.18-0.35] for T2DM PRS_n, and 0.18 [0.07-0.23] for MDD PRS_n.

Chronologically adding score PRS_n in order of publication into $PRSMix_AOI_n$ led to progressively stronger associations with the outcome. This is most evident for PRS_n from the past five publications, where the C-statistics associated with the top 10% risk classification using $PRSMix_AOI_n$ vs. PRS_n ranged from 0.545-0.546 vs. 0.515-0.536 for CAD, 0.543-0.547

vs. 0.520-0.546 for T2DM, and 0.521-0.522 vs. 0.500-0.521 for MDD, respectively.(Figure 3) The ORs were significantly stronger for the top 10% of individuals using PRSMix_AOI$_n$ vs. PRS$_n$ respectively, with median [IQR] ORs for CAD (1.99 [1.77-2.27] vs. 1.68 [1.48-1.87], $P_{heterogeneity}$<0.001), T2DM (2.46 [2.45-2.50] vs. 1.86 [1.60-1.96], $P_{heterogeneity}$<0.001), and MDD (1.47 [1.37-1.50] vs. 1.30 [1.17-1.40], $P_{heterogeneity}$<0.001). (Figure 4A) The Nagelkerke's pseudo-R^2 (median [IQR]) associated with top 10% risk classification for PRSMix_AOI$_n$ vs. PRS$_n$ was higher for CAD (0.008 [0.005-0.012] vs. 0.004 [0.002-0.006], $P_{heterogeneity}$<0.001), T2DM (0.017 [0.017-0.018] vs. 0.008 [0.004-0.009], $P_{heterogeneity}$<0.001) and MDD (0.0033 [0.0022-0.0038] vs. 0.0014 [0.0005-0.0023], $P_{heterogeneity}$<0.001). (Figure 4B)

More importantly, using PRSMix_AOI$_n$ resulted in more congruent high-risk classification compared to the respective PRS$_n$ itself over time. Using PRSMix_AOI$_n$ the Jaccard index improved to median [IQR] of 0.39 [0.24-0.53] for CAD PRSMix_AOI$_n$, 0.64 [0.45-0.85] for T2DM PRSMix_AOI$_n$, and 0.44 [0.15-0.80] for MDD PRSMix_AOI$_n$. (Figure 5) When restricting to PRS$_n$ from the five most recent publications, we observed significantly higher congruence in high-risk classifications between PRSMix_AOI$_n$ scores, with median Jaccard index estimates of 0.92 [0.91-0.93] for CAD PRSMix_AOI$_n$, 0.78 [0.78-0.78] for T2DM PRSMix_AOI$_n$, and 0.80 [0.69-0.95] for MDD PRSMix_AOI$_n$. (Extended Figure 3) These findings of improved congruence for high genetic risk of CAD, T2DM, and MDD generalize across ancestral subgroups.(Extended Figure 4)

As a result of the greater congruence in high-risk classification, using PRSMix_AOI$_n$ led to more consistent estimation of genetic risk percentile for individual participants. Individuals classified in top 10% risk by the PRSMix$_{all}$ for CAD have median percentiles across PRS$_n$ of 86 [79-91.5] vs. 93 [87-97] for PRSMix_AOI$_n$.(Figure 6A) Similarly, individuals classified in the top 10% risk by the PRSMix$_{all}$ for T2DM have median percentiles across PRS$_n$ of 87 [81-93] vs. 95 [91-98] for PRSMix_AOI$_n$.(Figure 6B) Additionally, individuals classified in the top 10% risk by the PRSMix$_{all}$ for MDD have median [IQR] percentiles across PRS$_n$ of 86 [79-92] vs. 95 [92-98] for PRSMix_AOI$_n$. (Figure 6C) These findings of improved individual-level classification for high genetic risk of CAD, T2DM, and MDD generalize across ancestral subgroups.(Extended Figure 5). For any given PRS$_n$, high-risk individuals fall to a lower, wider range of percentiles across all PRS$_n$ when compared with their paired PRSMix_AOI$_n$ scores. For the PRS$_n$ among the last five published PRS$_n$ for each trait that had the most inconsistent high-risk classification relative to PRSMix$_{all}$, the median [IQR] percentiles of individuals classified in top 10% risk by PRS$_n$ vs. PRSMix_AOI$_n$ were 78 [55-92] vs. 95 [92-98] for CAD, 87 [71-95] vs. 95 [93-98] for T2DM and 59 [33-81] vs. 95 [93-98] for MDD.(Extended Figure 6).

In a comprehensive assessment of published PRS for three complex diseases, high PRS classification of individual participants was highly inconsistent across individual PRS. Because the strengths of association of individual PRS were significant at the population level, each score captures complementary predictive information as a by-product of its training process. This can be a function of the size of its GWAS, the composition of the training dataset, the exact phenotype definition, and assumptions of the polygenic score method, among other factors. We observed instability in risk estimates even between scores generated from the same GWAS or trained in the same cohort study. This instability builds on the inherent large variances in individual PRS estimates due to propagation of error as a function of GWAS sample size, number of causal SNPs and SNP-heritability – collectively resulting in instability of prediction across scores.[22]

Discrepancy in classification of high genetic risk is problematic in current clinical implementation efforts based on single PRS. As GWAS continue to grow and PRS methods continue to improve, newer scores will be published. While we observe a high degree of correlation of scores across population-level metrics, we show that each subsequent score has poor agreement of high-risk (as defined as being in the top 10%) status. As single PRS continue to demonstrate iterative improvement, adopting new single PRS empirically will lead to continual reassignment of high-risk, which will yield confusion and lack of confidence. In addition to reporting population-based association, prediction, and calibration metrics, benchmarking individual scores in hold-out datasets to understand these implications is critical prior to implementation.

Use of PRSMix_AOI to sequentially incorporate new PRS is a framework that can be deployed by biobanks to provide a more robust and stable classification of high-risk as well as improved overall PRS population-based metrics. This method factors out highly correlated scores and utilizes complementary information of available PRS to help predict risk more accurately in the target population. When a new score is published, it can be incorporated into the PRSMix_AOI model yielding an improved PRS with less marked variability in updating the classification of high-risk for individuals. Importantly, stability is accompanied by more accurate high-risk classification. Very significant gains in predictive performance for an additional PRS would cause more substantial and likely appropriate classification changes with this approach. There is nominal additional computational effort needed in calculating the input scores needed to calculate PRSMix_AOI. Moreover, this method's stable and improved prediction generalizes across ancestral groups, helping mitigate previously characterized disparities in PRS performance.[23] The better accuracy and inter-PRSMix_AOI$_n$ consistency make it an ideal framework for incorporation into clinical workflows for PRS reporting and updating.

This study has several limitations. To mirror the development of the majority of published PRS, and given the recency of and middle-aged baseline of AOU, we performed association analyses with prevalent disease and still found significant instability in prediction. Future efforts on instability of incident disease prediction with follow-up beginning early in life will be additionally informative. We also used hospitalization and procedural codes to classify disease instead of clinically adjudicated outcomes. In this manuscript, we focus on common complex diseases for well-powered analyses. While some reports use OR for high risk classification, we show wide variability in population-level metrics and percentile classification leading to wide variability in individual-level OR estimation by PRS$_n$. Thus the proposed PRSMix_AOI approach provides stability for both percentile and OR-based reporting.

In conclusion, in a comprehensive assessment of published PRS for three complex diseases, high PRS classification was highly inconsistent. Using PRSMix_AOI to sequentially incorporate new PRS enables standardization toward more stable and reliable classification of high risk.

Online Methods:

Study population

The *All of Us* (AOU) Research Program is a cohort study focused on recruiting individuals traditionally underrepresented in biomedical research. Since 2018, AOU has enrolled people aged 18 and older from over 730 sites across the United States. The program has consented more than 800,000 participants, of whom over 560,000 have completed the basics of enrollment including collection of health questionnaires and biospecimens. For these participants, there is ongoing linkage to electronic health record (EHR) data, including ICD-9/ICD-10, SNOMED, and CPT codes. Genetic data comprises array samples from 315,000 participants and whole-genome sequencing (WGS) from 245,394 participants. This study used WGS data from the Controlled Tier Dataset version 7 release.

Outcome ascertainment

CAD was defined based on self-report, occurrences of at least 2 diagnosis codes for myocardial infarction or a single procedure code for coronary revascularization. T2DM was defined based on diagnosis codes, laboratory results and medication prescriptions, as previously described by the eMERGE consortium.[24] MDD was defined based on the presence of at least two diagnosis codes for major depressive disorder. All phenotype definitions are detailed in supplementary tables 2-4.

Genotyping and quality control

Participants in AOU were genotyped using the Illumina Global Diversity Array at AOU genome centers. Central quality control measures included filtering for sex concordance, a cross-individual contamination rate below 3%, and a call rate above 98%. Further quality control performed by AOU included filtering for variants with population-specific allele frequency greater than 1% or a population-specific allele count greater than 100 in any AOU-computed ancestry subpopulations. Ancestry was inferred based on genetic similarity with projections of 20 principal components of genetic ancestry using 1000 Genomes as a reference panel. Inferred genetic ancestry in AOU was estimated in high-quality WGS samples that were restricted to bi-allelic sites, a minor allele frequency above 0.1%, a call rate above 99%, and a linkage disequilibrium-pruned threshold $r^2 = 0.1$.

Polygenic score calculation

The variant effect sizes for 61 scores for CAD, 133 scores for T2DM and 18 scores for MDD were downloaded from Polygenic Score Catalog on May 28, 2024. Of these, we excluded 4 CAD scores and 4 T2DM scores that were already developed using PRSMix to prevent potential overfitting. This left 57 CAD scores, 129 T2DM scores, and 18 MDD scores for use in further analysis, none of which were previously trained in AOU or developed using AOU GWAS data. For score accession numbers see supplementary tables 5-7. All scores were harmonized to the AOU reference dataset using PRSMix. Scores with an OR/SD < 1 in the entire AOU cohort were considered anti-correlated and were thus inverted in all further analyses. All scoring in AOU was done using PLINK2 software. All polygenic scores were adjusted for enrollment age, sex, and the first ten principal components of genetic ancestry and then standardized to have a mean of 0 and a standard deviation of 1. Individuals were next binned into 100 groupings according to percentile of score, and individuals in the top 10% of the score distribution were deemed to have high genetic risk.

PRSMix and PRSMix_AOI

PRSMix is a framework that evaluates and leverages the data from a group of PRS for a target trait to generate a new score with improved prediction accuracy. PRSMix uses an elastic net model to produce a weighted linear combination of all the input PRS. The PRSMix framework was used first to harmonize all scores from the PGS Catalog to the AOU reference dataset via the *harmonize_snpeffect_toALT* function. In addition, PRSMix was used to combine PRS via the *combine_PRS* function. PRSMix was run using all default parameters. This includes the use of age, sex, and the first 10 principal components as covariates, as well as an 80% vs. 20% split of the cohort into the training and testing cohorts, respectively. For each trait, we generated a $PRSMix_{all}$ score defined as the PRSMix of all published PRS for a given trait.

Within each trait of interest, only a subset of all the trait-specific scores were selected to be used as a PRS_n for subsequent paired $PRSMix_AOI_n$. For each unique PGS Catalog Publication (PGP) ID found within the group of all trait-specific scores, the score with the highest OR/SD was selected as the representative PRS_n for that publication. This selection mechanism resulted in 40 PRS_n out of 57 total CAD scores, 39 PRS_n out of 129 total T2DM scores, and 7 PRS_n out of 18 total MDD scores. Each group of trait-specific PRS_n were then ordered by publication date, using the number of variants included in the score as the tiebreaker with smaller variant sizes coming first. For each of the three trait-specific groups of ordered PRS_n, we generated a parallel group of sequential add-one-in (AOI) $PRSMix_AOI_n$ scores. For each PRS_n, we used PRSMix to combine the PRS_n with all other PRS_n published prior in its group to produce its corresponding $PRSMix_AOI_n$ score.

Statistical analysis

The association of polygenic scores with outcome of interest were assessed using logistic regression with covariates of enrollment age, sex, and the first 10 principal components of genetic ancestry. The discrimination of each of these polygenic scores was assessed using Harrell's C-statistic. The proportion of phenotypic variance explained by the polygenic score on the observed scale was determined using Nagelkerke's pseudo-R^2 metric via the rcompanion R package. This calculation involved finding the R^2 for the complete model, which included the variable of interest and baseline model covariates, and then subtracting the R^2 for the baseline covariates alone. Significance in differences between distributions of PRS_n and $PRSMix_AOI_n$ were determined through subgroup heterogeneity meta-analysis. Standard errors for Nagelkerke's pseudo-R^2 were determined via standard bootstrapping procedure. Congruence of risk estimates was assessed with the Jaccard similarity index, which measures the amount of overlap or similarity between two sample sets.[25] All analyses were two-sided. Statistical analyses were performed using R version 4.3.1.

Data availability

All data are made available from the *All of Us* Research Study to researchers from universities and other institutions with genuine research inquiries following institutional review board and *All of Us* approval. This research was approved by the Mass General Brigham institutional review board. The weights of the polygenic scores analyzed in this study are publicly available for download from the Polygenic Score Catalog.[20] The mixing weights of the $PRSMix_AOI_n$ are available in the supplementary tables 5-7.

Software and Code Availability

No custom software or algorithms were developed for data collection. We used the following softwares for handling and scoring the genetic data: Plink 2.0, R 4.3.1 (including package PRSmix), and Python 3.11.5 (including packages pandas 2.1.4 and NumPy 1.6.2). For data analysis and visualization, we used R 4.3.1 (including packages PRSmix, ggplot2, meta, class, rcompanion, survminer) and Python 3.11.5 (including packages pandas 2.1.4, NumPy 1.6.2, and Matplotlib 3.7.2). Analyses were performed on the AOU Researcher Workbench. Results are reported in compliance with the AoU Data and Statistics Dissemination Policy.

