Stockholm3 validation study in a multi-**E**thnic cohort for **P**ros**TA**te cancer (SEPTA): Statistical Analysis Plan

SEPTA co-authors:

Hari Vigneswaran[a,b], Martin Eklund[a], Andrea Discacciati[a], Tobias Nordström[a], Rebecca Hubbard[c], Nathan Perlis[d], Michael Abern[b,e], Daniel Moreira[b], Scott Eggener[f], Paul Yonover[g], Alexander Chow[h], Kara Watts[i], Michael Liss[j], Gregory Thoreson[k], Andre Abreu[l], Geoffrey A. Sonn[m], Thorgerdur Palsdottir[a], Anna Plym[a], Fredrik Wiklund[a], Henrik Grönberg[a] Adam Murphy[n], on behalf of the SEPTA STHLM3 Study Group

[a] Karolinska Institutet, Department of Medical Epidemiology and Biostatistics
[b] University of Illinois at Chicago, Department of Urology
[c]University of Pennsylvania Perelman School of Medicine, Center for Clinical Epidemiology and Biostatistics
[d]University of Toronto, University Health Network, Department of Surgery
[e]Duke University School of Medicine, Department of Surgery
[f]University of Chicago, Department of Surgery
[g]Uropartners, LLC
[h]Rush University Medical Center, Department of Urology
[i]Montefiore Medical Center, Department of Urology
[j]University of Texas Health Science Center at San Antonio, Department of Urology
[k]Urology Clinics of North Texas
[l]University of Southern California Keck School of Medicine, Department of Urology
[m]Stanford University School of Medicine, Department of Urology
[n]Northwestern University Feinberg School of Medicine, Department of Urology

Revisions

Date	Authors	Reason	Main changes
190917	HV/ME/HG	Statistical analysis plan formulated (SAP)	
230412	HV/ME/AD/TN/RH/TP/AP/FW/HG	Additional aims	Additional aims, data structure, power analysis clarified
230512	HV/ME/AD/TN/RH/NP/MA/DM/SE/PY/AC/KW/ML/GT/AA/GS/TP/AP/FW/HG/AM	Review/consensus	

Contents

Abbreviations

SAP	Statistical analysis plan
SEPTA	Stockholm3 Validation Study in a multi-ethnic cohort for prostate cancer
PSA	Prostate specific antigen
DRE	Digital rectal examination
MRI	Magnetic resonance imaging
HIPAA	Health Insurance Portability and Accountability Act
PIPEDA	Personal Information Protection and Electronic Documents Act
GDPR	General Data Protection Regulation
SNP	single nucleotide polymorphism
csPC	clinically significant prostate cancer
ISUP	International Society of Urological Pathology
ROC	Receiver Operating Characteristics
AUC	Area under the curve
ERSPC	European Randomized study of Screening for Prostate Cancer
PBCG	Prostate Biopsy Collaborative Group
ICPCG	International Consortium for Prostate Cancer Genetics
PRACTICAL	Prostate Cancer Association Group to Investigate Cancer Associated Alterations in the Genome
GWAS	Genome wide association studies
SDI	social deprivation index
CIMD	Canadian Index of Multiple Deprivation
PI-RADS	Prostate imaging reporting and data system
PP	Per protocol
TPR	True positive rate (sensitivity)
TNR	True negative rate (specificity)

Preface

This statistical analysis plan (SAP) describes the planned analyses for the Stockholm3 Validation Study in a multi-ethnic cohort for prostate cancer (NCT04583072) (herein referred to as SEPTA for short). SEPTA is a validation study of the Stockholm3 test in a multi-ethnic cohort from North America. The proposed study is an analysis of men that are prospectively identified (in standard clinical care) as screened positive to undergo a prostate biopsy for suspicion of prostate cancer. The men will undergo Stockholm3 testing prior to biopsy for retrospective validation of Stockholm3. The planned analysis in this SAP will be included in future manuscripts, exploratory analyses not necessarily identified in this SAP may also be performed. The SAP may be updated throughout the course of the study; however, it will be finalized before the database lock, or any comparative analyses are performed.

Design

At each site, men will be scheduled to undergo a prostate biopsy for suspicion of prostate cancer based on abnormal prostate specific antigen (PSA), digital rectal exam (DRE), magnetic resonance imaging (MRI) or other suspicious clinical findings. Men with a prior prostate cancer diagnosis will be excluded. The aim of this study is to include 500 men of four self-described race/ethnicity groups as defined by the US Office of Management and Budget Standards including: White/Caucasian non-Hispanic/non-Latino, Black/African American, White/Caucasian Hispanic/Latino and Asian for a total of 2,000 men in the entire SEPTA trial. Men will be assigned to one of the four study groups based on their self-reported race/ethnicity. Prospective collection specifically for SEPTA will be guided by the recruitment aims to include 500 men in each of the four self-described race/ethnicities; each respective subgroup will close when the respective groups reach this milestone. There are 14 sites included (Northwestern University, Cook County Hospital, Jessie Brown Veterans Affairs in Chicago, University of Illinois at Chicago, University of Chicago, Uropartners, Rush University Medical Center, Montefiore Medical Center, University of Texas Health Science Center at San Antonio, University Health Network in Toronto, Urology Clinics of North Texas, University of Southern California, Los Angeles County Hospital, Stanford University) all located in North America (United States and Canada). After identification and before biopsy sampling, the men will undergo whole blood venipuncture. Clinical data and blood samples will be collected. The transportation of the de-identified blood samples for Stockholm3 and ancestry informative marker analysis will be sent to A3P lab in Uppsala, Sweden. All participants' samples will be treated in accordance with the HIPAA compliance in the United States and PIPEDA compliance in Canada during specimen and data collection and the GDPR of the European Union in Sweden and at A3P lab. Outcome data regarding the biopsy results will be blinded to those performing the lab assays and reporting Stockholm3 and ancestry informative marker results. The Stockholm3 test, as described previously[1-3], includes predictors: age, first-degree family history of prostate cancer [yes/no/don't know], and previous biopsy [yes/no/don't know], total PSA, free PSA, hK2 (aka KLK2), MIC1 (aka

GDF15), MSMB (aka PSP94), and a genetic risk score based on single nucleotide polymorphisms (SNP). The model will be performed with and without the inclusion of prostate volume.

The design of the study allows for evaluation of prostate cancer detection with Stockholm3, Stockholm3 with contemporary race and ethnicity-specific prostate risk SNPs, self-identified race and ethnicity, calculated ancestry, PSA, free to total PSA ratio (percent free PSA), PSA density, DRE, and clinical risk calculators such as the European Randomized study of Screening for Prostate Cancer (ERSPC) or Prostate Biopsy Collaborative Group (PBCG) in an ethnically enriched North American cohort. These men are conditionally enrolled after referral to biopsy for prostate cancer suspicion based on PSA, MRI, or other clinical risk. Subgroup analysis allows for the same comparisons in the four racial/ethnic groups as well as in sub cohorts simulating real world racial and ethnic demographic population distributions. For clarity, ancestry is based on objective genetic markers and race/ethnicity is based on self-report.[4, 5]

Study populations

1. *Study population* includes all men who:
    a. meet the criteria for a diagnostic prostate biopsy as a part of routine practice,
    b. signed the written informed consent to participate in SEPTA or for use of previously stored bio-banked material,
    c. are self-identified Black/African American, White/Caucasian Hispanic/Latino, Asian or White/Caucasian Non-Hispanic/non-Latino men and
    d. aged 45-75 years old.
2. Conditions for excluding patients from the study population, based on deviations from the study protocol:
    a. Previous prostate cancer diagnosis,
    b. men that have undergone DRE within 5 days of blood draw or
    c. men who in the three months prior to blood draw, start new treatment for benign prostatic hyperplasia or received any invasive urologic procedure such as thermotherapy, microwave therapy, laser therapy, transurethral resection of the prostate, urethral catheterization, and lower genitourinary tract endoscopy (cystoscopy).

