Title

Clinical consequences of a genetic predisposition toward higher benign prostate-specific antigen levels.

Permalink

https://escholarship.org/uc/item/2fw799vh

Authors

Shi, Mingjian
Shelley, John
Schaffer, Kerry
et al.

Publication Date

2023-10-19

DOI

10.1016/j.ebiom.2023.104838

Peer reviewed

Clinical consequences of a genetic predisposition toward higher benign prostate-specific antigen levels

Mingjian Shi,[a,g] John P. Shelley,[a,g] Kerry R. Schaffer,[b,c] Jeffrey J. Tosoian,[c,d] Minoo Bagheri,[a,b] John S. Witte,[e,f] Linda Kachuri,[e] and Jonathan D. Mosley[a,b,*]

[a]Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, TN, USA
[b]Department of Medicine, Vanderbilt University Medical Center, Nashville, TN, USA
[c]Vanderbilt–Ingram Cancer Center, Vanderbilt University Medical Center, Nashville, TN, USA
[d]Department of Urology, Vanderbilt University Medical Center, Nashville, TN, USA
[e]Department of Epidemiology and Population Health, Stanford University, Stanford, CA, USA
[f]Department of Biomedical Data Science and Genetics (by Courtesy), Stanford University, Stanford, CA, USA

Summary
Background Prostate-specific antigen (PSA) levels are influenced by genetic variation unrelated to prostate cancer risk. Whether a genetic predisposition to a higher PSA level predisposes to a diagnostic work-up for prostate cancer is not known.

Methods Participants were 3110 men of African and European ancestries ages 45–70, without prostate cancer and with a baseline PSA < 4 ng/mL, undergoing routine clinical PSA screening. The exposure was a polygenic score (PGS) comprising 111 single nucleotide polymorphisms associated with PSA level, but not prostate cancer. We tested whether the PGS was associated with a: 1) PSA value > 4 ng/mL, 2) International Classification of Diseases (ICD) code for an elevated PSA, 3) encounter with a urologist, or 4) prostate biopsy. Multivariable Cox proportional hazards models were adjusted for age and genetic principal components. Analyses were stratified by age (45–59 years, and 60–70 years old). Association estimates are per standard deviation change in the PGS.

Findings The median age was 56.6 years, and 2118 (68%) participants were 45–59 years. The median (IQR) baseline PSA level was 1.0 (0.6–1.7) ng/mL. Among men ages 45–59, the PGS was associated with a PSA > 4 (hazard ratio [HR] = 1.35 [95% CI, 1.17–1.57], p = 4.5 × 10^{-5}), an ICD code for elevated PSA (HR = 1.30 [1.12–1.52], p = 8.0 × 10^{-4}), a urological evaluation (HR = 1.34 [1.14–1.57], p = 4.8 × 10^{-4}), and undergoing a prostate biopsy (HR = 1.35 [1.11–1.64], p = 0.002). Among men ages 60–70, association effect sizes were smaller and not significant.

Interpretation A predisposition toward higher PSA levels was associated with clinical evaluations of an elevated PSA among men ages 45–59 years.

Funding National Institutes of Health (NIH).

Keywords: Prostate specific antigen (PSA); Polygenic score (PGS); Prostate cancer; Associations; Cox proportional hazards regression

Introduction

Screening biomarkers are used to identify early disease.[1] For instance, men ages 45–70 years may undergo measurement of prostate-specific antigen (PSA) levels in blood, and those with levels ≥4 ng/mL may receive a referral to a urologist for a diagnostic prostate biopsy to identify prostate cancer.[2] Importantly, PSA is not a mediator of disease, and its clinical use is strictly as a biomarker.[3,4] An important limitation of screening biomarkers is that their levels may be influenced by factors unrelated to the target disease.[5,6] Variability unrelated to disease may lead to escalations in clinical care with negative consequences, such as overdiagnosis of clinically insignificant disease or initiation of diagnostic evaluations that are unlikely to identify disease.[7] For instance, the positive predictive value for identifying

---

Research in context

Evidence before this study

We searched Pubmed, Google scholar and medRxiv for the terms "polygenic score" and ("PSA" or "prostate-specific antigen" or "prostate specific antigen") to identify studies examining clinical applications of a polygenic score (PGS) for PSA levels. Genome-wide association studies have identified 128 single nucleotide polymorphism (SNPs) associated with PSA levels, and methods have been developed to identify PSA-associated SNPs that do not contribute to prostate cancer risk. One study examined a SNP-based PGS for PSA and found that it could recalibrate measured PSA levels and reclassify some individuals to lower PSA levels that no longer met screening thresholds. Adjustment led to a net reclassification of a higher proportion of individuals who had a negative prostate biopsy, as compared to those with a positive biopsy. These findings suggest that a genetic predisposition toward elevated PSA levels could increase the risk of undergoing a diagnostic evaluation for an elevated PSA levels.

Added value of this study

This study examined a real-world retrospective cohort of men undergoing PSA screening as part of routine clinical care. We measured men's genetically determined PSA levels using a PSA PGS comprising SNPs associated with PSA levels that are unlikely to be associated with prostate cancer. We confirmed that this PGS was not associated with identifying prostate cancer on biopsy. This predisposition was associated with an increased risk for a range of clinical endpoints including referral to a urologist and undergoing a prostate biopsy. This study demonstrates that a genetic predisposition toward higher benign PSA levels associates with escalations in diagnostic evaluations but not an elevated prostate cancer risk.

