STHLM3 MR Phase II

Randomized diagnostic trial

STATISTICAL ANALYSIS PLAN

Authors

Tobias Nordström[a,b]
Andrea Discacciati[a]
Martin Eklund[a]

[a] Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden;
[b] Department of Clinical Sciences at Danderyd Hospital, Karolinska Institutet, Stockholm, Sweden.

Revisions

Date	Authors	Reason	Main changes
180217	ME	First draft of SAP	
180515	ME	Feedback after initial DSMB meeting	Hypotheses and aims clarified.
191115	AD	Updated sample size calculations.	Updated sample size calculations to support decision to expand study in order to have sufficient power for within-arm comparisons of Stockholm3 vs. PSA in the experimental arm.
191127	ME/AD/ TN	Publication plan	Adjustment definition TBx. Introducing publication plan with 2 main manuscripts. Clarification and specification of the wording of the hypotheses. Clarification of follow-up time.
200324	ME	Covid-19	Allow for potential analysis of the trial in two phases due to the Covid19 pandemic.

201223	ME	Description of post-hoc analyses	Description of analyses where biopsies and biopsy results on men with negative MRI but Stockholm≥25% are ignored.

Contents

Abbreviations

PSA	Prostate specific antigen
S3M	Stockholm3 model
SBx	Systematic biopsy
TBx	Targeted biopsy
TPF	True positive fraction
FPF	False positive fraction
DP	Detection Probability
rTPF	Relative true positive fraction
rFPF	Relative false positive fraction
rDP	Relative Detection Probability
BR	Biopsy rate
rBR	Relative biopsy rate
ITT	Intention-to-treat
PP	Per protocol
MP	Multiple imputation
ICER	Incremental cost-effectiveness ratio
MRI	Magnetic resonance imaging
DSMB	Data safety and monitoring committee
SAP	Statistical analysis plan
PI-RADS	Prostate imaging reporting and data system
ISUP	International Society of Urological Pathology
DRE	Digital rectal examination
SEK	Swedish Kronor
QALY	Quality-adjusted life-years

Preface

This statistical analysis plan (SAP) describes the planned analyses for STHLM3MR Phase 2 (NCT03377881) (herein referred to as STHLM3MR for short). STHLM3MR is a study with first a paired design step for the blood-tests PSA and Stockholm3 and a second randomized step where men with increased risk in the first step (based on either PSA or Stockholm3) are randomized to systematic or MRI+systematic+targeted biopsies, respectively (Figure 1). The planned analyses identified in this SAP will be included in future manuscripts. Exploratory analyses not necessarily identified in this SAP may be performed to support planned analyses. Any post-hoc exploratory or unplanned analyses not specified in this SAP will be identified as such in manuscripts for publication, and added as addenda to the SAP. To ensure blinding, arm allocations are stored in a separate location accessible only by an unblinded statistician. The SAP may be updated during the course of the trial but will be finalized before database lock or any comparative analyses.

Design

The study design is outlined in detail in the study protocol. Figure 1 shows a schematic overview of the study design. The design combines a paired step (PSA vs. Stockholm3) and a randomized step (systematic biopsies vs. MRI plus targeted and systematic biopsies on MRI positive men). The rationale behind the paired first step is that it maximizes statistical power (for a given sample size) without any clear risk of introducing biased comparisons[1]. The different biopsy strategies could also be compared in a paired design (like in e.g. Ahdoot et al.[2], Rouvière et al.[3], and Grönberg et al.[4]), however this markedly increases risk of introducing bias (since bleeding artefacts would interfere with performing the targeted biopsies if the systematic biopsies are performed first, and vice versa, even in the case of different urologists performing the two biopsy techniques within the same man, like in Rouvière et al.[3]). We therefore chose to have a randomized second step to enable higher quality in the comparison of biopsy strategies (systematic vs. MRI plus targeted and systematic biopsies for MRI positive men, defined as PIRADS ≥ 3). For safety reasons, all men with very high risk (Stockholm3 test ≥25%) will be recommended systematic biopsies. It should be noted that we by the Stockholm3 test in this study refer to the Stockholm3 test as described in Ström et al.[5] ***without*** the inclusion of prostate volume and digital rectal examination (DRE) as predictors (i.e., the set of predictors include age, first-degree family history of prostate cancer [yes/no], and previous biopsy [yes/no], total PSA, free PSA, ratio of free/total PSA, hK2, MIC1, MSMB, and a genetic score).

