ORIGINAL ARTICLE

MRI-Targeted or Standard Biopsy in Prostate Cancer Screening

Martin Eklund, Ph.D., Fredrik Jäderling, M.D., Ph.D., Andrea Discacciati, Ph.D., Martin Bergman, M.D., Magnus Annerstedt, M.D., Markus Aly, M.D., Ph.D., Axel Glaessgen, M.D., Ph.D., Stefan Carlsson, M.D., Ph.D., Henrik Grönberg, M.D., Ph.D., and Tobias Nordström, M.D., Ph.D., for the STHLM3 consortium*

From the Departments of Medical Epidemiology and Biostatistics (M.E., A.D., M.B., H.G., T.N.) and Molecular Medicine and Surgery (F.J., M. Aly, S.C.), and the Department of Clinical Sciences at Danderyd Hospital (T.N.), Karolinska Institutet, the Department of Diagnostic Radiology (F.J.), the Department of Surgery (M.B., H.G.) and the Department of Clinical Pathology and Cytology, Unilabs (A.G.), Capio St. Göran's Hospital, C-Medical Urology Odenplan (M. Annerstedt), and the Department of Urology, Karolinska University Hospital Solna (M. Aly, S.C.) — all in Stockholm. Address reprint requests to Dr. Eklund at the Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, SE-17177 Stockholm, or at martin.eklund@ki.se.

*A list of the members of the STHLM3 consortium is provided in the Supplementary Appendix, available at NEJM.org.

This article was published on July 9, 2021, at NEJM.org.

N Engl J Med 2021;385:908-20.
DOI: 10.1056/NEJMoa2100852

CME at NEJM.org

ABSTRACT

BACKGROUND

High rates of overdiagnosis are a critical barrier to organized prostate cancer screening. Magnetic resonance imaging (MRI) with targeted biopsy has shown the potential to address this challenge, but the implications of its use in the context of organized prostate cancer screening are unknown.

METHODS

We conducted a population-based noninferiority trial of prostate cancer screening in which men 50 to 74 years of age from the general population were invited by mail to participate; participants with prostate-specific antigen (PSA) levels of 3 ng per milliliter or higher were randomly assigned, in a 2:3 ratio, to undergo a standard biopsy (standard biopsy group) or to undergo MRI, with targeted and standard biopsy if the MRI results suggested prostate cancer (experimental biopsy group). The primary outcome was the proportion of men in the intention-to-treat population in whom clinically significant cancer (Gleason score ≥7) was diagnosed. A key secondary outcome was the detection of clinically insignificant cancers (Gleason score 6).

RESULTS

Of 12,750 men enrolled, 1532 had PSA levels of 3 ng per milliliter or higher and were randomly assigned to undergo biopsy: 603 were assigned to the standard biopsy group and 929 to the experimental biopsy group. In the intention-to-treat analysis, clinically significant cancer was diagnosed in 192 men (21%) in the experimental biopsy group, as compared with 106 men (18%) in the standard biopsy group (difference, 3 percentage points; 95% confidence interval [CI], –1 to 7; P<0.001 for noninferiority). The percentage of clinically insignificant cancers was lower in the experimental biopsy group than in the standard biopsy group (4% [41 participants] vs. 12% [73 participants]; difference, –8 percentage points; 95% CI, –11 to –5).

CONCLUSIONS

MRI with targeted and standard biopsy in men with MRI results suggestive of prostate cancer was noninferior to standard biopsy for detecting clinically significant prostate cancer in a population-based screening-by-invitation trial and resulted in less detection of clinically insignificant cancer. (Funded by the Swedish Research Council and others; STHLM3-MRI ClinicalTrials.gov number, NCT03377881.)

Organized prostate cancer screening in which prostate-specific antigen (PSA) testing is followed by standard systematic, ultrasonography-guided, transrectal biopsy of the prostate in men with elevated PSA levels reduces prostate cancer mortality.[1,2] However, PSA-based screening also leads to high rates of overdiagnosis and overtreatment of clinically insignificant prostate cancer and to many unnecessary biopsies.[3,4] As a result, no country except Lithuania has instituted organized prostate cancer screening programs.[5]

Magnetic resonance imaging (MRI) has generated interest as a method for improving prostate cancer diagnostics.[6] MRI can identify areas of the prostate suggestive of cancer, which allows prostate biopsies to be targeted toward those areas while unnecessary biopsies can be avoided in men with no visible lesions.[7,8] In studies involving men referred for biopsy because of clinical suspicion of prostate cancer, targeted biopsy in men with positive findings on MRI (i.e., suggestive of prostate cancer) resulted in less detection of clinically insignificant cancers than standard biopsy while showing similar or better ability to detect clinically significant cancers.[9-12] We aimed to assess the performance of a strategy of MRI-targeted biopsy in population-based prostate cancer screening.

The STHLM3-MRI trial compared several different screening strategies that used combinations of risk prediction, MRI-targeted biopsy, and standard biopsy in a population-based, organized, screening-by-invitation design.[13,14] Here, we report the results of a strategy of combined MRI-targeted and standard biopsy in men with positive results on MRI, as compared with a standard biopsy strategy, among participating men with elevated PSA levels (≥3 ng per milliliter).

Methods

Trial Design and Oversight

STHLM3-MRI was a prospective, randomized, population-based trial in men 50 to 74 years of age that evaluated various screening strategies for prostate cancer detection. Here, we report the findings of prespecified analyses in which we evaluated whether MRI followed by targeted and standard biopsy in participants in whom MRI indicated the presence of prostate cancer (experimental biopsy group) was noninferior to standard biopsy (standard biopsy group) for detecting clinically significant prostate cancer in men undergoing prostate cancer screening. Details of the trial design have been published previously[13,14]; details of the trial design and statistical analysis are provided in the Supplementary Appendix, which, along with the protocol (which adheres to the SPIRIT 2013 statement[15]), is available with the full text of this article at NEJM.org.

A Quick Take is available at NEJM.org

The STHLM3-MRI trial used a design that combined a paired-screen-positive step (in which two screening tests were used for all participants) and random assignment to either the experimental biopsy group or the standard biopsy group for all participants who had positive results on either of the two screening tests.[16-18] In the paired step, we used a PSA test and the Stockholm3 test to assess the risk of prostate cancer among enrolled participants. The Stockholm3 test is a risk-prediction model that is based on clinical variables (age, first-degree family history of prostate cancer, and previous biopsy), blood biomarkers (total PSA, free PSA, ratio of free PSA to total PSA, human kallikrein 2, macrophage inhibitory cytokine-1, and MSMB), and a polygenic risk score for predicting the risk of prostate cancer with a Gleason score of 7 or higher.[19,20] Participants with elevated PSA levels (≥3 ng per milliliter) or Stockholm3 scores (≥11%) were randomly assigned, in a 2:3 ratio, to the standard biopsy group or the experimental biopsy group (with the use of computer-generated blocks of five and stratified according to six cancer-risk strata, defined according to Stockholm3 risk distribution). This design establishes an analytic framework in which multiple different screening workflows (or strategies) can be compared according to combinations of conditions for biopsy referral (i.e., PSA only, Stockholm3 score only, or PSA or Stockholm3 or both) and biopsy method (i.e., standard, MRI-targeted, or MRI-targeted plus standard [with biopsy in the experimental group performed only in men with MRI results suggestive of cancer]).

