

1 An integrated polygenic score stratifies risk of peripheral artery disease and adverse  
2 limb events in ancestrally diverse cohorts

3 Short title: Integrated polygenic risk prediction for PAD

4  
5 Alyssa M. Flores, MD<sup>a-c\*</sup>; Yunfeng Ruan, PhD<sup>b,c\*</sup>; Anika Misra, BS<sup>b,c</sup>; Margaret S.  
6 Selvaraj, PhD<sup>b,c</sup>; Tiffany R. Bellomo, MD<sup>a-c</sup>; Tetsushi Nakao, MD, PhD<sup>b,c</sup>; Kenneth  
7 Rosenfield, MD<sup>d,e</sup>; Matthew Eagleton, MD<sup>a,e</sup>; Whitney Hornsby, PhD<sup>b,c</sup>; Aniruddh P.  
8 Patel, MD<sup>b-e#</sup>; Pradeep Natarajan, MD, MSc<sup>b-e#</sup>

9  
10 <sup>a</sup>Division of Vascular and Endovascular Surgery, Department of Surgery,  
11 Massachusetts General Hospital, Boston, USA. <sup>b</sup>Cardiovascular Disease Initiative,  
12 Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA. <sup>c</sup>Center for  
13 Genomic Medicine, Department of Medicine, Massachusetts General Hospital, Boston,  
14 Massachusetts, USA. <sup>d</sup>Harvard Medical School, Boston, Massachusetts, USA. <sup>e</sup>Division  
15 of Cardiology, Department of Medicine, Massachusetts General Hospital, Boston,  
16 Massachusetts, USA.

17 *\*These authors equally contributed to the work. #These authors jointly supervised the*  
18 *work.*

19  
20 **Address for Correspondence:**

21 Aniruddh P. Patel, MD and Pradeep Natarajan, MD, MSc  
22 185 Cambridge Street, CPZN 3.184  
23 Boston, MA 02114

1 Phone: 617-643-3451 | Fax: 617-726-2203

2 Email: [apatel32@mgh.harvard.edu](mailto:apatel32@mgh.harvard.edu) | Twitter: @AniruddhPatelMD

3 Email: [pnatarajan@mgh.harvard.edu](mailto:pnatarajan@mgh.harvard.edu) | Twitter: @pnatarajanmd

4

5

6

7

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

## 1 **Abbreviations**

- 2 All of Us, AoU.
- 3 Ankle-brachial index, ABI.
- 4 Body mass index, BMI.
- 5 Coronary artery disease, CAD.
- 6 Current Procedural Terminology, CPT.
- 7 Diastolic blood pressure, DBP.
- 8 Electronic health record, EHR.
- 9 Genome-wide association study, GWAS.
- 10 Glomerular filtration rate, GFR.
- 11 High-density lipoprotein cholesterol, HDL-C.
- 12 International Classification of Disease, ICD.
- 13 Least Absolute Shrinkage and Selection Operator, LASSO.
- 14 Low-density lipoprotein cholesterol, LDL-C.
- 15 Major adverse limb events, MALE.
- 16 Mass General Brigham Biobank, MGBB.
- 17 Office of Population Censuses and Surveys, OPCS.
- 18 Peripheral artery disease, PAD.
- 19 Polygenic risk scores, PRS.
- 20 Principal components, PCs.
- 21 Research Patient Data Registry, RPDR.
- 22 Systolic blood pressure, SBP.
- 23 United States Preventive Services Task Force, USPSTF.

## 1 **Abstract**

2 Background and Aims: Peripheral artery disease (PAD) is a heritable atherosclerotic  
3 condition that is underdiagnosed and undertreated. With growing knowledge of the  
4 genetic basis for PAD and related risk factors, this study sought to construct a new  
5 polygenic score for PAD (GPS<sub>PAD</sub>).

6  
7 Methods: GPS<sub>PAD</sub> was constructed by integrating multi-ancestry summary statistics for  
8 PAD and related traits. GPS<sub>PAD</sub> was trained in a UK Biobank dataset of 96,239  
9 individuals and validated in a holdout UK Biobank dataset (N=304,294) and All of Us  
10 (AoU; N=237,173) and Mass General Brigham Biobank (MGBB, N=37,017).

11  
12 Results: GPS<sub>PAD</sub> was associated with an OR-per SD increase of 1.64 in the UK Biobank  
13 dataset (95% CI 1.60-1.68). Compared to previously published PAD polygenic scores,  
14 GPS<sub>PAD</sub> was more strongly associated with PAD in AoU and MGBB, including enhanced  
15 transferability to non-European subgroups. GPS<sub>PAD</sub> improved discrimination of incident  
16 PAD ( $\Delta$ C-statistic 0.030) that was nearly equivalent to the additive performances of  
17 diabetes ( $\Delta$ C-statistic 0.029) and smoking ( $\Delta$ C-statistic 0.034). GPS<sub>PAD</sub> was associated  
18 with reduced ankle-brachial index in the MGBB with the top 8% of individuals having a  
19 mean ABI  $\leq$ 0.90 when assessed. Among individuals with prevalent PAD, GPS<sub>PAD</sub>  
20 consistently identified individuals at high MALE-risk in the UK Biobank (HR 1.48; 95%  
21 CI 1.24-1.77), MGBB, (HR 1.34; 95% CI 1.12-1.60), and AoU (HR 1.33; 95% CI 1.12-  
22 1.58).

23

1 Conclusions: An integrated, multi-ancestry polygenic score for PAD predicts disease  
2 and adverse limb outcomes in three diverse cohorts. Incorporating polygenic risk into  
3 PAD care has the potential to guide screening and tailor management to prevent MALE.

4

## 5 **Key Words**

6 Peripheral artery disease, polygenic risk, precision medicine

7

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

## 1 Introduction

2 Peripheral artery disease (PAD) is an atherosclerotic vascular condition that  
3 affects a global population of 230 million adults with high resource utilization owing to  
4 both systemic and limb ischemic events.<sup>1,2</sup> While PAD shares risk factors with coronary  
5 artery disease (CAD), 32-54% of individuals presenting with PAD do not have clinically  
6 significant coronary or cerebrovascular disease.<sup>3,4</sup> In addition, there are differences in  
7 the major etiologies of acute events in CAD and PAD. Ischemic CAD events most  
8 commonly result from plaque rupture, whereas the cause of acute limb ischemia is most  
9 commonly owing to embolism or *in situ* thrombosis, regardless of atherosclerosis  
10 extent.<sup>5</sup> Indeed, the increasingly recognized role of thrombosis in PAD is also supported  
11 by the discovery of genetic variants in coagulation factors, including mutations in Factor  
12 V Leiden, which have been uniquely associated with PAD and not CAD.<sup>6</sup>

13 Clinical subsets of PAD include asymptomatic, claudication, and chronic-limb  
14 threatening ischemia (CLTI) resulting in tissue loss and major adverse limb events  
15 (MALE). MALE are a devastating complication and are often associated with critical  
16 illness, numerous resource-intensive attempts at revascularization, and prolonged  
17 hospital stays as part of the limb salvage effort. Despite its high morbidity, PAD is  
18 grossly underdiagnosed and has a lack of consensus on screening indications. For  
19 example, while European and American guidelines recommend consideration of  
20 screening based on age and risk factors, the United States Preventive Services Task  
21 Force (USPSTF) does not recommend screening regardless of risk factors.<sup>7-9</sup>  
22 Conflicting recommendations have implications on care implementation, for example, in  
23 the U.S. where preventive service coverage is guided by USPSTF recommendations.<sup>10</sup>

1 In addition, there is no standard tool to predict complications before advanced disease  
2 develops. Such a tool would be useful as there are contemporary antithrombotic  
3 treatments targeted to prevent MALE but remain underutilized in PAD management.<sup>11</sup>

4 Polygenic risk scores (PRS) provide a quantitative metric for the inherited  
5 component of disease risk by integrating common genetic variants discovered from  
6 genome-wide association studies (GWAS) into a single instrument. As PAD has an  
7 estimated heritability of 20-50%<sup>12,13</sup>, incorporation of genetic factors offers an  
8 opportunity to optimize risk stratification. While there have been advances in PRS for  
9 CAD, there has been limited progress thus far in developments for PAD owing to largely  
10 European-based data, lack of validation in external datasets, and unclear transferability  
11 across diverse ancestry groups.<sup>14-16</sup> Additionally, PRS utility for incident prediction of  
12 adverse PAD events has not been described.

13 To address these needs, in the present study, we develop a new genome-wide  
14 polygenic score for PAD ( $GPS_{PAD}$ ) that incorporates multi-ancestry GWAS data for PAD  
15 and related traits from over 2 million individuals (**Figure 1**). We assess the performance  
16 of  $GPS_{PAD}$  in predicting PAD among individuals of diverse ancestry in a 304,294 internal  
17 validation cohort and two independent study populations comprising 259,627  
18 individuals. Lastly, we apply  $GPS_{PAD}$  to predict MALE and identify individuals with  
19 clinically important increased risk.

20

## 21 **Methods**

### 22 Study populations

1 GPS<sub>PAD</sub> was developed using data from the UK Biobank, a longitudinal cohort study of  
2 approximately 500,000 individuals from the United Kingdom aged 40-69 years.<sup>17</sup>  
3 Participants were followed for outcomes based on International Classification of  
4 Disease 9<sup>th</sup> and 10<sup>th</sup> revisions (ICD-9/ICD-10) and Office of Population Censuses and  
5 Surveys versions 3 and 4 (OPCS-3/OPCS-4). Ancestry was based on self-identified  
6 ethnicity and country of origin with African, East Asian, European, Latino, Middle  
7 Eastern/North African (MENA), and South Asian categories (Supplemental Tables 1-2).  
8 Coding of comorbidities and lifestyle factors is detailed in the Supplemental Materials  
9 (Supplemental Table 3).