Acknowledgements:
This work was supported by grants K08HL161448 (to A.C.F.), R01HL164629 (to A.C.F.), K08HL168238 (to A.P.P.), R01HL169015 (to A.P.P.), R01HL1427 (to P.N.), R01HL148565 (to P.N.), R01HL148050 (to P.N.) from the National Heart, Lung, and Blood Institute and grants T32HG010464 (to S.M.U.), R01HG012354 (to A.P.P.) and U01HG011719 (to A.P.P. and P.N.) from the National Human Genome Research Institute. This research has been conducted using the *All of Us* cohort study. We gratefully acknowledge *All of Us* participants for their contributions, without whom this research would not have been possible. We also thank the National Institutes of Health's *All of Us* Research Program for making available the participant data examined in this study.

Disclosures:
A.C.F. reports being co-founder of Goodpath, serving as scientific advisor to MyOme and HeartFlow, and receiving a research grant from Foresite Labs. J.W.S. is a member of the Scientific Advisory Board of Sensorium Therapeutics (with options), has received grant support from Biogen, Inc. and is PI of a collaborative study of the genetics of depression and bipolar disorder sponsored by 23andMe for which 23andMe provides analysis time as in-kind support but no payments. P.N. reports research grants from Allelica, Amgen, Apple, Boston Scientific, Genentech / Roche, and Novartis, personal fees from Allelica, Apple, AstraZeneca, Blackstone Life Sciences, Creative Education Concepts, CRISPR Therapeutics, Eli Lilly & Co, Esperion Therapeutics, Foresite Labs, Genentech / Roche, GV, HeartFlow, Magnet Biomedicine, Merck, Novartis, TenSixteen Bio, and Tourmaline Bio, equity in Bolt, Candela, Mercury, MyOme, Parameter Health, Preciseli, and TenSixteen Bio, and spousal employment at Vertex Pharmaceuticals, all unrelated to the present work. The remaining authors declare no competing interests.

References:

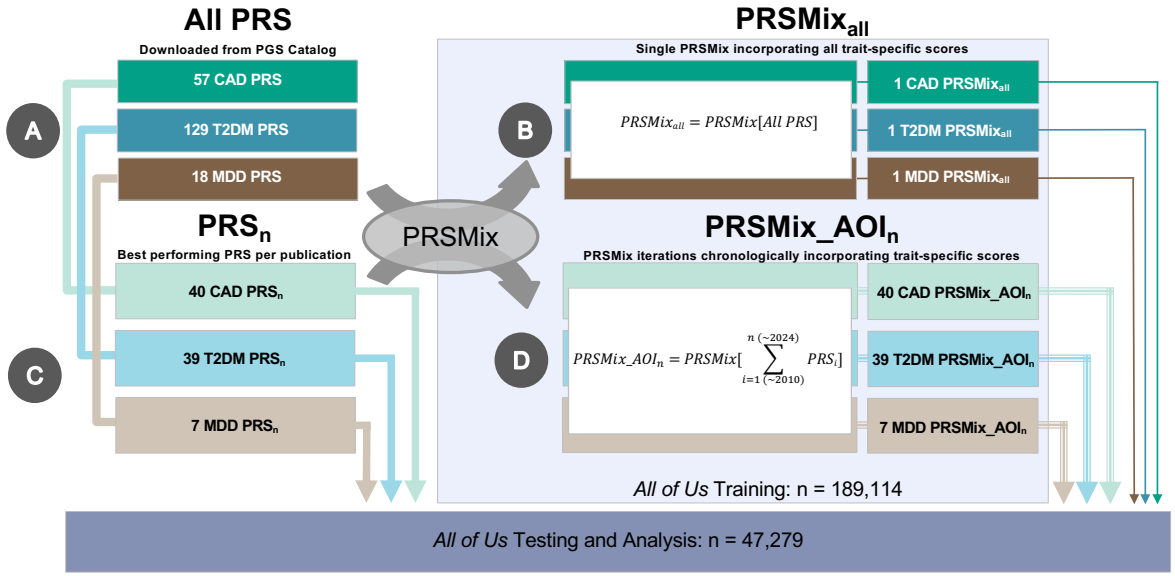
1. Patel, A. P. & Khera, A. V. Advances and Applications of Polygenic Scores for Coronary Artery Disease. *Annu. Rev. Med.* **74**, 141–154 (2023).
2. Privé, F., Arbel, J. & Vilhjálmsson, B. J. LDpred2: better, faster, stronger. *Bioinforma. Oxf. Engl.* btaa1029 (2020) doi:10.1093/bioinformatics/btaa1029.
3. Loh, P.-R. *et al.* Efficient Bayesian mixed-model analysis increases association power in large cohorts. *Nat. Genet.* **47**, 284–290 (2015).
4. Ruan, Y. *et al.* Improving polygenic prediction in ancestrally diverse populations. *Nat. Genet.* **54**, 573–580 (2022).
5. Patel, A. P. & Fahed, A. C. Pragmatic Approach to Applying Polygenic Risk Scores to Diverse Populations. *Curr. Protoc.* **3**, e911 (2023).
6. Yengo, L. *et al.* A saturated map of common genetic variants associated with human height. *Nature* **610**, 704–712 (2022).
7. Aragam, K. G. *et al.* Discovery and systematic characterization of risk variants and genes for coronary artery disease in over a million participants. *Nat. Genet.* **54**, 1803–1815 (2022).
8. Patel, A. P. *et al.* A multi-ancestry polygenic risk score improves risk prediction for coronary artery disease. *Nat. Med.* (2023) doi:10.1038/s41591-023-02429-x.
9. Multhaup, M. L. *et al.* The science behind 23andMe's Type 2 Diabetes report: Estimating the likelihood of developing type 2 diabetes with polygenic models. (2019).
10. Busby, G., Bolli, A., Di Domenico, P. & Botta, G. Development and Validation of a Polygenic Risk Score for Coronary Artery Disease.
11. Lennon, N. J. *et al.* Selection, optimization and validation of ten chronic disease polygenic risk scores for clinical implementation in diverse US populations. *Nat. Med.* **30**, 480–487 (2024).
12. Linder, J. E. *et al.* Returning integrated genomic risk and clinical recommendations: The eMERGE study. *Genet. Med. Off. J. Am. Coll. Med. Genet.* **25**, 100006 (2023).

13. McCarty, C. A. *et al.* The eMERGE Network: a consortium of biorepositories linked to electronic medical records data for conducting genomic studies. *BMC Med. Genomics* **4**, 13 (2011).
14. Hao, L. *et al.* Development of a clinical polygenic risk score assay and reporting workflow. *Nat. Med.* **28**, 1006–1013 (2022).
15. Karlson, E. W., Boutin, N. T., Hoffnagle, A. G. & Allen, N. L. Building the Partners HealthCare Biobank at Partners Personalized Medicine: Informed Consent, Return of Research Results, Recruitment Lessons and Operational Considerations. *J. Pers. Med.* **6**, (2016).
16. O'Sullivan, J. W. *et al.* Polygenic Risk Scores for Cardiovascular Disease: A Scientific Statement From the American Heart Association. *Circulation* **146**, e93–e118 (2022).
17. Kullo, I. J. *et al.* Incorporating a Genetic Risk Score Into Coronary Heart Disease Risk Estimates: Effect on Low-Density Lipoprotein Cholesterol Levels (the MI-GENES Clinical Trial). *Circulation* **133**, 1181–1188 (2016).
18. All of Us Research Program Investigators *et al.* The 'All of Us' Research Program. *N. Engl. J. Med.* **381**, 668–676 (2019).
19. Truong, B. *et al.* Integrative polygenic risk score improves the prediction accuracy of complex traits and diseases. *Cell Genomics* **4**, 100523 (2024).
20. Lambert, S. A. *et al.* The Polygenic Score Catalog as an open database for reproducibility and systematic evaluation. *Nat. Genet.* **53**, 420–425 (2021).
21. All of Us Research Program Genomics Investigators. Genomic data in the All of Us Research Program. *Nature* **627**, 340–346 (2024).
22. Ding, Y. *et al.* Large uncertainty in individual polygenic risk score estimation impacts PRS-based risk stratification. *Nat. Genet.* **54**, 30–39 (2022).
23. Martin, A. R. *et al.* Current clinical use of polygenic scores will risk exacerbating health disparities. *Nat. Genet.* **51**, 584–591 (2019).

24. Kho, A. N. *et al.* Electronic medical records for genetic research: results of the eMERGE consortium. *Sci. Transl. Med.* **3**, 79re1 (2011).

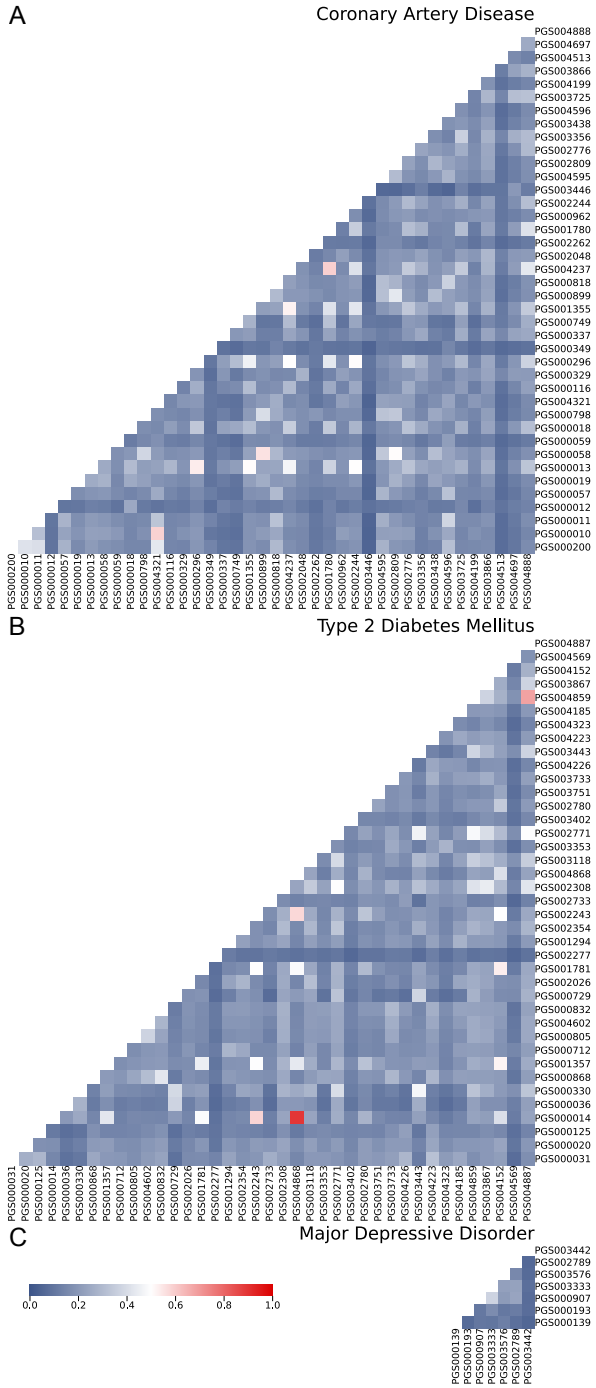
25. Tan, P.-N., Steinbach, M. & Kumar, V. *Introduction to Data Mining, (First Edition)*. (Addison-Wesley Longman Publishing Co., Inc., USA, 2005).

Figure 1: Overview of analysis



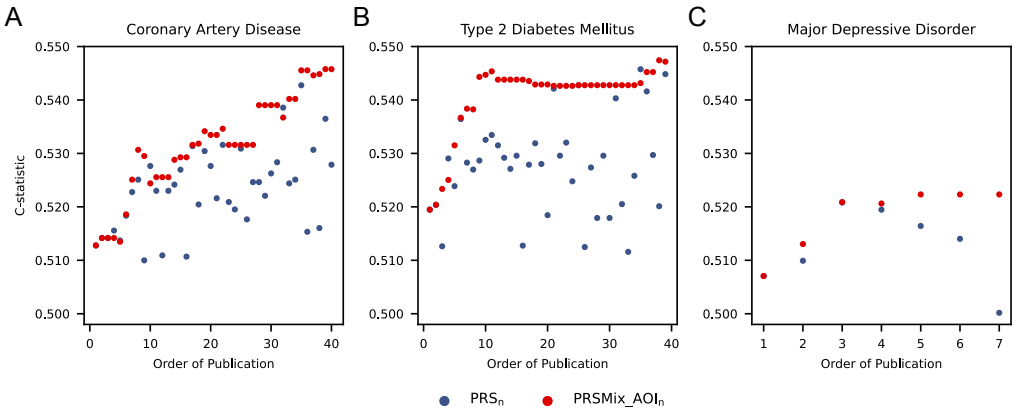
A) Polygenic scores that were not developed using PRSMix and do not include *All of Us* participants in their development were downloaded from the PGS catalog and calculated in the *All of Us* cohort for three traits of interest, including 57 scores for coronary artery disease (CAD), 129 scores for type 2 diabetes (T2DM), and 18 scores for major depressive disorder (MDD). B) PRSMix, a tool that incorporates information across several PRS for a given trait to improve prediction accuracy for a target population, was used to generate a $PRSMix_{all}$ for each trait incorporating all available scores in a training dataset of 189,114 individuals. Being in the top 10% of these trait-specific $PRSMix_{all}$ score distributions was used as the benchmark for classifying high genetic risk. C) For publications that had shared multiple scores in the PGS Catalog for a trait, the score with the strongest association with the trait of interest was identified and chronologically ordered as PRS_n. D) PRSMix was used to generate a set of chronological add-one-in (AOI) $PRSMix_AOI_n$ scores corresponding to each set of PRS_n, wherein we combined each PRS_n with all prior published PRS_n scores for the same trait. These PRS_n, corresponding $PRSMix_AOI_n$, and the single $PRSMix_{all}$ scores for each trait were calculated and further analyzed in a holdout population of 47,279 participants in *All of Us*.