The analyses reported here examined the safety and efficacy of standard biopsy as compared with a strategy that used MRI-targeted biopsy in a screening workflow in which the only condition for referral for MRI or standard biopsy was a PSA level of 3 ng per milliliter or greater, in accordance with the condition used in the European Randomized Study of Screening for Prostate Can-

cer (ERSPC). ERSPC provided level 1 evidence of lower prostate cancer mortality among men who were invited to undergo organized PSA screening than among men who were not invited to undergo screening.[1,2] In other words, although either an elevated PSA level (≥3 ng per milliliter) or a positive result on the Stockholm3 test was used as the condition for random assignment and subsequent biopsy, the analysis presented here includes only participants who underwent randomization and who had PSA levels of 3 ng per milliliter or greater, irrespective of their Stockholm3 results. Results of analyses of other workflows, including the use of PSA as compared with the Stockholm3 test for biopsy referral, are not shown here.

The trial was approved by the regional ethics review board in Stockholm and monitored by an independent data and safety monitoring board (Section S2 of the Supplementary Appendix). Reporting adhered to START (Standards of Reporting for MRI-targeted Biopsy Studies) and CONSORT (Consolidated Standards of Reporting Trials) guidelines.[21,22] The trial was designed by the authors, and data were collected by trial consortium members. The authors assume responsibility for the accuracy and completeness of the data and for the fidelity of the trial to the protocol. No one who is not an author contributed to the writing of the manuscript. The Swedish Research Council and the Swedish Cancer Society funded the trial but had no role in protocol development, data analysis or interpretation, or manuscript preparation.

PARTICIPANTS

Men 50 to 74 years of age living in Stockholm County, Sweden, were randomly selected by Statistics Sweden and invited by mail to participate. Men with a previous diagnosis of prostate cancer, a prostate biopsy within 60 days before the invitation, a contraindication to MRI, or severe illness (e.g., metastatic cancer, severe cardiovascular disease, or dementia) were not eligible to participate. Men who had undergone a previous prostate biopsy more than 60 days before the invitation as well as men who had never undergone a prostate biopsy were eligible to participate. Assessment of eligibility, documentation of informed consent, and evaluation of baseline characteristics were conducted through a secure Web portal, and digital laboratory referrals were created automatically.

The prespecified intention-to-treat population for this analysis included all participants with PSA levels of 3 ng per milliliter or greater who underwent randomization. The per-protocol population included participants with PSA levels of 3 ng per milliliter or greater who underwent randomization, adhered to their assigned intervention, and had complete data (MRI and pathology reports). Full details regarding the per-protocol population are provided in the Supplementary Appendix.

BLOOD SAMPLING

Participants provided blood samples (12 ml of blood plasma in EDTA collection tubes) at one of 60 laboratories in Stockholm County; at each laboratory, a trial nurse verified that the participant did not meet any exclusion criteria and that he understood the informed consent form. PSA was analyzed (B.R.A.H.M.S Kryptor compact PLUS) for all participants. Participants with PSA levels of less than 1.5 ng per milliliter were considered to be at low risk for clinically significant prostate cancer and were recommended to repeat testing in 6 years. For participants with PSA levels of 1.5 ng per milliliter or greater, the Stockholm3 test was performed at the A23 Laboratory (Uppsala, Sweden) as described previously.[19,20] Men with PSA levels that were 1.5 ng per milliliter or higher but less than 3 ng per milliliter and who had Stockholm3 scores of less than 11% were judged to have nonelevated risk, did not undergo randomization, and were recommended to repeat testing in 2 years.

PROSTATE BIOPSIES

All biopsies were performed by experienced urologists (each of whom had performed >200 procedures) at one of four participating clinics. Men undergoing biopsy were given a prophylactic antibiotic (oral ciprofloxacin, 750 mg). Participants in the standard biopsy group underwent standard transrectal ultrasonography–guided prostate biopsies to obtain 10 to 12 biopsy cores from the peripheral zone of the prostate (apical, midgland, and base).

In the experimental biopsy group, T2- and diffusion-weighted images were obtained with the use of a biparametric (i.e., combined T2- and diffusion-weighted imaging without contrast enhancement) MRI protocol developed for high-throughput screening (<16 minutes), with 1.5T

Magnetom Aera (Siemens) and 3T SIGNA Architect (GE Healthcare) scanners, without endorectal coil (details of the MRI protocol and quality control are provided in Section S3). Radiology readings were performed at Capio St. Göran's Hospital, Stockholm, by three uroradiologists; consensus by at least two radiologists was required for each case. Regions suggestive of prostate cancer were scored according to Prostate Imaging Reporting and Data System (PI-RADS), versions 2.0 and 2.1, on a scale of 1 to 5, with higher scores indicating more clinically suspicious lesions; scores of 3 to 5 defined a positive MRI. A maximum of three clinically significant lesions were identified per participant and delineated for targeted biopsy with the use of dedicated software (MIM Symphony DX, MIM Software). For quality-control purposes, an external uroradiologist, who was unaware of the PI-RADS scores assigned by the study uroradiologists, reviewed 99 of the biparametric MRIs, randomly sampled by PI-RADS score. If no clinically significant lesions were identified, biopsies were not performed except in cases of Stockholm3 test scores of 25% or greater (which indicated a high risk of clinically significant cancer despite a negative MRI).[23] Otherwise, we used the MRI-fusion technique (bkFusion, BK Medical) to perform transrectal sampling of 3 to 4 biopsy cores targeting each significant lesion. The urologist also obtained a standard 10-to-12–core biopsy specimen immediately after the targeted biopsy.

Pathological assessments were performed at Unilabs pathology unit (Capio St. Göran's Hospital, Stockholm) by one of four experienced uropathologists. Gleason score and number of millimeters of cancer in each biopsy core were reported for each core according to International Society of Urological Pathology 2014 guidelines.[24] The overall Gleason score was reported for each case and for each biopsy method; the reported score for the combined biopsy was the highest overall score across the two biopsy methods.

OUTCOMES

The primary outcome was the probability of detection of clinically significant prostate cancer, defined as the percentage of participants in each group who received a diagnosis of cancer with a Gleason score of 3+4 or greater (International Society of Urological Pathology grade ≥2). The Gleason score is composed of a primary (most predominant) grade plus a secondary (highest nonpredominant) grade; the sum is reached by adding the primary and secondary grades. The Gleason sum ranges from 6 to 10, with higher scores indicating a more aggressive form of prostate cancer. Secondary outcomes included the detection probabilities (i.e., proportions) of benign biopsies, clinically insignificant cancer (defined as a Gleason score of 3+3 or International Society of Urological Pathology grade 1 cancer), cancers with Gleason scores of 4+3 or greater (International Society of Urological Pathology grade ≥3), and serious adverse events (infections treated with antibiotics, hospitalization, or death within 30 days after the biopsy procedure) in each group.