Hypotheses

Overarching hypothesis

The overarching hypothesis of the SEPTA trial is that the Stockholm3 shows non-inferior sensitivity in detection of clinically significant prostate cancer (csPC) (defined as International Society of Urological Pathology (ISUP) Gleason grade group ≥2 prostate cancer (ISUP ≥2)) and superior specificity (i.e., reduction of positive Stockholm3 tests versus positive PSA tests in men with negative biopsies or detected with ISUP Gleason grade group 1 only (ISUP 1) tumors) in a multi-ethnic North

American setting (the SEPTA cohort) compared to using PSA as it has shown in other validation studies in Northern Europe.[6-8]

Additional hypotheses

1. Novel genetic variants associated with prostate cancer from contemporary multiethnic genome wide association studies incorporated into the genetic risk component of Stockholm3 will improve the Stockholm3 model in the entire SEPTA cohort and within self-described race/ethnicity subgroups (evaluated analogously to the primary hypothesis)
2. Within the pre-defined four racial/ethnic subgroups, the Stockholm3 shows non-inferior sensitivity in detection of csPC and superior specificity (analogous to the primary hypothesis).
3. Stockholm3 may perform differently in terms of calibration over the different self-identified race/ethnicity groups and may have different cutoffs at fixed relative sensitivities to PSA with respective specificities.
4. Within a subgroup of SEPTA that simulate the US demographic, the Stockholm3 shows non-inferior sensitivity in detection of csPC and superior specificity (analogous to the primary hypothesis).
5. Within a subgroup of SEPTA that did/did not undergo MRI, the Stockholm3 shows non-inferior sensitivity in detection of csPC and superior specificity (analogous to the primary hypothesis).
6. Stockholm3 will have improved discrimination over total PSA in detection of csPC in the overall SEPTA cohort and over the different self-described race/ethnicity groups.
7. Self-described race/ethnicity will have decreased impact on the Stockholm3 sensitivity and specificity for prediction of csPC, after adjusting for socioeconomic factors (median household income and social deprivation), age, and PSA.
8. Self-described race/ethnicity will have a decreased effect on the presence or absence of csPC after adjusting for socioeconomic factors (median household income and social deprivation), age, and PSA.
9. A granular SNP-derived ancestry may be more predictive than self-described Race/ancestry for the detection of csPC, after adjusting for socioeconomic factors (median household income and social deprivation), age, and PSA.
10. The Stockholm3 model and Stockholm3 with contemporary SNPs has superior discrimination for detection of csPC compared to free to total PSA ratio (precent free PSA), PSA density, DRE, and clinical risk calculators such as the European Randomized study of Screening for Prostate Cancer (ERSPC) or Prostate Biopsy Collaborative Group (PBCG).

Aims

Primary aim:

The primary aim is a combined aim to compare both sensitivity and specificity:

- To evaluate within SEPTA (n = 2000) both non-inferior sensitivity of Stockholm3 (at a threshold of ≥15) in detection of csPC compared to using PSA (at a threshold of ≥4 ng/mL) (if non-inferiority holds true, superiority will be evaluated) AND
- To evaluate within SEPTA (n = 2000) superior specificity (I.e., reduction of positive Stockholm3 tests versus positive PSA tests in men with negative biopsies or detected with ISUP 1 tumors) of Stockholm3 (at a threshold of ≥15) compared to using PSA (at a threshold of ≥4 ng/mL).

PSA cut-off of ≥4 ng/mL was based on the American Cancer Society guideline for the early detection of prostate cancer and the commonly accepted referral pattern for biopsy in the US.[9] Stockholm3 cut-off of ≥15 was based on the clinical application for increased specificity in the STHLM3-MRI trial.[10]

Additional aims

1. To evaluate input and replacement of a novel polygenetic risk score with SNPs based on a contemporary multiethnic genome wide association study for the detection of csPC with Stockholm3 (analogous analysis to primary aim performed)
2. To use SEPTA to show non-inferior sensitivity of Stockholm3 (at a threshold of ≥11%) in detection of csPC and superior specificity (I.e., reduction of positive Stockholm3 tests versus positive PSA tests in men with negative biopsies or detected with ISUP 1 tumors) compared to using PSA (at a threshold of ≥3 ng/mL). If non-inferiority holds true, superiority will be evaluated.
3. To evaluate the primary aim, and additional aim #1 in the sub-cohorts of the four self-identified race/ethnicity groups
4. If Stockholm3 performs differently in terms of calibration over the different self-identified race/ethnicity groups evaluate different cutoffs to optimize sensitivity and specificity (i.e., fixed relative sensitivities to PSA).
5. To evaluate the primary aim, additional aim #1 and #2 in the sub-cohorts of SEPTA weighted to resemble the US demographic distribution.
6. To evaluate the primary aim and additional aim #1 and #2 in the sub-cohorts of SEPTA in which men did/did not undergo MRI prior to biopsy.
7. To show superior AUC in detection of csPC for Stockholm3 compared to PSA in the SEPTA cohort as a whole and in all four self-identified racial/ethnic subgroups.
8. To evaluate the effects of socioeconomic variables (median household income, social deprivation index), self-identified race/ethnicity, PSA, and age on Stockholm3 sensitivity and specificity for detection of csPC and/or to assess mediation of the self-identified race/ethnicity effect on Stockholm3 sensitivity and specificity for detection of csPC by socioeconomic variables, age, and PSA.
9. To evaluate the effects of socioeconomic variables (median household income, social deprivation index), self-identified race/ethnicity, PSA, and age on the presence or absence of csPC and/or to assess mediation of the self-identified race/ethnicity effect on the presence or absence of csPC by socioeconomic variables, age, and PSA.
10. To calculate a granular genetic ancestry using a validated ancestry SNP model and evaluate the impact of calculated ancestry, socioeconomic variables (median household income, social deprivation index), PSA and age in a logistic regression model with the outcome or detection probability of csPC or no csPC.
11. To compare the AUC and operating characteristics of the original Stockholm3 without volume, Stockholm3 with contemporary SNPs (additional aim #1) without volume, Stockholm3 with volume, free to total PSA ratio (percent free PSA), PSA density, DRE, and risk calculators such as the European Randomized study of Screening for Prostate Cancer (ERSPC) or Prostate Biopsy Collaborative Group (PBCG).

Other aims

Other additional aims can be added at a later stage of the study.

Endpoints

The definitions of study endpoints and variables (independent variables and outcome variables) are described below.