Figure 2: Jaccard index heatmaps for high genetic risk across trait-specific polygenic score (PRS_n) pairs



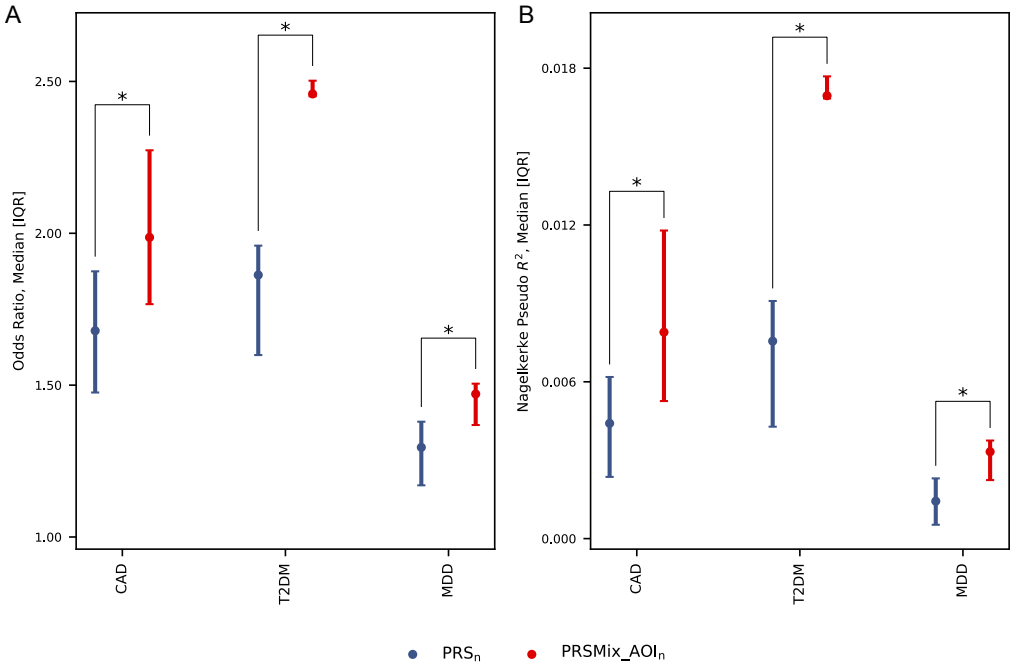
Jaccard indices of similarity calculated for pairs of classification of top 10% risk determined by polygenic scores (PRS_n) for coronary artery disease, type 2 diabetes mellitus, and major depressive disorder. Scores are ordered chronologically based on date of publication, advancing left to right and bottom to top. The median [interquartile range] Jaccard index for top 10% classification was: A) 0.17 [0.13-0.22] for coronary artery disease; B) 0.18 [0.15-0.22] for type 2 diabetes mellitus; and C) 0.11 [0.08-0.17] for major depressive disorder.

Figure 3: Chronologic trends in C-statistic estimates for PRS_n vs. $PRSMix_AOI_n$ across three traits



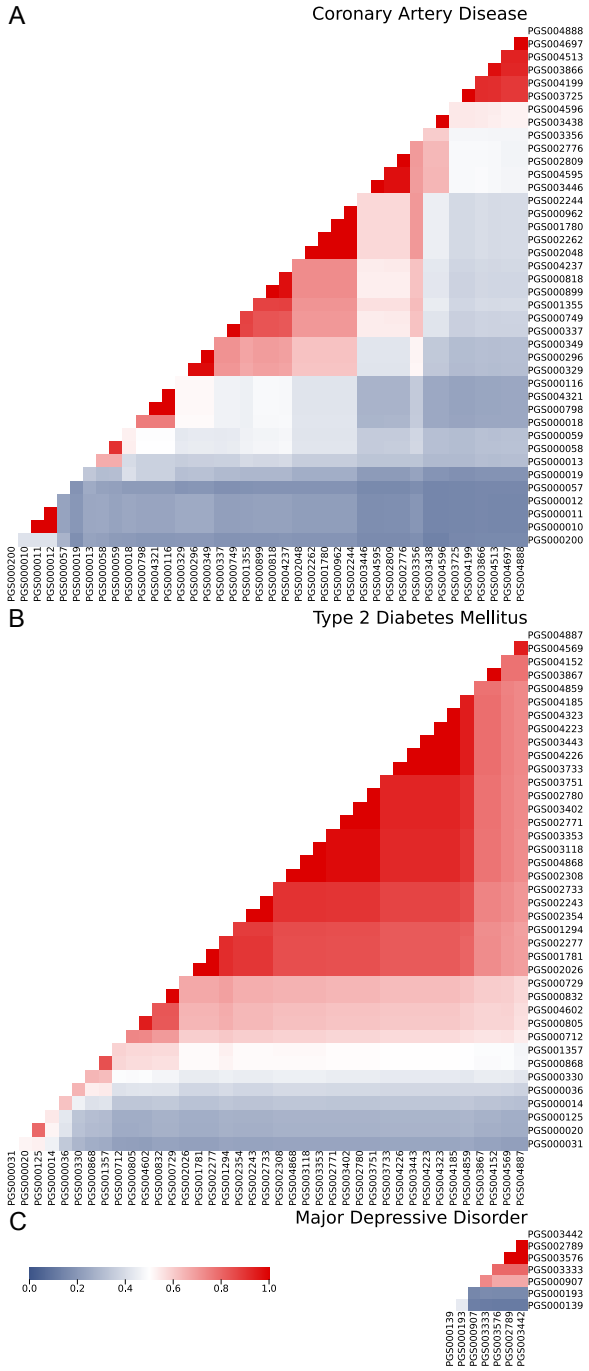
C-statistics are based on logistic regression models predicting disease of interest, with high genetic risk as the only variable. High genetic risk is defined as the top 10% of each respective polygenic score. PRS_n indicates a chosen subset of individual trait-specific PRS ordered based on date of publication. $PRSMix_AOI_n$ indicates polygenic scores generated using the PRSMix framework, wherein PRS_n is combined with all individual PRS_n published before it.

Figure 4: Association strength and variance explained for PRS_n vs. $PRSMix_AOI_n$ across three traits



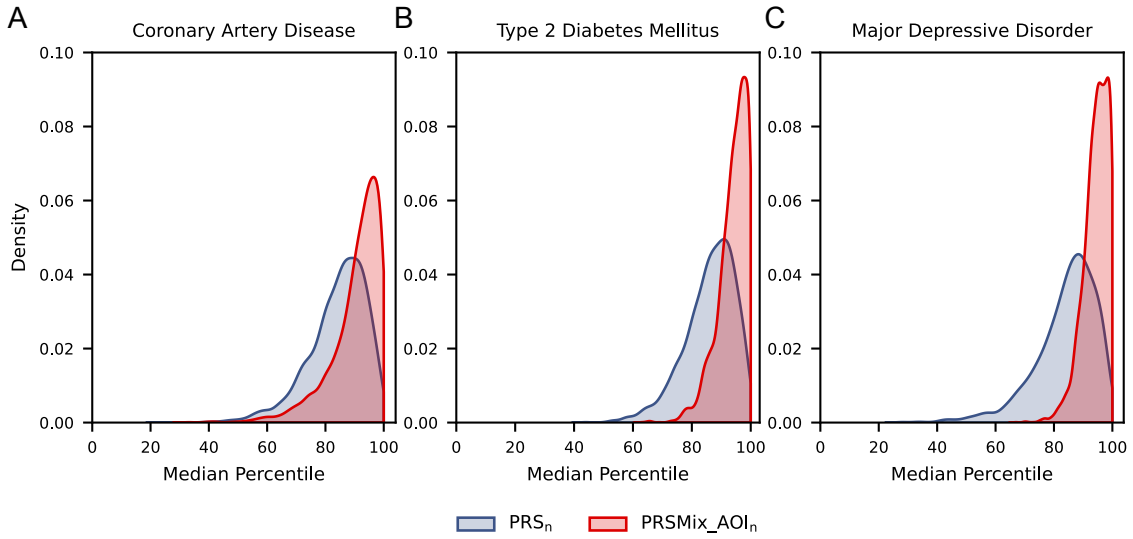
A) Median and interquartile range for odds ratios across scores are based on logistic regression models predicting disease of interest, with variables of high genetic risk, age, sex, and first ten principal components of genetic ancestry. High genetic risk is defined as the top 10% of each respective polygenic score. PRS_n indicates individual trait-specific PRS ordered based on date of publication. $PRSMix_AOI_n$ indicates polygenic scores generated using the PRSMix framework, PRS_n was combined with all individual PRS_n published before it. The proportion of phenotypic variance explained by the high polygenic risk classification by each polygenic score for each respective disease. B) Median and interquartile range for incremental Nagelkerke's pseudo-R^2 metric across scores, as the difference of the full model inclusive of the polygenic score plus age, sex, and the first ten principal components of ancestry minus R^2 for the covariates alone. CAD: coronary artery disease. T2DM: type 2 diabetes. MDD: major depressive disorder. * designates $P_{heterogeneity}<0.001$.

Figure 5: Jaccard index heatmaps for high genetic risk across trait-specific PRSMix_AOI$_n$ pairs



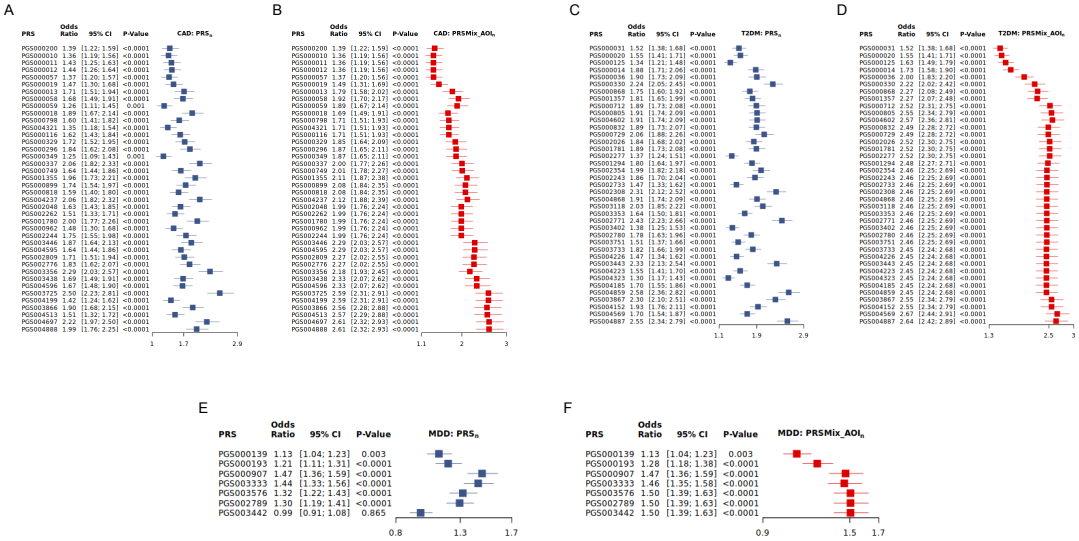
Jaccard indices of similarity calculated for pairs of classification of top 10% risk determined by PRSMix_AOI$_n$ - polygenic scores generated by chronologically adding in PRS$_n$ using PRSMix framework for coronary artery disease, type 2 diabetes mellitus, and major depressive disorder. Scores are ordered chronologically based on date of publication, advancing left to right and bottom to top. The median [interquartile range] Jaccard index for top 10% classification was: A) 0.39 [0.24-0.53] for coronary artery disease; B) 0.64 [0.45-0.85] for type 2 diabetes mellitus; and C) 0.44 [0.15-0.80] for major depressive disorder.

Figure 6: Median percentile distributions of PRS_n vs. $PRSMix_AOI_n$ across three traits for individuals identified as having high genetic risk using $PRSMix_{all}$