All participants were followed for a minimum of 200 days after receiving PSA test results. Men who underwent biopsy were followed for at least 30 days after the biopsy for monitoring of adverse events, and participants who underwent radical prostatectomy before October 22, 2020, were followed until prostatectomy pathology results were available.

STATISTICAL ANALYSIS

We planned to invite 50,000 men to participate in screening, assuming 25% participation and 13% of participants having PSA results of 3 ng per milliliter or greater.[19] This number would yield 1625 participants with PSA levels of 3 ng per milliliter undergoing randomization. Using a noninferiority margin of 4 percentage points and an alpha of 2.5% and assuming a relative detection probability of clinically significant cancer of 1.3 in favor of the experimental biopsy group (on the basis of previous studies[25]), 80% adherence to the assigned intervention, and 18% detection probability of clinically significant cancer in the standard biopsy group, we estimated that the trial would have more than 90% power to show the noninferiority of the experimental biopsy strategy to the standard biopsy strategy. The noninferiority margin was agreed upon at a consensus group meeting that included urologists, oncologists, and statisticians.

For the primary and secondary outcomes, absolute differences in detection probabilities and 95% two-sided Wald confidence intervals were computed (without adjustment for the variable used for stratification at randomization). If the lower boundary for the two-sided 95% confidence

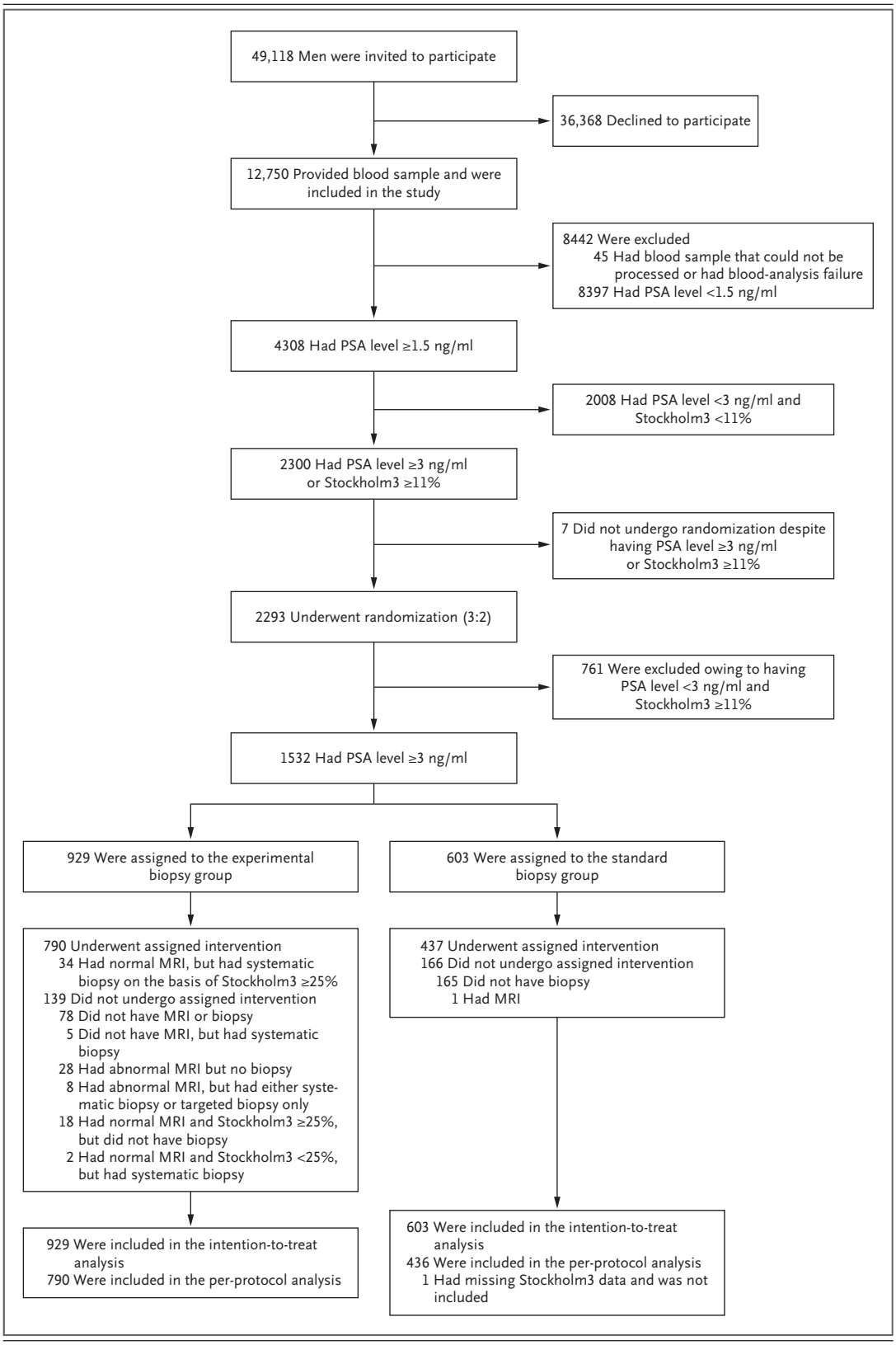


Figure 1 (facing page). Enrollment, Randomization, and Follow-up of the Trial Populations.

The protocol-defined analyses compared the safety and efficacy of standard biopsy with those of magnetic resonance imaging (MRI) and targeted and systematic biopsy in men with positive results on MRI. Participants assigned to the standard biopsy group underwent standard transrectal ultrasonography–guided biopsy to obtain 10 to 12 cores. Participants in the experimental biopsy group with a Prostate Imaging Reporting and Data System (PI-RADS) score of 3 or greater (on a scale of 1 to 5, with higher scores indicating more clinically suspicious lesions) underwent MRI-guided biopsy and standard biopsy; participants with no suspicious lesions on MRI (PI-RADS score of ≤2) did not undergo biopsy unless they were at very high risk for clinically significant prostate cancer (Stockholm3 score [a risk-prediction score based on clinical variables, blood biomarkers, and a polygenic risk score] of ≥25%). PSA denotes prostate-specific antigen.

interval in the absolute difference of clinically significant cancer between the experimental biopsy group and the standard biopsy group was greater than −4 percentage points, the experimental strategy would be deemed to be noninferior; if the lower boundary was greater than 0, the experimental strategy would be deemed to be superior. Prespecified subgroup analyses were performed according to age strata (50 to 59, 60 to 69, and 70 to 74 years), PSA strata (3 to 3.9, 4 to 9.9, and ≥10 ng per milliliter), and previous biopsy (yes or no). Analyses were performed in the intention-to-treat population (analyses of results in the per-protocol population were also performed for the primary outcome). In a prespecified sensitivity analysis, we used model-based multiple imputation to impute the outcome status with respect to clinically significant cancer for participants who did not undergo MRI or biopsy examinations (details regarding the imputation procedure are provided in Section S4). The imputation procedure was constructed to take into account the effect of the MRI result on missing outcome status.[26]

To further assess the effect of missing outcome status owing to participants not undergoing recommended MRI or biopsy procedures, we conducted two post hoc analyses. First, we assessed whether the results from the multiple imputation analysis were robust to deviations from the missing-at-random assumption by allowing the primary outcome to be missing-not-at-random. Second, to further account for incomplete adherence to protocol being dependent on baseline covariates — and in the experimental biopsy group dependent on MRI result — we estimated the difference in detection probabilities of clinically significant and insignificant prostate cancer and benign biopsy findings using inverse probability weighting.

