Predictive models for abdominal aortic aneurysms using polygenic scores and PheWAS-derived risk factors[a]

Jacklyn N. Hellwege[†]
Division of Genetic Medicine, Department of Medicine, Vanderbilt Genetics Institute
Vanderbilt University Medical Center
2525 West End Ave. Ste 700, Nashville, TN, 37203, USA
Email: jacklyn.hellwege@vumc.org

Chad Dorn
Department of Biomedical Informatics, Vanderbilt University Medical Center
2525 West End Ave. Ste 1500, Nashville, TN, 37203, USA
Email: chad.a.dorn@vumc.org

Marguerite R. Irvin
Department of Epidemiology, University of Alabama at Birmingham,
Birmingham, AL 35233, USA
Email: irvinr@uab.edu

Nita A. Limdi
Department of Neurology, University of Alabama at Birmingham,
Birmingham, AL 35233, USA
Email: nlimdi@uabmc.edu

James Cimino
Informatics Institute, University of Alabama at Birmingham,
Birmingham, AL 35233, USA
Email: jamescimino@uabmc.edu

T. Mark Beasley
Department of Biostatistics, University of Alabama at Birmingham,

[*]Some of the datasets used for analyses were obtained from Vanderbilt University Medical Center's BioVU which is supported by institutional funding, 1S10RR025141-01, and by the CTSA grant UL1TR000445 from NCATS/NIH. The eMERGE Network is funded by the following grants: U01HG8657 (Kaiser Washington/University of Washington); U01HG8685 (Brigham and Women's Hospital); U01HG8672 (Vanderbilt University Medical Center); U01HG8666 (Cincinnati Children's Hospital Medical Center); U01HG6379 (Mayo Clinic); U01HG8679 (Geisinger Clinic); U01HG8680 (Columbia University Health Sciences); U01HG8684 (Children's Hospital of Philadelphia); U01HG8673 (Northwestern University); U01HG8701 (Vanderbilt University Medical Center serving as the Coordinating Center); U01HG8676 (Partners Healthcare/Broad Institute); and U01HG8664 (Baylor College of Medicine).
[†]Work partially supported by K12 HD04348.

Birmingham, AL, 35233, USA.
Email: mbeasley@uab.edu

Philip S. Tsao
VA Palo Alto Health Care System
Stanford Cardiovascular Institute, Department of Medicine, Stanford University School of Medicine, Stanford, CA, 94305, USA
Email: ptsao@stanford.edu

Scott M. Damrauer
Corporal Michael J. Crescenz VA Medical Center
Department of Genetics, Department of Surgery, University of Pennsylvania Perelman School of Medicine Philadelphia, PA, 19104, USA
Email: Scott.Damrauer@pennmedicine.upenn.edu

Dan M. Roden
Division of Clinical Pharmacology, Department of Medicine; Department of Pharmacology; Department of Biomedical Informatics; Vanderbilt Genetics Institute; Vanderbilt University Medical Center
2215 Garland Ave., Nashville, TN, 37203, USA
Email: dan.roden@vumc.org

Digna R. Velez Edwards
Division of Quantitative Science, Department of Obstetrics and Gynecology, Department of Biomedical Informatics, Vanderbilt Genetics Institute, Vanderbilt University Medical Center
2525 West End Ave. Ste 600, Nashville, TN, 37203, USA
Email: digna.r.velez.edwards@vumc.org

Wei-Qi Wei
Department of Biomedical Informatics, Vanderbilt University Medical Center
2525 West End Ave. Ste. 1500, Nashville, TN, 37203, USA
Email: wei-qi.wei@vumc.org

Todd L. Edwards
Division of Epidemiology, Vanderbilt Genetics Institute, Vanderbilt University Medical Center
2525 West End Ave. Ste 600, Nashville, TN, 37203, USA
Email: todd.l.edwards@vumc.org

Abdominal aortic aneurysms (AAA) are common enlargements of the abdominal aorta which can grow larger until rupture, often leading to death. Detection of AAA is often by ultrasonography and screening recommendations are mostly directed at men over 65 with a smoking history. Recent large-scale genome-wide association studies have identified genetic loci associated with AAA risk. We combined known risk factors, polygenic risk scores (PRS) and precedent clinical diagnoses from electronic health records (EHR) to develop predictive models for AAA, and compared performance against screening recommendations. The PRS included genome-wide summary statistics from the Million Veteran Program and FinnGen (10,467 cases, 378,713 controls of European ancestry), with optimization in Vanderbilt's BioVU and validated in the eMERGE Network, separately across both White and Black participants. Candidate diagnoses were identified through a temporally-oriented

Phenome-wide association study in independent EHR data from Vanderbilt, and features were selected via elastic net. We calculated C-statistics in eMERGE for models including PRS, phecodes, and covariates using regression weights from BioVU. The AUC for the full model in the test set was 0.883 (95% CI 0.873-0.892), 0.844 (0.836-0.851) for covariates only, 0.613 (95% CI 0.604-0.622) when using primary USPSTF screening criteria, and 0.632 (95% CI 0.623-0.642) using primary and secondary criteria. Brier scores were between 0.003 and 0.023 for our models indicating good calibration, and net reclassification improvement over combined primary and secondary USPSTF criteria was 0.36-0.60. We provide PRS for AAA which are strongly associated with AAA risk and add to predictive model performance. These models substantially improve identification of people at risk of a AAA diagnosis compared with existing guidelines, with evidence of potential applicability in minority populations.