The design permits a large number of comparative contrasts to be performed (constructed of combinations of using either PSA or Stockholm3, using MRI or not, and using targeted biopsies or systematic biopsies or both can be compared). Specifically, the following diagnostic strategies can be compared:

1. PSA+SBx
2. PSA+MRI+TBx
3. PSA+MRI+TBx+SBx
4. S3M+SBx
5. S3M+MRI+TBx
6. S3M+MRI+TBx+SBx

In addition, the following diagnostic strategies are also possible:

7. (PSA | S3M)+SBx
8. (PSA | S3M)+MRI+TBx

9. (PSA | S3M)+MRI+TBx+SBx,

where (PSA | S3M) denotes positive on either PSA or S3M screening tests, SBx denotes systematic biopsies, and TBx denotes targeted biopsy. To be clear, only MRI positive men (PIRADS ≥ 3) are biopsied in the strategies that include MRI (apart from men with S3M ≥ 25%). This is true irrespectively of the biopsy procedure (i.e. SBx is also only performed in MRI positive men in these strategies) (Figure 1). ***The above strategies will in the rest of this document be referred to as Strategy 1, 2, 3, 4, 5, 6, 7, 8, and 9, respectively.***

Randomization stratified on disease risk [sextiles of the Stockholm3 test] will be used. Further, block allocation will be performed, with each block consisting of five men, two of which will be randomized to systematic biopsies and three will be randomized to the MRI arm. This means that randomization guarantees a proportional number of men in each arm and with more evenly distributed characteristics in terms of disease risk and test concordance.

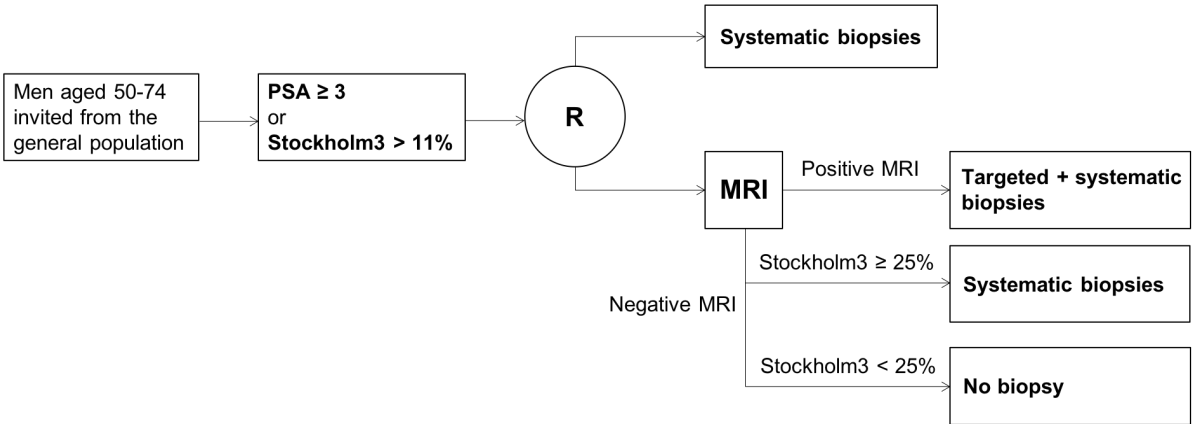

Figure 1: *Overview of the STHLM3 MR Phase 2 study design. Men aged 50-74 are invited to the study from the general population. Blood is sampled from study participants and PSA as well as Stockholm3 are measured. Men with elevated prostate cancer risk (PSA ≥ 3 ng/ml or Stockholm3 ≥ 11%) are randomized to be referred to either systematic biopsies (control arm) or undergo MRI and targeted plus systematic biopsies in case the MRI indicates areas of the prostate suggestive of prostate cancer (PIRADS ≥ 3). The design thus combines a paired step where PSA and Stockholm3 can be compared (paired screen positive design) and a randomized step where systematic biopsies can be contrasted to MRI and subsequent targeted and systematic biopsies for MRI positive men.*

Hypotheses

Overarching primary hypothesis

The overarching primary hypothesis of the STHLM3MRI trial is that a diagnostic pathway using the Stockholm3 test to select men for further workup using MRI followed by targeted biopsies and systematic biopsies in MRI positive men (S3M+MRI+TBx+SBx; Strategy 6) has non-inferior sensitivity for detecting clinically significant cancer (ISUP grade ≥2) and shows superior specificity (reduction in number of performed biopsy procedures and detected ISUP 1 tumours) compared with the diagnostic pathway using systematic biopsies in men with PSA ≥ 3 ng/mL (PSA+SBx; Strategy 1).