Primary endpoint

Variable	Measure	Comment
Clinically significant prostate cancer (csPC)	Yes/No	ISUP ≥ 2 (if there are multiple grades of PC reported from prostate biopsy, the grade is defined as the highest of all grades reported), see additional endpoints for alternative definition of csPC

Key secondary endpoint

Variable	Measure	Comment
Non-clinically significant prostate cancer **OR** non-cancer (benign) biopsy performed	Yes/No	ISUP = 1 (if there are multiple grades of PC reported from prostate biopsy, the grade is defined as the highest of all grades reported, non-clinically significant prostate cancer requires ISUP 1 detection in the absence of other PC grades) **OR** no cancer is detected after prostate biopsy

Additional endpoints

Variable	Measure	Comment
Non-clinically significant prostate cancer	Yes/No	ISUP = 1 (if there are multiple grades of PC reported from prostate biopsy, the grade is defined as the highest of all grades reported, non-clinically significant prostate cancer requires ISUP 1 detection in the absence of other PC grades)
Non-cancer (benign) biopsy performed	Yes/No	No cancer is detected after prostate biopsy
ISUP ≥ 3 prostate cancer	Yes/No	ISUP ≥ 3 (if there are multiple grades of PC reported from prostate biopsy, the grade is defined as the highest of all grades reported)
Cancer length	mm	Total mm of highest-grade cancer (if there are multiple grades of PC reported from prostate biopsy, the grade is defined as the highest of all grades reported) reported on prostate biopsy
Percentage of positive biopsy cores on systematic biopsy	Percentage	Number of reported biopsy cores positive for cancer in men with biopsy results of ISUP ≥2 cancer in relation to the number of total cores taken on systematic biopsy only, excluded if targeted only biopsies were performed
Suspicious MRI findings	Yes/No	If available, PIRADS ≥3 and/or PIRADS ≥4

Independent Variables

Variable	Measure	Comment
Age	Years	At blood sampling date

Race	1. Black/African American (Hispanic/non-Hispanic/Afro-Caribbean/Central-South African/West African/unknown) 2. White/Caucasian Hispanic/Latino 3. Asian (South/East/Southeast/Indian subcontinent/unknown) 4. White/Caucasian non-Hispanic/non-Latino	Patient self-reported race at study start
SNP-based Ancestry	1. American 2. East Asian 3. North Asian 4. South Central Asian 5. Southwest Asian 6. Pacific 7. Sub-Saharan African 8. Northeast African 9. North African 10. South European 11. North European	Calculated using principal component analysis and clustering by continental regions[11]
ZIP code based median income	Continuous (numeric)	Patient self-reported residential zip code-based income using 2021 US and Canada census/tax data. Income data may be separated into terciles based on low, middle, and upper class based on Pew Research Center 2021 survey data[12]
Zip code based social deprivation index (SDI)	Integer (0-100)	Calculated using zip code (USA), SDI is a composite measure of area level deprivation based on seven demographic characteristics.[13] Calculated using zip code (Canada), Canadian Index of Multiple Deprivation based on four associated dimensions of deprivation, updated CAN-Marg.[14] Increasing index correlates with increased deprivation. Index outcomes may be reported on respective quintile or terciles
Previous prostate biopsy	Yes/No	Patient self-reported
Family history of prostate cancer	Yes/No/Don't know	Self-reported, Any first degree relative with prostate cancer
PSA	ng/ml	At blood test
Stockholm3 risk score	Integer	At blood test
prostate volume	ml	MRI or ultrasound defined, can be represented as PSA density defined as the PSA quotient of prostate volume
PI-RADS (If available)	1–5 (integer)	Maximum PI-RADS score
DRE status	positive/negative/not measured	At visit to urologist, “not measured” may be considered normal for purposes of binary evaluation
Ongoing use of 5-alpha reductase inhibitors	Yes/No/Don't know	Patient self-reported within 3 months of study start

Statistical analysis

A description of how and on which data statistical testing will be performed is specified below. A3P lab personnel and Karolinska Institutet personnel that will calculate Stockholm3 risk score, ancestry, and the contemporary genetic risk scores will not have access to any outcome data or any extraneous patient characteristics. Analyses will begin after recruitment is closed and the last outcome data is input into the database. All statistics, including tables, figures, and listings, will be performed using R version >4.1.

The analyses will be performed and reported on the Per Protocol (PP) population. PP population includes men who:

1. have a valid PSA value and Stockholm3 score,
2. have a primary diagnosis from a systematic and/or targeted prostate biopsy, and
3. have a registered self-reported race/ethnicity.

Data structure and Analysis

The tables below lay out the general data structure for the SEPTA trial. The data will be structured in the following format but will not be limited to this structure.

Patients' characteristics

Patients' characteristics will be presented with descriptive statistics, overall, by race/ethnicity, and/or by screening test (positive/negative), as appropriate. Continuous variables will be summarized using measures of central tendency and variability. Categorical variables will be summarized using absolute and relative frequencies. No formal statistical testing will be performed (Table 1).

Table 1

	All (N = …)	*Caucasian/ White (n = …)*	*African American/ Black (n = …)*	*Hispanics/ Latino (n = …)*	*Asian (n = …)*
Age, years (median, IQR)					
Self-identified Race/ethnicity: *Caucasian/ White (N, %), African American/ Black (N, %), Hispanics/ Latino (N, %), Asian (N, %)*		-	-	-	-
SNP-based Ancestry *American (N, %), East Asian (N, %), North Asian (N, %), South Central Asian (N, %), Southwest Asian (N, %), Pacific (N, %), Sub-Saharan African (N, %), Northeast African (N, %), North African (N, %), South European (N, %), North European (N, %)*					
PSA (ng/ml) (median, IQR)					

DRE abnormal (N, %)					
Family history of prostate cancer (N, %)					
Previous negative prostate biopsy (N, %)					
Stockholm3 (median, IQR)					
Underwent MRI before biopsy (N, %)					
PSA density (ng/ml^2) (median, IQR)					
Free/total PSA (%) (median, IQR)					
***European Randomized study of Screening for Prostate Cancer** (ERSPC) Risk Calculator (risk of csPC)*					
Prostate Biopsy Collaborative Group (PBCG) Risk Calculator (risk of csPC)					
Median household income (Terciles based on Pew Research Center 2021 survey) **Low (less than $52,000) (N, %)** **Middle ($52,200-156,600) (N, %)** **High (More than $156,600) (N, %)**					
Measure of increasing deprivation: Social deprivation index (SDI)/ Canadian Index of Multiple Deprivation (CIMD) tercile (n, %) **1st tercile: % (n), 2nd tercile: % (n), 3rd tercile: % (n)**					
Underwent MRI (N, %) **PIRADS ≥3 (N, %)** **PIRADS ≥4 (N, %)** **PIRADS score missing (N, %)**					
Benign biopsy (N, %)					
ISUP 1 Prostate Cancer (N, %)					
ISUP ≥2 Prostate Cancer (N, %)					
ISUP ≥3 Prostate Cancer (N, %)					

Primary analysis

Stockholm3 sensitivity (at a threshold of ≥15) to detect csPC in the multi-ethnic SEPTA cohort will be compared to PSA sensitivity (at a threshold of ≥4 ng/mL) in relative terms (relative sensitivity, with PSA as the referent test). Non-inferiority of Stockholm3 sensitivity will be declared if the p-value obtained from a one-sided test with non-inferiority margin for relative sensitivity equal to 0.8 is smaller than alpha=0.025 (see hypothesis set “Power analysis 1”). Equivalently, non-inferiority will be declared if the lower bound of the two-sided 95% confidence interval around the relative sensitivity estimate is above 0.8.