10

11 GPS<sub>PAD</sub> performance was evaluated in two external datasets: Mass General Brigham  
12 Biobank (MGBB) and All of Us (AoU). The MGBB is a New England-healthcare based  
13 cohort that includes >145,000 patients treated at seven regional hospitals and clinics,  
14 including ~56,000 have been genotyped.<sup>18</sup> Baseline phenotypes are linked to the  
15 electronic health record (EHR) with available data from clinical notes, ICD-9/ICD-10,  
16 and Current Procedural Terminology (CPT) codes.

17

18 The AoU Research program is a United States-based cohort study that aims to recruit  
19 individuals who have been historically underrepresented in biomedical research.<sup>19</sup> The  
20 program has recruited ~400,000 individuals and includes health questionnaires,  
21 biospecimen collection, and longitudinal EHR data.

22

23 Genetic data and quality control



1 In the UK Biobank, individuals were genotyped using UK BiLEVE Axiom Array or UK  
2 Biobank Axiom Array and centrally imputed to the 1000 Genomes (1000G) Panel,  
3 Haplotype Reference Consortium, or UK10K Panel.<sup>17,20</sup> After performing quality control  
4 (Details Methods) and excluding individuals of Latino ancestry due to low population  
5 representation (N=11 PAD cases) and unreported/mixed ethnicity, 304,294 individuals  
6 were included for internal validation.

7  
8 MGBB samples were genotyped on Illumina Multi-Ethnic Genotyping Array or Global  
9 Screening Array.<sup>21</sup> Imputation was performed to the multi-ancestry TOPMED r2  
10 reference panel. Given lack of detailed self-reported ethnicity or ancestry data, ancestry  
11 was genetically predicted using a K-nearest neighbor model trained with principal  
12 components (PCs) from the 1000G reference panels for European, African, Latino/Ad  
13 Mixed American, East Asian, and South Asian (Detailed Methods).<sup>18</sup> We excluded  
14 individuals who were in prior discovery GWAS for CAD<sup>22</sup> and did not map to a single  
15 genetic ancestry, leaving 37,017 individuals for external validation.

16  
17 AoU participants were genotyped using the Illumina Global Diversity Array at AoU  
18 genome centers.<sup>23</sup> Similar to the MGBB, there was a lack of detailed reported ancestry  
19 data, thus ancestry was assigned based on genetic similarity. AoU assigns categorical  
20 ancestries to African, Latino/Ad Mixed American, East Asian, South Asian, European,  
21 MENA, and Other based on a random forest classifier trained using gnomAD, Human  
22 Genome Diversity Project, and 1000G reference labels (Detailed Methods).<sup>23</sup> After  
23 exclusion of individuals with missing/other ancestry, 237,173 individuals were included.

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22

Clinical endpoints

In the UK Biobank, PAD was defined based on self-reported history, ICD codes, OPCS codes for lower extremity revascularization or major amputation, and cilostazol prescription (Supplemental Table 4). In AoU, PAD was defined on self-report, occurrences of  $\geq 2$  ICD codes, or a single procedure code for revascularization, major amputation or supervised exercise therapy for PAD (Supplemental Table 5).<sup>24</sup>

In the MGBB, PAD was derived from phenotypes developed by the Mass General Brigham Research Patient Data Registry (RPDR) based on structured and unstructured clinical data from the EHR (Detailed Methods).<sup>25</sup> ABI were extracted from imaging reports retrieved from the RPDR. We selected each individual's minimum ABI and excluded values that were  $>1.4$  or non-compressible.

MALE was defined as a surrogate of major amputation and acute limb ischemia based on diagnosis and procedure codes (Supplemental Tables 6-8). Revascularization included thrombectomy, thrombolysis, and emergency lower extremity bypass. In the MGBB, CPT codes for catheter-directed thrombolysis/thrombectomy required concomitant coding for aortogram or lower extremity angiography to ensure arterial intervention.

GPS<sub>PAD</sub> Construction

1 PRS predictive performance is improved by the incorporation of data from increasingly  
2 diverse genetic ancestries and consideration of genetically correlated traits.<sup>26-28</sup> To  
3 leverage the common mechanistic pathways of PAD, other atherosclerotic conditions,  
4 and PAD risk factors, GWAS results of PAD and 14 candidate traits were considered in  
5 GPS<sub>PAD</sub> construction: PAD, CAD, ischemic stroke, glomerular filtration rate, diabetes  
6 mellitus, smoking, systolic blood pressure (SBP), diastolic blood pressure (DBP), low-  
7 density lipoprotein cholesterol (LDL-C), total cholesterol, high-density lipoprotein  
8 cholesterol (HDL-C), triglycerides, BMI, carotid plaque burden, and carotid intima-media  
9 thickness (Supplemental Table 9). After collection of GWAS summary statistics, GPS<sub>PAD</sub>  
10 was trained using target data from 96,239 European individuals in the UK Biobank.  
11 GWAS results from the UK Biobank were not used to construct GPS<sub>PAD</sub> such that data  
12 for score development and training were non-overlapping.  
13  
14 GPS<sub>PAD</sub> was developed in a two-layer process (**Figure 1**).<sup>26</sup> Layer 1 involved using  
15 ancestry-stratified GWAS data for each trait to calculate multi-ancestry polygenic scores  
16 that were optimized according to their PAD predictive performance. Separate scores  
17 were constructed using LDpred2, a widely used method that adjusts marginal single  
18 nucleotide polymorphism effect sizes for linkage disequilibrium (LD) patterns and  
19 selects a subset of variants with non-zero effects to calculate the polygenic score.<sup>29</sup>  
20 Using LDpred2-auto, scores were calculated using GWAS results stratified by African,  
21 East Asian, European, Latino, and South Asian ancestry (R bigsnpr v.11.4). This  
22 resulted in 100 scores across all ancestries and traits. For each trait, the ancestry-  
23 specific scores were combined into the best-performing multi-ancestry score per trait

1 using Least Absolute Shrinkage and Selection Operator (LASSO) regression in models  
2 predicting PAD (R glmnet v4.0-2). Feature selection was performed iteratively for all 15  
3 traits in layer 1, yielding 15 multi-ancestry trait-specific scores with mixing weights  
4 detailed in Supplemental Table 9.

5  
6 Layer 2 involved combining trait-specific scores from layer 1 to derive the integrated,  
7 multi-trait  $GPS_{PAD}$ . Trait scores from Layer 1 were input into a LASSO regression model  
8 predicting PAD to construct the final  $GPS_{PAD}$ . The final weight in  $GPS_{PAD}$  was derived by  
9 calculating the proportional weight from layer 1 in layer 2, normalized to 100%. Of the  
10 15 candidate trait scores, 11 traits contributed to  $GPS_{PAD}$  with mixing weights in  
11 Supplemental Table 10.  $GPS_{PAD}$  was then calculated in the validation UK Biobank  
12 dataset. See Detailed Methods for details on  $GPS_{PAD}$  construction.

#### 13 14 $GPS_{PAD}$ validation and benchmarking

15  $GPS_{PAD}$  was calculated in the MGBB and AoU for external validation.  $GPS_{PAD}$   
16 performance was compared with previously published PRS in the Polygenic Score  
17 Catalog<sup>30</sup> and a recently reported multi-ancestry polygenic score for CAD ( $GPS_{CAD}$ ).<sup>26</sup>  
18 We previously developed  $GPS_{CAD}$  utilizing a similar two-layer framework for CAD as in  
19 the present study with slight modifications including stepwise regression for feature  
20 selection.<sup>26</sup> From the Polygenic Score Catalog,  $PRS_{PLR}$  is a single-trait score for PAD  
21 developed using LASSO regression and individual-level European data from the UK  
22 Biobank.<sup>15</sup>  $PRS_{LDpred2}$  is another single-trait score for PAD trained using UK Biobank  
23 data, but was calculated with LDpred2-auto.<sup>15</sup>

1 We also calculated the lifetime risk of PAD in individuals in the UK Biobank based on  
2 clinical variables using the Johns Hopkins University PAD risk tool  
3 (<http://ckdpcrisk.org/padrisk>, Detailed Methods).<sup>31</sup>

4

#### 5 Statistical analysis

6 The unadjusted rate of PAD was calculated across percentiles of GPS<sub>PAD</sub>. Model  
7 calibration was assessed with the Hosmer–Lemeshow test (R ResourceSelection v0.3-  
8 6) and by comparing the observed and predicted prevalence across percentiles  
9 calculated using a logistic regression model with only GPS<sub>PAD</sub> as a predictor. We  
10 estimated PAD risk in the extremes of the GPS<sub>PAD</sub> distribution using logistic regression  
11 models. We also derived the proportion of the population with a given magnitude of risk  
12 by calculating the odds ratio (OR) of varying extremes of GPS<sub>PAD</sub> percentiles compared  
13 to the middle quintile group (40-59%).

14

15 The association of GPS<sub>PAD</sub> and other scores with all PAD (incident and prevalent cases)  
16 were assessed using logistic regression. Performance metrics included OR, area under  
17 the receiver operator characteristic curve (AUC, R pROC v1.17.0.1), and phenotypic  
18 variance explained (Nagelkerke-R<sup>2</sup>). Incident event analyses were performed using Cox  
19 proportional hazards models with metrics including hazard ratio (HR) and C-statistic (R  
20 survival v3.5-7). MALE analyses were restricted to individuals with prevalent PAD  
21 diagnoses. Time-to-MALE curves were estimated using the Kaplan-Meier method,  
22 standardized to mean age and gender (R survminer v0.4.9). Linear regression was  
23 used to evaluate the relationship between polygenic scores and minimum ABI.