Additional hypotheses

1. When compared with performing systematic biopsies for men with elevated risk of prostate cancer in prostate cancer screening, targeted and systematic prostate biopsies performed on MRI positive men has non-inferior sensitivity for detecting clinically significant cancer (ISUP grade group ≥2) and reduces the number of performed biopsy procedures, which also translates to lower proportion of men with elevated risk who experience severe post-biopsy infections. Elevated risk can here be defined using PSA or S3M – we clarify the exact contrasts for testing this hypothesis below.
2. When compared with performing systematic biopsies for men with elevated risk of prostate cancer in prostate cancer screening, targeted biopsies only performed on MRI positive men has non-inferior sensitivity for detecting clinically significant cancer (ISUP grade group ≥2) and reduces the number of performed biopsy procedures, which also translates to lower proportion of men with elevated risk who experience severe post-biopsy infections. Elevated risk can here be defined using PSA or S3M – we will clarify the exact contrasts for testing this hypothesis below.
3. A diagnostic chain consisting of Stockholm3 followed by MRI and targeted+systematic biopsies (S3M+MRI+TBx+SBx) versus a diagnostic chain based on PSA ≥3 ng/ml followed by MRI and targeted+systematic biopsies (PSA+MRI+TBx+SBx) will lead to: non-inferior sensitivity for detecting clinically significant cancer (ISUP grade group ≥ 2); an inferior sensitivity for ISUP1 cancers (i.e. reduced overdiagnosis); and a reduction in the number of MRI examinations and performed biopsies.
4. A diagnostic pathway using the Stockholm3 test to select men for further workup using MRI followed by ONLY targeted biopsies (S3M+MRI+TBx) has non-inferior sensitivity for detecting clinically significant cancer (ISUP grade group ≥2) and reduces the number of performed biopsy procedures compared with a diagnostic pathway using systematic biopsies in men with PSA ≥3 ng/mL (PSA+SBx).
5. Biopsy compliance is higher after biopsy is recommended based on MRI compared to recommended without MRI.
6. SBx in the MRI arm has superior sensitivity than SBx in the non-MRI arm (due to cognitive fusion).
7. A diagnostic chain consisting of Stockholm3 followed by MRI and targeted+systematic biopsies (S3M+MRI+TBx+SBx) is cost-effective (ICER < 750 000 SEK per QALY gained) compared to a diagnostic chain based on PSA ≥3 ng/ml followed by MRI and targeted+systematic biopsies (PSA+MRI+TBx+SBx) due to reductions in number of performed procedures (men undergoing MRI and biopsy).
8. A diagnostic chain using the Stockholm3 test to select men for further workup using MRI and targeted+systematic biopsies (S3M+MRI+TBx+SBx) is cost-effective compared to a diagnostic chain using systematic biopsies in men with PSA ≥3 ng/ml (PSA+SBx).
9. Adding prostate volume as a variable in the diagnostic chain with Stockholm3 test (i.e. using the full Stockholm3 model described in Ström et al.[5]) and MRI/Fusion biopsies improves model precision, leading to further improvements in specificity compared to the use of the Stockholm3 test without the inclusion of prostate volume.

Publications strategy of main study results

A large number of publications will likely be written based on the data collected within the STHLM3MRI trial. The main study results from the trial will be reported in the two first publications:

Publication 1

Publication 1 will as primary contrast report the comparison of Strategy 3 vs. 1 (PSA+MRI+TBx+SBx vs PSA+SBx). Additional analysis in the report will include contrasts of Strategies 2 vs 1.

This comparison uses only the randomized step of the trial design and is motivated by the need to provide level 1 evidence about the performance of MRI+targeted+systematic biopsies versus systematic biopsies alone in men with PSA ≥3 ng/ml in a screening-by-invitation context, where such data is lacking entirely. The choice of using systematic biopsies without MRI as the comparator hinges on the fact that, presently, this is the typical diagnostic strategy offered to men with an elevated PSA (Ahmed et al.[6]). Furthermore, level 1 evidence about a mortality benefit from early detection of prostate cancer is available only for a diagnostic strategy based on PSA+systematic biopsies (Schröder et al. 2009[7] and 2014[8]). Thus, Publication 1 will assess whether introducing PSA+MRI+TBx+SB into prostate cancer screening can diagnose clinically significant prostate cancer with non-inferior sensitivity to PSA+SB, for which there is level 1 evidence of a reduction in prostate cancer mortality. Publication 1 will cover additional hypotheses 1 and 2.