Stockholm3 specificity (at a threshold of ≥15) to detect benign biopsies and ISUP1 cancers will be compared to PSA specificity (at a threshold of ≥4 ng/mL) in relative terms (relative specificity, with PSA as the referent test). Superiority of Stockholm3 specificity will be declared if the p-value obtained from a one-sided test (superiority margin equal to 1.0 for relative specificity) is smaller than alpha=0.025 (see hypothesis set "Power analysis 2"). Equivalently, superiority will be declared if the lower bound of the two-sided 95% confidence interval around the relative specificity estimate is above 1.0.

Asymptotic standard errors for relative sensitivity and specificity estimates will be derived according to formulas in Alonzo et al.[15]

The proportion of men with a positive Stockholm3 test (at a threshold of ≥15) or a positive PSA test (at a threshold of ≥4 ng/mL) among those diagnosed with csPC in the multi-ethnic SEPTA cohort will be calculated (sensitivity). The proportion of men with a negative Stockholm3 test or a negative PSA test among those with a benign biopsy or ISUP 1 tumors diagnosed will be calculated (specificity). (Table 3).

Table 2: Primary aim evaluating relative sensitivity and relative specificity of Stockholm3 to PSA

Strategy	***Threshold***	***Performed Biopsies***		***Cancer Detection***			
				ISUP Grade Group ≥2		*ISUP Grade Group 1 or benign biopsies*	
		n	*% (95% CI)*	*n*	*Relative Sensitivity (95% CI)*	*n*	Relative Specificity (95% CI)
PSA	*≥4*						
Stockholm3	*≥15*						
Stockholm3 with contemporary SNPs	*≥15*						

Table 3: Operating and performance characteristics of Stockholm3 and PSA

	Threshold	*Avoided ISUP grade 1 and benign biopsies, n (%)*	*Avoid ISUP grade 1 detection, n (%)*	*Specificity*	*NPV*	*Detected ISUP grade ≥2, n (%)*	*Missed ISUP ≥2 PC, n (%)*	*Missed ISUP ≥3 PC, n (%)*	*Sensitivity*	*PPV*
All	*None*	*0, (0%)*	*0, (0%)*	*0%*	-	*n, (100%)*	*0, (0%)*	*0, (0%)*	*100%*	-
PSA	*3*									
	4									
	...									
Stockholm3	*11%*									
	15									
	...									
Stockholm3 with contemporary SNPS	*11%*									
	15									
	...									

Additional Analysis

Stratified analysis by self-identified race and ethnicity will be conducted using analogous methods to the primary analysis. In addition, stratified analysis with different clinical diagnostic thresholds, in the men without MRI, and the men that were specifically recruited for SEPTA (not bio-banked) may be conducted using analogous methods to the primary analysis.

Novel genetic variants associated with prostate cancer and aggressive prostate cancer will be investigated based on contemporary multi-ethnic genome wide association studies. For these novel variants, an allelic odds ratio (OR) and allele frequency will contribute to a genetic risk score. The genetic risk score will be calculated by summing the number of risk alleles (0,1, or 2) at each of the SNPs multiplied by the logarithm of that SNP's OR based on the most recent GWAS study. These novel variants may first be investigated in an independent cohort for validation to be used in the SEPTA cohort.

Operating and performance characteristics (sensitivity, specificity, positive and negative predictive values, as well as avoided biopsies) may be computed at different thresholds for Stockholm3 and PSA (Table 3). Relative positive and negative predictive values will be calculated in the overall SEPTA cohort and the four multi-ethnic subgroups using various thresholds. Thresholds may be determined from fixed sensitivity points such as 80%, 90%, and 95% relative sensitivity. We will perform ROC curve analysis and compute the AUC in the SEPTA cohort and within SEPTA subgroups. We will compare the AUCs for Stockholm3 and PSA using DeLong test[16] to test for differences and may also compare PSA density, free/total PSA ratio, DRE, and clinical risk calculators such as the European Randomized study of Screening for Prostate Cancer (ERSPC) or Prostate Biopsy Collaborative Group (PBCG). Graphical calibration analyses as well as calibration-in-the-large to model the detection probability, true positive fraction given covariates may be performed.[17] Decision curve analysis may be performed.[18] We will employ standard generalized linear models or marginal models,[19] as appropriate to measure impact of subgroups on operating characteristics.

Ancestry will be calculated using a validated ancestry SNP-based model.[11] The SNP allele frequencies will be determined and principal component analysis (PCA) with clustering by continental region will be performed. Based on validated instruments, at least 11 geographic clusters will be evaluated, but these 11 clusters may be further condensed to fewer but broader clusters.

If differences in sensitivity or specificity of Stockholm3 or the prevalence of presence/absence of csPC are observed across self-identified race/ethnicity subgroups, we will investigate factors contributing to these differences using logistic regression and causal mediation analysis. Analyses of sensitivity will be conducted using the outcome of Stockholm3 ≥15 in the subset of men with ISUP ≥2; analyses of specificity will be conducted using the outcome of Stockholm3 < 15 in the subset of men with ISUP = 1 or no cancer; analyses of presence/absence of csPC will be conducted in the full cohort using the outcome of ISUP ≥2. For each of these three analyses, we will first estimate the

association between the outcome and race/ethnicity, age, PSA, median household income, and SDI using logistic regression. This analysis will provide an estimate of the independent effect of each variable. We will then conduct a causal mediation analysis to decompose the effect of self-identified race/ethnicity into the direct effect and the indirect effect mediated by median household income and SDI.[20] This will provide an estimate of the proportion of the total race/ethnicity effect that is mediated by these covariates.

Power analysis

Four analyses were used to evaluate the power:

1. Non-inferior sensitivity (TPR) of Stockholm3 ≥15 versus PSA ≥4 ng/ml to detect csPC (Appendix, Figure 1);
2. Superior specificity (TNR) of Stockholm3 ≥15 versus PSA ≥4 ng/ml (Appendix, Figure 2);
3. Jointly non-inferior TPR and superior TNR of Stockholm3 ≥15 versus PSA ≥4 ng/ml (Appendix, Figure 3);
4. Heterogeneity in relative TPR across the four race/ethnicity groups (Appendix, Table 1).

Power analysis 1

We evaluated the power to detect a non-inferior TPR for Stockholm3 ≥15 versus PSA≥4 ng/ml to detect csPC.

The non-inferiority margin for the lower bound confidence interval of the relative TPR (rTPR) was set to 0.8 with an alpha set to 0.025. The non-inferiority margin was based on an analogous relative margin to the PRECISION trial, with clinically and statistically important differences in csPC detection. The sample size was determined with cost and feasibility consideration. The number of enrolled participants included 2000 for the Overall analysis (ie, the four race/ethnicity groups combined) and 500 for each of the four race/ethnicity groups.

The null (H_0) and alternative hypothesis (H_a) were:

$$H_0: rTPR \leq 0.8$$
$$H_a: rTPR > 0.8$$

We explored 180 scenarios *(Supplementary Figure 1).* These scenarios were characterized by a constant PSA ≥4 TPR and rTPR across the four race/ethnicity groups. The csPC detection rate was group dependent. For the Overall analysis, the csPC detection rate was given by the average of the race/ethnicity group-specific detection rates.

Power analysis 2

We evaluated the power to detect a superior TNR for Stockholm3 ≥15 versus PSA ≥4 ng/ml.