1  
2 Logistic, linear, and Cox models including polygenic scores were adjusted for age, sex,  
3 genotyping array, and the first ten PCs. Cox models were used to estimate the 10-year  
4 incidence of PAD across lifestyle and genetic risk groups, standardized to the mean of  
5 covariates in each group. In the UK Biobank, we performed sensitivity analyses to test  
6 the strength of  $GPS_{PAD}$  associations with PAD and incident MALE after additionally  
7 accounting for clinical variables (current smoking, hypertension, diabetes,  
8 hyperlipidemia, and chronic kidney disease); and continuous variables for individuals  
9 with available physical and laboratory data (SBP, DBP, hemoglobin A1c, HDL-C,  
10 estimated untreated LDL-C<sup>32</sup>, GFR, and anti-hypertensive use).

11  
12  $GPS_{PAD}$  and other continuous variables were scaled to a mean of zero and one  
13 standard deviation (SD) such that effect sizes indicate OR- or HR-per SD. Statistical  
14 significance was defined as  $P < 0.05$  or 95% confidence interval (CI) that excluded the  
15 null value. Statistical analyses were performed using R-4.1.0.

## 16 17 **Results:**

18 Given that complex diseases such as PAD share genetic backgrounds with related  
19 traits, we leveraged this genetic correlation to more fully capture human genetic  
20 architecture and construct  $GPS_{PAD}$ .<sup>6,33,34</sup>  $GPS_{PAD}$  included 603,595 variants across 11  
21 traits and 5 ancestries, including PAD, smoking, CAD, ischemic stroke, diabetes, SBP,  
22 LDL-C, glomerular filtration rate, BMI, carotid IMT, and total cholesterol. While  $GPS_{PAD}$   
23 was predominantly derived from European discovery data, the score incorporated

1 genetic variation discovered in African, East Asian, South Asian, and Latino populations  
2 (Supplemental Table 10). Of the non-European populations, African GWAS discovery  
3 data contributed the most to  $GPS_{PAD}$ .

4

#### 5 Association of $GPS_{PAD}$ with PAD risk

6 Within the UK Biobank,  $GPS_{PAD}$  was associated with a OR 1.77 (95% CI: 1.70-1.86) for  
7 PAD in the European training sample (Supplemental Table 12). The effect size was  
8 mildly attenuated in the multiethnic validation cohort including 304,294 individuals, but  
9 remained strongly associated with PAD (OR 1.63; 95% CI 1.60-1.68). The holdout  
10 validation cohort included 164,108 females (53.9%) and 286,356 participants of  
11 European (94%), 7,680 South Asian (2.5%), 6,939 African (2.3%), 1,761 East Asian  
12 (0.6%), and 1,558 (0.51%) MENA ancestry. There was a decrement, yet persistently  
13 significant association with adjustment for clinical risk factors (OR 1.35; 95% CI 1.31-  
14 1.39) and in a model adjusted for continuous covariates (OR 1.37; 95% CI: 1.33-1.41).

15

16 We found significant differences in PAD rates across the  $GPS_{PAD}$  percentile distribution,  
17 ranging from 0.78% in the bottom percentile to 7.91% in the top percentile (**Figure 2A**,  
18 Supplemental Table 13). Predicted PAD prevalence was overall consistent with  
19 observed prevalence, excluding the  $\geq 98\%$  percentile where there was slight risk  
20 underestimation by  $GPS_{PAD}$  alone driven by a sharp increase in the observed PAD  
21 prevalence (Supplemental Figure 1).  $GPS_{PAD}$  stratified individuals in low and high-risk  
22 groups with up to 4.71-fold increased PAD risk in the top 1% of  $GPS_{PAD}$  (OR 4.71, 95%  
23 CI 4.07-5.42) compared to the middle quintile (**Figure 2B**).

1  
2 Compared to previously published scores,  $GPS_{PAD}$  resulted in an improvement in effect  
3 estimate, discrimination, and proportion of phenotypic variance explained compared to  
4  $PRS_{PLR}$ . In a model adjusted for age, sex, genotyping array, and the first ten PCs of  
5 ancestry,  $GPS_{PAD}$  had an OR of 1.63, AUC of 0.74, and Nagelkerke- $R^2$  of 0.087, while  
6  $PRS_{PLR}$  had an OR of 1.20, AUC 0.71, and Nagelkerke- $R^2$  of 0.064.  $PRS_{LDpred2}$  achieved  
7 greater performance in the UK Biobank than  $GPS_{PAD}$  with OR 2.46 and nearly  
8 equivalent discriminative ability with AUC 0.75, however Nagelkerke- $R^2$  and effect sizes  
9 in the extremes of  $PRS_{LDpred2}$  percentiles strongly suggested overfitting. For example,  
10  $PRS_{LDpred2}$  Nagelkerke- $R^2$  was 0.17 in the UK Biobank validation cohort and OR for the  
11 top 1%  $PRS_{LDpred2}$  was 71.9 (95% CI 64.9-79.7).

12  
13  $GPS_{PAD}$  identified individuals with distinct risk compared to an integrated polygenic  
14 score for CAD ( $GPS_{CAD}$ ).<sup>26</sup> In the UK Biobank,  $GPS_{PAD}$  was more strongly associated  
15 with PAD (OR 1.63; 95% CI 1.60-1.68) than  $GPS_{CAD}$  (OR 1.48; 95% CI 1.44-1.52) and  
16 identified a greater number of high-risk individuals (Supplemental Table 13). For  
17 example, the proportion of the UK Biobank population at 3-fold and 4-fold greater odds  
18 for PAD was 5.7% (N=17,345) and 1.90% (N=5,782) when classified by  $GPS_{PAD}$ ,  
19 respectively. There was a comparatively small fraction of individuals at similar PAD risk  
20 according to  $GPS_{CAD}$  with only 1.90% (N=5,782) at 3-fold greater risk and 0.20%  
21 (N=609) at 4-fold greater risk.

22



1 We then compared polygenic score performance in ancestral subgroups in the UK  
2 Biobank (**Figure 2C**). Across each ancestral population,  $GPS_{PAD}$  predicted disease risk  
3 in African, European, East Asian, MENA, and South Asian subgroups with effect  
4 estimates of 1.26, 1.66, 1.60, 1.64, and 1.25 respectively. While  $PRS_{LDpred2}$  and  $PRS_{PLR}$   
5 predicted PAD risk in European individuals, both were not associated with PAD in non-  
6 European subgroups (Supplemental Table 14). In clinical subgroups,  $GPS_{PAD}$   
7 performance was fairly stable, with OR ranges of 1.33-1.72 depending on demographics  
8 and the presence of risk factors (Supplemental Figure 2).

9

#### 10 Validation of $GPS_{PAD}$ in distinct external cohorts

11  $GPS_{PAD}$  was externally validated in two independent multi-ancestry cohorts and showed  
12 improved cross-population predictive performance relative to previously published  
13 scores.  $GPS_{PAD}$  was significantly associated with an increased risk for PAD (OR 1.26;  
14 95% CI 1.18-1.37, **Figure 3A**) in the MGBB, which included 5.3% African (N=1,945),  
15 87% European (N=32,139), and 7.9% Latino individuals (N=2,933).  $GPS_{PAD}$  performed  
16 similarly in AoU (OR 1.21; 95% CI 1.19-1.23), which embodied 22% African (N=51,691),  
17 2.2% East Asian (N=5,268), 50% European (N=119,167), 17% Latino (N=39,576),  
18 0.21% MENA (N=509), and 0.97% South Asian individuals (N=2,289). Of note, while  
19  $PRS_{LDpred2}$  outperformed  $GPS_{PAD}$  in the UK Biobank where it was derived,  $PRS_{LDpred2}$   
20 was not significantly associated with PAD risk in the MGBB (OR 1.02; 95% CI 0.95-  
21 1.08;  $P=0.06$ ) and was more weakly associated with PAD than  $GPS_{PAD}$  in AoU (OR  
22 1.10; 95% CI 1.08-1.12).  $GPS_{PAD}$  consistently had stronger effect sizes than  $PRS_{PLR}$  in  
23 each external cohort (**Figure 3A**).

1  
2 When stratifying by ancestry, GPS<sub>PAD</sub> demonstrated improved transferability in most  
3 subgroups (Supplemental Table 14). Among African (N=51,691), European  
4 (N=119,167), Latino (N=39,576), and South Asian (N=2,289) individuals in AoU,  
5 GPS<sub>PAD</sub> was the most strongly predictive of PAD compared to previously published  
6 scores (**Figure 3B**). GPS<sub>PAD</sub> was also transferable to Europeans in the MGBB with a  
7 larger effect size (OR 1.26; 95% CI 1.17-1.34) than PRS<sub>PLR</sub> (OR 1.09; 95% CI 1.02-  
8 1.16) and PRS<sub>LDpred2</sub>, which was not associated with disease risk (OR 1.01; 95% CI  
9 0.94-1.08). No polygenic scores were associated with PAD in AoU East Asians, AoU  
10 MENA, and non-European subgroups in the MGBB (Supplemental Figure 3).

11

#### 12 Modeling incident PAD according to polygenic risk and clinical risk factors

13 We then analyzed 301,932 participants of the UK Biobank who were free of PAD at  
14 study enrollment. Over a median 12.1 years follow-up (interquartile range (IQR) 10.6-  
15 13.5 years), incident PAD was observed in 4,202 individuals (1.39%). Within the UK  
16 Biobank, the baseline model of age, sex, and ten PCs of ancestry had a C-statistic of  
17 0.731 (**Figure 4A**). The addition of GPS<sub>PAD</sub> resulted in improvement in discrimination to  
18 0.761, which was greater than the improvement of most risk factors also in isolation,  
19 and was approximately equivalent to the additive benefit of diabetes (C-statistic 0.760)  
20 and smoking (C-statistic 0.765). The greatest improvement in discrimination was  
21 observed with the baseline model, clinical variables, and GPS<sub>PAD</sub> (C-statistic 0.087).