Publication 2

Publication 2 will report the contrasts of Strategy 6 vs. 3 (S3M+TBx+SBx vs. PSA+TBx+SBx) and Strategy 6 vs. 1 (S3M+TBx+SBx vs. PSA+SBx), where the latter contrast corresponds to assessing the performance of the entire diagnostic chain of using Stockholm3 test to select men for further workup using MRI followed by targeted biopsies and systematic biopsies compared to using PSA followed by systematic biopsies (i.e. testing the overarching primary hypothesis of the trial). Sensitivity analyses will include contrast of Strategy 5 vs. 2 and Strategy 5 vs. 1. In addition, for the contrast of Strategy 6 vs. 1, we will – if Stockholm3 ≥ 11% is more sensitive than PSA ≥ 3 – also report results at the operating point of Stockholm3 (i.e. the Stockholm3 cutoff) that gives equal sensitivity as PSA ≥ 3 within the experimental arm (analogously to how we performed the STHLM3 trial, see Grönberg et al.[9]), see the “Additional analysis” section in this document for more information.

This comparison is motivated by the fact that MRI+targeted biopsy will be used more and more frequently, with the possibility of them eventually replacing systematic biopsies as the de-facto standard diagnostic tool[1]. For example, the National Institute for Health Care and Excellence (NICE) in the UK already recommends MRI as “the first-line investigation for people with suspected clinically localised prostate cancer”[2]. Therefore, we aim to compare the Stockholm3 test with the PSA test – in terms of sensitivity (ISUP≥2), specificity (ISUP1), number of biopsies and number of MRI scans – as a tool to select men for further workup, as well as compare the entire diagnostic chain of using Stockholm3 followed by MRI and TBx+SBx to the traditional diagnostic chain of using PSA followed by SBx (for which there exists level 1 evidence of reduced prostate cancer specific mortality when used for prostate cancer screening (Schröder et al. 2009[7] and 2014[8])). Publication 2 will cover the primary overarching hypothesis, as well as additional hypotheses 3 and 4.

Another way to motivate the order of these two publications is that the first publication will assess whether MRI and TBx+SBx improves diagnostic accuracy in a population based

---

[1] As also pointed out in by Professor Mark Emberton and Professor Caroline M. Moore in the review of the study protocol [Nordström et al, BMJ Open 2019]. The comments are now available at the journal’s website.

[2] See https://www.nice.org.uk/guidance/NG131/chapter/recommendations#multiparametric-mri-and-protocol-for-active-surveillance

screening-by-invitation setting, as it seems to do in clinical cohorts (Kasivisvanathan et al.[10] and Ahdoot et al.[2], Rouvière et al.[3], and Grönberg et al.[4]), whereas the second publication will assess whether Stockholm3 can improve selection of men to undergo MRI (an key point since MRI is an expensive and scarce resource and population-based screening involving MRI will lead to large number of MRI examinations).

The results presented in Publications 1 and 2 will thus together cover the testing of the overarching primary hypothesis, as well as additional hypothesis 1, 2, 3, 4, and 5.

Aims and endpoints

The primary aim and key secondary aims of this trial are described below (corresponding with the overarching primary hypothesis and additional hypothesis 1-4 above), together with definitions of study variables (independent variables and outcome variables). A description of how and on which data statistical testing will be performed is specified in the Statistical Analysis section.

Primary aim

To test whether a diagnostic pathway using the Stockholm3 test to select men for further workup using MRI followed by targeted biopsies and systematic biopsies (Strategy 6; S3M+MRI+TBx+SBx) has non-inferior sensitivity for detecting clinically significant cancer (ISUP grade ≥2) and shows superior specificity (reduction in number of performed biopsy procedures and detected ISUP 1 tumours) compared with the diagnostic pathway using systematic biopsies in men with PSA ≥ 3 ng/mL (Strategy 1; PSA+SBx).