We performed a prespecified analysis that ignored results of the standard biopsy in men with positive MRI results to estimate results of performing only targeted biopsy, and a post hoc analysis that ignored biopsy outcome for participants who had negative MRI results but who were at high risk (Stockholm3 score of ≥25%) in order to estimate results if these participants had not undergone biopsies. No adjustment for multiplicity was made. P values are reported only for the primary outcome.[27] For secondary outcomes and subgroup analyses, the reported two-sided 95% confidence intervals for the individual contrasts have not been adjusted for multiplicity and should be interpreted with caution. The analysis plan was approved by the data and safety monitoring board.

RESULTS

TRIAL POPULATION

From February 2018 through March 2020, a total of 49,118 men were invited to participate, and 12,750 men consented to screening and provided blood samples; 2293 had PSA results of 3 ng per milliliter or higher or Stockholm3 scores of 11% or greater and underwent randomization. For analyses presented here, we included participants with PSA levels of 3 ng per milliliter or higher, irrespective of Stockholm3 score (1532 participants); 929 were randomly assigned to the experimental biopsy group and 603 to the standard biopsy group (Fig. 1 and Section S1). The characteristics of the patients at baseline were similar in the two groups (Table 1). In the experimental biopsy group, 338 participants (36%) underwent biopsies — 297 on the basis of PI-RADS scores of 3 or higher, 34 on the basis of Stockholm3 results of 25% or greater despite negative MRI findings, and 7 on the basis of the judgment of the study physician to overrule the trial protocol — with a median of 15 cores obtained (Table S1 and Fig. S2). In the standard biopsy group, 438

Table 1. Characteristics of the Participants at Baseline.*

Characteristic	Enrolled Population (N=12,750)	Experimental Biopsy Group (N=929)	Standard Biopsy Group (N=603)
Age — yr			
Median	61	66	66
Interquartile range	55–67	61–71	61–71
Previous biopsy procedure — no. (%)			
Yes	533 (4)	154 (17)	97 (16)
No	11,736 (92)	737 (79)	488 (81)
Missing data	481 (4)	38 (4)	18 (3)
PSA at enrollment — ng/ml			
Median	1.03	4.25	4.29
Interquartile range	0.60–1.88	3.50–5.94	3.50–5.61
Prostate volume — cm^3†			
Median		45	44
Interquartile range		33–62	34–60
Score on PI-RADS — no. (%)			
≤2		521 (56)	0
3		175 (19)	0
4		85 (9)	0
5		65 (7)	1 (0.2)
No MRI performed		83 (9)	602 (100)

* Of the 12,750 who were enrolled, 1532 men with prostate-specific antigen (PSA) levels of 3 ng per milliliter or higher were randomly assigned to the experimental biopsy group or to the standard biopsy group. Participants assigned to the experimental biopsy group underwent magnetic resonance imaging (MRI). If the MRI screening indicated the presence of prostate cancer (a score of ≥3 as categorized according to the Prostate Imaging Reporting and Data System [PI-RADS], on which scores range from 1 to 5, with higher scores indicating more clinically suspicious lesions), the participant underwent targeted biopsy and standard biopsy (combined biopsy). Men whose MRI screenings did not indicate prostate cancer (PI-RADS score of ≤2) were not recommended to undergo biopsy unless they had a very high risk of clinically significant prostate cancer (Stockholm3 score [a risk-prediction score based on clinical variables, blood biomarkers, and a polygenic risk score] of ≥25%). Participants who were assigned to the standard biopsy group underwent standard transrectal ultrasonography–guided biopsy, with 10 to 12 core samples obtained. The characteristics of the participants at baseline were similar in the two groups. Percentages may not total 100 because of rounding.

† Prostate volume was measured with the use of MRI in the experimental biopsy group and ultrasonography in the standard biopsy group.

participants (73%) underwent biopsy, with a median of 12 cores obtained. In the intention-to-treat population, the absolute between-group difference in the percentage of men who underwent biopsy was −36 percentage points (95% confidence interval [CI], −41 to −32); in the inverse probability weighting analysis, the absolute between-group difference was −55 percentage points (95% CI, −58 to −51) (Table S2).

BIOPSY OUTCOMES

In the intention-to-treat analysis, clinically significant cancer was diagnosed in 192 of 929 participants (21%) in the experimental biopsy group, as compared with 106 of 603 participants (18%) in the standard biopsy group, a difference of 3 percentage points (95% CI, −1 to 7) (Table 2 and Fig. 2). Since the lower boundary of the two-sided 95% confidence interval was greater than −4 percentage points, the experimental strategy was deemed noninferior to the standard strategy for detecting clinically significant cancer (P<0.001). In the per-protocol analysis, clinically significant cancer was detected in 183 of 790 participants (23%) in the experimental biopsy group, as compared with 105 of 436 participants (24%)

Table 2. Comparison of Cancer Detection in the Trial Groups.

Variable	Analysis*	Experimental Biopsy Group (N=929)	Standard Biopsy Group (N=603)	Difference (95% CI)†	P value‡
Biopsy procedures — no. (%)	ITT	338 (36)	438 (73)	−36 (−41 to −32)	
Biopsy outcome — no. (%)§	ITT				
Benign		105 (11)	259 (43)	−32 (−36 to −27)	
Gleason score 6		41 (4)	73 (12)		
Gleason score 3+4		125 (13)	63 (10)		
Gleason score 4+3		29 (3)	21 (3)		
Gleason score ≥4+4		38 (4)	22 (4)		
Referred for biopsy, biopsy not performed¶		124 (13)	165 (27)		
Detection of clinically insignificant cancer: Gleason score 6 — no. (%)§	ITT	41 (4)	73 (12)	−8 (−11 to −5)	
Detection of clinically significant cancer: Gleason score ≥7 — no./total no. (%)§	ITT	192/929 (21)	106/603 (18)	3 (−1 to 7)	<0.001
	MBI	231/929 (25)	130/603 (22)	3 (−1 to 8)	<0.001
	IPW	230/929 (25)	133/603 (22)	3 (−2 to 8)	0.004
	PP	183/790 (23)	105/436 (24)	−1 (−6 to 4)	0.11
Detection of Gleason Score ≥4+3 cancer — no. (%)§	ITT	67 (7)	43 (7)	0.1 (−3 to 3)	
No biopsy performed given negative result on MRI and Stockholm3 score of <25% — no. (%)		467 (50)	0		

* The intention-to-treat (ITT) analysis included all the participants who underwent randomization. The per-protocol (PP) analysis included participants with PSA levels of 3 ng per milliliter or higher who underwent randomization and completed their assigned intervention as specified in the protocol. In addition, we performed a prespecified model-based multiple imputation (MBI) analysis, in which biopsy outcomes for participants who underwent randomization but who did not undergo biopsy were imputed on the basis of Stockholm3 risk score for the standard biopsy group, and Stockholm3 risk score in combination with PI-RADS score for the experimental biopsy group. We also performed an inverse probability weighting (IPW) analysis to account for incomplete adherence to the protocol.