Keywords: Abdominal Aortic Aneurysm, Polygenic Scores, Prediction, Precision Medicine

1. Introduction

Abdominal aortic aneurysms (AAA) is a common and life-threatening condition in which enlargement of the abdominal aorta can lead to a deadly rupture. Rupture is associated with a mortality rate as high as 81%, including mortality of over 50% even among individuals that rupture in a hospital setting[1]. Current estimates suggest that approximately 4% of the US population over 65 has an AAA, and 41,000 deaths a year are attributed to AAA complications[2,3]. Based on AHA 2019 Heart Disease and Stroke statistics, the prevalence of AAA ranges from 1.3% in males 45-54 years old to 12.5% in males 75-84 years old[4]. For females, the prevalence ranges from 0% in the youngest to 5.2% in the oldest age groups[4].

Common risk factors for AAA risk are race, age, sex, smoking behavior, atherosclerosis, hypertension, and hyperlipidemia[5-7]. A family history of AAA is associated with an adjusted OR of 2.17[8]. Factors associated with aorta diameter from Mendelian randomization studies include pulse pressure, triglycerides, and height[9]. An estimate of SNP-based heritability for AAA is not available, however, heritability of AAA is estimated to be as high as 70%[10]. Multiple genome-wide association studies have been conducted and have detected 24 distinct loci[11-15]. These observations provide a basis for including genetic information in prediction of future AAA events.

There are no currently available pharmacological therapies for prevention or treatment of AAA. When discovered, AAA cases are monitored using periodic ultrasounds, where the goal is to observe AAA expansion until the risk of rupture is deemed to be larger than the risks posed by surgical repair[16], which for many patients is when the diameter reaches 5.5 cm[17]. AAA cases are most often either discovered incidentally by abdominal imaging for some other indication, or by screening programs that target specific high-risk groups.

Current US Preventative Services Task Force (USPSTF) guidelines focus on screening men between 65 and 75 years of age with a history of smoking[18]. In a recent large retrospective study of almost 291,850 AAA hospitalizations, 23% were women, and over 60% were not between 65 and 75 years of age[19]. USPSTF recommendations are not derived from statistical models and may underserve understudied groups or individuals who are at unusually high risk for their demographic category due to an accumulation of known and unknown risk factors.

Strong racial disparities have been observed in prevalence, risk, and response to surgical treatments in AAA patients[20,21]. These important and poorly understood aspects of AAA epidemiology are often neglected in screening guidelines. Because effective AAA management depends on detection, this opportunity for improving the screening strategy has the potential to save lives, many of whom are in underserved groups. In this paper, we leverage prior GWAS of AAA and electronic health records (EHR) linked to genetic information to develop predictive models that outperform the USPSTF guidelines in identifying high-risk individuals and evaluating the performance of polygenic predictors in multiple ancestral groups.

2. Methods

2.1. *Synthetic Derivative*

The Synthetic Derivative (SD) is a deidentified mirror of EHR at Vanderbilt University Medical Center (VUMC) with records for >3 million patients dating to January 1990 and updated regularly.

2.2. *BioVU*

The BioVU DNA Repository is a subset of the SD at VUMC with linkage to individuals' DNA samples. A detailed description of the database and how it is maintained has been published elsewhere[22,23]. BioVU participant DNA samples were genotyped on a custom Illumina Multi-Ethnic Genotyping Array (MEGA-ex; Illumina Inc., San Diego, CA, USA). Quality control included excluding samples or variants with missingness rates above 2%. Samples were also excluded if consent had been revoked, sample was duplicated, or failed sex concordance checks. Imputation was performed on the Michigan Imputation Server (MIS) v1.2.4[24] using Minimac4 and the Haplotype Reference Consortium (HRC) panel v1.1[25]. AAA cases were identified using phecodes[26,27]: 2 or more instances of an International Classification of Diseases (ICD) version 9 or 10 diagnostic code for AAA, while controls were those without any ICD codes for AAA or phecodes in range 440-449.9 (Diseases of Arteries, Arterioles, and Capillaries). Individuals with one AAA ICD code were excluded. Smoking status was defined using ICD codes.