Alpha was set to 0.025, and the number of enrolled participants to 500 for each of the four race/ethnicity groups and 2000 for the Overall analysis.

The null (H_0) and alternative hypothesis (H_a) were:

$$H_0: rTNR \leq 1$$
$$H_a: rTNR > 1$$

We explored 240 scenarios *(Supplementary Figure 2)*. These scenarios were characterized by a constant PSA ≥4 TNR and rTNR across the four race/ethnicity groups. The complement to 1 of csPC detection rate (ie, the probability of a negative or ISUP=1 biopsy result) was group dependent. For the Overall analysis, the complement to one of the csPC detection rate was given by the average of the race/ethnicity group-specific rates.

Power analysis 3

We evaluated the power to jointly detect a non-inferior TPR and a superior TNR for Stockholm3 ≥15 versus PSA ≥4 ng/ml.

The alpha value for each of the two statistical tests used to test this hypothesis was set to 0.025. The number of enrolled participants was set to 500 for each of the four race/ethnicity groups and 2000 for the Overall analysis.

The null (H_0) and alternative hypothesis (H_a) were:

$$H_0: rTPR \leq 0.8 \text{ or } rTNR \leq 1$$
$$H_a: rTPR > 0.8 \text{ and } rTNR > 1$$

Given the independence of the data on which rTPR and rTNR are estimated, this joint power is given by the product of the power from Analysis 1 and Analysis 2. The type I error rate of rejecting the null hypothesis if the null hypothesis was true (ie, incorrectly rejecting either or both null hypotheses for Analysis 1 and 2) is $1 - (1 - 0.025)^2 \approx 0.05$.

We explored 2880 scenarios *(Supplementary Figure 3)*, obtained by forming all pairwise combinations within race/ethnic groups of the scenarios defined for Analysis 1 and 2 and then discarding those with incompatible csPC detection rates.

Power analysis 4

We evaluated the power to detect heterogeneity in TPR across the four race/ethnicity groups when present.

Alpha was set to 0.05. The number of enrolled participants was set to 500 for each of the four race/ethnicity groups and 2000 for the Overall analysis.

The null (H_0) and alternative hypothesis (H_a) were:

$$H_0 \quad : rTPR^{B/AA} = rTPR^{W/C} = rTPR^{H/L} = rTPR^{A}$$
$$H_a \quad : \text{at least two } rTPR \text{ are different}$$

Where (B/AA), Black/African American; (W/C), White/Caucasian; (H/L), Hispanic/Latino Caucasian; (A), Asian.

We also assessed the power to detect a non-inferior TPR (non-inferiority margin = 0.8) for the Overall analysis (same hypothesis set as for Analysis 1) in the presence of rTPR heterogeneity.

We explored 30 scenarios generated by combining one scenario per race/ethnicity group randomly selected from those used in Analysis 1.

5 Other statistical and computational considerations

Power was assessed via statistical simulation (1500 simulations for each scenario) *(Supplementary Table 1)*. The R code to reproduce the analyses is available at: https://github.com/anddis/septa.

The number of simulations was chosen so that 95% asymptotic (Wald) confidence intervals for the estimated rejection proportions (power) for Analysis 1 and Analysis 2 are never wider than the arbitrary threshold of $2 \times 1.96 \times \sqrt{0.5 \times 0.5/1500} \approx 0.05$.

For each simulation, the total number of subjects in the table for the joint classification of the two biomarkers given biopsy outcome (n_D or $n_{\bar{D}}$; Pepe 2003, Table 3.2) was randomly sampled from a binomial distribution and considered fixed. The parameter of the binomial distribution was set equal to the assumed csPC detection rate (or its complement to one, as appropriate). The number of subjects in the four cells of the table ($\{a, b, c, d\}$ or $\{e, f, g, h\}$; ibid.) was then sampled from a multinomial distribution. The parameter vector for the multinomial distribution was derived from the assumed TPR (TNR) for PSA≥4 ng/ml, rTPR (rTNR), and concordance probability TPPR (TNNR). In particular, TPPR (TNNR) was always chosen so as to guarantee conservative power estimates [15].

For Analysis 1 and 2, p-values were computed according to formulas for paired study designs (Pepe, 2003[15]; Section 3.3). For Analysis 4, p-values for heterogeneity were computed using Wald tests with 3 degrees of freedom.

All relative TPRs and relative TNRs are intended for Stockholm3≥15 versus PSA≥4 ng/ml (referent).

Sensitivity analysis

A subgroup analysis may be conducted for men recruited specifically for SEPTA versus men recruited for other studies with bio-banked material. We will evaluate sensitivity and specificity of Stockholm3 ≥15 and PSA ≥4 ng/mL for detection of csPC and will perform ROC curve analysis and

compute the AUC in these SEPTA subgroups. We will compare the AUCs for Stockholm3 and PSA using DeLong test to test for differences and may also compare PSA density, free/total PSA ratio.

Handling of missing data

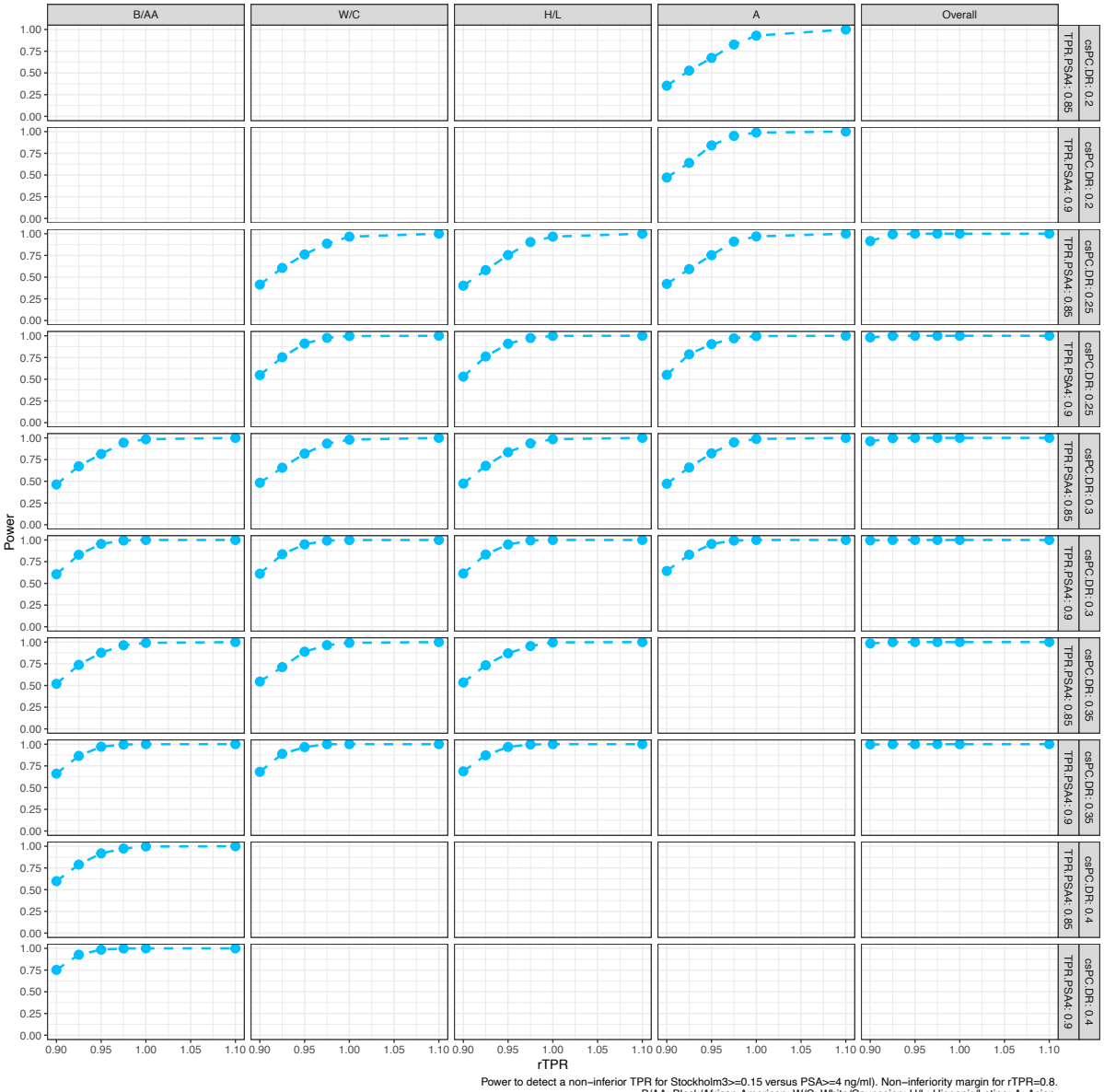
Analysis will be performed based on the per protocol population (requiring input data for Stockholm3 and the outcome data including the primary and key secondary endpoint), however absence of other variable information does not preclude analysis. We assume missing data is completely at random, meaning there is no relationship between the missingness of the data and any values, observed or missing.