† Between-group differences are shown in percentage points.

‡ P values are for a test of the noninferiority of the experimental biopsy strategy to the standard biopsy strategy, at a noninferiority margin of −4 percentage points, with respect to detection of clinically significant cancers.

§ For purposes of the primary analysis, clinically significant cancer was defined as Gleason score 3+4 (Gleason sum of 7) or greater, and clinically insignificant cancer was defined as Gleason score of 3+3 (Gleason sum of 6). In a prespecified additional analysis, we also compared the detection probabilities of an alternative definition of significant prostate cancer (Gleason score ≥4+3).

¶ Of the participants who were referred for biopsy but did not undergo biopsy, 2% in the experimental biopsy group and 18% in the standard biopsy group chose not to undergo biopsy after discussions with trial urologists. The remaining participants did not undergo biopsy within the 200-day follow-up period after they received the serum PSA test result.

in the standard biopsy group. The results with model-based multiple imputation were consistent with the intention-to-treat results and were robust to departures from the missing-at-random assumption. In addition, the results with inverse probability weighting were consistent with the intention-to-treat results (Table 2 and Fig. 2).

In the intention-to-treat analysis, clinically insignificant cancer was diagnosed in fewer participants in the experimental biopsy group than in the standard biopsy group (41 [4%] vs. 73 [12%]), representing a difference of −8 percentage points (95% CI, −11 to −5). Fewer participants had benign biopsy findings in the experimental biopsy group than in the standard biopsy group (105 [11%] vs. 259 [43%]), representing a difference of −32 percentage points (95% CI, −36 to −27). These differences were more marked in the inverse probability weighting analysis, where the difference between the experimental biopsy group and the standard biopsy group in the diagnosis of clinically insignificant cancer was −11 percentage points (95% CI, −15 to −7) and

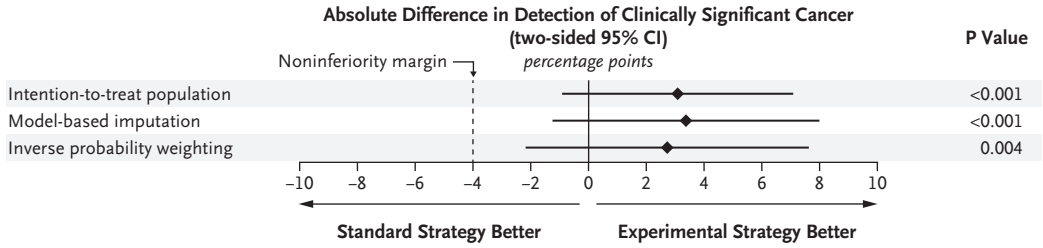


Figure 2. Intention-to-treat, Imputation, and Inverse Probability Weighted Analyses of the Primary Outcome for the Detection Probability of Clinically Significant Prostate Cancer.

Black diamonds indicate the absolute differences in detection probability of clinically significant prostate cancer (Gleason score ≥7) between the standard biopsy and experimental biopsy groups. The intention-to-treat analysis included all participants with a PSA level of 3 ng per milliliter or higher who underwent randomization. We performed a prespecified model-based multiple imputation analysis in which biopsy outcomes for participants who underwent randomization but did not undergo biopsy were imputed according to Stockholm3 risk score for the standard biopsy group and according to Stockholm3 risk score in combination with the PI-RADS score for the experimental biopsy group. We also performed an inverse probability weighting analysis to account for incomplete adherence to protocol. If the lower boundary of the two-sided 95% confidence interval for the difference in detection probability (experimental biopsy group minus standard biopsy group) was greater than −4 percentage points (dashed vertical line), MRI with targeted and standard biopsy in men in whom MRI indicated the presence of prostate cancer (PI-RADS score of ≥3) was deemed to be noninferior. If the lower boundary was greater than 0 (solid vertical line), MRI with targeted and standard biopsy was deemed superior. P values are for noninferiority of the experimental biopsy strategy to the standard biopsy strategy.

in a biopsy with benign findings was −46 percentage points (95% CI, −52 to −41). Figure 3 summarizes the comparative number of procedures and clinically significant and insignificant cancers detected by means of each diagnostic strategy, normalized to a population of 10,000 screened men.

Subgroup analyses showed that in the intention-to-treat population, the differences in detection probabilities of significant and insignificant cancer in the experimental and standard biopsy groups were consistent across PSA and age strata and between men who had undergone a previous negative prostate biopsy and men who had not undergone a previous biopsy (details regarding subgroup analyses provided in Tables S3 through S5).

Ignoring biopsy outcomes for participants who had negative MRI results but who were at high risk (Stockholm3 score ≥25%) resulted in 6 fewer clinically significant cancers detected in the experimental biopsy group and 5 fewer insignificant cancers detected (Table S6). In this case, the experimental strategy remained noninferior in the intention-to-treat population.

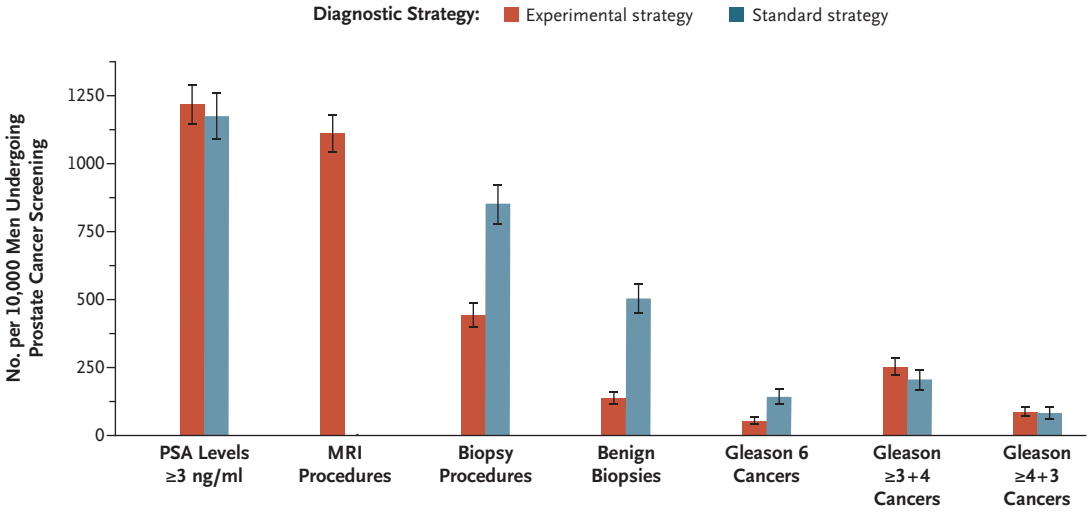
Ignoring the supplemental standard biopsy, the MRI-targeted biopsy alone in the experimental biopsy group reduced the detection probability of significant cancer to 17% in the intention-to-treat population (162 of 929 participants), and the experimental strategy no longer met the noninferiority criterion at the 2.5% alpha level (difference, −0.1 percentage point; 95% CI, −4 to 4) (Table S7).