2.3. *eMERGE*

The eMERGE Network is a consortium of several EHR-linked biorepositories formed with the goal of developing approaches for the use of the EHR in genomic research[28,29]. Consortium membership has evolved over eMERGE's 11-year history, with many sites contributing data: Group Health/University of Washington, Marshfield Clinic, Mayo Clinic, Northwestern University, Vanderbilt University, Children's Hospital of Philadelphia (CHOP), Boston Children's Hospital (BCH), Cincinnati Children's Hospital Medical Center (CCHMC), Geisinger Health System, Mount Sinai School of Medicine, Harvard University and Columbia University. The eMERGE study was approved by the Institutional Review Board at each site and all methods were performed in accordance with the relevant guidelines and regulations. Participants at all sites provided written informed consent. AAA cases and controls were defined as in BioVU.

2.4. *Genome-wide Summary Statistics*

We combined genome-wide summary statistics for AAA from the Million Veteran Program[11] and FinnGen[30] for a total of 10,467 cases and 378,713 controls of European ancestry) using fixed-effects inverse-variance weighted meta-analysis implemented in METAL[31].

2.5. *Polygenic Score Development*

PRSs were constructed using PRS-CS[32] software and PLINK2[33], followed by p-value thresholding (range: p=1 - $5x10^{-8}$) as in Ref[34]. Optimal p-value thresholds were 1.0 in Whites and p<$5x10^{-3}$ in Blacks, as determined by maximal variance explained in BioVU (0.76% and 0.59%, respectively).

2.6. *Identification of phecode risk factors*

We extracted all diagnostic codes from individuals in the SD who were not part of the BioVU MEGA genotyped set who classified as either a case or control for AAA status. Codes for AAA cases were censored following the earliest AAA diagnosis code – i.e. all diagnoses post-AAA were removed, in order to capture only those diagnoses which preceded AAA diagnosis and represent potential risk factors for subsequent diagnosis of AAA. We performed a phenome-wide association study[35] (PheWAS) on this temporally-censored dataset with AAA as the outcome with each phecode status used as predictor, adjusted for age and sex, stratified by self-reported race/ethnicity. Bonferroni correction was used to set significance thresholds to identify significant phecodes.

2.7. *Selection of independent components with elastic nets*

We used elastic net models with 10-fold cross validation in BioVU to estimate feature weights, implemented in the glmnet R package[36,37] for selection of candidate risk features derived from the temporal PheWAS in the SD. Among the variables considered were 196 candidate phecodes (significant in at least one temporal PheWAS), age, sex, BMI, smoking status, race, and ethnicity. Individuals missing status (with only one AAA ICD code or an exclusion code) were classified with controls (using probit linkages) in a case-cohort design to allow simultaneous modeling of phecodes.

2.8. *Predictive models*

Prediction of AAA diagnoses in eMERGE data used logistic regression implemented in R, and evaluated area under the receiver operator curve (pROC package), net reclassification index (nricens package), and Brier scores. Phecodes selected from the elastic net were included alongside age, sex, BMI, smoking status, polygenic scores, and principal components of ancestry.

3. Results

3.1. *Polygenic risk score development, performance, and association with AAA*

We performed meta-analysis of MVP and FinnGen summary statistics for AAA using a fixed-effects inverse-variance weighted method in METAL. Polygenic scores were constructed using PRS-CS to generate weights, followed by p-value thresholding. The optimal p-value threshold was 1.0 in non-Hispanic Whites (NHW), while the optimal threshold in non-Hispanic Blacks (NHB) was p<$5x10^{-3}$ as determined by maximal variance explained in BioVU (0.76% and 0.59%, respectively; Table 1); at these thresholds, the PRSs contained 1,118,966 and 12,314 SNPs, respectively.

Table 1. Variance explained across PRS p-value thresholds in BioVU Non-Hispanic Whites and Blacks

RACE	1	0.5	5.0E-02	5.0E-03	5.0E-04	5.0E-05	5.0E-06	5.0E-07	5.0E-08
NHW	**0.0076**	0.0070	0.0072	0.0062	0.0056	0.0042	0.0039	0.0031	0.0014
NHB	0.0018	0.0021	0.0018	**0.0059**	0.0032	0.0021	0.0023	0.0012	0.00001

We observed increasing odds of AAA in eMERGE by PRS of both scores when modeled adjusting for age, sex, body mass index (BMI), and 10 principal components (Figure 1). In NHW, the scores were both significant (p-value = < 2e-16) and each explained 0.014% of the variance, while in NHB only the p<5e-3 score (PRS-B) was significant (p-value = 0.0028). When modeled as deciles, associations trended toward higher odds ratios at higher deciles for both PRS in NHW, but more consistently in NHB with the p=5e-3 PRS (Figure 1). The 95th and higher percentile vs. the rest odds ratios were 2.45 (95% Confidence Interval [CI]: 2.09-2.88; p-value <2x10^{-16}) and 2.11 (95% CI 0.84-5.31; p-value = 0.11) for NHW and NHB subsets, respectively, for the p=1 score (Table 2). For the p=5x10^{-3} PRS, the odds ratios were 2.2 (95% CI: 1.87-2.59; p-value <2x10^{-16}) and 3.34 (95% CI 1.49-7.47; p-value = 0.003) for NHW and NHB subsets, respectively.