Implications of all the available evidence

The implications of these findings are that accounting for measurable benign genetic variation could avoid escalations in clinical care that are apt to cause more harm than benefit. These findings have direct applications to a broad range of screening biomarkers, which frequently have a significant polygenic component.

prostate cancer on prostate biopsy with a PSA $\geq$ 4 ng/mL is approximately 30%.[8] Since this non-disease related variability cannot be directly measured using traditional epidemiological methods, its clinical significance is not well-characterized.

One source of biomarker variability is heritable genetic variation that is unrelated to disease.[9] Genome wide association studies (GWAS) have identified over 128 common single nucleotide polymorphisms (SNPs) that explain 8% of PSA variation.[10] Many of these SNPs are not related to prostate cancer risk but rather modulate homeostatic physiological mechanisms.[11] An individual's genetic predisposition can be quantitated using a polygenic score (PGS) which measures the aggregative effects of SNPs carried by the individual.[12] We hypothesized that a PGS measuring an individual's genetically-determined baseline PSA level could be used to measure the clinical consequences of a genetic predisposition toward higher benign PSA levels.

We examined a real-world retrospective clinical cohort of men of African and European ancestries receiving routine PSA screening for prostate cancer to define the impacts of PSA genetic variation unrelated to prostate cancer on clinical care. Specifically, we determined the association between a PSA polygenic score (PGS) comprising 111 SNPs that were not associated with prostate cancer and the occurrence of incident outcomes relevant to diagnostic evaluations of an elevated PSA level. Incorporating methods to account for measurable benign genetic variability could improve the quality and value of cancer screening markers.

Methods

Study populations

The study populations were from the Vanderbilt University Medical Center (VUMC) BioVU resource, a DNA biobank comprising approximately ~270,000 consented participants and linked to a de-identified electronic health record (EHR).[13] Two populations were constructed among participants with existing genome-wide single nucleotide polymorphism (SNP) genotyping and who were evaluated between 1997 and 2022. The first cohort (the "biopsy cohort") was used to verify that the PSA PGS did not associate with a prostate cancer diagnosis and was derived from a previously described multi-ancestry cohort of 655 men who underwent a prostate biopsy for the first time.[14] Briefly, that study identified men of European or African ancestry ages 40–80 years without a prior history of prostate cancer who were referred to a urologist and underwent a subsequent prostate biopsy (the "biopsy cohort"). Men with a PSA value $\geq$ 25 ng/mL, a prior prostate biopsy, a history of organ transplant, or who were on testosterone replacement were excluded. Detailed information on the urology visit and biopsy results were manually extracted for each patient. These analyses were restricted to a subset of 468 men (407 European ancestry [EA] and 61 African ancestry [AA]) aged 45–70 years who underwent a first prostate cancer biopsy by a urologist for a clinical concern of an elevated PSA.

The second cohort (the "longitudinal cohort") was used to evaluate incident clinical outcomes relevant to PSA screening and evaluation in a population of men of

European and African ancestries without baseline prostate cancer who were undergoing annual prostate cancer screening. This cohort was derived from 10,305 men with at least 1 available PSA measurement. Participants outside the standard screening age range (age 45 and 70 years) were excluded. To remove men with known or likely prostate cancer at baseline and men who were referred to VUMC for a concerning PSA level measured at an outside institution, participants were excluded if they had a prostate cancer diagnosis within 30 days of their first recorded PSA measurement, a urology visit within 14 days, a first PSA measurement ≥4 ng/mL or a history of organ transplant. To exclude men likely undergoing active or high risk surveillance, rapidly lost to follow-up or likely receiving care at outside institutions, participants were excluded if they had ≥3 PSA measurements per year (suggestive of surveillance for a known prostate cancer), or had only 1 PSA measurement within their first 3 years of follow-up. After exclusions, there were 3110 participants. There were 138 participants who were in both cohorts.

These studies were evaluated and determined to be non-human subjects research by the VUMC Institutional Review Board.

Genetic data

All participants had prior genotyping on Illumina's Expanded Multi-Ethnic Genotyping Array (MEGA^{EX}) platform. Participants were excluded if genetic quality control analyses indicated a female sex, >4% missing genetic data, or excessive heterozygosity. One of each pair of related individuals (pi-hat > 0.2 for both EA and AA participants) were excluded. Genetic ancestry was determined by principal components analysis in conjunction with HAPMAP reference populations and individuals who did not fall within European or African ancestral groups were excluded. Prior to imputation, data were pre-phased using Eagle v2.4.1.[15] Imputation was performed using the Michigan Imputation Server in conjunction with the 10/2014 release of the 1000 Genomes cosmopolitan reference haplotypes. Imputed data were filtered for a sample missingness rate <2%, a SNP missingness rate <4% and SNP deviation from Hardy-Weinberg $p < 10^{-6}$. Genetic principal components (PCs) to adjust for population stratification were calculated using the SNPRelate package.[16]

PSA and prostate cancer polygenic predictors

A PSA polygenic score (PGS) comprising SNPs not associated with prostate cancer was developed using data derived from a GWAS of 95,768 men without prostate cancer.[10] That study identified 128 independent SNPs significantly associated with PSA levels. Some of these SNPs were at least nominally associated with prostate cancer in prior GWAS,[17] in many instances due to index event bias whereby an elevated PSA level

initiated a work-up for prostate cancer. Index event bias-corrected association estimates for the 128 SNPs were provided in supplementary table 10 of the PSA GWAS paper.[10] In our analysis, we excluded SNPs which were nominally associated with prostate cancer in unadjusted analyses (p < 0.001) and had the same direction of association with prostate cancer as with PSA levels with a nominal association (p < 0.05) after index event bias adjustment. After exclusions, there were 111 SNPs that were used in the PGS for this study. The effect size of the association with PSA for each SNP was based on a subset of the PSA GWAS that did not include BioVU participants (Supplementary Table S1). The PGS-PSA$_{111}$ was computed for each participant as the sum of the product of the genotype dosage and the SNP effect on PSA.