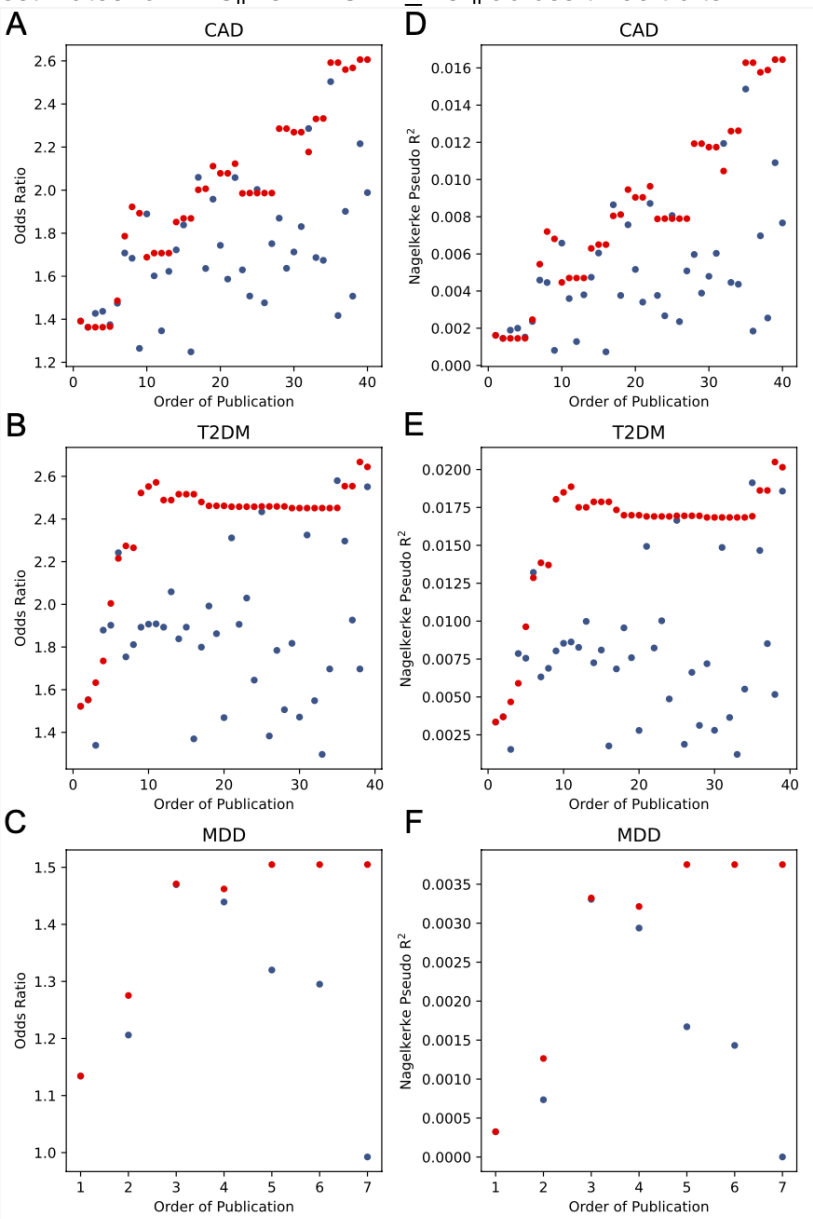
Distributions of the median percentile per polygenic score type for individuals classified in top 10% risk by the $PRSMix_{all}$ for each disease of interest. PRS_n indicates individual trait-specific PRS ordered based on date of publication. $PRSMix_AOI_n$ indicates polygenic scores generated using the PRSMix framework, PRS_n was combined with all individual PRS published before it. A) Individuals classified in top 10% risk by the $PRSMix_{all}$ for CAD have median percentiles across PRS_n of 86 [79-91.5] vs. 93 [87-97] for $PRSMix_AOI_n$. B) Similarly, individuals classified in the top 10% risk by the $PRSMix_{all}$ for T2DM have median percentiles across PRS_n of 87 [81-93] vs. 95 [91-98] for $PRSMix_AOI_n$. C) Additionally, individuals classified in the top 10% risk by the $PRSMix_{all}$ for MDD have median percentiles across PRS_n of 86 [79-92] vs. 95 [92-98] for $PRSMix_AOI_n$.

Extended Figure 1: Odds ratios for high genetic risk using PRS_n vs. $PRSMix_AOI_n$ with three traits



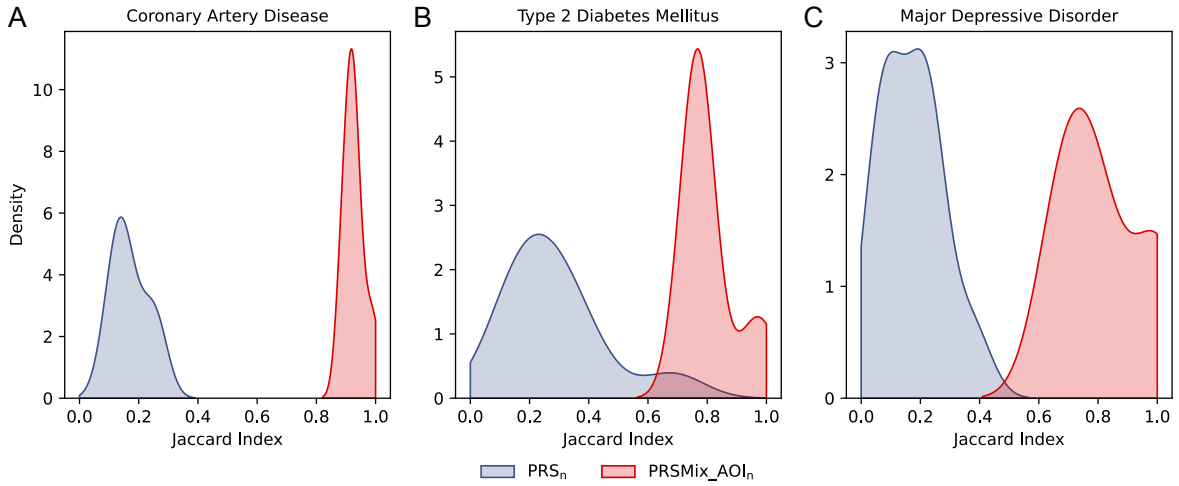
Odds ratios are based on logistic regression models predicting disease of interest, with variables of high genetic risk, age, sex, and first ten principal components of genetic ancestry. High genetic risk is defined as the top 10% of each respective polygenic score. PRS_n indicates a specific subset of individual trait-specific PRS ordered chronologically based on date of publication. $PRSMix_AOI_n$ indicates polygenic scores generated using the PRSMix framework wherein each PRS_n was combined with all individual PRS_n published before it. CAD: coronary artery disease; T2DM: type 2 diabetes; MDD: major depressive disorder.

Extended Figure 2: Chronologic trends in association strength and variance explained estimates for PRS_n vs. $PRSMix_AOI_n$ across three traits



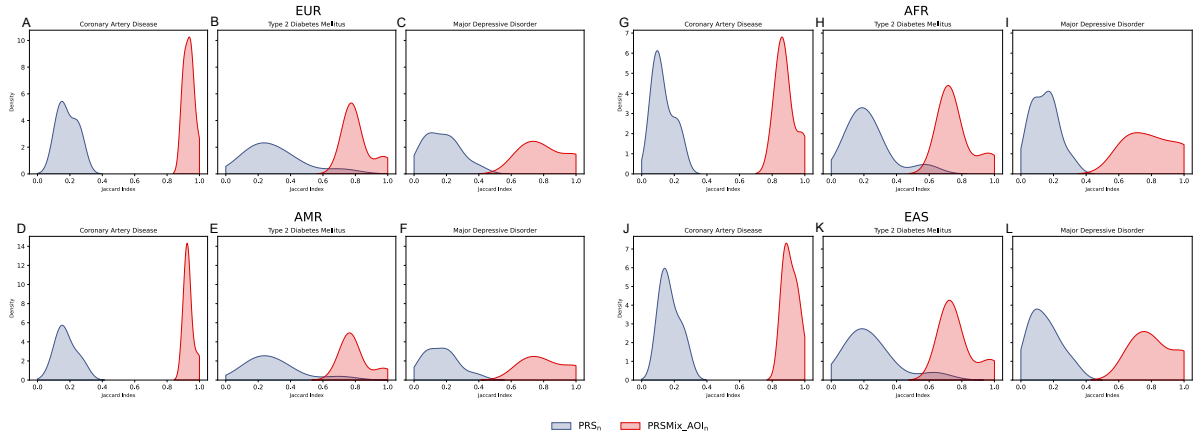
A) Odds ratios are based on logistic regression models predicting disease of interest, with variables of high genetic risk, age, sex, and first ten principal components of genetic ancestry. High genetic risk is defined as the top 10% of each respective polygenic score. PRS_n indicates individual trait-specific PRS ordered chronologically based on date of publication. $PRSMix_AOI_n$ indicates polygenic scores generated using the PRSMix framework, PRS_n was combined with all individual PRS_n published before it. The proportion of phenotypic variance is explained by the high polygenic risk classification by each polygenic score for each respective disease: B) Nagelkerke's pseudo-R^2 metric, as the difference of the full model inclusive of the polygenic score plus age, sex, and the first ten principal components of ancestry minus R^2 for the covariates alone. CAD: coronary artery disease; T2DM: type 2 diabetes; MDD: major depressive disorder.

Extended Figure 3: Jaccard index distributions for high genetic risk determined by trait-specific PRS_n vs. $PRSMix_AOI_n$ from five most recent publications



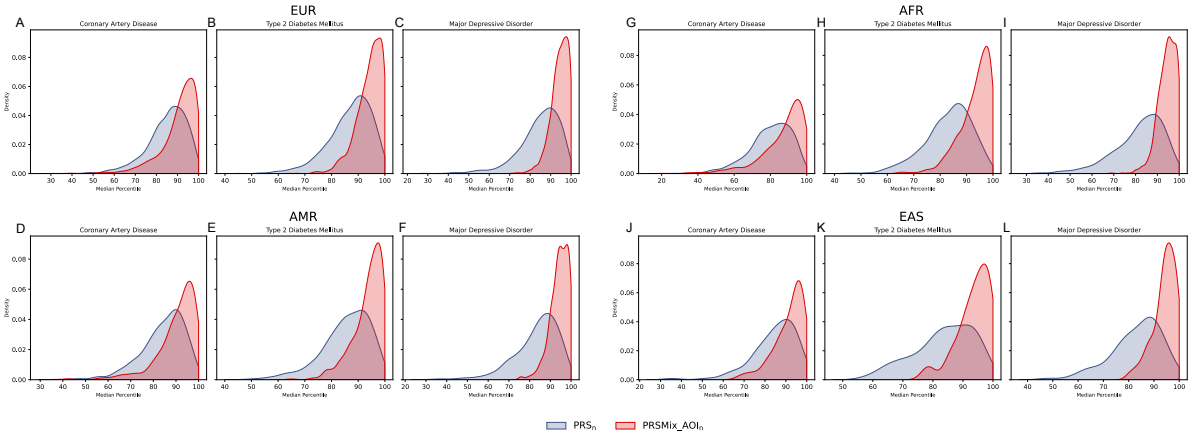
Jaccard indices of similarity calculated for pairs of classification of top 10% risk determined by trait-specific polygenic scores from the last five respective publications. PRS_n indicates individual trait-specific PRS ordered based on date of publication. $PRSMix_AOI_n$ indicates polygenic scores generated using the PRSMix framework, wherein each PRS_n was combined with all individual PRS_n published before it.

Extended Figure 4: Jaccard index distributions for high genetic risk determined by trait-specific PRS_n vs. $PRSMix_AOI_n$ from five most recent publications across genetically predicted ancestry groups



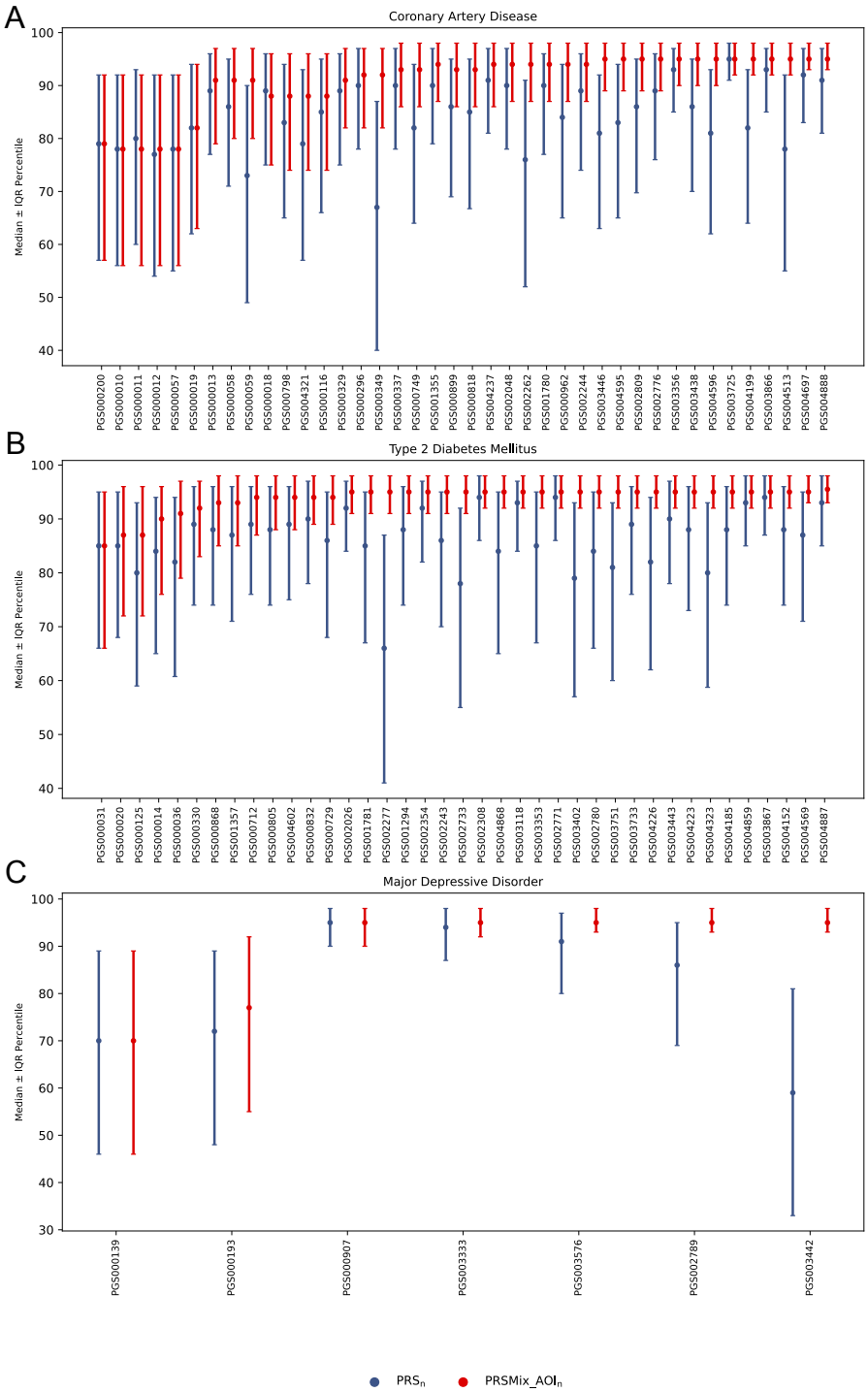
Jaccard indices of similarity calculated for pairs of classification of top 10% of risk determined by trait-specific polygenic scores from the last five respective publications. PRS_n indicates individual trait-specific PRS ordered based on date of publication. $PRSMix_AOI_n$ indicates polygenic scores generated using the PRSMix framework wherein each PRS_n was combined with all individual PRS_n published before it. Genetically inferred ancestry based on k-nearest neighbor approach as European (EUR), African (AFR), admixed American (AMR), and East Asian (EAS).