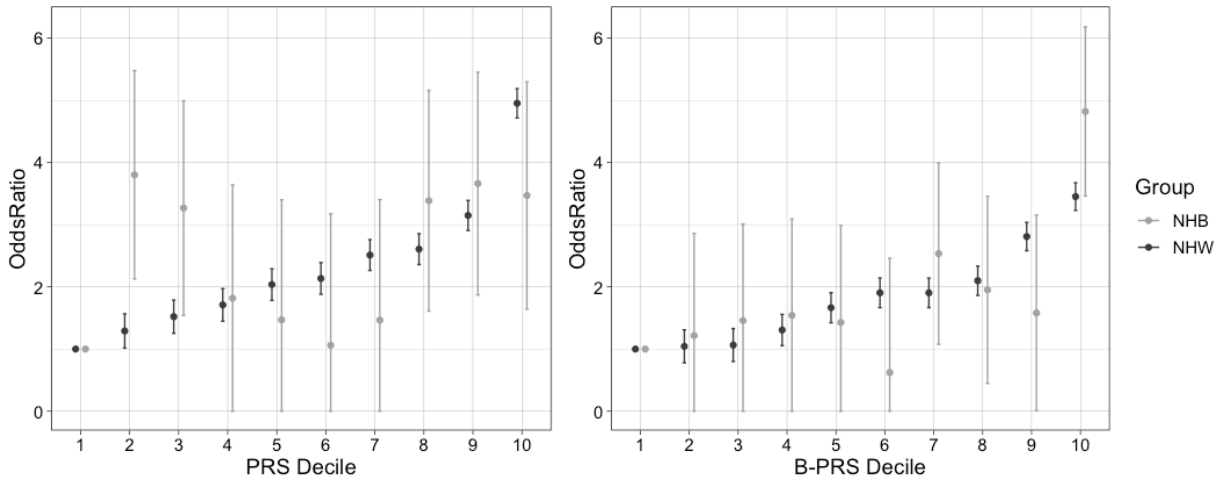

Figure 1. Odds ratios for AAA with p=1 PRS (A) and p=5e-3 PRS (B) deciles in eMERGE

Table 2. Association between AAA PRS and AAA outcome in eMERGE

RACE	CASES / CONTROLS	P=1 PRS OR (95% CI)	P=1 PRS P-VALUE	P=5E-3 PRS-B OR (95% CI)	P=5E-3 PRS-B P-VALUE
NHW	2,165 / 42,843	2.45 (2.09-2.88)	**<2.0x10^{-16}**	2.20 (1.87-2.59)	**<2.0x10^{-16}**
NHB	42 / 4,492	2.11 (0.84-5.31)	0.11	3.34 (1.49-7.47)	**0.003**

Each PRS modeled as top 5% of distribution compared to remainder. Covariates included age, sex, BMI and 10 principal components of ancestry.

3.2. *Identification of phecode diagnosis risk factors*

In order to identify risk-associated diagnoses which precede AAA diagnosis/events, we performed a temporally-censored PheWAS. Within the Vanderbilt Synthetic Derivative dataset, we censored any diagnosis codes occurring after an ICD code for AAA, and performed a PheWAS using AAA

as the outcome and each phecode as the predictor. Atherosclerosis phecodes were broadly significant, while Kawasaki disease was significant only in NHB individuals. In total, 192 phecodes were significant in analyses of NHW, 10 in NHB, 3 in Hispanic and none in non-Hispanic Asian (NHA) (Table 3). In total, 196 phecodes were significantly associated in at least one analysis. All significant phecodes were included as components in the elastic net.

Table 3. Feature-identifying PheWAS in Vanderbilt Synthetic Derivative

RACE	CASES	CONTROLS	PHECODES ANALYZED	SIGNIFICANT PHECODES
NHW	4,416	1,202,332	1866	192
NHB	292	166,170	1860	10
NHA	23	23,490	1802	0
Hispanic	31	47,003	1843	3

Of 202 variables (196 Phecodes) included in the elastic net, 87 were retained in the model- four *a priori* variables (smoking status, median BMI, age, and gender), and 83 Phecode diagnoses. 67 of 87 features were negatively associated, that is, diagnosis of a preceding Phecode was associated with a reduced risk of AAA diagnosis. Chromosomal abnormalities and genetic disorders diagnoses (phecode 758) had the largest weighting in the elastic net model, despite being generally uncommon in the population studied (0.04%). Evaluation of the 83 phecodes indicated several hierarchical codes which were collapsed to select independent features, resulting in a final set of 68 phecodes.