We also examined a previously validated multi-ancestry polygenic risk score for prostate cancer comprising 269 common SNP variants (PRS-PCA$_{269}$).[17]

Clinical data

For the biopsy cohort, the age at biopsy and whether the biopsy identified any prostate cancer or high grade cancer (Gleason Grade [GG] ≥ 2) were extracted, as previously described.[14] Two cancer outcomes were defined. The first outcome was the presence of any prostate cancer on biopsy with controls defined as participants without cancer. The second outcome was the presence of a GG ≥ 2 cancer with controls defined as participants with either no cancer or low grade (GG = 1) cancer.

For the longitudinal cohort, race and gender, as recorded in the EHR, and all PSA measurements were extracted from data tables in the EHR. The following incident outcomes were defined: (1) time to a first PSA measurement >4 ng/mL; (2) time to first assignment of an International Classification of Disease [ICD]-9 (790.93) or ICD-10 (R97.20) code for an elevated PSA level; (3) time to a first clinic appointment with a urologist in conjunction with a PSA > 4; (4) time to a first incident prostate biopsy in conjunction with a PSA > 4. For the urologist and biopsy incident outcomes, individuals who had an event, but a PSA ≤ 4 at the time of the event were not counted as having an outcome and were censored at the time of the visit. For these outcomes, a sensitivity analysis was also performed whereby individuals were defined as having the outcome if they had an ICD code for an elevated PSA level on or before the event.

Secondary incident outcomes examined were benign prostatic hyperplasia (BPH) (defined by codes from ICD-9: 600.0, 600.00, 600.01, 600.1, 600.10, 600.11, 600.2, 600.20, 600.21, 600.9, 600.90, 600.91 and ICD-10: N40.0, N40.1, N40.2, N40.3) and prostatitis (defined by codes from ICD-9: 601, 601.0, 601.00, 601.1, 601.2, 601.3, 601.4, 601.8, 601.9, 98.12, 98.32, 131.03 and ICD-10: N41.0, A18.14, A54.22, A59.02, B38.81).

For each outcome examined, subjects were censored from the study at the time of the event or 12 months after their last PSA measurement. If there was a duration of greater than 3 years between any two PSA measurements, participants were censored 12 months after the start of the gap.

Analysis

The PGS-PSA$_{111}$ and PRS-PCA$_{269}$ were normalized to have a mean of 0 and standard deviation of 1. Association statistics represent the change per standard deviation change in the PGS. As genetic effects are often more apparent among younger individuals,[17] results are presented for all participants and are stratified by age: participants aged 45–59 years old and 60–70 years old.

To determine whether the PGS-PSA$_{111}$ and PRS-PCA$_{269}$ were associated with cancer risk, a multivariable logistic regression model was used to test the association between each predictor and the outcomes of 1) any cancer and a 2) GG > 2 cancer in the biopsy cohort. All models were adjusted for age at biopsy and 10 PCs.

Baseline characteristics for the entire longitudinal cohort and for age strata were computed. We also present the characteristics of participants excluded from this cohort (Supplementary Table S5). For continuous variables, median values and interquartile ranges were computed.

In the longitudinal cohort, the partial correlation (adjusted for age and 10 PCs) between the PGS-PSA$_{111}$ and log-transformed baseline PSA level was calculated. The distribution of the variance explained by each individual SNP is presented in Supplementary Fig. S2.

Cox proportional hazards regression was used to determine the association between both the PGS-PSA$_{111}$ and PRS-PCA$_{269}$ and incident events of 1) PSA > 4, 2) an elevated PSA ICD code, 3) a urology visit, and 4) undergoing a prostate biopsy. Each model was adjusted for age at baseline and 10 PCs. To ensure appropriate model fit, Schoenfeld residuals were examined to test proportional hazards assumptions and Martingale residuals were examined to test nonlinearity. Results were visualized using Kaplan–Meier plots, stratifying the cohort into low (>1 standard deviation below the median), middle (≤1 s.d. below and ≥1 s.d. above the median) and high (>1 s.d. above the median) PGS groups.

For all analyses, two-sided statistical tests were used and a nominal p < 0.05 was considered significant. Analyses used the RStudio v. 4.0.0 tools and packages. These analyses followed the STREGA guidelines.[18]

A simulation is presented in Supplementary Methods demonstrating how a genetic predictor of a biomarker (like the PGS-PSA$_{111}$) that is not associated with cancer risk in the general population can manifest as an inverse association with cancer among individuals with an elevated biomarker level.

Role of funders

The funding organizations did not have a role in the design, analysis or interpretation of this study.