22

1 Compared to the individual performance of other scores, GPS<sub>PAD</sub> resulted in greater  
2 benefit in discrimination of incident PAD than PRS<sub>PLR</sub> (C-statistic 0.735) and the lifetime  
3 PAD risk score<sup>31</sup> (C-statistic 0.675). PRS<sub>LDpred2</sub> achieved a larger improvement in C-  
4 statistic (C-statistic 0.774) from the baseline model compared to GPS<sub>PAD</sub>, but the HR  
5 was drastically large in the UK Biobank where it was wholly derived, and consistent with  
6 overfitting (HR 199; 95% CI 172-231), as observed in cross-sectional analyses.

7  
8 In the baseline model, GPS<sub>PAD</sub> HR for incident PAD was 1.66 (95% CI 1.61-1.71) and  
9 HR 1.37 (95% CI 1.30-1.46) in a multivariable model adjusted for clinical risk factors.  
10 When examining the interaction between GPS<sub>PAD</sub> and clinical variables, we observed a  
11 weak interaction with levels of LDL-C (GPS<sub>PAD</sub>\*LDL-C HR 1.05; 95% CI 1.01-1.08;  
12  $P_{interaction}=0.005$ ) and hemoglobin A1c (GPS<sub>PAD</sub>\*hemoglobin A1c HR 1.02; 95% CI 1.00-  
13 1.03;  $P_{interaction}=0.01$ ). Lifestyle factors influenced PAD risk across each stratum of  
14 genetic risk (Supplemental Table 15). For the high genetic risk group, the standardized  
15 10-year incidence of PAD of 3.17% for smokers and decreased to 1.05% for those who  
16 refrained from smoking (**Figure 4A**). There were similar patterns of an offset of genetic  
17 risk by modifiable lifestyle factors according to dietary patterns (**Figure 4B**).

#### 18 19 Integration of polygenic risk with clinical guidelines

20 We then assessed whether GPS<sub>PAD</sub> could identify individuals at equivalent risk for  
21 incident disease as the high-risk groups that current guidelines recommend screening  
22 for PAD.<sup>8</sup> We first quantified the 10-year incidence of PAD among subgroups of  
23 individuals proposed to be candidates for ABI screening according to the AHA/ACC

1 Guidelines.<sup>8</sup> Individuals above 65y had a 10-year incidence of 2.28% (95% CI 2.16-  
2 2.41), individuals who were age 50-64y with one risk factor for atherosclerosis had an  
3 incidence of 1.23% (95% CI 1.17-1.30%), individuals <50y with diabetes and 1  
4 additional risk factor had an incidence of 2.90% (95% CI 1.62-3.21%), and those with  
5 known atherosclerosis in another vascular bed had the greatest incidence of 5.95%  
6 (95% CI 5.48-6.42). As a comparison, incidence rates ranged from 2.81-4.40% in the  
7 top 1-5% GPS<sub>PAD</sub> compared to the middle quintile distribution (Supplemental Table 16).  
8 We also looked at age-based combinations of each individual risk factor and compared  
9 event rates to subgroups with the combination of risk factors and high polygenic risk.  
10 This analysis revealed that the combination of high GPS<sub>PAD</sub> with one atherosclerosis  
11 risk factor identified more individuals who subsequently developed PAD than many of  
12 the subgroups defined by the combination of similar risk factors with age (**Figure 5A**).  
13 For example, 7.96% of individuals in the top 5% GPS<sub>PAD</sub> and diabetes developed PAD,  
14 as compared to a 4.02% incident rate among individuals aged 50-64y with diabetes.  
15 Similarly, there was a relative increase in PAD cases among individuals in the top 5%  
16 GPS<sub>PAD</sub> with dyslipidemia (5.63%; 95% CI 4.88-6.36) compared to individuals 50-64y  
17 with dyslipidemia (1.97%; 95% CI 1.80-2.15). The addition of the top 5% GPS<sub>PAD</sub> also  
18 increased the proportion of additional cases of PAD when combined with other  
19 atherosclerotic conditions such as CAD and carotid stenosis compared to consideration  
20 of the conditions alone (Supplemental Table 16).

21

22 Predicting ankle-brachial indices based on polygenic risk

1 In addition to predicting binary PAD presence, we evaluated whether  $GPS_{PAD}$  was  
2 associated with minimum ABI for a subset of individuals in the MGBB (N=883,  
3 Supplemental Figure 4). The minimum ABI was first used to validate the PAD  
4 phenotype in the MGBB. PAD cases indeed had a lower mean ABI compared to  
5 controls (0.78 vs. 1.10;  $P < 2 \times 10^{-16}$ , Supplemental Figure 5). After adjusting for age, sex,  
6 and the first ten PCs,  $GPS_{PAD}$  was significantly associated with lower ABI (**Figure 5B**),  
7 while previously published scores were not (all  $P > 0.1$ , Supplemental Table 17). Each  
8 SD increase in  $GPS_{PAD}$  was associated with an -0.088 reduction in ABI (95% CI -0.161  
9 to -0.015;  $P = 0.019$ ). Based on the percentile distribution in the MGBB, the group in the  
10 top 8%  $GPS_{PAD}$  had a mean  $ABI \leq 0.90$  diagnostic for PAD (95% CI 0.83-0.97, **Figure**  
11 **5C**).<sup>35</sup>

### 13 Association of $GPS_{PAD}$ with incident major adverse limb events

14 Lastly, we assessed whether polygenic background could identify individuals at  
15 increased risk for MALE after PAD diagnosis. In external validation cohorts,  $GPS_{PAD}$   
16 was significantly associated with incident MALE and provided improved risk stratification  
17 than previously published scores. Individuals with high polygenic risk (top 20% of  
18  $GPS_{PAD}$ ) in the UK Biobank had a 75% higher relative risk of incident MALE (HR 1.75;  
19 95% CI 1.18-2.57;  $P = 0.005$ ) over a median of 12.9 years (IQR 11.7-14.1) compared to  
20 the remainder of the PAD population. The top 20%  $GPS_{PAD}$  also had an increased risk  
21 of MALE in the MGBB (HR 1.56; 95% CI 1.06-2.30;  $P = 0.02$ ) during 5.19 years of follow-  
22 up (IQR 2.63-7.75) and AoU cohort (HR 1.51; 95% CI 1.03-2.24;  $P = 0.03$ ) over 2.32  
23 years of follow-up (IQR 0.35-4.29, **Figure 6A**).

1  
2 In the overall PAD cohort in the UK Biobank,  $GPS_{PAD}$  was significantly associated with  
3 MALE (HR 1.48; 95% CI 1.24-1.77; **Figure 6B**), which was diminished but remained  
4 associated after multivariable adjustment (HR 1.30; 95% CI 1.07-1.59;  $P=0.009$ ).  
5 Associations of  $GPS_{PAD}$  with incident MALE remained significant in the prevalent PAD  
6 cohorts of the MGBB (HR 1.34; 95% CI 1.12-1.60) and AoU (HR 1.33; 95% CI 1.11-  
7 1.58; **Figure 6B**), while other scores demonstrated weaker or non-significant risk  
8 estimates.

9

## 10 **Discussion:**

11 Using ancestry-stratified GWAS results from multiple populations, we developed  
12 a polygenic score that aggregates the effects of variants in PAD and related traits  
13 across five ancestries.  $GPS_{PAD}$  shows strong disease prediction and is transferable to  
14 external cohorts with ancestrally diverse populations, demonstrating associations with  
15 lower ABI and greater effect sizes for PAD association in non-European populations.  
16 This new score also identifies more high-risk individuals than a polygenic score for  
17 CAD, highlighting the differences in genetic risk for these atherosclerotic conditions.<sup>28</sup>  
18 Moreover, among individuals with diagnosed PAD, high  $GPS_{PAD}$  predicts future risk of  
19 MALE, defined as a surrogate of major amputation and acute limb ischemia.

20 Prior polygenic scores for PAD were solely derived from European data or  
21 employed GWAS data from a single population and were not validated in external  
22 cohorts.<sup>14-16</sup> Our computational framework considered variants discovered from >2  
23 million ancestrally diverse individuals and leveraged the pleiotropy between PAD and

1 related risk factors to enhance genetic risk prediction.<sup>26</sup> Such approaches that  
2 incorporate genome-wide genetic correlation increase the power of polygenic  
3 association analyses and improve prediction accuracy.<sup>36,37</sup> GPS<sub>PAD</sub> was indeed the most  
4 strongly predictive of PAD risk in external validation cohorts. Unlike previously published  
5 polygenic scores, GPS<sub>PAD</sub> was able to predict PAD risk in Latino individuals in All of Us  
6 and both African and South Asian individuals in two distinct cohorts (All of Us and the  
7 UK Biobank).

8 In prospective analysis in the UK Biobank, GPS<sub>PAD</sub> was associated with incident  
9 PAD with similar effect sizes as cross-sectional analysis. The inclusion of GPS<sub>PAD</sub>  
10 resulted in a comparable benefit in discriminative capacity as those afforded by smoking  
11 and diabetes.<sup>3</sup> We also note a significant interaction between GPS<sub>PAD</sub> and LDL-C or  
12 hemoglobin A1c in models predicting incident PAD. This suggests that individuals with  
13 higher GPS<sub>PAD</sub> may have a greater benefit from dietary changes, statin therapy, or  
14 glycemic control. Indeed, post-hoc analyses of statin therapy trials demonstrate that  
15 statins confer a greater cardiovascular risk reduction in those with high genetic risk  
16 compared to the general population.<sup>38</sup>

17 Current European guidelines more broadly recommend consideration of  
18 screening based on age and cardiovascular risk factors, while American cardiovascular  
19 society guidelines recommend screening for PAD with ABI in four groups: age<sub>≥</sub>65  
20 years, age 50-64 years with atherosclerotic risk factors, age <50 years with diabetes  
21 and 1 risk factor, and history of atherosclerosis in other vascular beds.<sup>8,9</sup> Addition of the  
22 GPS<sub>PAD</sub> has the potential to further streamline these screening recommendations and  
23 identify higher risk populations. When evaluating the combination of polygenic risk and

1 clinical risk assessment, incorporation of the top 5% polygenic risk with conventional  
2 risk factors identified more individuals who went on to develop PAD compared to the  
3 combination of risk factors with age. These results suggest the top 5% genetic risk and  
4 presence of risk factors could be considered a criteria for screening with ABIs to  
5 establish PAD diagnosis. Indeed, when leveraging ABIs, our analysis revealed an ABI  
6 diagnostic of PAD associated with high  $GPS_{PAD}$  in the MGBB, although replication in  
7 other cohorts is needed.