References

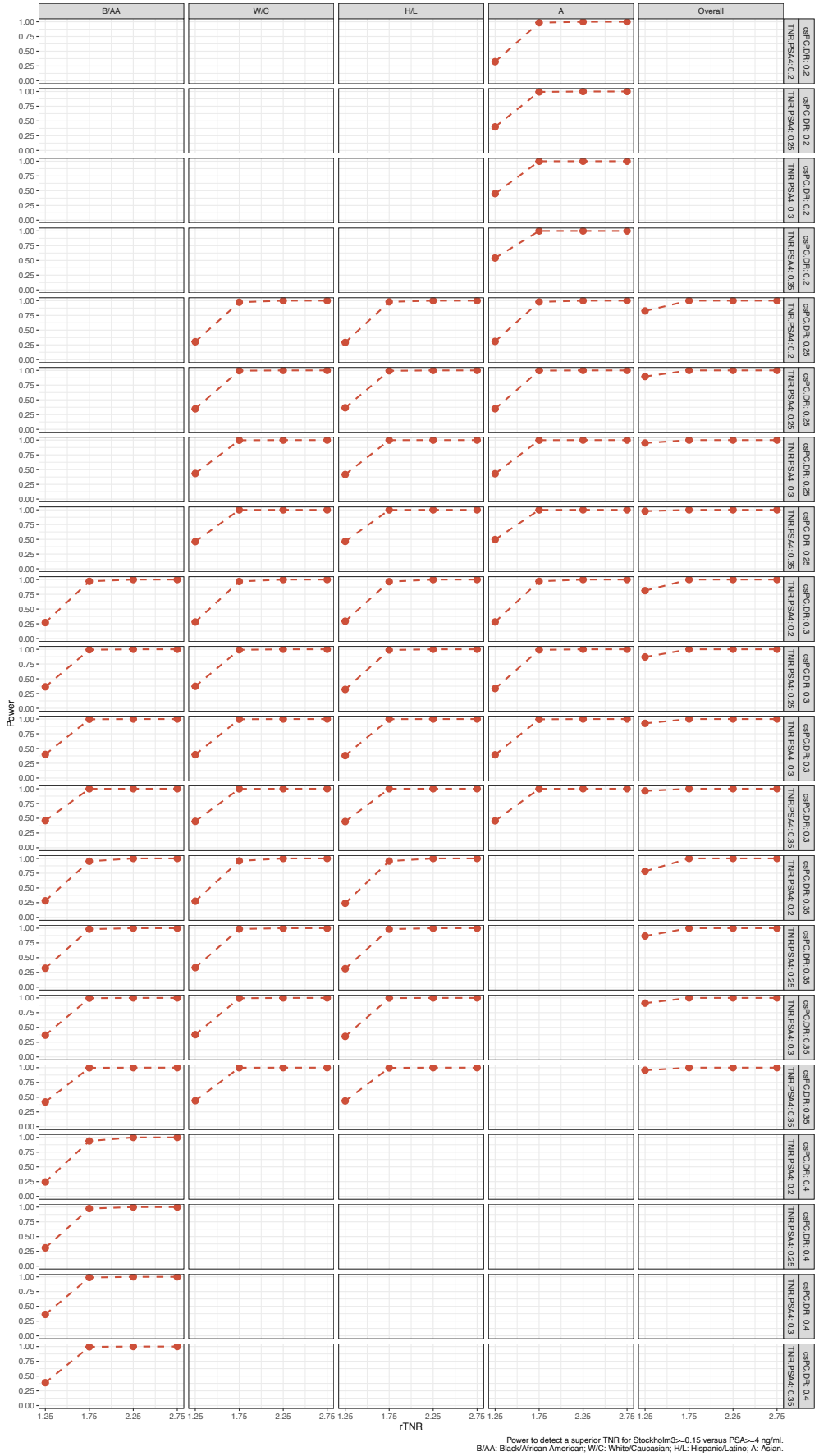
1. Gronberg, H., et al., *Prostate cancer screening in men aged 50-69 years (STHLM3): a prospective population-based diagnostic study.* Lancet Oncol, 2015. **16**(16): p. 1667-76.
2. Gronberg, H., et al., *Prostate Cancer Diagnostics Using a Combination of the Stockholm3 Blood Test and Multiparametric Magnetic Resonance Imaging.* Eur Urol, 2018. **74**(6): p. 722-728.
3. Strom, P., et al., *The Stockholm-3 Model for Prostate Cancer Detection: Algorithm Update, Biomarker Contribution, and Reflex Test Potential.* Eur Urol, 2018. **74**(2): p. 204-210.
4. Conti, D.V., 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**(1): p. 65-75.
5. Plym, A., et al., *Evaluation of a Multiethnic Polygenic Risk Score Model for Prostate Cancer.* J Natl Cancer Inst, 2022. **114**(5): p. 771-774.
6. Moller, A., et al., *The Stockholm3 blood-test predicts clinically-significant cancer on biopsy: independent validation in a multi-center community cohort.* Prostate Cancer Prostatic Dis, 2019. **22**(1): p. 137-142.
7. Viste, E., et al., *Effects of replacing PSA with Stockholm3 for diagnosis of clinically significant prostate cancer in a healthcare system - the Stavanger experience.* Scand J Prim Health Care, 2020. **38**(3): p. 315-322.
8. Walden, M., et al., *A Head-to-head Comparison of Prostate Cancer Diagnostic Strategies Using the Stockholm3 Test, Magnetic Resonance Imaging, and Swedish National Guidelines: Results from a Prospective Population-based Screening Study.* Eur Urol Open Sci, 2022. **38**: p. 32-39.
9. Wolf, A.M., et al., *American Cancer Society guideline for the early detection of prostate cancer: update 2010.* CA Cancer J Clin, 2010. **60**(2): p. 70-98.
10. Nordstrom, T., et al., *Prostate cancer screening using a combination of risk-prediction, MRI, and targeted prostate biopsies (STHLM3-MRI): a prospective, population-based, randomised, open-label, non-inferiority trial.* Lancet Oncol, 2021. **22**(9): p. 1240-1249.
11. Kidd, K.K., et al., *Progress toward an efficient panel of SNPs for ancestry inference.* Forensic Sci Int Genet, 2014. **10**: p. 23-32.
12. *How the American middle class has changed in the past five decades.* Pew Research Center, Washington, D.C. p. https://www.pewresearch.org/fact-tank/2022/04/20/how-the-american-middle-class-has-changed-in-the-past-five-decades.
13. Butler, D.C., et al., *Measures of social deprivation that predict health care access and need within a rational area of primary care service delivery.* Health Serv Res, 2013. **48**(2 Pt 1): p. 539-59.
14. Matheson, F.I., et al., *Development of the Canadian Marginalization Index: a new tool for the study of inequality.* Can J Public Health, 2012. **103**(8 Suppl 2): p. S12-6.