Results

Association with biopsy outcomes

To verify that the PGS-PSA$_{111}$ was not associated with increased cancer risk, we tested the association with incident biopsy findings in a cohort of 468 men who underwent a first prostate biopsy for an elevated PSA. There were 236 (50%) men with any prostate cancer identified on biopsy and 108 (23%) men with a high grade (GG ≥ 2) cancer on biopsy (Supplementary Table S2). The PGS-PSA$_{111}$ was significantly inversely associated with an outcome of any prostate cancer (OR = 0.77 [95% CI, 0.62–0.95], p = 0.02) and with high grade cancer (OR = 0.70 [0.55–0.89], p = 0.004) (Fig. 1 and Supplementary Table S3). In contrast, the PRS-PCA$_{269}$ was positively associated with these outcomes (OR = 1.96 [1.60–2.44], p = 3.3×10^{-10}) (Supplementary Table S3).

Thus, a genetic predisposition to higher PSA levels (i.e., a higher score on PGS-PSA$_{111}$) was associated with a decreased likelihood of finding prostate cancer on a biopsy.

Association with incident outcomes

To determine whether the PGS-PSA$_{111}$ was associated with incident events related to prostate cancer screening and diagnosis, we identified a cohort of 3110 male participants with a baseline PSA < 4 undergoing routine clinical screening of PSA (Fig. 2). The cohort was derived from 7936 genotyped males ages 45–70 years old with a PSA measurement. Characteristics of participants excluded from this cohort are presented in Supplementary Table S5. The median age was 56.6
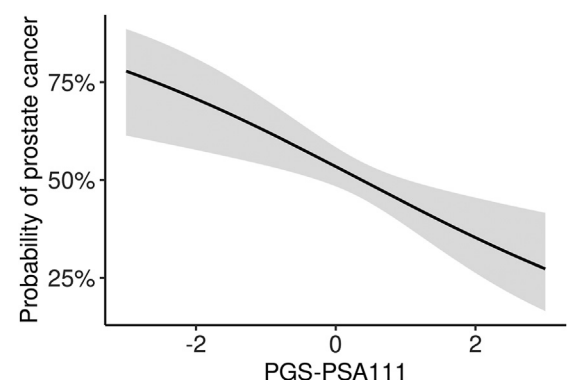

Fig. 1: Predicted probabilities of prostate cancer on biopsy by PGS-PSA$_{111}$ value. Predicted probabilities are from a multivariable logistic regression model, adjusted for age at biopsy and 10 genetic principal components, among 468 men who underwent a first biopsy.
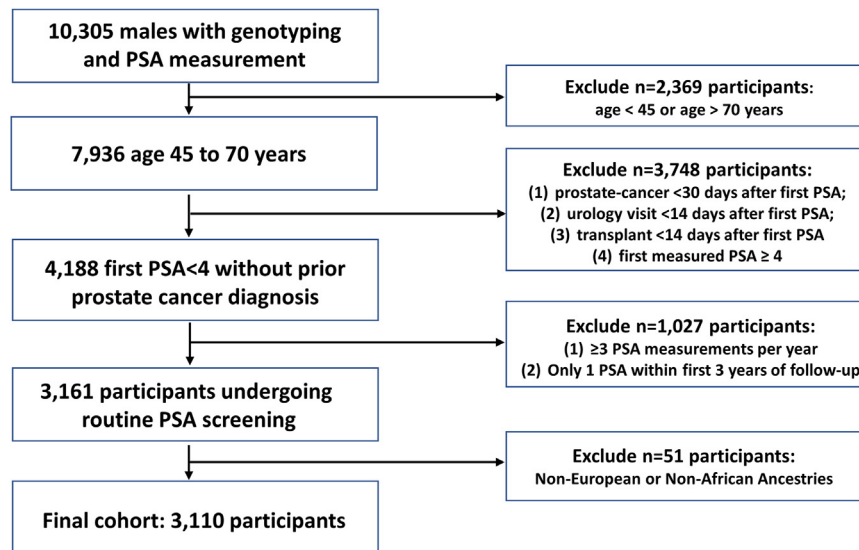
```
┌─────────────────────────────┐
│ 10,305 males with genotyping │
│ and PSA measurement          │
└─────────────────────────────┘
              │                      ┌──────────────────────────────────────────┐
              │─────────────────────▶│ Exclude n=2,369 participants:            │
              ▼                      │ age < 45 or age > 70 years               │
┌─────────────────────────────┐     └──────────────────────────────────────────┘
│ 7,936 age 45 to 70 years     │
└─────────────────────────────┘     ┌──────────────────────────────────────────┐
              │                      │ Exclude n=3,748 participants:            │
              │─────────────────────▶│ (1) prostate-cancer <30 days after first PSA; │
              ▼                      │ (2) urology visit <14 days after first PSA; │
┌─────────────────────────────┐     │ (3) transplant <14 days after first PSA  │
│ 4,188 first PSA<4 without prior │  │ (4) first measured PSA ≥ 4               │
│ prostate cancer diagnosis    │     └──────────────────────────────────────────┘
└─────────────────────────────┘
              │                      ┌──────────────────────────────────────────┐
              │─────────────────────▶│ Exclude n=1,027 participants:            │
              ▼                      │ (1) ≥3 PSA measurements per year         │
┌─────────────────────────────┐     │ (2) Only 1 PSA within first 3 years of follow-up │
│ 3,161 participants undergoing │    └──────────────────────────────────────────┘
│ routine PSA screening        │
└─────────────────────────────┘     ┌──────────────────────────────────────────┐
              │                      │ Exclude n=51 participants:               │
              │─────────────────────▶│ Non-European or Non-African Ancestries   │
              ▼                      └──────────────────────────────────────────┘
┌─────────────────────────────┐
│ Final cohort: 3,110 participants │
└─────────────────────────────┘
```

Fig. 2: Selection of the PSA longitudinal cohort.