8         Secondary prevention efforts in PAD focus on limb salvage. MALE are  
9 associated with poor prognosis for PAD patients and represent index events for  
10 subsequent hospitalizations, reinterventions, major adverse cardiac events, and even  
11 mortality.<sup>39</sup> Once acute limb ischemia occurs, numerous returns to the operating room  
12 for revascularization with combined exposure to anticoagulation and thrombolytic  
13 therapy and subsequent reperfusion is a major physiologic stress in an already  
14 comorbid patient population. Staging systems such as the Wound, Ischemia, and foot  
15 Infection are used to determine amputation risk in the CLTI population, but there are no  
16 current vetted clinical tools for predicting MALE early in the disease course.<sup>3</sup> Our  
17 analysis found that  $GPS_{PAD}$  stratified MALE risk specifically in the PAD population and  
18 remained consistently predictive of incident MALE in the UK Biobank, Mass General  
19 Brigham Biobank, and All of Us.

20         These findings have several implications. First, an individual's genetic risk in  
21 combination with their cardiovascular risk factor profile may be employed as a criteria  
22 for consideration of screening with non-invasive vascular testing. PAD is an  
23 underdiagnosed and undertreated condition with guideline-directed therapies that lag



1 that of CAD.<sup>8</sup> Incorporation of polygenic risk may be used as an adjunct to clinical risk  
2 assessment to identify a greater proportion of high-risk individuals for medical  
3 optimization. In addition, as with studies of polygenic risk and CAD<sup>38,40-42</sup>, there may be  
4 early benefits for targeted intensive lifestyle and risk factor optimization in PAD with high  
5 genetic risk. However, there is need for further testing the effect of polygenic scores on  
6 hard cardiovascular outcomes in prospective studies, particularly among younger age  
7 groups. Given that high GPS<sub>PAD</sub> was strongly associated with MALE in three distinct  
8 PAD cohorts, there is promise for implementation of polygenic scores into secondary  
9 prevention measures for vascular practitioners, such as in frequency of surveillance or  
10 guiding more aggressive or prolonged antithrombotic therapies shown to reduce MALE  
11 incidence.<sup>11,39</sup>

12 Our study has several limitations. First, GPS<sub>PAD</sub> was trained using data from  
13 largely European individuals due to limitations in sample size of other ancestries.  
14 Nevertheless, we demonstrated evidence of portability to non-European groups, not  
15 previously achieved by prior PAD polygenic scores. Second, this study was restricted to  
16 individuals with genetic similarities to single ancestries and had more limited cases of  
17 non-European groups, which could have limited power to detect associations especially  
18 among the East Asian and MENA groups. Race alone was importantly not used to infer  
19 ancestry, however such broad ancestral categorizations can mask heterogeneity in  
20 polygenic score performance in some groups. Tailoring of polygenic scores to non-  
21 European target data, continued recruitment of underrepresented populations in  
22 biobanks, and development of methods to account for admixture should be a priority to  
23 improve polygenic score performance. Lastly, PAD and MALE phenotyping were based

1 on diagnosis and procedures codes from the EHR. There could have been variation in  
2 the quality of definitions between cohorts that influenced analyses towards null  
3 associations. However, we used non-invasive studies in the MGBB as another  
4 surrogate for PAD and observed similar associations of polygenic risk with binary PAD  
5 classifications in both the MGBB and AoU, where the latter is comprised of several U.S.  
6 healthcare institutions.

7

## 8 **Ethics Approval**

9 Individuals in the UK Biobank, All of Us, and Mass General Brigham Biobank underwent  
10 signed consent for genetic sequencing, storage of biological specimens, and access to  
11 electronic health record data. This research was conducted using the UK Biobank  
12 resource under application number 7089. Secondary analyses of the UK Biobank,  
13 MGBB, and AoU was approved by the Massachusetts General Hospital Institutional  
14 Review Board.

15

## 16 **Data Availability**

17 Individual-level data from the UK Biobank and All of Us are available upon request from  
18 researchers to each organization. This study used Controlled Tier data from All of Us  
19 which is available to authorized users on the All of Us Researcher Workbench. GPS<sub>PAD</sub>  
20 constructed in this paper will be made available in the Polygenic Score Catalog  
21 following publication. Polygenic scores used for comparison of GPSPAD performance  
22 are available in the Polygenic Score Catalog through accession ID PGS001843 and  
23 PGS002055.

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23

## **Acknowledgments**

We thank the participants and investigators in the UK Biobank, All of Us, and Mass General Brigham Biobank for their contributions to this study.

## **Funding**

This work was supported by the Harvard Medical School LaDue Fellowship in Cardiovascular Medicine (to A.M.F.) and National Institutes of Health (grants R01HL127564 and U01HG011719 to P.N. and K08HL168238 to A.P.P.).

## **Disclosure of Interest**

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, Foresite Labs, Genentech / Roche, GV, HeartFlow, Magnet Biomedicine, Merck, and Novartis, scientific advisory board membership of Esperion Therapeutics, Preciseli, TenSixteen Bio, and Tourmaline Bio, scientific co-founder of TenSixteen Bio, equity in MyOme, Preciseli, and TenSixteen Bio, and spousal employment at Vertex Pharmaceuticals, all unrelated to the present work. All other authors have no relevant disclosures.

1 **References:**

- 2 1. Song P, Rudan D, Zhu Y, *et al.* Global, regional, and national prevalence and risk  
3 factors for peripheral artery disease in 2015: an updated systematic review and  
4 analysis. *Lancet Glob Health* 2019;**7**:e1020-e1030. doi: 10.1016/S2214-  
5 109X(19)30255-4
- 6 2. Mahoney EM, Wang K, Keo HH, *et al.* Vascular hospitalization rates and costs in  
7 patients with peripheral artery disease in the United States. *Circ Cardiovasc Qual*  
8 *Outcomes* 2010;**3**:642-651. doi: 10.1161/CIRCOUTCOMES.109.930735
- 9 3. Aday AW, Matsushita K. Epidemiology of Peripheral Artery Disease and  
10 Polyvascular Disease. *Circ Res* 2021;**128**:1818-1832. doi:  
11 10.1161/CIRCRESAHA.121.318535
- 12 4. Hackler EL, 3rd, Hamburg NM, White Solaru KT. Racial and Ethnic Disparities in  
13 Peripheral Artery Disease. *Circ Res* 2021;**128**:1913-1926. doi:  
14 10.1161/CIRCRESAHA.121.318243
- 15 5. Narula N, Olin JW, Narula N. Pathologic Disparities Between Peripheral Artery  
16 Disease and Coronary Artery Disease. *Arterioscler Thromb Vasc Biol* 2020;**40**:1982-  
17 1989. doi: 10.1161/ATVBAHA.119.312864
- 18 6. Klarin D, Lynch J, Aragam K, *et al.* Genome-wide association study of peripheral  
19 artery disease in the Million Veteran Program. *Nat Med* 2019;**25**:1274-1279. doi:  
20 10.1038/s41591-019-0492-5
- 21 7. Mills J, Duffy M. Screening for Peripheral Artery Disease and Cardiovascular  
22 Disease Risk Assessment with the Ankle-Brachial Index. *Am Fam Physician*  
23 2018;**98**:754-755. doi:

- 1 8. Gornik HL, Aronow HD, Goodney PP, *et al.* 2024
- 2 ACC/AHA/AACVPR/APMA/ABC/SCAI/SVM/SVN/SVS/SIR/VESS Guideline for the
- 3 Management of Lower Extremity Peripheral Artery Disease: A Report of the American
- 4 College of Cardiology/American Heart Association Joint Committee on Clinical Practice
- 5 Guidelines. *Circulation* 2024;**149**:e1313-e1410. doi: 10.1161/CIR.0000000000001251
- 6 9. Mazzolai L, Teixido-Tura G, Lanzi S, *et al.* 2024 ESC Guidelines for the
- 7 management of peripheral arterial and aortic diseases. *Eur Heart J* 2024. doi:
- 8 10.1093/eurheartj/ehae179
- 9 10. Internal Revenue Service, Department of the Treasury; Employee Benefits
- 10 Security Administration, Department of Labor; Centers for Medicare & Medicaid
- 11 Services, Department of Health and Human Services. Coverage of Certain Preventive
- 12 Services Under the Affordable Care Act.
- 13 [https://www.federalregister.gov/documents/2015/07/14/2015-17076/coverage-of-](https://www.federalregister.gov/documents/2015/07/14/2015-17076/coverage-of-certain-preventive-services-under-the-affordable-care-act)
- 14 [certain-preventive-services-under-the-affordable-care-act](https://www.federalregister.gov/documents/2015/07/14/2015-17076/coverage-of-certain-preventive-services-under-the-affordable-care-act) (December 21, 2023)
- 15 11. Bonaca MP, Hamburg NM, Creager MA. Contemporary Medical Management of
- 16 Peripheral Artery Disease. *Circ Res* 2021;**128**:1868-1884. doi:
- 17 10.1161/CIRCRESAHA.121.318258
- 18 12. Carmelli D, Fabsitz RR, Swan GE, *et al.* Contribution of genetic and
- 19 environmental influences to ankle-brachial blood pressure index in the NHLBI Twin
- 20 Study. National Heart, Lung, and Blood Institute. *Am J Epidemiol* 2000;**151**:452-458.
- 21 doi: 10.1093/oxfordjournals.aje.a010230