ADVERSE EVENTS

In the intention-to-treat population, 20 participants (2%) in the experimental biopsy group and 23 (4%) in the standard biopsy group had infections after the biopsy procedure that were treated with antibiotics (difference, −2 percentage points; 95% CI, −4 to 0.1) (Table S8). The incidence of hospitalization was 1% (13 participants) in the experimental biopsy group and 3% (17 participants) in the standard biopsy group (difference, −1 percentage point; 95% CI, −3 to 0.1). No deaths were reported in either group during the trial period.

QUALITY CONTROL

In the quality-control review of MRI results, the central study radiology team and the external



	PSA Levels ≥3 ng/ml	MRI Procedures	Biopsy Procedures	Benign Biopsies	Gleason 6 Cancers	Gleason ≥3+4 Cancers	Gleason ≥4+3 Cancers
Experimental Strategy — no.	1221	1112	444	138	54	252	88
Standard Strategy — no.	1175	2	853	504	142	206	84
Absolute Difference (95% CI) — percentage points	0.5 (−0.7 to 1.6)	—	−4.1 (−5.0 to −3.2)	−3.7 (−4.3 to −3.0)	−0.9 (−1.2 to −0.5)	0.5 (−0.1 to 1.0)	0.0 (−0.3 to 0.4)
Relative Difference (95% CI)	1.04 (0.94 to 1.14)	—	0.52 (0.45 to 0.60)	0.27 (0.22 to 0.34)	0.38 (0.26 to 0.55)	1.22 (0.97 to 1.55)	1.05 (0.72 to 1.54)

Figure 3. Elevated PSA Levels, Performed Procedures, and Detected Cancers.

Bars indicate comparative results from the STHLM3-MRI trial normalized to a population of 10,000 men 50 to 74 years of age undergoing prostate cancer screening. Red bars represent results from the MRI-enhanced screening strategy used in the experimental biopsy group (PSA testing followed by MRI for participants with PSA levels of ≥3 ng per milliliter and combined biopsy in men with MRI results suggestive of prostate cancer). Blue bars represent results of the screening strategy used in the standard biopsy group (PSA testing followed by standard biopsy in men with PSA levels of ≥3 ng per milliliter). The Gleason score is composed of a primary (most predominant) grade plus a secondary (highest nonpredominant) grade; the sum is reached by adding the primary and secondary grades. The Gleason sum ranges from 6 to 10, with higher scores indicating a more aggressive form of prostate cancer. The definition of the Gleason score suggests that Gleason score 4+3 generally indicates worse prognosis than Gleason score 3+4, although both sum to 7. I bars represent 95% confidence intervals.

radiologist were in concordance with regard to biopsy referral in 83% of cases reviewed (82 of 99). As calculated with use of the linearly weighted Cohen's kappa statistic for PI-RADS, version 2.1, concordance was 0.78 (Table S9). A total of 59 patients in the standard biopsy group underwent radical prostatectomy during follow-up, as compared with 124 patients in the experimental biopsy group. The use of MRI and combined biopsy did not result in significantly higher agreement between biopsy Gleason score and whole-mount histopathological analysis of the prostatectomy samples than standard biopsies in the standard biopsy group (linear weighted Cohen's kappa 0.6 and 0.5, respectively; difference, 0.1; 95% CI, −0.1 to 0.3) (Table S10).

DISCUSSION

High rates of overdiagnosis and unnecessary biopsies are the primary reasons that organized prostate cancer screening programs have not been implemented.[3-5] We showed, in a population-based, screening-by-invitation trial, that detection of clinically insignificant tumors and benign findings on biopsy were lower (by 64% and 74%, respectively) among men with elevated PSA levels (≥3 ng per milliliter) when biopsy was performed only in men with positive MRI results than when biopsy was performed according to the standard strategy. These results were achieved without compromising detection rates of clinically significant cancers. In addition, fewer infections after the biopsy procedure were noted in

the experimental biopsy group than in the standard biopsy group (although this difference was not significant at the 5% significance level), because fewer participants underwent biopsy in the experimental biopsy group.

The use of a PSA of 3 ng per milliliter or higher as a condition for biopsy in this analysis was designed to reflect the conditions for biopsy referral used in ERSPC. After 16 years of follow-up, ERSPC has shown that there were 20% fewer deaths among men invited to undergo prostate cancer screening than among men not invited, providing level 1 evidence of mortality reduction from organized screening. In ERSPC, the standard biopsy method was used. Although the MRI-targeted biopsy approach appears to offer important improvements to standard biopsy in the STHLM3-MRI population, longer follow-up would be needed to estimate the effect on mortality.

Most studies of MRI-targeted biopsy have shown greater sensitivity to detect clinically significant cancers than standard biopsy.[12] These study populations were restricted to men referred for biopsy on the basis of clinical suspicion of prostate cancer, and the results are thus challenging to interpret in the context of population-based screening, in which the majority of men referred for biopsy would be expected to be at lower risk for clinically significant prostate cancer. For example, in the studies by Kasivisvanathan et al.,[9] Ahdoot et al.,[10] Grönberg et al.,[23] and Klotz et al.,[28] the median PSA among participants was 6.2 to 6.7 ng per milliliter, and the percentage of negative MRIs was 19 to 38%. In our trial, the median PSA was 4.3 ng per milliliter among participants referred for prostate biopsy on the basis of a PSA level of 3 ng per milliliter or higher, and the percentage of negative MRIs was 61%. This contrast illustrates the marked differences between clinical cohorts (patients referred to urologists for prostate biopsy) and screening cohorts and emphasizes the need to improve the selection of men for biopsy referral to minimize overdiagnosis and unnecessary biopsies in organized screening.