(interquartile range [IQR], 51.4–61.5) years and 2118 participants were age 45–59. The median follow-up duration was 6.2 (3.7–10.9) years, and the median PSA value was 1.0 (IQR, 0.6–1.7) ng/mL (Table 1).

The PGS-PSA$_{111}$ was positively correlated with baseline PSA levels (partial correlation [r] = 0.23, p < 2.2×10^{-16}, Supplementary Fig. S1). Partial correlation point estimates were higher among men 45–59 years, as compared to older men (r = 0.26 versus r = 0.18).

There were 440 (14%) participants who developed a PSA > 4 and 530 (17%) participants assigned an ICD diagnosis code for an elevated PSA level. Among participants ages 45–59, the PGS-PSA$_{111}$ was significantly

associated with developing a PSA > 4 (hazard ratio [HR] = 1.35 [95% CI, 1.17–1.57], p = 4.5×10^{-5}) and being assigned an ICD code for elevated PSA (HR = 1.30 [95% CI, 1.12–1.52], p = 8.0×10^{-4}), respectively (Figs. 3 and 4, Table 2). Associations were in the same direction but not significant for the 60–69 year old age group (HR = 1.14 [95% CI, 0.99–1.31], p = 0.07 and HR = 1.08 [95% CI, 0.92–1.26], p = 0.35).

There were 332 (11%) participants who saw a urologist and had a PSA > 4. The PGS-PSA$_{111}$ was significantly associated with this outcome among the younger (hazard ratio [HR] = 1.34 [95% CI, 1.14–1.57], p = 4.8×10^{-4}) but not older participants (HR = 1.13 [95% CI, 0.96–1.33], p = 0.14) (Table 2, Fig. 4). There

Characteristic	All participants	Participants age 45–59	Participants age 60–69
n	3110	2118	992
Age at first PSA (years)	56 (51–62)	53 (50–56)	61 (58–65)
PSA (ng/mL)	1 (0.58–1.7)	0.9 (0.6–1.5)	1.1 (0.6–2.0)
Follow-up duration (years)	6.2 (3.7–10.9)	6.3 (3.7–11.5)	6.1 (3.7–9.9)
PSA measurements per participant	5 (3–9)	5 (3–10)	5 (3–9)
Incident outcomes:			
PSA >4 ng/mL	440 (14.1%)	236 (11.1%)	204 (20.6%)
Elevated PSA ICD code	530 (17.0%)	309 (14.6%)	221 (22.3%)
Urology visit with PSA > 4 ng/mL	332 (10.7%)	182 (8.6%)	150 (15.1%)
Prostate biopsy with PSA > 4 ng/mL	226 (7.3%)	133 (6.3%)	93 (9.4%)
Benign prostatic hyperplasia (BPH)	806 (25.9%)	486 (22.9%)	320 (32.3%)
Prostatitis	144 (4.6%)	92 (4.3%)	52 (5.2%)

[a]For continuous values, values are median (interquartile range). For categorical values, values are n (%).

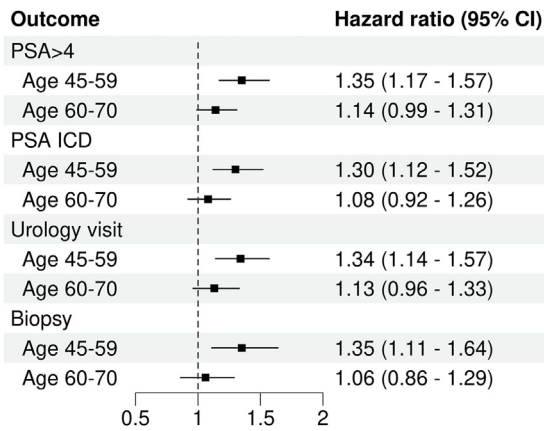*Table 1:* Population characteristics by participant age.[a]

Outcome		Hazard ratio (95% CI)
PSA>4		
Age 45-59		1.35 (1.17 - 1.57)
Age 60-70		1.14 (0.99 - 1.31)
PSA ICD		
Age 45-59		1.30 (1.12 - 1.52)
Age 60-70		1.08 (0.92 - 1.26)
Urology visit		
Age 45-59		1.34 (1.14 - 1.57)
Age 60-70		1.13 (0.96 - 1.33)
Biopsy		
Age 45-59		1.35 (1.11 - 1.64)
Age 60-70		1.06 (0.86 - 1.29)

Fig. 3: **Associations between the PGS-PSA_{111} and incident outcomes.** Hazard-ratio estimates are from a multivariable Cox regression model, adjusted for age and 10 genetic principal components, among 3110 men undergoing PSA screening. Hazard ratios are per standard deviation change in the PGS-PSA_{111}.

were 226 (7%) participants who underwent a prostate biopsy with a PSA > 4. Similarly, the PGS-PSA_{111} was significantly associated with this outcome among the younger (hazard ratio [HR] = 1.35 [95% CI, 1.11–1.64], p = 2.3×10^{-3}) but not older participants (HR = 1.06 [95% CI, 0.86–1.29], p = 0.60) (Fig. 4). Similar results were seen when the outcome was either a urology visit or biopsy in conjunction with a history of having an ICD code for an elevated PSA (Table 2). The magnitude and

direction of all associations were similar when stratified by genetic ancestry (Supplementary Table S6). There were not significant associations between the PGS-PSA_{111} and outcomes of BPH or prostatitis (Supplementary Table S7).