- 1 13. Murabito JM, Guo CY, Fox CS, D'Agostino RB. Heritability of the ankle-brachial  
2 index: the Framingham Offspring study. *Am J Epidemiol* 2006;**164**:963-968. doi:  
3 10.1093/aje/kwj295
- 4 14. Wang F, Ghanzouri I, Leeper NJ, Tsao PS, Ross EG. Development of a  
5 polygenic risk score to improve detection of peripheral artery disease. *Vasc Med*  
6 2022;**27**:219-227. doi: 10.1177/1358863X211067564
- 7 15. Prive F, Aschard H, Carmi S, *et al.* Portability of 245 polygenic scores when  
8 derived from the UK Biobank and applied to 9 ancestry groups from the same cohort.  
9 *Am J Hum Genet* 2022;**109**:12-23. doi: 10.1016/j.ajhg.2021.11.008
- 10 16. Zhu K, Qian F, Lu Q, *et al.* Modifiable Lifestyle Factors, Genetic Risk, and  
11 Incident Peripheral Artery Disease Among Individuals With Type 2 Diabetes: A  
12 Prospective Study. *Diabetes Care* 2024. doi: 10.2337/dc23-1503
- 13 17. Bycroft C, Freeman C, Petkova D, *et al.* The UK Biobank resource with deep  
14 phenotyping and genomic data. *Nature* 2018;**562**:203-209. doi: 10.1038/s41586-018-  
15 0579-z
- 16 18. Koyama S, Wang Y, Paruchuri K, *et al.* Decoding Genetics, Ancestry, and  
17 Geospatial Context for Precision Health. *medRxiv* 2023. doi:  
18 10.1101/2023.10.24.23297096
- 19 19. All of Us Research Program I, Denny JC, Rutter JL, *et al.* The "All of Us"  
20 Research Program. *N Engl J Med* 2019;**381**:668-676. doi: 10.1056/NEJMs1809937
- 21 20. Genomes Project C, Auton A, Brooks LD, *et al.* A global reference for human  
22 genetic variation. *Nature* 2015;**526**:68-74. doi: 10.1038/nature15393

- 1 21. Boutin NT, Schechter SB, Perez EF, *et al.* The Evolution of a Large Biobank at  
2 Mass General Brigham. *J Pers Med* 2022;**12**. doi: 10.3390/jpm12081323
- 3 22. Aragam KG, Jiang T, Goel A, *et al.* Discovery and systematic characterization of  
4 risk variants and genes for coronary artery disease in over a million participants. *Nat*  
5 *Genet* 2022;**54**:1803-1815. doi: 10.1038/s41588-022-01233-6
- 6 23. All of Us All of Us Genomic Quality Report  
7 [https://support.researchallofus.org/hc/en-us/articles/4617899955092-All-of-Us-](https://support.researchallofus.org/hc/en-us/articles/4617899955092-All-of-Us-Genomic-Quality-Report-)  
8 [Genomic-Quality-Report-](https://support.researchallofus.org/hc/en-us/articles/4617899955092-All-of-Us-Genomic-Quality-Report-)
- 9 24. Yu S, Ma Y, Gronsbell J, *et al.* Enabling phenotypic big data with PheNorm. *J Am*  
10 *Med Inform Assoc* 2018;**25**:54-60. doi: 10.1093/jamia/ocx111
- 11 25. Castro VM, Gainer V, Wattanasin N, *et al.* The Mass General Brigham Biobank  
12 Portal: an i2b2-based data repository linking disparate and high-dimensional patient  
13 data to support multimodal analytics. *J Am Med Inform Assoc* 2022;**29**:643-651. doi:  
14 10.1093/jamia/ocab264
- 15 26. Patel AP, Wang M, Ruan Y, *et al.* A multi-ancestry polygenic risk score improves  
16 risk prediction for coronary artery disease. *Nat Med* 2023;**29**:1793-1803. doi:  
17 10.1038/s41591-023-02429-x
- 18 27. Ruan Y, Lin YF, Feng YA, *et al.* Improving polygenic prediction in ancestrally  
19 diverse populations. *Nat Genet* 2022;**54**:573-580. doi: 10.1038/s41588-022-01054-7
- 20 28. Patel AP, Khera AV. Advances and Applications of Polygenic Scores for  
21 Coronary Artery Disease. *Annu Rev Med* 2023;**74**:141-154. doi: 10.1146/annurev-med-  
22 042921-112629

- 1 29. Prive F, Arbel J, Vilhjalmsson BJ. LDpred2: better, faster, stronger.  
2 *Bioinformatics* 2021;**36**:5424-5431. doi: 10.1093/bioinformatics/btaa1029
- 3 30. Lambert SA, Gil L, Jupp S, *et al.* The Polygenic Score Catalog as an open  
4 database for reproducibility and systematic evaluation. *Nat Genet* 2021;**53**:420-425. doi:  
5 10.1038/s41588-021-00783-5
- 6 31. Matsushita K, Sang Y, Ning H, *et al.* Lifetime Risk of Lower-Extremity Peripheral  
7 Artery Disease Defined by Ankle-Brachial Index in the United States. *J Am Heart Assoc*  
8 2019;**8**:e012177. doi: 10.1161/JAHA.119.012177
- 9 32. Fahed AC, Wang M, Patel AP, *et al.* Association of the Interaction Between  
10 Familial Hypercholesterolemia Variants and Adherence to a Healthy Lifestyle With Risk  
11 of Coronary Artery Disease. *JAMA Netw Open* 2022;**5**:e222687. doi:  
12 10.1001/jamanetworkopen.2022.2687
- 13 33. Klarin D, Tsao PS, Damrauer SM. Genetic Determinants of Peripheral Artery  
14 Disease. *Circ Res* 2021;**128**:1805-1817. doi: 10.1161/CIRCRESAHA.121.318327
- 15 34. van Zuydam NR, Stiby A, Abdalla M, *et al.* Genome-Wide Association Study of  
16 Peripheral Artery Disease. *Circ Genom Precis Med* 2021;**14**:e002862. doi:  
17 10.1161/CIRCGEN.119.002862
- 18 35. Criqui MH, Matsushita K, Aboyans V, *et al.* Lower Extremity Peripheral Artery  
19 Disease: Contemporary Epidemiology, Management Gaps, and Future Directions: A  
20 Scientific Statement From the American Heart Association. *Circulation* 2021;**144**:e171-  
21 e191. doi: 10.1161/CIR.0000000000001005



- 1 36. van Rheenen W, Peyrot WJ, Schork AJ, Lee SH, Wray NR. Genetic correlations  
2 of polygenic disease traits: from theory to practice. *Nat Rev Genet* 2019;**20**:567-581.  
3 doi: 10.1038/s41576-019-0137-z
- 4 37. Truong B, Hull LE, Ruan Y, *et al.* Integrative polygenic risk score improves the  
5 prediction accuracy of complex traits and diseases. *Cell Genom* 2024;**4**:100523. doi:  
6 10.1016/j.xgen.2024.100523
- 7 38. Natarajan P, Young R, Stitzel NO, *et al.* Polygenic Risk Score Identifies  
8 Subgroup With Higher Burden of Atherosclerosis and Greater Relative Benefit From  
9 Statin Therapy in the Primary Prevention Setting. *Circulation* 2017;**135**:2091-2101. doi:  
10 10.1161/CIRCULATIONAHA.116.024436
- 11 39. Anand SS, Caron F, Eikelboom JW, *et al.* Major Adverse Limb Events and  
12 Mortality in Patients With Peripheral Artery Disease: The COMPASS Trial. *J Am Coll*  
13 *Cardiol* 2018;**71**:2306-2315. doi: 10.1016/j.jacc.2018.03.008
- 14 40. Kullo IJ, Jouni H, Austin EE, *et al.* Incorporating a Genetic Risk Score Into  
15 Coronary Heart Disease Risk Estimates: Effect on Low-Density Lipoprotein Cholesterol  
16 Levels (the MI-GENES Clinical Trial). *Circulation* 2016;**133**:1181-1188. doi:  
17 10.1161/CIRCULATIONAHA.115.020109
- 18 41. Muse ED, Chen SF, Liu S, *et al.* Impact of polygenic risk communication: an  
19 observational mobile application-based coronary artery disease study. *NPJ Digit Med*  
20 2022;**5**:30. doi: 10.1038/s41746-022-00578-w
- 21 42. Khera AV, Emdin CA, Drake I, *et al.* Genetic Risk, Adherence to a Healthy  
22 Lifestyle, and Coronary Disease. *N Engl J Med* 2016;**375**:2349-2358. doi:  
23 10.1056/NEJMoa1605086

## 1 **Legends**

### 2 **Figure 1. Schematic of GPS<sub>PAD</sub> construction and study design.**

3 1A. Overview of GPS<sub>PAD</sub> development using GWAS summary statistics for PAD and 14  
4 related traits stratified by African, East Asian, European, Latino, or South Asian  
5 ancestry.

6 1B. GPS<sub>PAD</sub> was then evaluated in three validation datasets consisting of distinct,  
7 ancestrally diverse individuals, with outcomes including PAD (prevalent and incident),  
8 ankle-brachial indices, and MALE. Diagram designed with FreePik. AFR, African. EAS,  
9 East Asian. EUR, European. LAT, Latino. LD<sub>ref</sub>, ancestry-matched LD reference panel.  
10 MENA, Middle Eastern/North African. SAS, South Asian.

11

### 12 **Figure 2: GPS<sub>PAD</sub> is an integrated polygenic score that stratifies PAD risk in** 13 **diverse ancestries**

14 2A. Prevalence of PAD across GPS<sub>PAD</sub> percentiles in the UK Biobank validation dataset  
15 (N = 304,294).

16 2B. Estimated PAD risk in the top and bottom of GPS<sub>PAD</sub> percentile distributions relative  
17 to the middle quintile (40-59%), as assessed in logistic regression models adjusted for  
18 age, sex, genotyping array, and ten principal components of ancestry. Results are  
19 shown on log-axis.