3.3. *Predictive models*

We validated our AAA risk prediction models developed in BioVU using external data to evaluate its discrimination and calibration. We benchmarked our models to the performance of the USPTF screening criteria. A sparse model containing age, sex, BMI, smoking status and principal components of ancestry performed substantially better than USPTF screening criteria, with AUCs over 0.8 in all three groups compared to AUCs ranging from 0.55-0.63 for USPTF primary and secondary criteria (Table 4, Figure 2). The AUCs when including PRS and covariates were 0.846

Table 4. AUC (CI) for predictive models fit in BioVU and applied to eMERGE

MODEL	ALL	NHW	NHB
USPTF-B	0.613 (0.604-0.622)	0.614 (0.605-0.623)	0.545 (0.504-0.586)
USPTF-C	0.632 (0.623-0.642)	0.632 (0.622-0.642)	0.594 (0.539-0.650)
COV	0.844 (0.836-0.851)	0.838 (0.830-0.845)	0.819 (0.765-0.873)
PHE	0.859 (0.849-0.870)	0.853 (0.842-0.864)	0.807 (0.732-0.883)
PHE+COV	0.883 (0.874-0.893)	0.877 (0.868-0.887)	0.758 (0.659-0.857)
PRS	0.494 (0.484-0.505)	0.598 (0.586-0.610)	0.531 (0.448-0.613)
PRS+COV	0.836 (0.829-0.844)	0.846 (0.838-0.854)	0.820 (0.766-0.874)
FULL	**0.883 (0.874-0.893)**	**0.877 (0.868-0.887)**	**0.758 (0.659-0.857)**
PRS-B	0.533 (0.522-0.544)	0.601 (0.589-0.613)	0.580 (0.498-0.662)
PRS-B+COV	0.846 (0.839-0.854)	0.846 (0.838-0.853)	0.830 (0.776-0.874)
FULL-B	**0.883 (0.873-0.892)**	**0.880 (0.870-0.890)**	**0.758 (0.659-0.857)**

PRS: Best performing PRS overall; PRS-B/FULL-B: models including p<5e-3 optimal PRS in NHB

(0.839-0.854), 0.846 (0.838-0.853) and 0.830 (0.776-0.884) for the entire dataset, in NHW, and in NHB respectively. Adding phecode predictors to the models improved AUCs further: 0.883 (0.873-0.892), 0.880 (0.870-0.890) in the entire data and NHW set, respectively, but not in NHB (AUC = 0.758 (0.659-0.857).

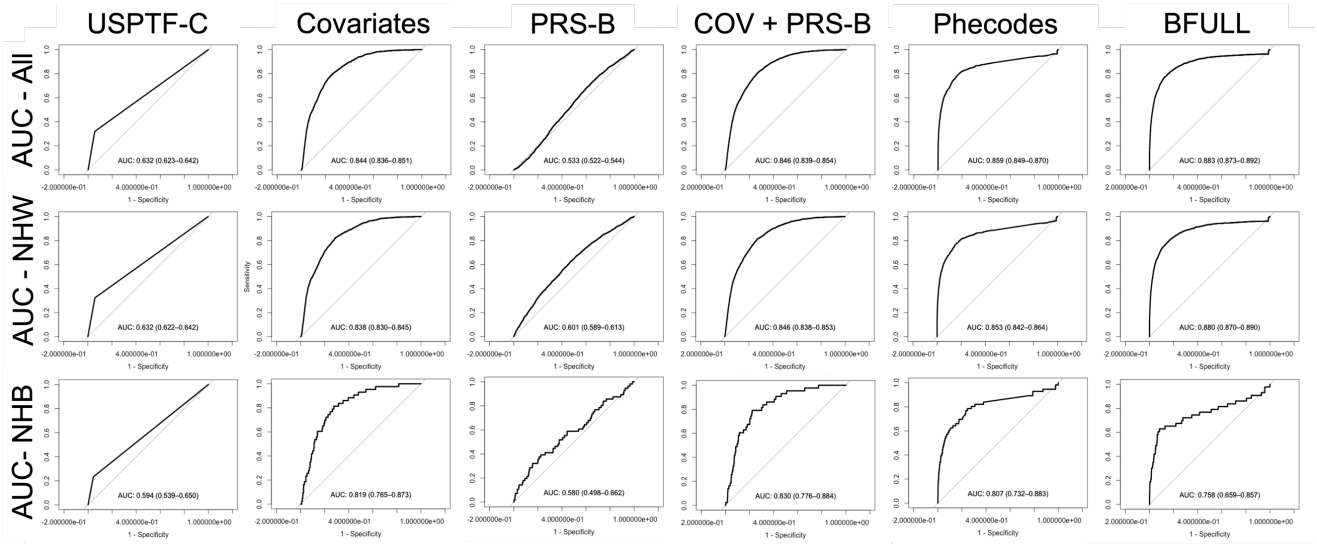

Figure 2. Receiver-operator curve plots using models applied in (top to bottom:) eMERGE overall, NHW and NHB for (left to right:) USPTF primary+secondary guidelines, covariates only, PRS-B only, covariates + PRS-B, phecodes only, and full models (covariates, PRS-B, and phecodes).