Similar to the PGS-PSA_{111}, the PRS-PCA_{269} demonstrated positive associations with all incident outcomes (Supplementary Table S4). Furthermore, these associations were significant in both the younger and older age groups.

In sum, among individuals ages 45–59, a genetic predisposition unrelated to prostate cancer to higher PSA levels was associated with an increased risk of a diagnosis of an elevated PSA level and undergoing a urological evaluation and prostate biopsy.

Discussion

Many screening biomarkers, such as PSA, are often paired with a clinical decision threshold that is used to trigger a diagnostic evaluation. A one-size-fits-all approach to defining decision thresholds results in some individuals undergoing escalations in diagnostic care due to an intrinsic predisposition to elevated biomarker levels. One approach to address this is to develop genetically-informed reference ranges that account for an individual's intrinsic baseline. An essential first step toward implementing such an approach is to develop instruments that measure an individual's genetic predisposition and demonstrate that these instruments associate with clinical activity in the absence
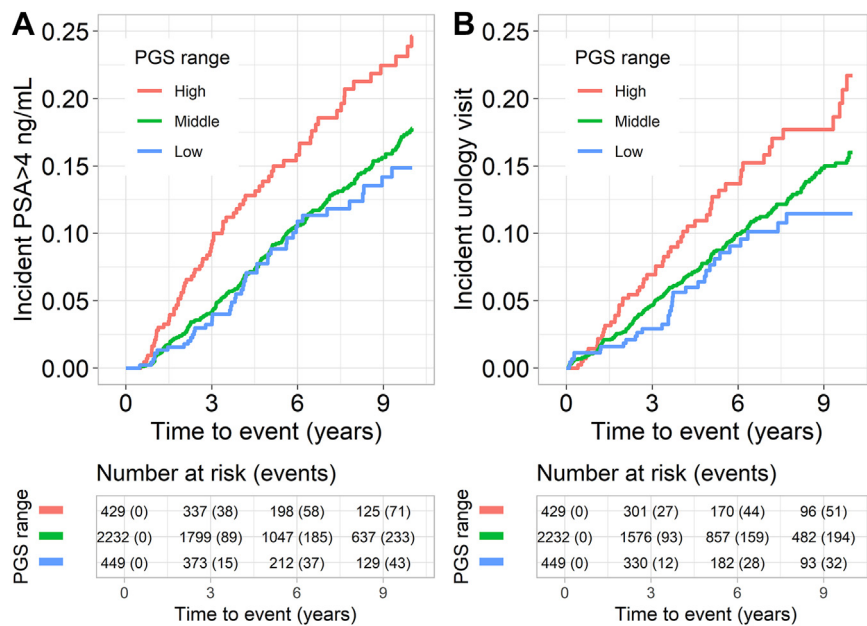

Fig. 4: **Kaplan–Meier curves of incident outcomes by PGS-PSA_{111} strata for participants 45–59 years old.** Kaplan–Meier curves for development of (a) a PSA > 4 ng/mL and (b) seeing a urologist with a PSA > 4 ng/mL among 2118 men. The PGS-PSA_{111} strata are: Low (<1 s.d. below the mean), Middle (≥−1 s.d. and ≤1 s.d.), and High (>1 s.d.).

Outcome	Age group[a]	Hazard ratio[b]	95% CI	p-value
PSA > 4 ng/mL	All	1.22	(1.11–1.35)	7.6×10^{-5}
	45–59	1.35	(1.17–1.57)	4.5×10^{-5}
	60–70	1.14	(0.99–1.31)	0.07
ICD code for elevated PSA	All	1.18	(1.06–1.31)	2.9×10^{-3}
	45–59	1.30	(1.12–1.52)	8.0×10^{-4}
	60–70	1.08	(0.92–1.26)	0.35
Urology visit[c]				
Urology visit (PSA > 4)	All	1.23	(1.09–1.38)	5.2×10^{-4}
	45–59	1.34	(1.14–1.57)	4.8×10^{-4}
	60–70	1.13	(0.96–1.33)	0.14
Urology visit (PSA ICD)	All	1.16	(1.01–1.32)	0.03
	45–59	1.22	(1.02–1.46)	0.03
	60–70	1.12	(0.91–1.37)	0.28
Prostate biopsy[c]				
Prostate biopsy (PSA > 4)	All	1.20	(1.04–1.38)	0.01
	45–59	1.35	(1.11–1.64)	2.3×10^{-3}
	60–70	1.06	(0.86–1.29)	0.60
Prostate biopsy (PSA ICD)	All	1.13	(0.99–1.28)	0.07
	45–59	1.22	(1.03–1.46)	0.02
	60–70	1.02	(0.84–1.25)	0.82

[a]Results are shown for all participants and strata representing men ages 45–59 years and 60–70 years old. [b]Cox proportional hazards models are adjusted for age at first PSA measurement and 10 PCs. [c]Outcomes are for a urology visit or prostate biopsy and having a PSA > 4 or having an ICD code for an elevated PSA level.