20 2C. Risk of PAD associated with GPS<sub>PAD</sub> in ancestral subgroups in the UK Biobank.  
21 Results were calculated from a logistic regression model with age, sex, genotyping  
22 array, and ten principal components as covariates.

23

1 **Figure 3. GPS<sub>PAD</sub> shows improved transferability to external cohorts.**

2 4A. OR per-SD change of polygenic scores in external validation cohorts, AoU (N =  
3 237,173) and MGBB (N = 37,017).

4 4B. Predictive performance for each score is also shown in ancestry subgroups: African  
5 individuals in AoU (N = 51,691) and MGBB (N = 1,945); East Asians in AoU (N = 5,268),  
6 European individuals in AoU (N = 119,167) and MGBB (N = 32,139); Latino individuals  
7 in AoU (N = 39,576) and MGBB (N = 2,933); and South Asian individuals in AoU (N =  
8 2,289).

9 OR were calculated from a logistic regression model adjusted for age, sex, genotyping  
10 array, and the first ten principal components. Asterisks represent P-value for  
11 associations, \*=P<0.05. \*\*=P<0.01. \*\*\*P<0.0001.

12 EAS, East Asian. EUR, European. LAT, Latino. SAS, South Asian.

13

14 **Figure 4. Performance of GPS<sub>PAD</sub> in predicting incident PAD in the UK Biobank.**

15 4A. Changes in C-statistic of GPS<sub>PAD</sub> (red point) and clinical risk factors (blue points)  
16 relative to the baseline model including age, sex, and the first ten principal components  
17 of ancestry (gray point), as calculated from Cox proportional hazards models in the UK  
18 Biobank holdout validation dataset. Lines represent confidence intervals of the C-  
19 statistic change.

20 4B-C. 10-year rates of PAD according to lifestyle and genetic categories, showing the  
21 influence of smoking (4B) and dietary patterns (4C). Shown are standardized  
22 cumulative rates of incident PAD standardized to the population averages for each  
23 covariate in the Cox models.

1

2 **Figure 5. Incorporation of GPS<sub>PAD</sub> into screening criteria and PAD diagnosis.**

3 5A. 10-year incidence of first-occurring PAD events from prospective analyses of  
4 individuals without prior PAD in the UK Biobank. Cox proportional hazards regression  
5 models were used to calculate the 10-year event risk among groups for which current  
6 clinical guidelines recommend ABI screening<sup>8</sup> (blue) and individuals in the top 5%  
7 GPS<sub>PAD</sub> with individual risk factors (red). Error bars represent 95% CI and are not  
8 completely shown for groups in the age<50y with diabetes + CKD or top 5% GPS<sub>PAD</sub> +  
9 CKD, but can be found in Supplemental Table 16.

10 5B. Loess regression plot of GPS<sub>PAD</sub> and minimum ABI among a subset of individuals in  
11 the Mass General Brigham Biobank (N = 763).

12 5C. Mean ABI for individuals in the top percentiles of GPS<sub>PAD</sub>. \* indicates percentile  
13 group with mean ABI<sub>≤</sub>0.9.

14 AAA, abdominal aortic aneurysm. CKD, chronic kidney disease.

15

16 **Figure 6. GPS<sub>PAD</sub> stratifies risk for incident major adverse limb events among**  
17 **individuals with PAD.**

18 6A. Kaplan-Meier curves of MALE according to high GPS<sub>PAD</sub> compared to the remainder  
19 of the PAD population, standardized to the mean age and sex of each cohort.

20 6B. Risk associated with polygenic scores for incident MALE among individuals with  
21 diagnosed PAD in the UK Biobank (N = 2,094), Mass General Brigham Biobank (N =  
22 867), and All of Us (N = 8,022). HRs were calculated from adjusted Cox regression


1 models including age, sex, genotyping array, the first ten principal components of  
2 ancestry as covariates.  
3  
4  
5  
6  
7  
8  
9  
10  
11

Figure 1

**A**


**GPS<sub>PAD</sub> fine-tuning in 96,239 individuals from the UK Biobank**

**Layer 1**



- Calculate ancestry-specific polygenic scores for PAD and 14 related traits from ancestry-stratified summary statistics and LD<sub>ref</sub>
- Derive multi-ancestry score for each trait by combining ancestry-specific scores in a LASSO regression model predicting PAD

**Layer 2**



- Input multi-ancestry, trait-specific scores from Layer 1 as covariates in LASSO regression model predicting PAD
- Derive GPS<sub>PAD</sub> based on optimal combination of multi-ancestry trait scores for PAD prediction


**B**

**GPS<sub>PAD</sub> validation in multi-ancestry cohorts**

**Validate PAD prediction performance in 3 distinct populations**

UK Biobank		Mass General Brigham	All of Us	
6,939 AFR	1,335 MENA	1,945 AFR	51,691 AFR	39,576 LAT
1,761 EAS	7,680 SAS	32,139 EUR	5,268 EAS	509 MENA
286,356 EUR		2,933 LAT	119,167 EUR	2,289 SAS

**Assess utility of GPS<sub>PAD</sub> in guiding screening and secondary prevention**



- Assess magnitude of risk discrimination ability for incident PAD
- Evaluate relationship of GPS<sub>PAD</sub> with ankle-brachial indices
- Test association of GPS<sub>PAD</sub> with incident MALE in prevalent PAD