An important question is whether men who have positive MRI results should undergo standard biopsy in addition to targeted biopsy. In our trial, 30 fewer clinically significant cancers would have been detected among the 929 men in the experimental biopsy group if the additional standard biopsy had not been performed, while 18 fewer clinically insignificant cancers would have been found; thus, detection of 1.7 clinically significant cancers would be delayed for each clinically insignificant cancer avoided. Our results therefore support the use of standard biopsy in addition to targeted biopsy for men who have positive MRI results, an observation that is in line with previous findings.[29]

Key strengths of this trial include the screening-by-invitation design, the large number of participants, the random assignment of biopsy techniques (which avoided the risk of incorporation bias arising from comparing different techniques applied to the same patient), and the pragmatic design with a short biparametric MRI protocol suitable for high-volume, population-based screening. However, although biparametric protocols in several studies have performed on par with multiparametric protocols,[30-32] the use of biparametric 1.5T and 3T MRI protocols could potentially have contributed to the somewhat lower relative detection probability of targeted biopsies than that observed in previous studies. The STHLM3-MRI trial was performed in Stockholm, Sweden, with centralized radiologic and pathological assessment, which may limit generalizability to other health care settings. Furthermore, only a single round of screening was performed, so whether the reduction in overdiagnosis will be retained through multiple rounds of screening is unknown. However, participants in the STHLM3-MRI trial will be invited for subsequent screening, and men with negative MRI results will be followed to ensure that clinically significant cancers were not overlooked. We also cannot draw definitive conclusions regarding the equivalency of MRI-targeted and standard biopsy approaches with respect to prostate cancer mortality, although equivalency seems plausible since the Gleason score distributions across clinically significant cancers were similar in the two trial groups. Although consensus is lacking on the definition of clinically significant prostate cancer, we used Gleason scores of 7 or higher, a common definition used in previous studies,[9-11,25] while also reporting outcomes for other scores (Gleason Score ≥4+3) (Table 2). In the STHLM3-MRI trial, men with negative MRI results but

Stockholm3 scores of 25% or greater were recommended to undergo standard biopsy as a safety mechanism. Recognizing that Stockholm3 testing is not widely available, we included sensitivity analyses in which the biopsy outcomes of these participants were omitted. The effect on the results was small. Other safety mechanisms could be used — for example, biopsy on the basis of PSA alone (e.g., PSA ≥10 ng per milliliter) or PSA combined with free PSA, prostate volume (i.e., PSA density), or both.

Overall adherence to trial recommendations was 85% in the experimental biopsy group and 72% in the standard biopsy group, which suggests the participants' increased willingness to undergo biopsy after identification of lesions visible on MRI. As a consequence, the difference in diagnosed insignificant prostate cancers, benign biopsy results, and overall biopsies was greater in results that were based on inverse probability weighting than in results of the intention-to-treat analysis, whereas the detection probability for significant prostate cancer was similar in the two analyses (Table 2 and Fig. 2). Participants in the standard biopsy group who did not adhere to biopsy recommendations had a lower risk of clinically significant prostate cancer than nonadherent participants in the experimental biopsy group (median Stockholm3 risk 11% vs. 22%). The reason for this difference is that, by design, nonadherent participants in the experimental biopsy group had to have either a positive MRI or Stockholm3 score of at least 25%. This explains why the results with model-based multiple imputation to impute the primary outcome status and results based on inverse probability weighting are closer to the intention-to-treat result than to the per-protocol result (Table 2): among nonadherent participants, more biopsy results were imputed with a clinically significant cancer finding in the experimental biopsy group than in the standard biopsy group, resulting in higher detection probability for significant cancer in the imputed data than in the per-protocol results.

In a trial of population-based screening by invitation, our results showed that among men with elevated PSA levels, combined biopsy performed only in men who had positive results on MRI was noninferior to standard biopsy for detecting clinically significant prostate cancer. The markedly reduced incidences of unnecessary biopsy and diagnosis of clinically insignificant cancer address key barriers impeding implementation of population-based screening for prostate cancer. When normalized to a population of 10,000 men 50 to 74 years of age in which those with elevated PSA levels (≥3 ng per milliliter) are referred for biopsy, the combined biopsy approach in men with positive MRI scans would result in 409 fewer men undergoing biopsy, 366 fewer biopsies with benign findings, and 88 fewer clinically insignificant cancers detected than with the standard biopsy approach. These numbers represent 48%, 73%, and 62% lower incidences, respectively, with the use of MRI and the combined biopsy approach (Fig. 3). The reduced biopsy rate and potential downstream savings that result from less overtreatment offer potential cost savings that may offset the additional costs of MRI.[33]

Supported by the Swedish Research Council (Vetenskapsrådet), the Swedish Cancer Society (Cancerfonden), the Percy Falk Foundation, the Magnus Bergvall Foundation, the Strategic Research Program on Cancer (StratCan) at Karolinska Institutet, the Hagstrand Memorial Fund (Hagstrandska Minnesfonden), Region Stockholm, Svenska Druidorden, Åke Wibergs Stiftelse, and Swedish e-Science Research Center (SeRC), Karolinska Institutet, and the Swedish Prostate Cancer Foundation (Prostatacancerförbundet).

Disclosure forms provided by the authors are available with the full text of this article at NEJM.org.

A data sharing statement provided by the authors is available with the full text of this article at NEJM.org.

We thank all the trial participants; the STHLM3 core management group (Carin Cavalli-Björkman, Astrid Björklund, and Britt-Marie Hune) for taking care of all contact with participants and organizing the databases; Ola Steinberg and Dr. Karl Andersson (A23 Laboratory) for organizing sample handling and performing analyses; all staff at Unilabs Odenplan MRI and radiologists Dr. Edward Meurling and Dr. Fredrik Wimmercranz for performing and interpreting MRI procedures; all staff at Unilabs Pathology department and Dr. Nada Majeed, Dr. Rihab Awad Elkarim, and Dr. Ing-Marie Fyhr for their work with pathological assessments; all staff at GHP Odenplan, Capio Solna, Capio St. Göran's Hospital, and Urologifocus, and the following urologists for performing prostate biopsies: Dr. Dag Sandström, Dr. Linda Waage, Dr. Otabek Imamov, Dr. Rafael Lantz, Dr. Andreas Thorstensson, Dr. Carl Stiernstedt, Dr. Dushaid Wande, Dr. Gunnar Trygg, Dr. Harald Söderbäck, Dr. Jerzy Michajlowski, Dr. Lars Leykamm, Dr. Nils-Erik Svedberg, Dr. Tommy Bergman, Dr. Raimundas Sabockis, and Dr. Sirvan Akrawi; Dr. Jonas Wallström for radiology reassessment; Dr. Mark Clements and Shuang Hao for important discussions on the study design; Dr. Jeff Matthews and Dr. Ashkan Mortezavi for detailed and critical review of the submitted manuscript; and the data and safety monitoring board (Prof. Ola Bratt, Prof. Lars Holmberg, and Dr. Hans Garmo).