Table 2: Associations between the PGS-PSA$_{111}$ and incident outcomes.

of disease. Toward this end, we found that a polygenic predisposition to higher benign PSA levels was associated with an increased risk for incident outcomes of a PSA > 4, a urology evaluation for an elevated PSA, and a prostate biopsy. In sum, these data suggest that a PSA polygenic score could enable personalized reference ranges that avoid futile diagnostic odysseys.

Twin studies have suggested that up to 45–55% of PSA variability may be due to heritable factors, and heritability estimates based on common variants are 30–40%.[10,19,20] The most significant associations are among SNPs near the gene encoding PSA itself (*KLK3*) and, in this study, these SNPs explained the largest portion of PSA variation, suggesting that at least a portion of this genetic variation is not associated with prostate cancer risk. However, many SNPs associated with PSA also associate with prostate cancer, likely due to index event (collider) bias, a form of detection bias that results from the analysis of prostate cancer cases that were initially identified by an elevated PSA.[21]

To avoid spurious associations between a PSA PGS and cancer risk, our analyses used PSA-associated SNPs which did not associate with prostate cancer after correction for index event bias.[10] To confirm that the PGS-PSA$_{111}$ was not measuring prostate cancer risk, we tested the association between the PGS-PSA$_{111}$ and the presence of cancer on biopsy. The PGS-PSA$_{111}$ was inversely associated with a finding of cancer, indicating that a polygenic predisposition toward higher PSA was not associated with cancer. In contrast, a prostate cancer polygenic risk score was positively associated with outcomes in this cohort. This pattern of association is due to a well-described collider bias related to ascertaining prostate cancer status among individuals with an elevated PSA level.[22,23] This phenomenon was demonstrated in the data simulation. Thus, an inverse association would be expected for a genetic predictor that measures benign PSA elevations as these elevations are not due to underlying cancer. They also suggest that genetically-determined PSA levels would adversely affect the positive predictive value of the PSA biomarker.[10]

Our findings suggest that males under 60 may be more adversely impacted by a benign predisposition to elevated PSA. The age-dependency of the genetic association is likely because polygenic variation unrelated to prostate cancer accounts for a larger portion of total PSA variation among younger males. Consistently, we found that the partial correlations between measured PSA levels and the PGS-PSA$_{111}$ were higher among younger males (0.26 versus 0.18).

The clinical consequences of a genetic predisposition to higher benign PSA levels may include more frequent PSA testing or as demonstrated here, an increased likelihood of undergoing a diagnostic evaluation. Regardless of outcome, increased diagnostic evaluations can lead to increased anxiety among those identified as elevated risk.[24,25] Unnecessary invasive procedures also increase the risk of adverse outcomes without an offsetting benefit.[26] These procedures may also lead to

overdiagnosis which can result in increased expenditures, serial biopsies, or nonbeneficial treatments for clinically insignificant disease.[27]

There are strengths to the current study. By utilizing over two decades of clinical data, we were able to examine real-world prospective outcomes among a multi-ancestry population undergoing prostate cancer screening with PSA. This work demonstrates that polygenic variation in a cancer screening biomarker is associated with important clinical consequences and may lead to inappropriate escalations in clinical care. Thus, we show that understanding genetic determinants of biomarker variation may be important for improving the performance of screening biomarkers. Applying this knowledge could improve the quality of care especially with respect to false positive PSA elevations, increasing the positive predictive value of invasive diagnostic testing and avoiding overdiagnosis of indolent prostate cancers.

This study also has limitations. Our retrospective cohort is from a single center. It is possible that patients also received care at other institutions which may result in under ascertainment of outcomes of interest. However, all providers and patients were blinded to genotype, which mitigates the likelihood that ascertainment of outcome would differ by genotype status. Standards of care with respect to an elevated PSA have evolved over the time periods examined in this study, which can limit the generalizability of these findings. However, the consistency of the findings across a range of relevant outcomes supports the overall merit of our findings. Our conservative approach to deriving the PGS-PSA$_{111}$ removed variants that were nominally associated with prostate cancer which reduced the predictive power of the score and excluded PSA-associated SNPs that may simply be more pleiotropic.

In conclusion, a genetic predisposition toward higher benign PSA levels was associated with escalations in clinical care related to evaluating an elevated PSA among men 45–59 years. Accounting for polygenic variation in a screening biomarker could avoid escalations in diagnostic evaluations.

Contributors
JDM, MS, and KRS conceptualized the design of the study. MS and JDM established the methodology for this study. MS curated the data, performed the primary analyses, and produced data visualizations. JDM acquired the primary funding for this study, validated the analyses, and supervised all steps involved. MS, JPS, and JDM wrote the first draft of the manuscript. MS, JPS, KRS, JJT, MB, JW, LK, and JDM critically reviewed the manuscript for important intellectual content, edited the manuscript for clarity, and provided final approval of the version to be published.

Data sharing statement
Upon publication, data sets of individual-level phenotype data and corresponding data dictionaries to replicate the primary findings presented here for research purposes will be made available upon request from the repository (biovu@vumc.org). Institutional IRB approval, data use agreements, and administrative and scientific reviews are required prior to using individual-level data from BioVU.