Extended Figure 5: Median percentile of PRS_n vs. $PRSMix_AOI_n$ across three traits for individuals identified as having high genetic risk using $PRSMix_{all}$, across genetically predicted ancestry groups

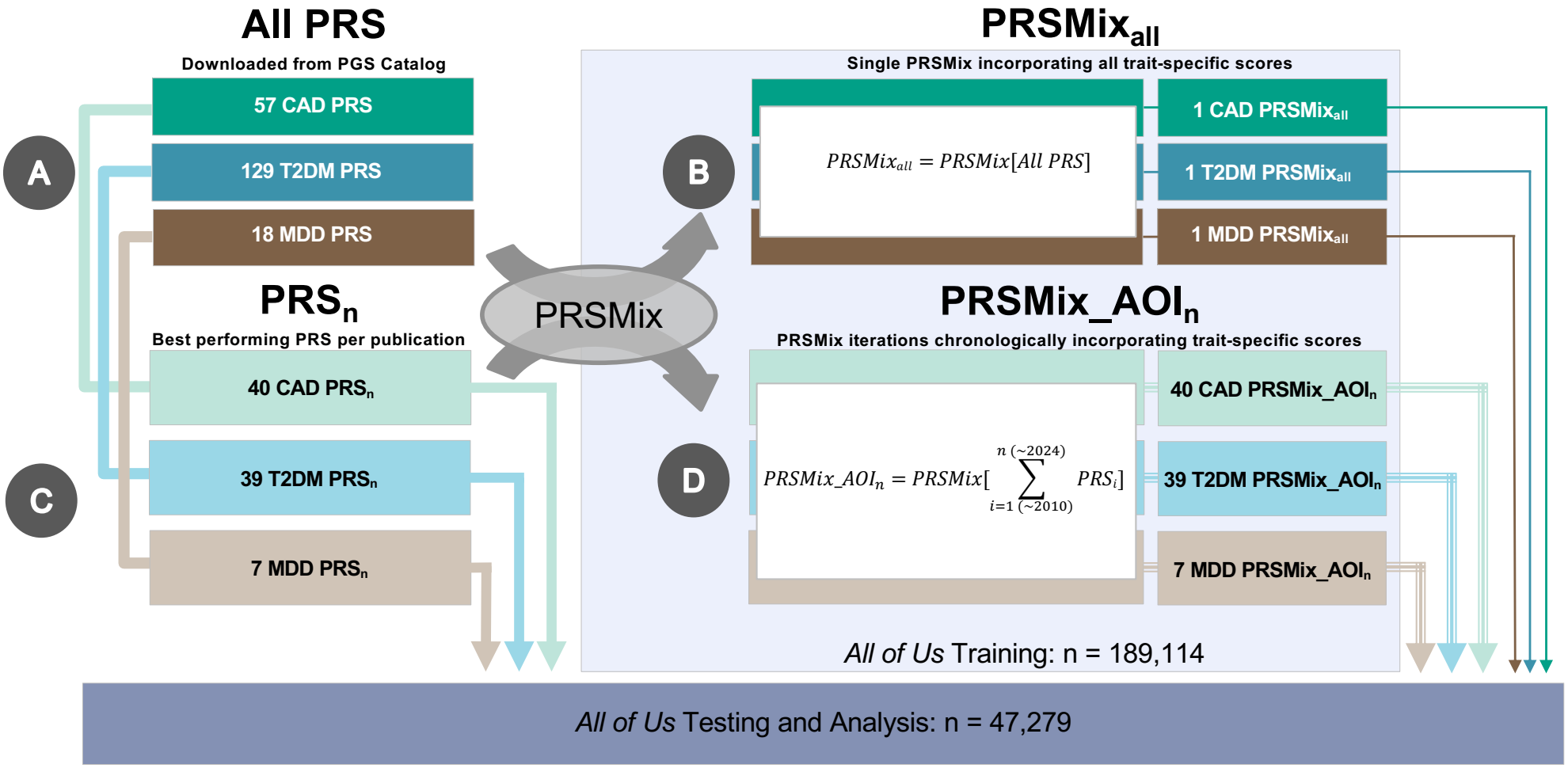


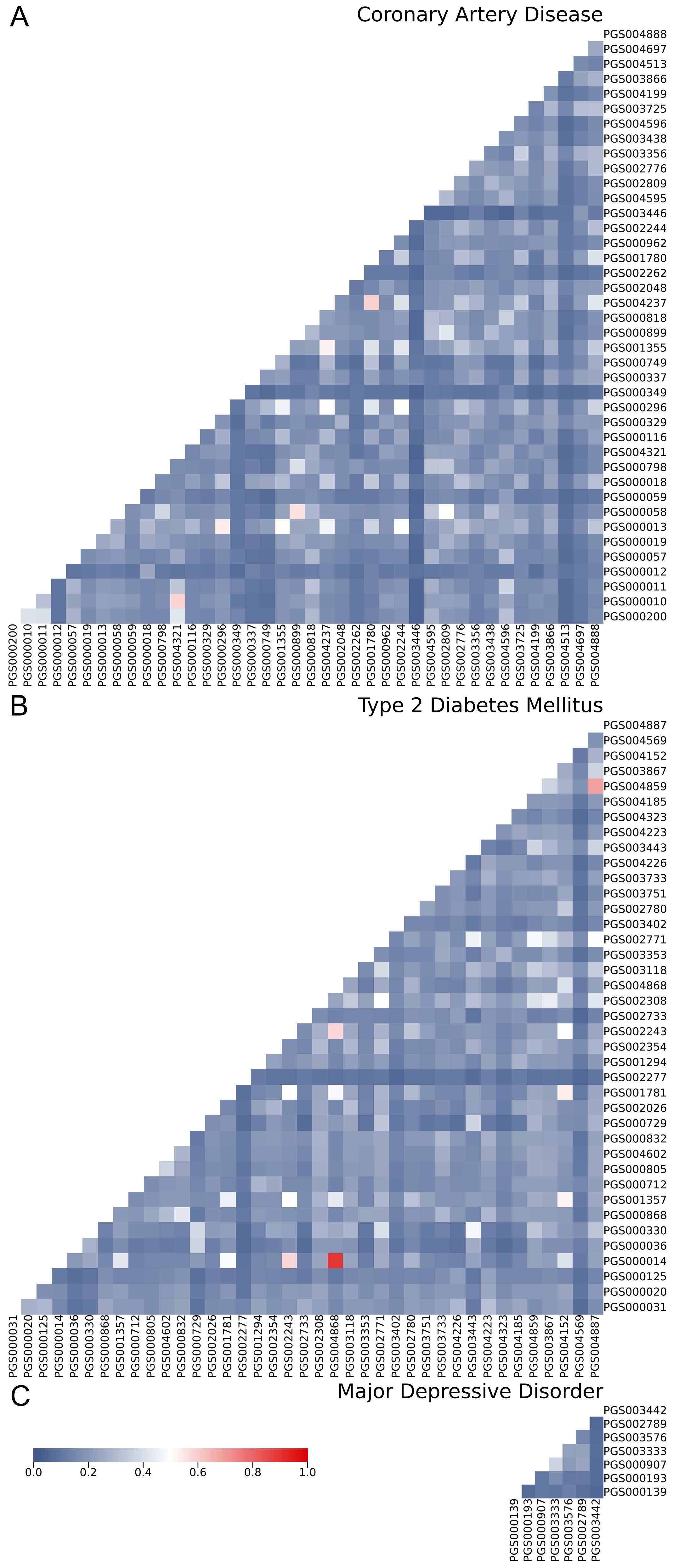
Distributions of the median percentile per polygenic score type for individuals classified in top 10% risk by the $PRSMix_{all}$ for each disease of interest. PRS_n indicates individual trait-specific PRS ordered based on date of publication. $PRSMix_AOI_n$ indicates polygenic scores generated using the PRSMix framework, PRS_n was combined with all individual PRS_n published before it. Genetically inferred ancestry based on k-nearest neighbor approach as European (EUR), African (AFR), admixed American (AMR), and East Asian (EAS).

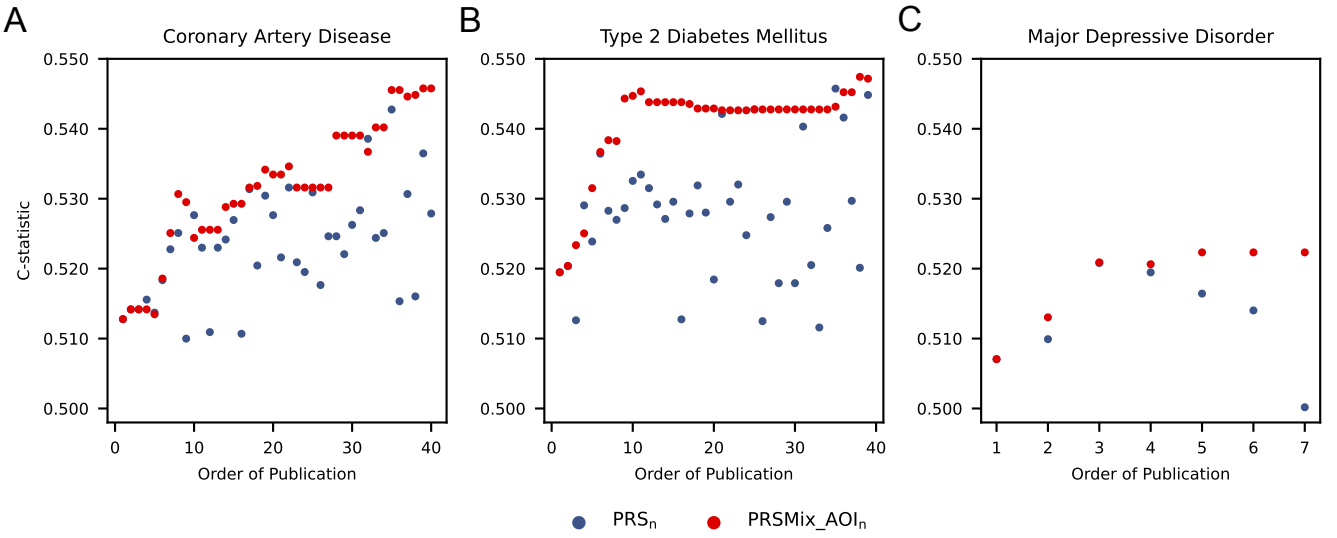
Extended Figure 6: Score-specific percentile distributions of PRS_n vs. $PRSMix_AOI_n$ across three traits for individuals identified as having high genetic risk using $PRSMix_{all}$

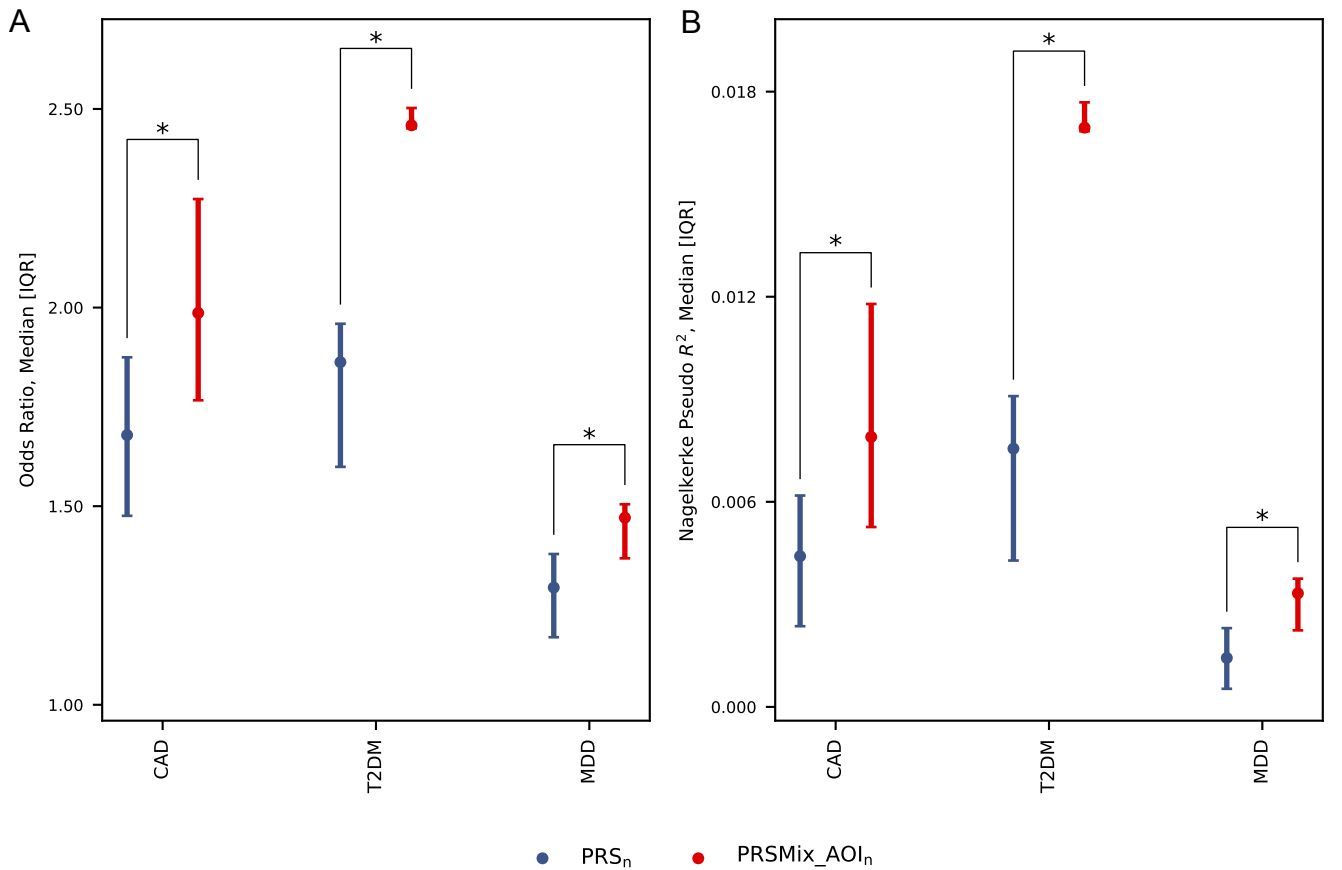


Median percentile and interquartile range per polygenic score type for individuals classified in top 10% risk by the $PRSMix_{all}$ for each disease of interest. PRS_n indicates individual trait-specific PRS ordered based on date of publication. $PRSMix_AOI_n$ indicates polygenic scores generated using the PRSMix framework, PRS_n was combined with all individual PRS_n published before it.