REFERENCES

1. Schröder FH, Hugosson J, Roobol MJ, et al. Screening and prostate cancer mortality: results of the European Randomised Study of Screening for Prostate Cancer (ERSPC) at 13 years of follow-up. Lancet 2014;384:2027-35.
2. Hugosson J, Roobol MJ, Månsson M, et al. A 16-yr follow-up of the European Randomized study of Screening for Prostate Cancer. Eur Urol 2019;76:43-51.
3. Welch HG, Albertsen PC. Reconsidering prostate cancer mortality — the future of PSA screening. N Engl J Med 2020;382:1557-63.
4. Shoag JE, Nyame YA, Gulati R, Etzioni R, Hu JC. Reconsidering the trade-offs of prostate cancer screening. N Engl J Med 2020;382:2465-8.
5. Roobol MJ. Screening for prostate cancer: are organized screening programs necessary? Transl Androl Urol 2018;7:4-11.
6. Drost F-JH, Osses DF, Nieboer D, et al. Prostate MRI, with or without MRI-targeted biopsy, and systematic biopsy for detecting prostate cancer. Cochrane Database Syst Rev 2019;4:CD012663.
7. Ahmed HU, El-Shater Bosaily A, Brown LC, et al. Diagnostic accuracy of multi-parametric MRI and TRUS biopsy in prostate cancer (PROMIS): a paired validating confirmatory study. Lancet 2017;389:815-22.
8. Sathianathen NJ, Omer A, Harriss E, et al. Negative predictive value of multiparametric magnetic resonance imaging in the detection of clinically significant prostate cancer in the Prostate Imaging Reporting and Data System era: a systematic review and meta-analysis. Eur Urol 2020;78:402-14.
9. Kasivisvanathan V, Rannikko AS, Borghi M, et al. MRI-targeted or standard biopsy for prostate-cancer diagnosis. N Engl J Med 2018;378:1767-77.
10. Ahdoot M, Wilbur AR, Reese SE, et al. MRI-targeted, systematic, and combined biopsy for prostate cancer diagnosis. N Engl J Med 2020;382:917-28.
11. Rouvière O, Puech P, Renard-Penna R, et al. Use of prostate systematic and targeted biopsy on the basis of multiparametric MRI in biopsy-naive patients (MRI-FIRST): a prospective, multicentre, paired diagnostic study. Lancet Oncol 2019;20:100-9.
12. Kasivisvanathan V, Stabile A, Neves JB, et al. Magnetic resonance imaging-targeted biopsy versus systematic biopsy in the detection of prostate cancer: a systematic review and meta-analysis. Eur Urol 2019;76:284-303.
13. Nordström T, Picker W, Aly M, et al. Detection of prostate cancer using a multistep approach with prostate-specific antigen, the Stockholm 3 test, and targeted biopsies: the STHLM3 MRI Project. Eur Urol Focus 2017;3:526-8.
14. Nordström T, Jäderling F, Carlsson S, Aly M, Grönberg H, Eklund M. Does a novel diagnostic pathway including blood-based risk prediction and MRI-targeted biopsies outperform prostate cancer screening using prostate-specific antigen and systematic prostate biopsies? Protocol of the randomised study STHLM3MRI. BMJ Open 2019;9(6):e027816.
15. Chan A-W, Tetzlaff JM, Altman DG, et al. SPIRIT 2013 statement: defining standard protocol items for clinical trials. Ann Intern Med 2013;158:200-7.
16. Pepe MS. The statistical evaluation of medical tests for classification and prediction. Oxford statistical science series no. 31. New York: Oxford University Press, 2003.
17. Pepe MS, Alonzo TA. Comparing disease screening tests when true disease status is ascertained only for screen positives. Biostatistics 2001;2:249-60.
18. Alonzo TA, Pepe MS, Moskowitz CS. Sample size calculations for comparative studies of medical tests for detecting presence of disease. Stat Med 2002;21:835-52.
19. Grönberg H, Adolfsson J, Aly M, et al. Prostate cancer screening in men aged 50-69 years (STHLM3): a prospective population-based diagnostic study. Lancet Oncol 2015;16:1667-76.
20. Ström P, Nordström T, Aly M, Egevad L, Grönberg H, Eklund M. The Stockholm-3 model for prostate cancer detection: algorithm update, biomarker contribution, and reflex test potential. Eur Urol 2018;74:204-10.
21. Moore CM, Kasivisvanathan V, Eggener S, et al. Standards of reporting for MRI-targeted biopsy studies (START) of the prostate: recommendations from an international working group. Eur Urol 2013;64:544-52.
22. Schulz KF, Altman DG, Moher D. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. BMJ 2010;340:c332.
23. Grönberg H, Eklund M, Picker W, et al. Prostate cancer diagnostics using a combination of the Stockholm3 blood test and multiparametric magnetic resonance imaging. Eur Urol 2018;74:722-8.
24. Epstein JI, Egevad L, Amin MB, Delahunt B, Srigley JR, Humphrey PA. The 2014 International Society of Urological Pathology (ISUP) consensus conference on Gleason grading of prostatic carcinoma: definition of grading patterns and proposal for a new grading system. Am J Surg Pathol 2016;40:244-52.
25. Schoots IG, Roobol MJ, Nieboer D, Bangma CH, Steyerberg EW, Hunink MGM. Magnetic resonance imaging-targeted biopsy may enhance the diagnostic accuracy of significant prostate cancer detection compared to standard transrectal ultrasound-guided biopsy: a systematic review and meta-analysis. Eur Urol 2015;68:438-50.
26. Hernán MA, Robins JM. Per-protocol analyses of pragmatic trials. N Engl J Med 2017;377:1391-8.
27. Wang R, Lagakos SW, Ware JH, Hunter DJ, Drazen JM. Statistics in medicine — reporting of subgroup analyses in clinical trials. N Engl J Med 2007;357:2189-94.
28. Klotz L, Chin J, Black PC, et al. Comparison of multiparametric magnetic resonance imaging-targeted biopsy with systematic transrectal ultrasonography biopsy for biopsy-naive men at risk for prostate cancer: a phase 3 randomized clinical trial. JAMA Oncol 2021;7:534-42.
29. Dell'Oglio P, Stabile A, Soligo M, et al. There is no way to avoid systematic prostate biopsies in addition to multiparametric magnetic resonance imaging targeted biopsies. Eur Urol Oncol 2020;3:112-8.
30. Kuhl CK, Bruhn R, Krämer N, Nebelung S, Heidenreich A, Schrading S. Abbreviated biparametric prostate MR imaging in men with elevated prostate-specific antigen. Radiology 2017;285:493-505.
31. van der Leest M, Israël B, Cornel EB, et al. High diagnostic performance of short magnetic resonance imaging protocols for prostate cancer detection in biopsy-naïve men: the next step in magnetic resonance imaging accessibility. Eur Urol 2019;76:574-81.
32. Bass EJ, Pantovic A, Connor M, et al. A systematic review and meta-analysis of the diagnostic accuracy of biparametric prostate MRI for prostate cancer in men at risk. Prostate Cancer Prostatic Dis 2020 November 20 (Epub ahead of print).
33. Callender T, Emberton M, Morris S, Pharoah PDP, Pashayan N. Benefit, harm, and cost-effectiveness associated with magnetic resonance imaging before biopsy in age-based and risk-stratified screening for prostate cancer. JAMA Netw Open 2021;4(3):e2037657.