We evaluated model reclassification and calibration using net reclassification indices (NRI) and Brier scores, respectively. Generally, although model calibration was very good for the full models (0.003-0.023; Table 5), inclusion of both PRS and phecode predictors to models using covariates had a moderate impact on reclassification indices (0.23) in combined datasets, with larger impacts in NHB (Table 6). The NRIs from these data compared to USPTF guidelines is striking, with covariates alone having an NRI of 0.20-0.37, and full models 0.46-0.83.

Table 5. Brier scores for various models in eMERGE

MODEL	ALL	NHW	NHB
FULL	0.021	0.023	0.0032
FULL-B	0.021	0.023	0.0030

4. Discussion

We have integrated a variety of data types to construct models for predicting AAA diagnoses across multiple EHR systems. Our polygenic scores for AAA, despite being developed using only European-ancestry genetic data, associated with AAA in NHB as well as NHW, and are being made available through the polygenic score catalog (pgscatalog.org). Addition of the PRS in the entire eMERGE dataset had a small negative effect on the model (ΔAUC = -0.008), however the model improved in the NHW and NHB strata separately, as did all PRS-B models.

Our study suggests an enhanced disease screening program of asymptomatic individuals who would otherwise be considered lower risk by USPTF guidelines would substantially improve AAA detection in the US population. Even covariates alone perform substantially better than the USPTF guidelines, similar to what has been shown in a recent UK Biobank study with a simple predictive

Table 6. NRI for predictive models in eMERGE compared with USPSTF screening criteria

MODELS	EMERGE	EMERGE NHW	EMERGE NHB
USPTF C : B	0	0	0
COV : USPTF B	0.37	0.31	0.20
COV : USPTF C	0.37	0.31	0.20
PRS+COV : COV	0.025	0.031	0.008
PRS-B+COV : COV	0.018	0.024	0.048
FULL : COV	0.23	0.25	0.61
BFULL : COV	0.23	0.24	0.63
FULL : USPTF B	0.60	0.50	0.82
FULL : USPTF C	0.60	0.50	0.82
BFULL : USPTF B	0.60	0.46	0.83
BFULL : USPTF C	0.60	0.46	0.83

PRS: Best performing PRS overall; PRS-B/FULL-B: models including p<5e-3 PRS (optimal in NHB).

model that lacked variables for genetics, sex, or race[38]. This demonstrates the principle that opportunities exist to substantially improve the public health impact of AAA. Clinical decision support tools for identifying patients for AAA screening based on USPTF guidelines have existed for over a decade[39-42], however, recent reports indicate that even those fitting USPTF criteria remain unlikely to receive screening (only 13% of eligible patients within ≥ two years)[43]. Importantly, these studies focused on male patients, while in both BioVU and eMERGE, females made up 23-25% of the AAA cases, higher than the 17% observed in the UK Biobank risk prediction study[38].

A critical aspect of implementing predictive models that rely on multiple structured data elements and complex calculations is scalability. Compared with the USPSTF guidelines, which are straightforward to incorporate into clinical practice, implementing the models we present here would require that calculations be integrated into EHR systems. Ideally risk determinations would be presented to the clinical practitioner in real time during an encounter with a patient. Given the significant discrimination improvement over USPSTF criteria, and examples of implementation for other traits[44], we believe that real-time risk evaluation is feasible. Enhanced screening seems unlikely to lead to unnecessary invasive clinical procedures, as previous meta-analyses indicate that repair of small unruptured aneurysms had no advantage over routine ultrasound surveillance[45].

Recent studies have explored integration of imaging-derived parameters in prediction of AAA growth, rupture and mortality[46-49]. While our analyses rely on diagnostic codes and demographic information, our overarching goal is to identify potentially high-risk individuals for AAA screening via imaging. The goals of these approaches are distinct: identification of who is likely to develop AAA and who among AAA patients requires intervention. Restriction to extant structured data in the EHR improves the likelihood and feasibility of implementation of models in the clinical setting.

Our study is most limited by sample counts for most diverse racial/ethnic groups being too small to include as separate strata. This is concerning due to racial/ethnic differences in screening prevalence but also in clinical presentation, treatment, and mortality following surgical repair[19-21,50,51]. We were able to include NHB individuals in all phases of this analysis, and confirmed that performance of USPTF criteria is lower in this group[19,43,51], but that clinically meaningful prediction (AUC>0.8) were attainable using either basic covariates or medical diagnoses.

Our results in eMERGE NHB participants incorporating phecodes suggested that despite use of cross-validation, our models from BioVU were likely overfit due to sparseness of NHB participants relative to the number of terms estimated. Larger numbers of NHB participants would facilitate improved models, however, we observed good discriminative performance compared with USPTF.

Predictive models including a PRS optimized in NHB individuals resulted in models that performed nearly equally as well in NHW but provided modest improvements in NHB. This is unusual for genetic studies based solely on European-ancestry participants[52] but suggests that risk variants may persist across diverse populations, making prediction of events easier. Although the PRS alone was little better than chance at predicting AAA diagnosis, including covariates was sufficient to yield clinical utility[53]. Future work evaluating scalability and incorporating sex-stratified estimates into models will enhance quality of prediction and clinical implementation.