Figure 2

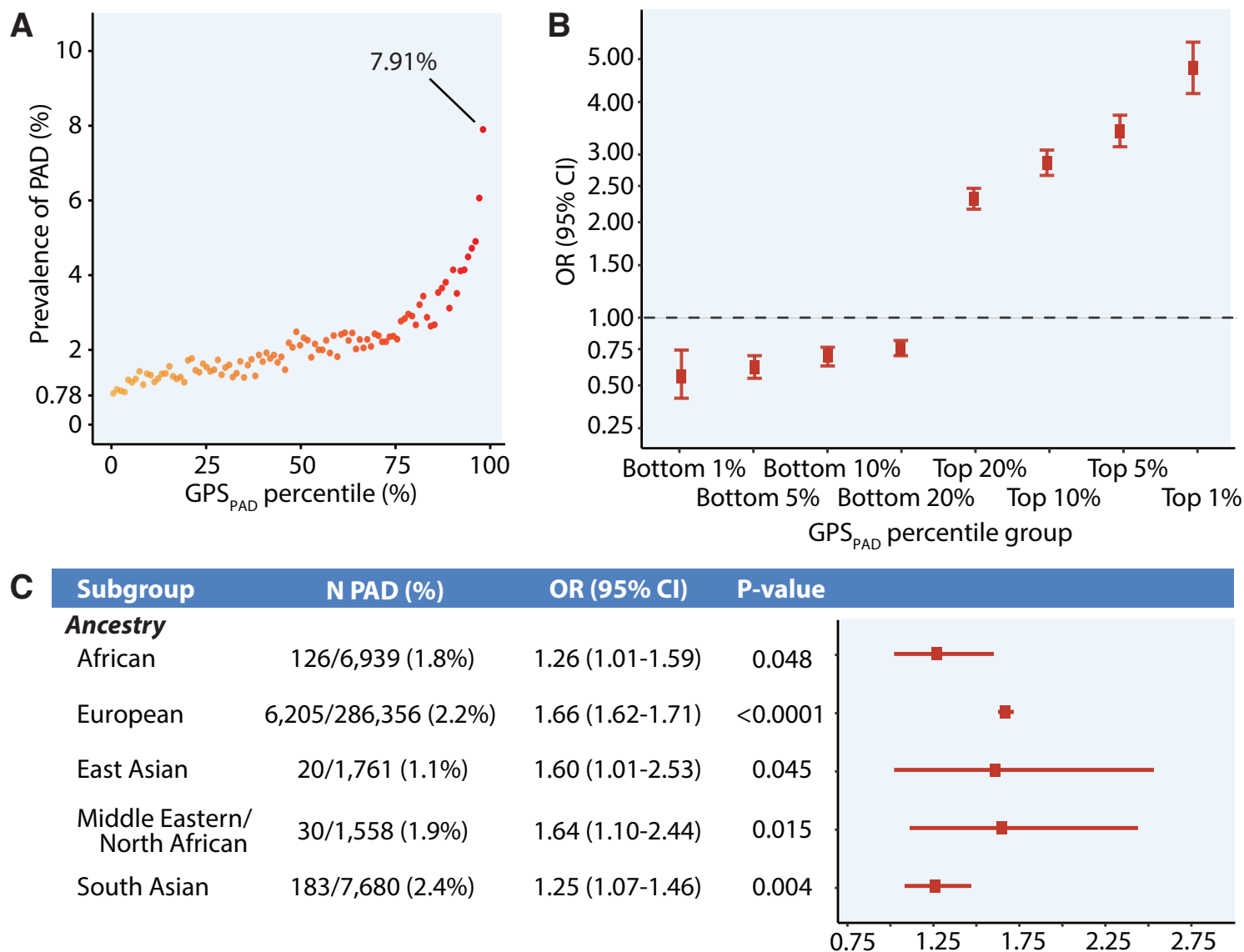


Figure 3

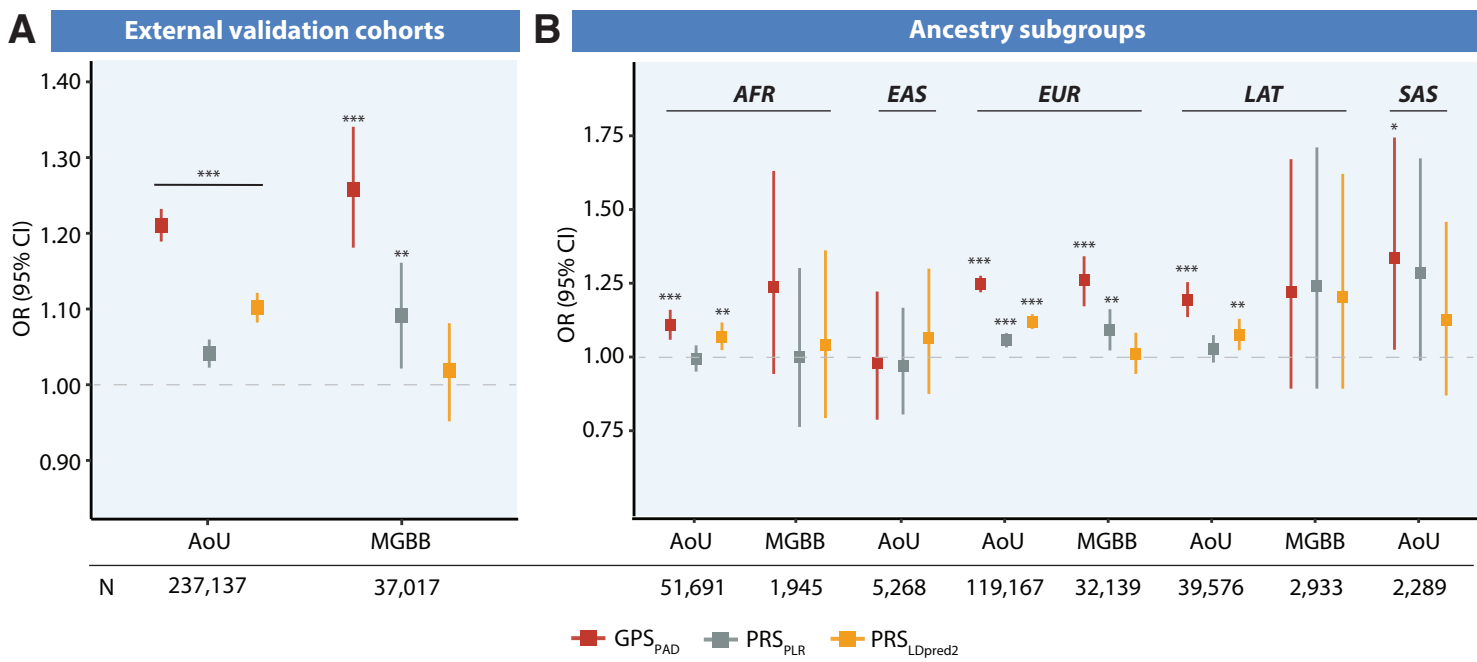




Figure 4

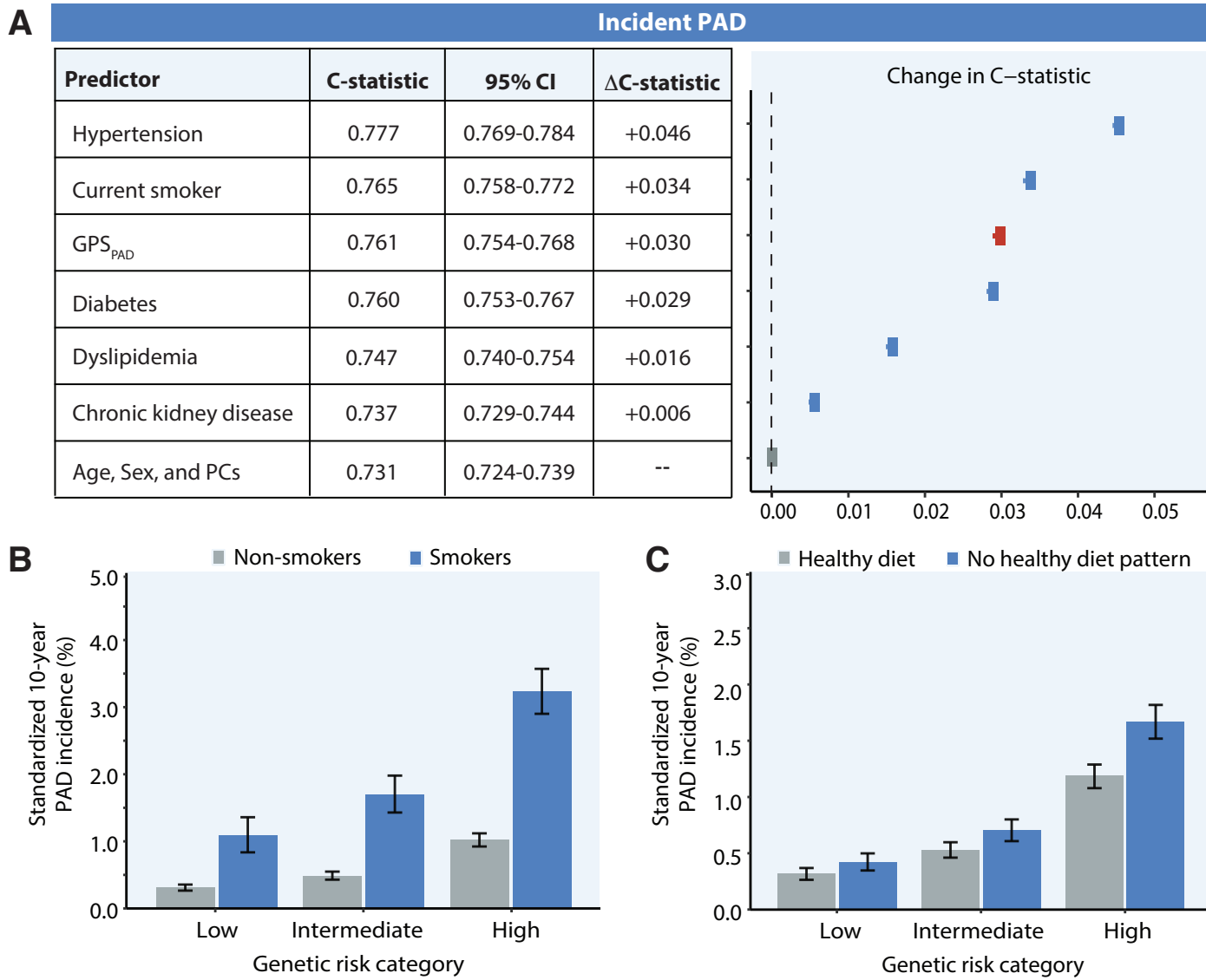


Figure 5

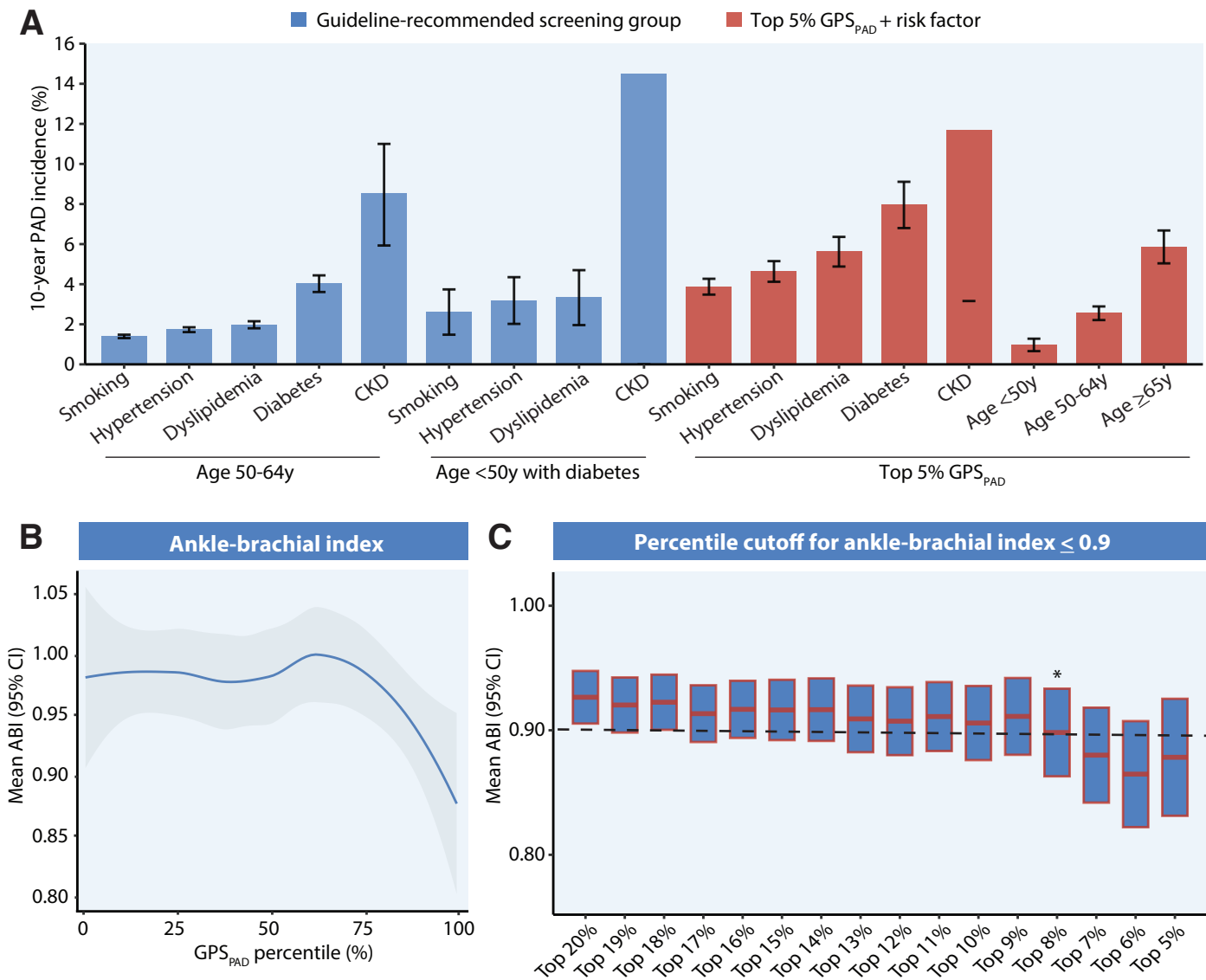


Figure 6