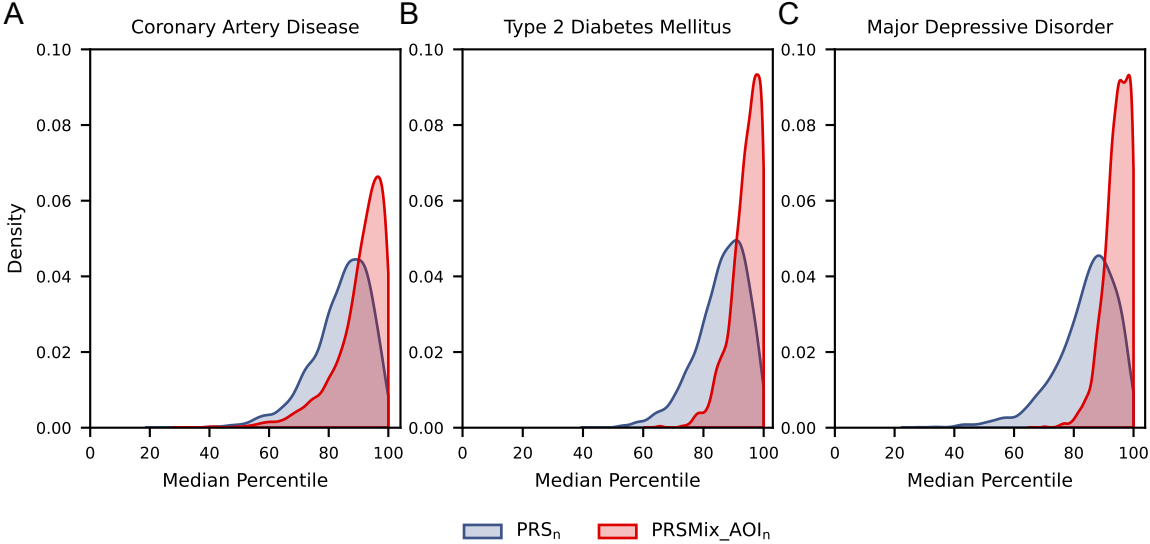

All PRS
Downloaded from PGS Catalog
57 CAD PRS
129 T2DM PRS
18 MDD PRS
A
PRSMix
PRSMix_all
Single PRSMix incorporating all trait-specific scores
B
$PRSMix_{all} = PRSMix[All\ PRS]$
1 CAD PRSMix_all
1 T2DM PRSMix_all
1 MDD PRSMix_all
PRS_n
Best performing PRS per publication
40 CAD PRS_n
39 T2DM PRS_n
7 MDD PRS_n
C
PRSMix_AOI_n
PRSMix iterations chronologically incorporating trait-specific scores
D
$PRSMix_AOI_n = PRSMix[\sum_{i=1\ (\sim 2010)}^{n\ (\sim 2024)} PRS_i]$
40 CAD PRSMix_AOI_n
39 T2DM PRSMix_AOI_n
7 MDD PRSMix_AOI_n
All of Us Training: n = 189,114
All of Us Testing and Analysis: n = 47,279


A
Coronary Artery Disease
PGS004888
PGS004697
PGS004513
PGS003866
PGS004199
PGS003725
PGS004596
PGS003438
PGS003356
PGS002776
PGS002809
PGS004595
PGS003446
PGS002244
PGS000962
PGS001780
PGS002262
PGS002048
PGS004237
PGS000818
PGS000899
PGS001355
PGS000749
PGS000337
PGS000349
PGS000296
PGS000329
PGS000116
PGS004321
PGS000798
PGS000018
PGS000059
PGS000058
PGS000013
PGS000019
PGS000057
PGS000012
PGS000011
PGS000010
PGS000200
B
Type 2 Diabetes Mellitus
PGS004887
PGS004569
PGS004152
PGS003867
PGS004859
PGS004185
PGS004323
PGS004223
PGS003443
PGS004226
PGS003733
PGS003751
PGS002780
PGS003402
PGS002771
PGS003353
PGS003118
PGS004868
PGS002308
PGS002733
PGS002243
PGS002354
PGS001294
PGS002277
PGS001781
PGS002026
PGS000729
PGS000832
PGS004602
PGS000805
PGS000712
PGS001357
PGS000868
PGS000330
PGS000036
PGS000014
PGS000125
PGS000020
PGS000031
C
Major Depressive Disorder
PGS003442
PGS002789
PGS003576
PGS003333
PGS000907
PGS000193
PGS000139
0.0
0.2
0.4
0.6
0.8
1.0


A
Coronary Artery Disease
B
Type 2 Diabetes Mellitus
C
Major Depressive Disorder
C-statistic
0.550
0.540
0.530
0.520
0.510
0.500
0
10
20
30
40
1
2
3
4
5
6
7
Order of Publication
PRS$_n$
PRSMix_AOI$_n$

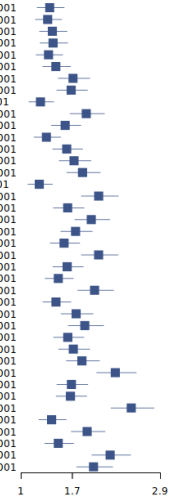

A
B
Odds Ratio, Median [IQR]
2.50
2.00
1.50
1.00
Nagelkerke Pseudo R^2, Median [IQR]
0.018
0.012
0.006
0.000
*
CAD
T2DM
MDD
PRS_n
$PRSMix_AOI_n$

A
Coronary Artery Disease
PGS004888
PGS004697
PGS004513
PGS003866
PGS004199
PGS003725
PGS004596
PGS003438
PGS003356
PGS002776
PGS002809
PGS004595
PGS003446
PGS002244
PGS000962
PGS001780
PGS002262
PGS002048
PGS004237
PGS000818
PGS000899
PGS001355
PGS000749
PGS000337
PGS000349
PGS000296
PGS000329
PGS000116
PGS004321
PGS000798
PGS000018
PGS000059
PGS000058
PGS000013
PGS000019
PGS000057
PGS000012
PGS000011
PGS000010
PGS000200
PGS000200
PGS000010
PGS000011
PGS000012
PGS000057
PGS000019
PGS000013
PGS000058
PGS000059
PGS000018
PGS000798
PGS004321
PGS000116
PGS000329
PGS000296
PGS000349
PGS000337
PGS000749
PGS001355
PGS000899
PGS000818
PGS004237
PGS002048
PGS002262
PGS001780
PGS000962
PGS002244
PGS003446
PGS004595
PGS002809
PGS002776
PGS003356
PGS003438
PGS004596
PGS003725
PGS004199
PGS003866
PGS004513
PGS004697
B
Type 2 Diabetes Mellitus
PGS004887
PGS004569
PGS004152
PGS003867
PGS004859
PGS004185
PGS004323
PGS004223
PGS003443
PGS004226
PGS003733
PGS003751
PGS002780
PGS003402
PGS002771
PGS003353
PGS003118
PGS004868
PGS002308
PGS002733
PGS002243
PGS002354
PGS001294
PGS002277
PGS001781
PGS002026
PGS000729
PGS000832
PGS004602
PGS000805
PGS000712
PGS001357
PGS000868
PGS000330
PGS000036
PGS000014
PGS000125
PGS000020
PGS000031
PGS000031
PGS000020
PGS000125
PGS000014
PGS000036
PGS000330
PGS000868
PGS001357
PGS000712
PGS000805
PGS004602
PGS000832
PGS000729
PGS002026
PGS001781
PGS002277
PGS001294
PGS002354
PGS002243
PGS002733
PGS002308
PGS004868
PGS003118
PGS003353
PGS002771
PGS003402
PGS002780
PGS003751
PGS003733
PGS004226
PGS003443
PGS004223
PGS004323
PGS004185
PGS004859
PGS003867
PGS004152
PGS004569
PGS004887
C
Major Depressive Disorder
PGS003442
PGS002789
PGS003576
PGS003333
PGS000907
PGS000193
PGS000139
PGS000139
PGS000193
PGS000907
PGS003333
PGS003576
PGS002789
PGS003442
0.0
0.2
0.4
0.6
0.8
1.0


A
Coronary Artery Disease
B
Type 2 Diabetes Mellitus
C
Major Depressive Disorder
Density
0.10
0.08
0.06
0.04
0.02
0.00
0
20
40
60
80
100
Median Percentile
PRS$_n$
PRSMix_AOI$_n$

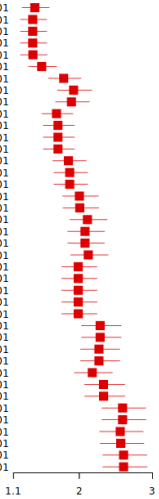
A

PRS	Odds Ratio	95% CI	P-Value
PGS000200	1.39	[1.22; 1.59]	<0.0001
PGS000010	1.36	[1.19; 1.56]	<0.0001
PGS000011	1.43	[1.25; 1.63]	<0.0001
PGS000012	1.44	[1.26; 1.64]	<0.0001
PGS000057	1.37	[1.20; 1.57]	<0.0001
PGS000019	1.47	[1.30; 1.68]	<0.0001
PGS000013	1.71	[1.51; 1.94]	<0.0001
PGS000058	1.68	[1.49; 1.91]	<0.0001
PGS000059	1.26	[1.11; 1.45]	0.001
PGS000018	1.89	[1.67; 2.14]	<0.0001
PGS000798	1.60	[1.41; 1.82]	<0.0001
PGS004321	1.35	[1.18; 1.54]	<0.0001
PGS000116	1.62	[1.43; 1.84]	<0.0001
PGS000329	1.72	[1.52; 1.95]	<0.0001
PGS000296	1.84	[1.62; 2.08]	<0.0001
PGS000349	1.25	[1.09; 1.43]	0.001
PGS000337	2.06	[1.82; 2.33]	<0.0001
PGS000749	1.64	[1.44; 1.86]	<0.0001
PGS001355	1.96	[1.73; 2.21]	<0.0001
PGS000899	1.74	[1.54; 1.97]	<0.0001
PGS000818	1.59	[1.40; 1.80]	<0.0001
PGS004237	2.06	[1.82; 2.32]	<0.0001
PGS002048	1.63	[1.43; 1.85]	<0.0001
PGS002262	1.51	[1.33; 1.71]	<0.0001
PGS001780	2.00	[1.77; 2.26]	<0.0001
PGS000962	1.48	[1.30; 1.68]	<0.0001
PGS002244	1.75	[1.55; 1.98]	<0.0001
PGS003446	1.87	[1.64; 2.13]	<0.0001
PGS004595	1.64	[1.44; 1.86]	<0.0001
PGS002809	1.71	[1.51; 1.94]	<0.0001
PGS002776	1.83	[1.62; 2.07]	<0.0001
PGS003356	2.29	[2.03; 2.57]	<0.0001
PGS003438	1.69	[1.49; 1.91]	<0.0001
PGS004596	1.67	[1.48; 1.90]	<0.0001
PGS003725	2.50	[2.23; 2.81]	<0.0001
PGS004199	1.42	[1.24; 1.62]	<0.0001
PGS003866	1.90	[1.68; 2.15]	<0.0001
PGS004513	1.51	[1.32; 1.72]	<0.0001
PGS004697	2.22	[1.97; 2.50]	<0.0001
PGS004888	1.99	[1.76; 2.25]	<0.0001



B

PRS	Odds Ratio	95% CI	P-Value
PGS000200	1.39	[1.22; 1.59]	<0.0001
PGS000010	1.36	[1.19; 1.56]	<0.0001
PGS000011	1.36	[1.19; 1.56]	<0.0001
PGS000012	1.36	[1.19; 1.56]	<0.0001
PGS000057	1.37	[1.20; 1.56]	<0.0001
PGS000019	1.49	[1.31; 1.69]	<0.0001
PGS000013	1.79	[1.58; 2.02]	<0.0001
PGS000058	1.92	[1.70; 2.17]	<0.0001
PGS000059	1.89	[1.67; 2.14]	<0.0001
PGS000018	1.69	[1.49; 1.91]	<0.0001
PGS000798	1.71	[1.51; 1.93]	<0.0001
PGS004321	1.71	[1.51; 1.93]	<0.0001
PGS000116	1.71	[1.51; 1.93]	<0.0001
PGS000329	1.85	[1.64; 2.09]	<0.0001
PGS000296	1.87	[1.65; 2.11]	<0.0001
PGS000349	1.87	[1.65; 2.11]	<0.0001
PGS000337	2.00	[1.77; 2.26]	<0.0001
PGS000749	2.01	[1.78; 2.27]	<0.0001
PGS001355	2.11	[1.87; 2.38]	<0.0001
PGS000899	2.08	[1.84; 2.35]	<0.0001
PGS000818	2.08	[1.84; 2.35]	<0.0001
PGS004237	2.12	[1.88; 2.39]	<0.0001
PGS002048	1.99	[1.76; 2.24]	<0.0001
PGS002262	1.99	[1.76; 2.24]	<0.0001
PGS001780	1.99	[1.76; 2.24]	<0.0001
PGS000962	1.99	[1.76; 2.24]	<0.0001
PGS002244	1.99	[1.76; 2.24]	<0.0001
PGS003446	2.29	[2.03; 2.57]	<0.0001
PGS004595	2.29	[2.03; 2.57]	<0.0001
PGS002809	2.27	[2.02; 2.55]	<0.0001
PGS002776	2.27	[2.02; 2.55]	<0.0001
PGS003356	2.18	[1.93; 2.45]	<0.0001
PGS003438	2.33	[2.07; 2.62]	<0.0001
PGS004596	2.33	[2.07; 2.62]	<0.0001
PGS003725	2.59	[2.31; 2.91]	<0.0001
PGS004199	2.59	[2.31; 2.91]	<0.0001
PGS003866	2.56	[2.28; 2.88]	<0.0001
PGS004513	2.57	[2.29; 2.88]	<0.0001
PGS004697	2.61	[2.32; 2.93]	<0.0001
PGS004888	2.61	[2.32; 2.93]	<0.0001