15. Alonzo, T.A., M.S. Pepe, and C.S. Moskowitz, *Sample size calculations for comparative studies of medical tests for detecting presence of disease.* Stat Med, 2002. **21**(6): p. 835-52.
16. DeLong, E.R., D.M. DeLong, and D.L. Clarke-Pearson, *Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach.* Biometrics, 1988. **44**(3): p. 837-45.
17. Van Calster, B., et al., *Calibration: the Achilles heel of predictive analytics.* BMC Med, 2019. **17**(1): p. 230.
18. Vickers, A.J. and E.B. Elkin, *Decision curve analysis: a novel method for evaluating prediction models.* Med Decis Making, 2006. **26**(6): p. 565-74.
19. Pepe, M.S., *The Statistical Evaluation of Medical Tests for Classification and Prediction*. 43 line illus. ed. 2003.
20. Bellavia, A. and L. Valeri, *Decomposition of the Total Effect in the Presence of Multiple Mediators and Interactions.* Am J Epidemiol, 2018. **187**(6): p. 1311-1318.

Supplementary Figure 1 Power to detect a non-inferior TPR (sensitivity) for Stockholm ≥0.15 versus PSA ≥4 ng/mL for clinically significant prostate cancer

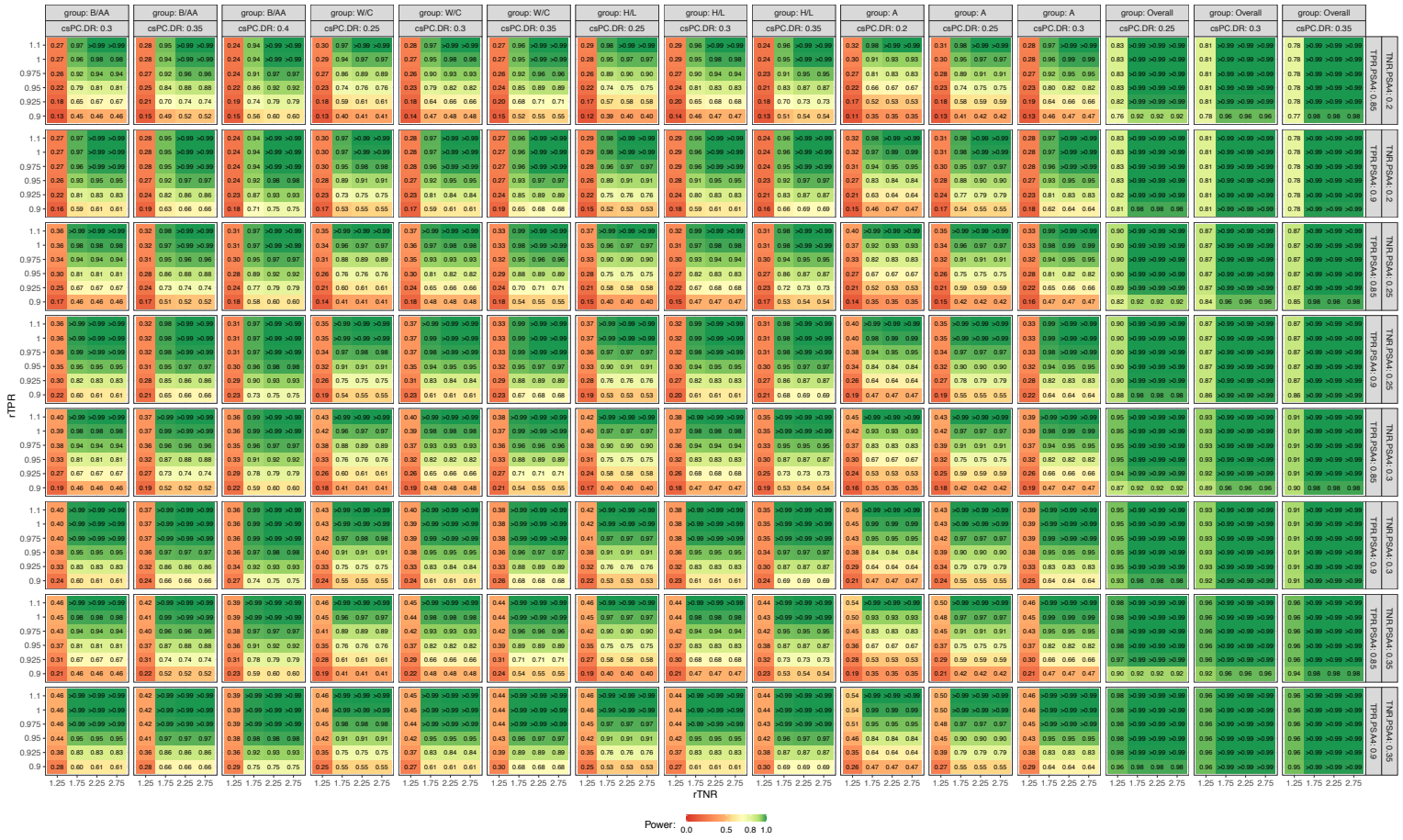


Supplementary Figure 2: Power to detect a superior TNR (specificity) for Stockholm ≥0.15 versus PSA ≥4 ng/mL for non-clinically significant prostate cancer or benign biopsies



Power to detect a superior TNR for Stockholm3>=0.15 versus PSA>=4 ng/ml.
B/AA: Black/African American; W/C: White/Caucasian; H/L: Hispanic/Latino; A: Asian.

Supplementary Figure 3: Power to jointly detect a non-inferior TPR and a superior TNR for Stockholm3 ≥0.15 versus PSA ≥4 ng/ml



Power to detect a non–inferiority TPR (non–inferiority margin for rTPR=0.8) and a superior TNR for Stockholm3>=0.15 versus PSA>=4 ng/ml.
B/AA: Black/African American; W/C: White/Caucasian; H/L: Hispanic/Latino; A: Asian.