Appendix A. Supplementary data
Supplementary data related to this article can be found at https://doi.org/10.1016/j.ebiom.2023.104838.

References
1 Califf RM. Biomarker definitions and their applications. *Exp Biol Med (Maywood)*. 2018;243:213–221.
2 Balk SP, Ko Y-J, Bubley GJ. Biology of prostate-specific antigen. *J Clin Oncol*. 2003;21:383–391.
3 Pezaro C, Woo HH, Davis ID. Prostate cancer: measuring PSA. *Intern Med J*. 2014;44:433–440.
4 Adhyam M, Gupta AK. A review on the clinical utility of PSA in cancer prostate. *Indian J Surg Oncol*. 2012;3:120–129.
5 Enroth S, Johansson A, Enroth SB, Gyllensten U. Strong effects of genetic and lifestyle factors on biomarker variation and use of personalized cutoffs. *Nat Commun*. 2014;5:4684.
6 McCormack JP, Holmes DT. Your results may vary: the imprecision of medical measurements. *BMJ*. 2020;368:m149.
7 Loeb S, Bjurlin MA, Nicholson J, et al. Overdiagnosis and overtreatment of prostate cancer. *Eur Urol*. 2014;65:1046–1055.
8 Wolf AMD, Wender RC, Etzioni RB, et al. American Cancer Society guideline for the early detection of prostate cancer: update 2010. *CA Cancer J Clin*. 2010;60:70–98.
9 Sinnott-Armstrong N, Tanigawa Y, Amar D, et al. Genetics of 35 blood and urine biomarkers in the UK Biobank. *Nat Genet*. 2021;53:185–194.
10 Kachuri L, Hoffmann TJ, Jiang Y, et al. Genetically adjusted PSA levels for prostate cancer screening. *Nat Med*. 2023;29:1412–1423.
11 Hoffmann TJ, Passarelli MN, Graff RE, et al. Genome-wide association study of prostate-specific antigen levels identifies novel loci independent of prostate cancer. *Nat Commun*. 2017;8:14248.
12 Choi SW, Mak TS-H, O'Reilly PF. Tutorial: a guide to performing polygenic risk score analyses. *Nat Protoc*. 2020;15:2759–2772.
13 Roden DM, Pulley JM, Basford MA, et al. Development of a large-scale de-identified DNA biobank to enable personalized medicine. *Clin Pharmacol Ther*. 2008;84:362–369.
14 Schaffer KR, Shi M, Shelley JP, et al. A polygenic risk score for prostate cancer risk prediction. *JAMA Intern Med*. 2023;183(4):386–388.
15 Pasaniuc B, Zaitlen N, Shi H, et al. Fast and accurate imputation of summary statistics enhances evidence of functional enrichment. *Bioinformatics*. 2014;30:2906–2914.
16 Zheng X, Levine D, Shen J, Gogarten SM, Laurie C, Weir BS. A high-performance computing toolset for relatedness and

principal component analysis of SNP data. *Bioinformatics.* 2012;28:3326–3328.

17 Conti DV, Darst BF, Moss LC, et al. Trans-ancestry genome-wide association meta-analysis of prostate cancer identifies new susceptibility loci and informs genetic risk prediction. *Nat Genet.* 2021;53:65–75.

18 Little J, Higgins JPT, Ioannidis JPA, et al. Strengthening the reporting of genetic association studies (STREGA): an extension of the STROBE Statement. *Hum Genet.* 2009;125:131–151.

19 Bansal A, Murray DK, Wu JT, Stephenson RA, Middleton RG, Meikle AW. Heritability of prostate-specific antigen and relationship with zonal prostate volumes in aging twins. *J Clin Endocrinol Metab.* 2000;85:1272–1276.

20 Pilia G, Chen W-M, Scuteri A, et al. Heritability of cardiovascular and personality traits in 6,148 Sardinians. *PLoS Genet.* 2006;2:e132.

21 Dahabreh IJ, Kent DM. Index event bias as an explanation for the paradoxes of recurrence risk research. *JAMA.* 2011;305:822–823.

22 Langston ME, Sfanos KS, Khan S, et al. Why do epidemiologic studies find an inverse association between intraprostatic inflammation and prostate cancer: a possible role for colliding bias? *Cancer Epidemiol Biomark Prev.* 2021;30:255–259.

23 Harrison S, Tilling K, Turner EL, et al. Systematic review and meta-analysis of the associations between body mass index, prostate cancer, advanced prostate cancer, and prostate-specific antigen. *Cancer Causes Control.* 2020;31:431–449.

24 McNaughton-Collins M, Fowler FJ, Caubet J-F, et al. Psychological effects of a suspicious prostate cancer screening test followed by a benign biopsy result. *Am J Med.* 2004;117:719–725.

25 Fowler FJ, Barry MJ, Walker-Corkery B, et al. The impact of a suspicious prostate biopsy on patients' psychological, socio-behavioral, and medical care outcomes. *J Gen Intern Med.* 2006;21:715–721.

26 Pinsky PF, Parnes HL, Andriole G. Mortality and complications after prostate biopsy in the prostate, lung, colorectal and ovarian cancer screening (PLCO) trial. *BJU Int.* 2014;113:254–259.

27 Etzioni R, Penson DF, Legler JM, et al. Overdiagnosis due to prostate-specific antigen screening: lessons from U.S. prostate cancer incidence trends. *J Natl Cancer Inst.* 2002;94:981–990.