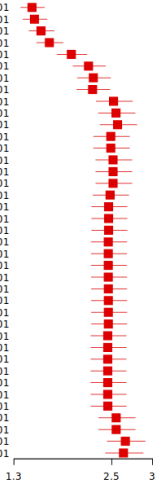


C

PRS	Odds Ratio	95% CI	P-Value
PGS000031	1.52	[1.38; 1.68]	<0.0001
PGS000020	1.55	[1.41; 1.71]	<0.0001
PGS000125	1.34	[1.21; 1.48]	<0.0001
PGS000014	1.88	[1.71; 2.06]	<0.0001
PGS000036	1.90	[1.73; 2.09]	<0.0001
PGS000330	2.24	[2.05; 2.45]	<0.0001
PGS000868	1.75	[1.60; 1.92]	<0.0001
PGS001357	1.81	[1.65; 1.99]	<0.0001
PGS000712	1.89	[1.73; 2.08]	<0.0001
PGS000805	1.91	[1.74; 2.09]	<0.0001
PGS004602	1.91	[1.74; 2.09]	<0.0001
PGS000832	1.89	[1.73; 2.07]	<0.0001
PGS000729	2.06	[1.88; 2.26]	<0.0001
PGS002026	1.84	[1.68; 2.02]	<0.0001
PGS001781	1.89	[1.73; 2.08]	<0.0001
PGS002277	1.37	[1.24; 1.51]	<0.0001
PGS001294	1.80	[1.64; 1.97]	<0.0001
PGS002354	1.99	[1.82; 2.18]	<0.0001
PGS002243	1.86	[1.70; 2.04]	<0.0001
PGS002733	1.47	[1.33; 1.62]	<0.0001
PGS002308	2.31	[2.12; 2.52]	<0.0001
PGS004868	1.91	[1.74; 2.09]	<0.0001
PGS003118	2.03	[1.85; 2.22]	<0.0001
PGS003353	1.64	[1.50; 1.81]	<0.0001
PGS002771	2.43	[2.23; 2.66]	<0.0001
PGS003402	1.38	[1.25; 1.53]	<0.0001
PGS002780	1.78	[1.63; 1.96]	<0.0001
PGS003751	1.51	[1.37; 1.66]	<0.0001
PGS003733	1.82	[1.66; 1.99]	<0.0001
PGS004226	1.47	[1.34; 1.62]	<0.0001
PGS003443	2.33	[2.13; 2.54]	<0.0001
PGS004223	1.55	[1.41; 1.70]	<0.0001
PGS004323	1.30	[1.17; 1.43]	<0.0001
PGS004185	1.70	[1.55; 1.86]	<0.0001
PGS004859	2.58	[2.36; 2.82]	<0.0001
PGS003867	2.30	[2.10; 2.51]	<0.0001
PGS004152	1.93	[1.76; 2.11]	<0.0001
PGS004569	1.70	[1.54; 1.87]	<0.0001
PGS004887	2.55	[2.34; 2.79]	<0.0001

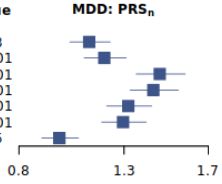
D

PRS	Odds Ratio	95% CI	P-Value
PGS000031	1.52	[1.38; 1.68]	<0.0001
PGS000020	1.55	[1.41; 1.71]	<0.0001
PGS000125	1.63	[1.49; 1.79]	<0.0001
PGS000014	1.73	[1.58; 1.90]	<0.0001
PGS000036	2.00	[1.83; 2.20]	<0.0001
PGS000330	2.22	[2.02; 2.42]	<0.0001
PGS000868	2.27	[2.08; 2.49]	<0.0001
PGS001357	2.27	[2.07; 2.48]	<0.0001
PGS000712	2.52	[2.31; 2.75]	<0.0001
PGS000805	2.55	[2.34; 2.79]	<0.0001
PGS004602	2.57	[2.36; 2.81]	<0.0001
PGS000832	2.49	[2.28; 2.72]	<0.0001
PGS000729	2.49	[2.28; 2.72]	<0.0001
PGS002026	2.52	[2.30; 2.75]	<0.0001
PGS001781	2.52	[2.30; 2.75]	<0.0001
PGS002277	2.52	[2.30; 2.75]	<0.0001
PGS001294	2.48	[2.27; 2.71]	<0.0001
PGS002354	2.46	[2.25; 2.69]	<0.0001
PGS002243	2.46	[2.25; 2.69]	<0.0001
PGS002733	2.46	[2.25; 2.69]	<0.0001
PGS002308	2.46	[2.25; 2.69]	<0.0001
PGS004868	2.46	[2.25; 2.69]	<0.0001
PGS003118	2.46	[2.25; 2.69]	<0.0001
PGS003353	2.46	[2.25; 2.69]	<0.0001
PGS002771	2.46	[2.25; 2.69]	<0.0001
PGS003402	2.46	[2.25; 2.69]	<0.0001
PGS002780	2.46	[2.25; 2.69]	<0.0001
PGS003751	2.46	[2.25; 2.69]	<0.0001
PGS003733	2.45	[2.24; 2.68]	<0.0001
PGS004226	2.45	[2.24; 2.68]	<0.0001
PGS003443	2.45	[2.24; 2.68]	<0.0001
PGS004223	2.45	[2.24; 2.68]	<0.0001
PGS004323	2.45	[2.24; 2.68]	<0.0001
PGS004185	2.45	[2.24; 2.68]	<0.0001
PGS004859	2.45	[2.24; 2.68]	<0.0001
PGS003867	2.55	[2.34; 2.79]	<0.0001
PGS004152	2.55	[2.34; 2.79]	<0.0001
PGS004569	2.67	[2.44; 2.91]	<0.0001
PGS004887	2.64	[2.42; 2.89]	<0.0001



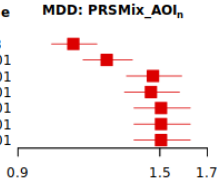
E

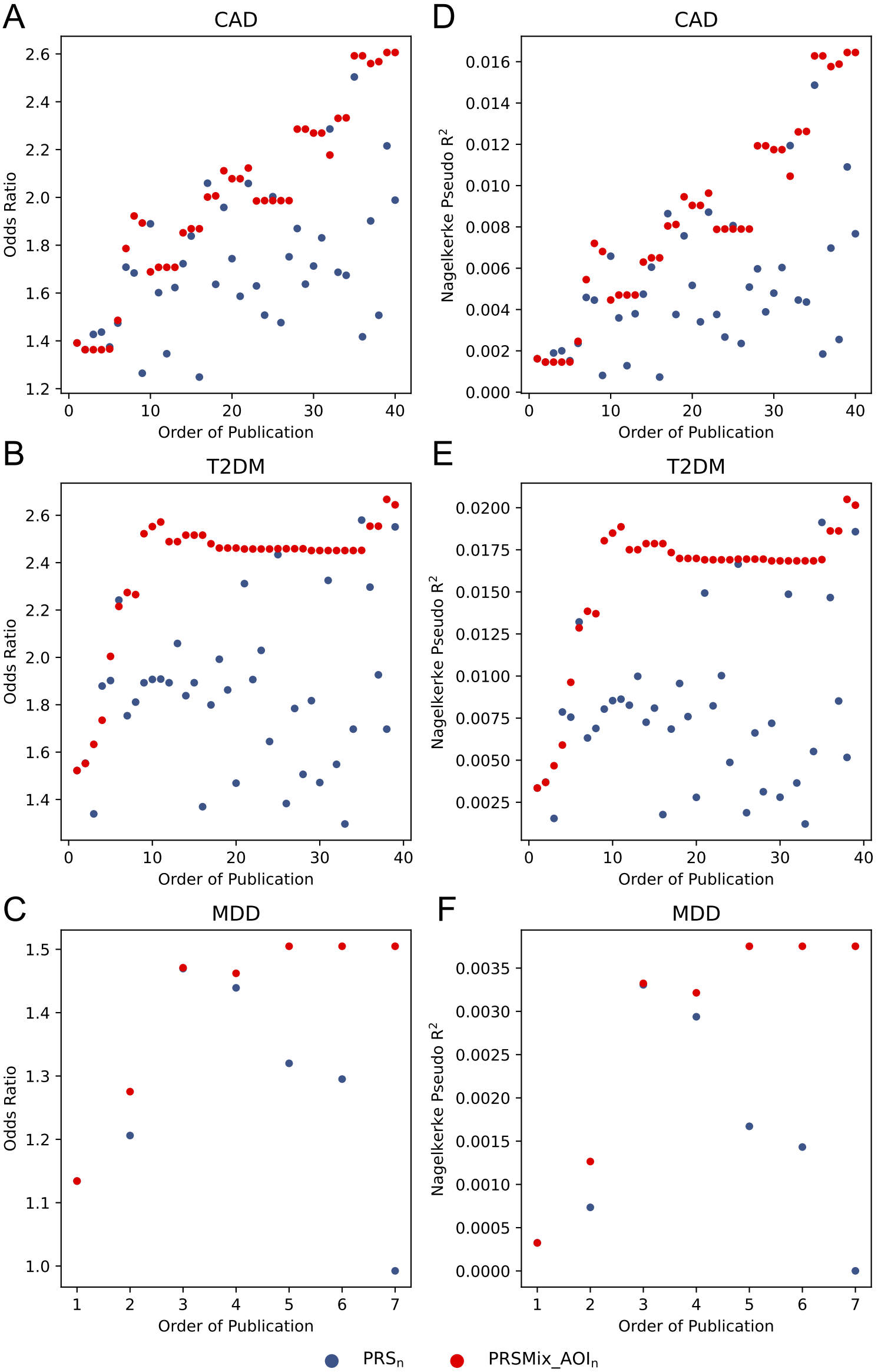
PRS	Odds Ratio	95% CI	P-Value
PGS000139	1.13	[1.04; 1.23]	0.003
PGS000193	1.21	[1.11; 1.31]	<0.0001
PGS000907	1.47	[1.36; 1.59]	<0.0001
PGS003333	1.44	[1.33; 1.56]	<0.0001
PGS003576	1.32	[1.22; 1.43]	<0.0001
PGS002789	1.30	[1.19; 1.41]	<0.0001
PGS003442	0.99	[0.91; 1.08]	0.865