In summary, we provide predictive models and polygenic scores for AAA which strongly associated with and predict AAA risk in multiple populations. These models substantially improve identification of people at risk of a AAA diagnosis compared with existing guidelines.

References

1. Dua A, et al. Epidemiology of aortic aneurysm repair in the United States from 2000 to 2010. *J Vasc Surg.* 2014;59(6):1512-1517.
2. Summers KL, et al. Evaluating the prevalence of abdominal aortic aneurysms in the United States through a national screening database. *J Vasc Surg.* 2021;73(1):61-68.
3. Stuntz M. Modeling the Burden of Abdominal Aortic Aneurysm in the USA in 2013. *Cardiology.* 2016;135(2):127-131.
4. Benjamin EJ, et al. Heart Disease and Stroke Statistics-2019 Update: A Report From the American Heart Association. *Circulation.* 2019;139(10):e56-e528.
5. Lo RC, et al. Abdominal aortic aneurysms in women. *J Vasc Surg.* 2016;63(3):839-844.
6. Jahangir E, et al. Smoking, sex, risk factors and abdominal aortic aneurysms: a prospective study of 18 782 persons aged above 65 years in the Southern Community Cohort Study. *Journal of epidemiology and community health.* 2015;69(5):481-488.
7. Pleumeekers HJ, et al. Aneurysms of the abdominal aorta in older adults. The Rotterdam Study. *Am J Epidemiol.* 1995;142(12):1291-1299.
8. Ye Z, et al. Family history of atherosclerotic vascular disease is associated with the presence of abdominal aortic aneurysm. *Vasc Med.* 2016;21(1):41-46.
9. Portilla-Fernandez E, et al. Genetic and clinical determinants of abdominal aortic diameter: genome-wide association studies, exome array data and Mendelian randomization study. *Hum Mol Genet.* 2022.
10. Wahlgren CM, et al. Genetic and environmental contributions to abdominal aortic aneurysm development in a twin population. *J Vasc Surg.* 2010;51(1):3-7; discussion 7.
11. Klarin D, et al. Genetic Architecture of Abdominal Aortic Aneurysm in the Million Veteran Program. *Circulation.* 2020;142(17):1633-1646.
12. Jones GT, et al. Meta-Analysis of Genome-Wide Association Studies for Abdominal Aortic Aneurysm Identifies Four New Disease-Specific Risk Loci. *Circ Res.* 2017;120(2):341-353.
13. Bradley DT, et al. A variant in LDLR is associated with abdominal aortic aneurysm. *Circ Cardiovasc Genet.* 2013;6(5):498-504.
14. Bown MJ, et al. Abdominal aortic aneurysm is associated with a variant in low-density lipoprotein receptor-related protein 1. *Am J Hum Genet.* 2011;89(5):619-627.

15. Gretarsdottir S, et al. Genome-wide association study identifies a sequence variant within the DAB2IP gene conferring susceptibility to abdominal aortic aneurysm. *Nat Genet.* 2010;42(8):692-697.
16. Chaikof EL, et al. The Society for Vascular Surgery practice guidelines on the care of patients with an abdominal aortic aneurysm. *J Vasc Surg.* 2018;67(1):2-77.e72.
17. Kent KC. Clinical practice. Abdominal aortic aneurysms. *N Engl J Med.* 2014;371(22):2101-2108.
18. Guirguis-Blake JM, et al. Primary Care Screening for Abdominal Aortic Aneurysm: Updated Evidence Report and Systematic Review for the US Preventive Services Task Force. *Jama.* 2019;322(22):2219-2238.
19. Li SR, et al. Epidemiology of age-, sex-, and race-specific hospitalizations for abdominal aortic aneurysms highlights gaps in current screening recommendations. *J Vasc Surg.* 2022.
20. Deery SE, et al. Racial disparities in outcomes after intact abdominal aortic aneurysm repair. *J Vasc Surg.* 2018;67(4):1059-1067.
21. Williams TK, et al. Disparities in outcomes for Hispanic patients undergoing endovascular and open abdominal aortic aneurysm repair. *Ann Vasc Surg.* 2013;27(1):29-37.
22. Pulley J, et al. Principles of human subjects protections applied in an opt-out, de-identified biobank. *Clinical and translational science.* 2010;3(1):42-48.
23. Roden DM, et al. Development of a large-scale de-identified DNA biobank to enable personalized medicine. *Clinical pharmacology and therapeutics.* 2008;84(3):362-369.
24. Das S, et al. Next-generation genotype imputation service and methods. *Nat Genet.* 2016;48(10):1284-1287.
25. McCarthy S, et al. A reference panel of 64,976 haplotypes for genotype imputation. *Nat Genet.* 2016;48(10):1279-1283.
26. Wu P, et al. Mapping ICD-10 and ICD-10-CM Codes to Phecodes: Workflow Development and Initial Evaluation. *JMIR Med Inform.* 2019;7(4):e14325.
27. Wei WQ, et al. Evaluating phecodes, clinical classification software, and ICD-9-CM codes for phenome-wide association studies in the electronic health record. *PLoS One.* 2017;12(7):e0175508.
28. Gottesman O, et al. The Electronic Medical Records and Genomics (eMERGE) Network: past, present, and future. *Genetics in medicine : official journal of the American College of Medical Genetics.* 2013;15(10):761-771.
29. McCarty CA, et al. The eMERGE Network: a consortium of biorepositories linked to electronic medical records data for conducting genomic studies. *BMC medical genomics.* 2011;4:13.
30. Kurki MI, et al. FinnGen: Unique genetic insights from combining isolated population and national health register data. *medRxiv.* 2022:2022.2003.2003.22271360.
31. Willer CJ, et al. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics.* 2010;26(17):2190-2191.
32. Ge T, et al. Polygenic prediction via Bayesian regression and continuous shrinkage priors. *Nature communications.* 2019;10(1):1776.
33. Chang CC, et al. Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience.* 2015;4:7.
34. Manca R, et al. The neural signatures of psychoses in Alzheimer's disease: a neuroimaging genetics approach. *European archives of psychiatry and clinical neuroscience.* 2022.