Supplementary Table 1: Power to detect heterogeneity in TPR (true positive rate, sensitivity) across the four race/ethnicity groups when present; B/AA, Black/African American; W/C, White/Caucasian; H/L, HIispanic/Latino Caucasian; A, Asian; csPC.DR, clinically significant prostate cancer detection rate; TPR.PSA4, true positive rate, sensitivity of PSA ≥4 ng/mL; rTPR, relative true positive rate of Stockholm3 ≥0.15; sim.power.het, simulatied power with heterogeneity with groups; sim.power.overall, simulated power overall with heterogeneity between groups

scenario	group	csPC.DR	TPR.PSA4	rTPR	n	sim.power.het	sim.power.overall
1	B/AA	0.30	0.90	0.950	500	0.185333333333333	1
1	W/C	0.25	0.85	0.975	500		
1	H/L	0.25	0.90	0.900	500		
1	A	0.25	0.85	1.000	500		
2	B/AA	0.35	0.85	0.925	500	0.058	0.996666666666667
2	W/C	0.30	0.85	0.900	500		
2	H/L	0.35	0.85	0.900	500		
2	A	0.30	0.90	0.925	500		
3	B/AA	0.40	0.90	0.900	500	0.974666666666667	1
3	W/C	0.35	0.90	0.900	500		
3	H/L	0.35	0.85	0.975	500		
3	A	0.20	0.90	1.100	500		
4	B/AA	0.40	0.90	1.100	500	0.796	1
4	W/C	0.25	0.85	0.900	500		
4	H/L	0.25	0.85	0.975	500		
4	A	0.25	0.90	1.100	500		
5	B/AA	0.30	0.90	0.975	500	0.0513333333333333	1

scenario	group	csPC.DR	TPR.PSA4	rTPR	n	sim.power.het	sim.power.overall
5	W/C	0.30	0.90	0.975	500		
5	H/L	0.30	0.85	0.975	500		
5	A	0.20	0.90	1.000	500		
6	B/AA	0.30	0.90	0.900	500	0.968	1
6	W/C	0.30	0.90	1.100	500		
6	H/L	0.25	0.90	0.975	500		
6	A	0.25	0.90	0.925	500		
7	B/AA	0.30	0.85	0.950	500	0.0793333333333333	0.992
7	W/C	0.35	0.85	0.900	500		
7	H/L	0.30	0.85	0.900	500		
7	A	0.25	0.90	0.900	500		
8	B/AA	0.35	0.90	1.100	500	0.891333333333333	1
8	W/C	0.30	0.90	1.000	500		
8	H/L	0.30	0.90	0.950	500		
8	A	0.20	0.90	0.925	500		
9	B/AA	0.30	0.85	0.900	500	0.758	1
9	W/C	0.25	0.85	1.000	500		
9	H/L	0.35	0.90	0.925	500		
9	A	0.25	0.85	1.100	500		
10	B/AA	0.35	0.90	1.000	500	0.585333333333333	1
10	W/C	0.25	0.90	0.925	500		
10	H/L	0.30	0.85	1.100	500		
10	A	0.25	0.85	0.975	500		
11	B/AA	0.40	0.85	1.100	500	0.801333333333333	1
11	W/C	0.30	0.90	0.925	500		
11	H/L	0.35	0.90	0.975	500		
11	A	0.30	0.90	0.950	500		
12	B/AA	0.35	0.90	0.925	500	0.820666666666667	1
12	W/C	0.25	0.85	1.100	500		
12	H/L	0.35	0.90	0.900	500		
12	A	0.30	0.90	1.000	500		
13	B/AA	0.40	0.85	0.975	500	0.12	1
13	W/C	0.25	0.85	0.950	500		
13	H/L	0.25	0.85	0.900	500		
13	A	0.25	0.85	0.925	500		
14	B/AA	0.30	0.85	1.000	500	0.559333333333333	1
14	W/C	0.35	0.90	0.975	500		
14	H/L	0.35	0.85	1.100	500		
14	A	0.30	0.85	1.100	500		
15	B/AA	0.35	0.85	1.000	500	0.142666666666667	1
15	W/C	0.30	0.85	1.000	500		
15	H/L	0.35	0.90	1.000	500		
15	A	0.30	0.85	0.925	500		
16	B/AA	0.40	0.85	0.925	500	0.929333333333333	1
16	W/C	0.25	0.90	1.100	500		
16	H/L	0.35	0.85	0.950	500		
16	A	0.30	0.85	0.900	500		
17	B/AA	0.40	0.85	1.000	500	0.259333333333333	1
17	W/C	0.35	0.90	1.000	500		
17	H/L	0.25	0.90	0.925	500		
17	A	0.20	0.85	0.900	500		
18	B/AA	0.35	0.90	0.975	500	0.696666666666667	1
18	W/C	0.35	0.85	1.000	500		

scenario	group	csPC.DR	TPR.PSA4	rTPR	n	sim.power.het	sim.power.overall
18	H/L	0.30	0.90	1.100	500		
18	A	0.30	0.85	0.975	500		
19	B/AA	0.40	0.90	0.925	500	0.0913333333333333	1
19	W/C	0.30	0.85	0.975	500		
19	H/L	0.25	0.90	0.950	500		
19	A	0.20	0.85	0.925	500		
20	B/AA	0.35	0.90	0.950	500	0.922666666666667	1
20	W/C	0.35	0.85	0.950	500		
20	H/L	0.35	0.90	0.950	500		
20	A	0.30	0.90	1.100	500		
21	B/AA	0.30	0.90	1.100	500	0.851333333333333	1
21	W/C	0.25	0.90	0.900	500		
21	H/L	0.25	0.85	1.100	500		
21	A	0.25	0.90	1.000	500		
22	B/AA	0.40	0.90	1.000	500	0.676	1
22	W/C	0.35	0.85	1.100	500		
22	H/L	0.30	0.90	0.975	500		
22	A	0.25	0.85	0.900	500		
23	B/AA	0.35	0.85	0.975	500	0.124666666666667	1
23	W/C	0.25	0.90	1.000	500		
23	H/L	0.35	0.85	0.925	500		
23	A	0.20	0.85	0.975	500		
24	B/AA	0.30	0.85	0.925	500	0.0806666666666667	0.998666666666667
24	W/C	0.30	0.85	0.950	500		
24	H/L	0.30	0.90	0.925	500		
24	A	0.20	0.90	0.900	500		
25	B/AA	0.30	0.90	1.000	500	0.142666666666667	1
25	W/C	0.30	0.85	0.925	500		
25	H/L	0.25	0.90	1.000	500		
25	A	0.20	0.90	0.975	500		
26	B/AA	0.40	0.90	0.950	500	0.943333333333333	1
26	W/C	0.35	0.90	0.950	500		
26	H/L	0.35	0.90	1.100	500		
26	A	0.30	0.90	0.975	500		
27	B/AA	0.30	0.85	1.100	500	0.648666666666667	1
27	W/C	0.35	0.85	0.925	500		
27	H/L	0.25	0.85	0.950	500		
27	A	0.30	0.85	1.000	500		
28	B/AA	0.40	0.85	0.900	500	0.942	1
28	W/C	0.35	0.90	1.100	500		
28	H/L	0.30	0.85	0.925	500		
28	A	0.20	0.85	1.100	500		
29	B/AA	0.30	0.90	0.925	500	0.0833333333333333	1
29	W/C	0.25	0.90	0.950	500		
29	H/L	0.30	0.90	0.900	500		
29	A	0.25	0.85	0.950	500		
30	B/AA	0.40	0.90	0.975	500	0.0753333333333333	1
30	W/C	0.25	0.90	0.975	500		
30	H/L	0.30	0.85	0.950	500		
30	A	0.25	0.90	0.950	500		