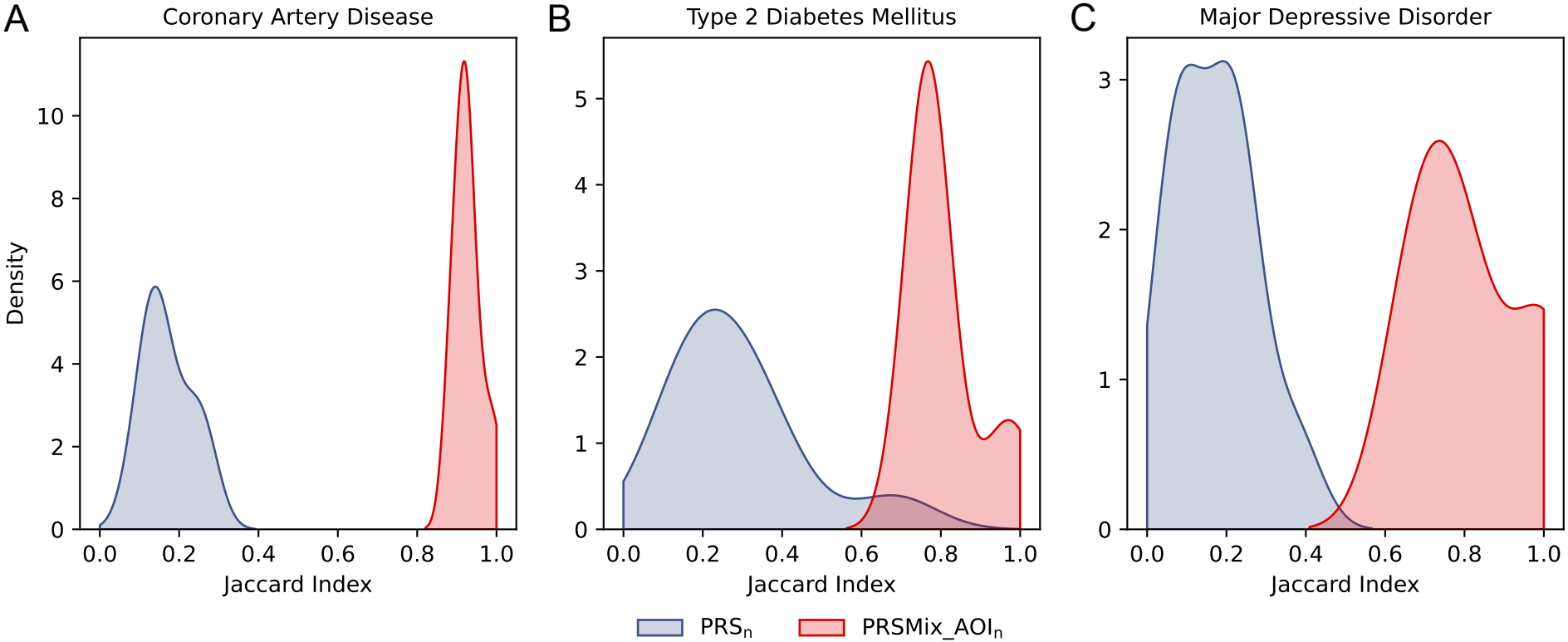


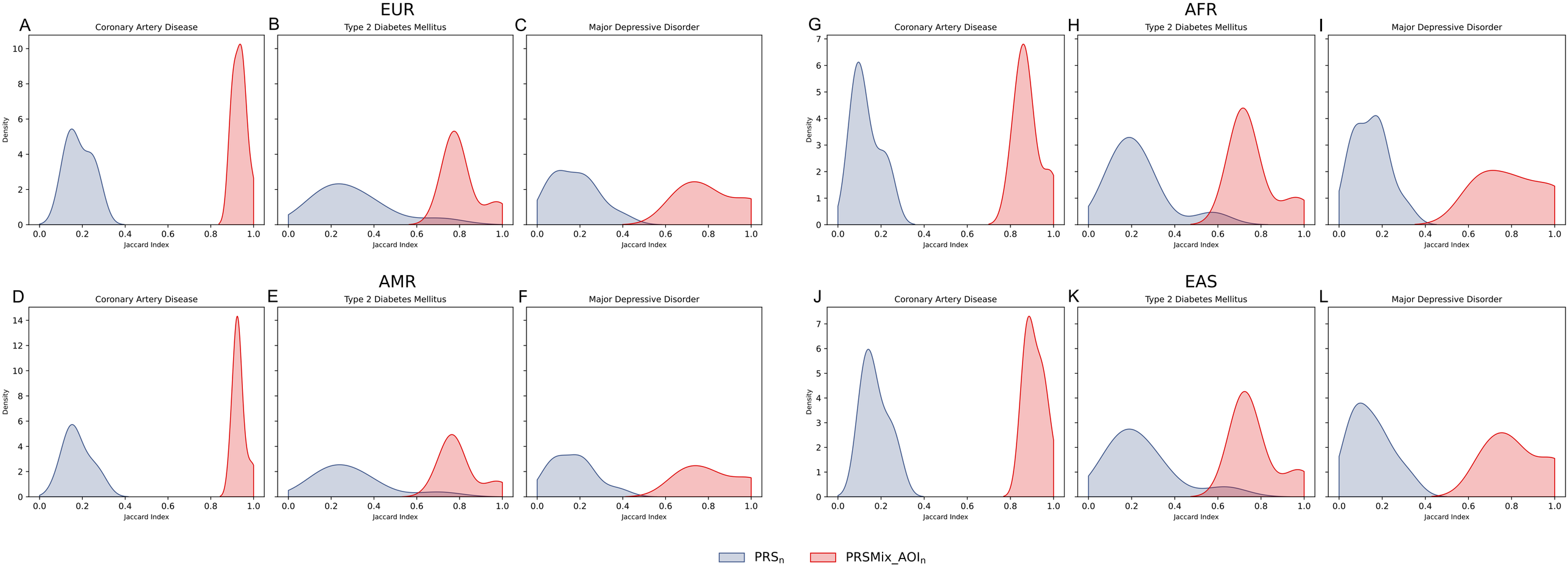
F

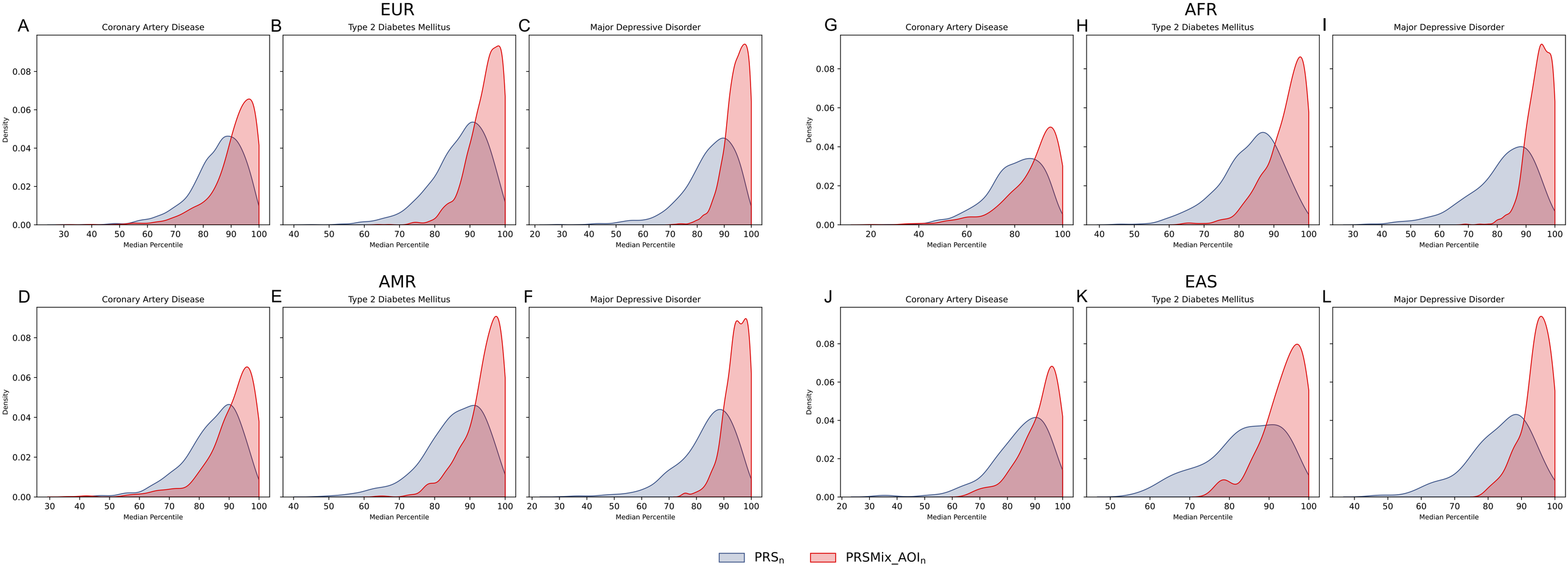
PRS	Odds Ratio	95% CI	P-Value
PGS000139	1.13	[1.04; 1.23]	0.003
PGS000193	1.28	[1.18; 1.38]	<0.0001
PGS000907	1.47	[1.36; 1.59]	<0.0001
PGS003333	1.46	[1.35; 1.58]	<0.0001
PGS003576	1.50	[1.39; 1.63]	<0.0001
PGS002789	1.50	[1.39; 1.63]	<0.0001
PGS003442	1.50	[1.39; 1.63]	<0.0001

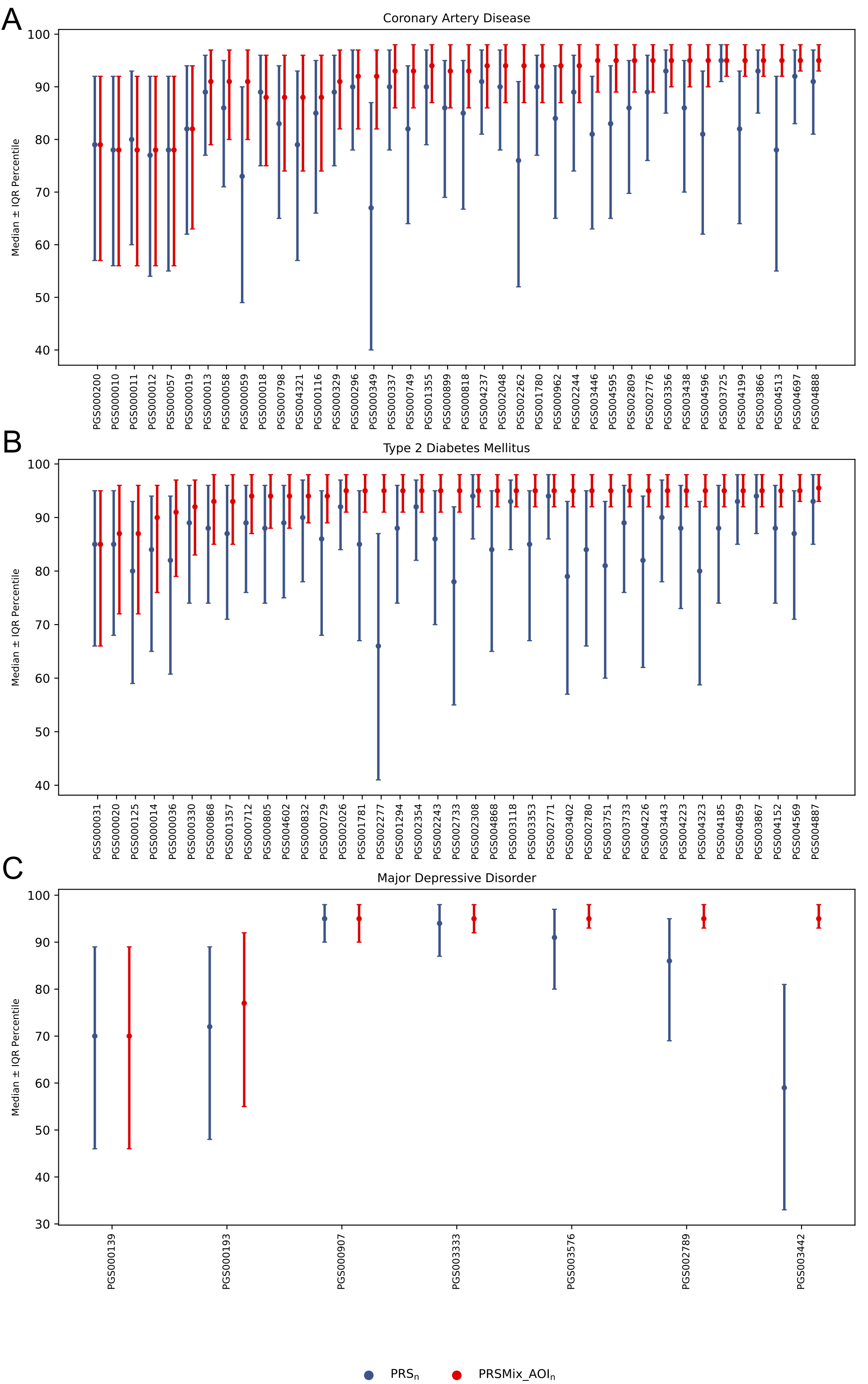



A
CAD
Odds Ratio
Order of Publication
D
CAD
Nagelkerke Pseudo R^2
Order of Publication
B
T2DM
Odds Ratio
Order of Publication
E
T2DM
Nagelkerke Pseudo R^2
Order of Publication
C
MDD
Odds Ratio
Order of Publication
F
MDD
Nagelkerke Pseudo R^2
Order of Publication
PRS_n
$PRSMix_AOI_n$


A
Coronary Artery Disease
B
Type 2 Diabetes Mellitus
C
Major Depressive Disorder
Density
Jaccard Index
PRS$_n$
PRSMix_AOI$_n$


EUR
AFR
AMR
EAS
A
Coronary Artery Disease
B
Type 2 Diabetes Mellitus
C
Major Depressive Disorder
G
Coronary Artery Disease
H
Type 2 Diabetes Mellitus
I
Major Depressive Disorder
D
Coronary Artery Disease
E
Type 2 Diabetes Mellitus
F
Major Depressive Disorder
J
Coronary Artery Disease
K
Type 2 Diabetes Mellitus
L
Major Depressive Disorder
Density
Jaccard Index
PRSn
PRSMix_AOIn


EUR
AFR
AMR
EAS
A
Coronary Artery Disease
B
Type 2 Diabetes Mellitus
C
Major Depressive Disorder
G
Coronary Artery Disease
H
Type 2 Diabetes Mellitus
I
Major Depressive Disorder
D
Coronary Artery Disease
E
Type 2 Diabetes Mellitus
F
Major Depressive Disorder
J
Coronary Artery Disease
K
Type 2 Diabetes Mellitus
L
Major Depressive Disorder
Density
Median Percentile
PRS$_n$
PRSMix_AOI$_n$


A
Coronary Artery Disease
Median ± IQR Percentile
100
90
80
70
60
50
40
PGS000200
PGS000010
PGS000011
PGS000012
PGS000057
PGS000019
PGS000013
PGS000058
PGS000059
PGS000018
PGS000798
PGS004321
PGS000116
PGS000329
PGS000296
PGS000349
PGS000337
PGS000749
PGS001355
PGS000899
PGS000818
PGS004237
PGS002048
PGS002262
PGS001780
PGS000962
PGS002244
PGS003446
PGS004595
PGS002809
PGS002776
PGS003356
PGS003438
PGS004596
PGS003725
PGS004199
PGS003866
PGS004513
PGS004697
PGS004888
B
Type 2 Diabetes Mellitus
Median ± IQR Percentile
100
90
80
70
60
50
40
PGS000031
PGS000020
PGS000125
PGS000014
PGS000036
PGS000330
PGS000868
PGS001357
PGS000712
PGS000805
PGS004602
PGS000832
PGS000729
PGS002026
PGS001781
PGS002277
PGS001294
PGS002354
PGS002243
PGS002733
PGS002308
PGS004868
PGS003118
PGS003353
PGS002771
PGS003402
PGS002780
PGS003751
PGS003733
PGS004226
PGS003443
PGS004223
PGS004323
PGS004185
PGS004859
PGS003867
PGS004152
PGS004569
PGS004887
C
Major Depressive Disorder
Median ± IQR Percentile
100
90
80
70
60
50
40
30
PGS000139
PGS000193
PGS000907
PGS003333
PGS003576
PGS002789
PGS003442
PRS$_n$
PRSMix_AOI$_n$