35. Carroll RJ, et al. R PheWAS: data analysis and plotting tools for phenome-wide association studies in the R environment. *Bioinformatics.* 2014;30(16):2375-2376.
36. Simon N, et al. Regularization Paths for Cox's Proportional Hazards Model via Coordinate Descent. *J Stat Softw.* 2011;39(5):1-13.
37. Friedman J, et al. Regularization Paths for Generalized Linear Models via Coordinate Descent. *J Stat Softw.* 2010;33(1):1-22.
38. Welsh P, et al. Derivation and Validation of a 10-Year Risk Score for Symptomatic Abdominal Aortic Aneurysm: Cohort Study of Nearly 500 000 Individuals. *Circulation.* 2021;144(8):604-614.
39. Chaudhry R, et al. Use of a Web-based clinical decision support system to improve abdominal aortic aneurysm screening in a primary care practice. *J Eval Clin Pract.* 2012;18(3):666-670.
40. Hye RJ, et al. Leveraging the electronic medical record to implement an abdominal aortic aneurysm screening program. *J Vasc Surg.* 2014;59(6):1535-1542.
41. Eaton J, et al. Effect of visit length and a clinical decision support tool on abdominal aortic aneurysm screening rates in a primary care practice. *J Eval Clin Pract.* 2012;18(3):593-598.
42. Lee ES, et al. Implementation of an aortic screening program in clinical practice: implications for the Screen For Abdominal Aortic Aneurysms Very Efficiently (SAAAVE) Act. *J Vasc Surg.* 2009;49(5):1107-1111.
43. Anjorin AC, et al. Underutilization of Guideline-based Abdominal Aortic Aneurysm Screening in an Academic Health System. *Ann Vasc Surg.* 2022;83:184-194.
44. Pasley J. Predicting blood clots before they happen in pediatric patients. *VUMC Reporter.* May 28, 2021, 2021. https://news.vumc.org/2021/05/26/predicting-blood-clots-before-they-happen-in-pediatric-patients/.
45. Ulug P, et al. Surgery for small asymptomatic abdominal aortic aneurysms. *Cochrane Database Syst Rev.* 2020;7(7):Cd001835.
46. Dong H, et al. MR Elastography of Abdominal Aortic Aneurysms: Relationship to Aneurysm Events. *Radiology.* 2022;304(3):721-729.
47. Lorandon F, et al. Scannographic Study of Risk Factors of Abdominal Aortic Aneurysm Rupture. *Ann Vasc Surg.* 2021;73:27-36.
48. Jalalzadeh H, et al. Estimation of Abdominal Aortic Aneurysm Rupture Risk with Biomechanical Imaging Markers. *J Vasc Interv Radiol.* 2019;30(7):987-994.e984.
49. Hirata K, et al. Machine Learning to Predict the Rapid Growth of Small Abdominal Aortic Aneurysm. *J Comput Assist Tomogr.* 2020;44(1):37-42.
50. Ribieras AJ, et al. Racial disparities in presentation and outcomes for endovascular abdominal aortic aneurysm repair. *J Vasc Surg.* 2022.
51. Barshes NR, et al. Racial and ethnic disparities in abdominal aortic aneurysm evaluation and treatment rates in Texas. *J Vasc Surg.* 2022;76(1):141-148.e141.
52. Martin AR, et al. Clinical use of current polygenic risk scores may exacerbate health disparities. *Nature genetics.* 2019;51(4):584-591.
53. Lambert SA, et al. Towards clinical utility of polygenic risk scores. *Hum Mol Genet.* 2019;28(R2):R133-r142.