Key secondary aims

1. To test whether targeted prostate and systematic biopsies performed in MRI positive men will lead to: non-inferior sensitivity for detecting clinically significant cancer (ISUP grade group ≥2); reduced number of performed biopsy procedures; and lower proportion of men with elevated risk who experience severe post-biopsy infections compared to systematic biopsies for men with elevated risk (defined as being positive on PSA and/or Stockholm3) of prostate cancer in prostate cancer screening.
2. To test whether targeted prostate biopsies performed in MRI positive men will lead to: non-inferior sensitivity for detecting clinically significant cancer (ISUP grade ≥2); reduced number of performed biopsy procedures; and lower proportion of men with elevated risk who experience severe post-biopsy infections compared to systematic biopsies for men with elevated risk (defined as being positive on PSA and/or Stockholm3) of prostate cancer in prostate cancer screening.
3. To test whether a diagnostic chain consisting of Stockholm3 followed by MRI and targeted+systematic biopsies versus a diagnostic chain based on PSA ≥3 ng/ml followed by MRI and targeted+systematic biopsies will lead to: non-inferior sensitivity for detecting clinically significant cancer (ISUP grade group ≥ 2); an inferior sensitivity for ISUP1 cancers (i.e. reduced overdiagnosis); and a reduction in the number of MRI examinations and performed biopsies.
4. To test whether a diagnostic pathway using the Stockholm3 test to select men for further workup using MRI followed by ONLY targeted biopsies (S3M+MRI+TBx) has non-inferior sensitivity for detecting clinically significant cancer (ISUP grade ≥2) and reduces the number of performed biopsy procedures compared with a diagnostic pathway using systematic biopsies in men with PSA ≥3 ng/mL (PSA+SBx).

Additional aims

Additional aims corresponding to hypotheses 4-8 above will be assessed.

Main endpoints

Primary endpoint:

1. Diagnosed ISUP grade ≥ 2 cancers

Key secondary endpoints

2. Diagnosed ISUP grade 1 cancers
3. Performed biopsies
4. Performed MRI examinations

Additional endpoints

The primary and secondary endpoints are reported in the tables below. All these endpoints can be used for comparisons between the nine diagnostic strategies listed in the ‘Design’ section.

Study variables

Primary endpoints

Variable	Measure	Comment
Clinically significant prostate cancer	Yes/No	ISUP ≥ 2

Key secondary endpoints

Variable	Measure	Comment
Non-clinically significant prostate cancer	Yes/No	ISUP = 1
Biopsy performed	Yes/No	Any biopsy procedure performed
MRI performed	Yes/No	MR procedure performed

Other secondary endpoints

Variable	Measure	Comment
Any cancerous finding	Yes/No	ISUP ≥ 1
ISUP ≥ 3 prostate cancer	Yes/No	ISUP ≥ 3
ISUP 2 through 5	Yes/No	Four separate endpoints:

		(1) ISUP=2, (2) ISUP=3, (3) ISUP=4, and (4) ISUP=5
Lesion volume	ml	Ellipsoid calc (4*pi*h*w*l /3)(4*pi*h*w*l /3) Total volume of all reported lesions PIRADS≥3
Number of biopsies	Integer	No of reported biopsy needles
Cancer length	mm	total mm of cancer in reported biopsy
% Gleason ≥4	%	summarized % Gleason ≥4 in needles with cancer
Maximum cancer core length	mm	maximum length of cancer in any core
Maximum GS ≥ 4 cancer core length	mm	maximum length of Gleason ≥ 4 cancer in any core
Serious adverse events	Yes/No	(1) Hospitalisation within 30 days after biopsy procedure, (2) infection treated with antibiotics within 30 days after biopsy procedure, or (3) death within 30 days after biopsy procedure

Independent variables

Variable	Measure	Comment
Age	Years	At referral creation
Previous prostate biopsy	Yes/No	Patient self-reported
Family history of prostate cancer	Yes/No	Any first degree relative with prostate cancer
PSA	ng/ml	At blood test
free PSA	ng/ml	At blood test

Stockholm3 risk score	% risk of ISUP≥2	At blood test
prostate volume	ml	MRI defined
PI-RADS	1–5 (integer)	Maximum PI-RADS score

Statistical analysis

All analyses will be performed after the study is completed and the database is released. All statistics, including tables, figures and listings, will be performed using R version >3.5.
Men creating referral before 2020-01-09 are included in main analysis. All participants with will be followed a minimum of 200 days after receiving the result from their blood sample (PSA and Stockholm3 score). In addition, biopsied men will be followed at least 30 days post-biopsy to monitor adverse events, and participants undergoing radical prostatectomy prior to database lock will be followed until pathology results from the prostatectomy are available.

Study populations

1. *ITT population* includes all men who:
    a. signed the written informed consent to participate in the study,
    b. fulfilled all inclusion criteria and none of the exclusion criteria,
    c. were randomised to either study arm.

Conditions for excluding patients from the ITT population, based on deviations from the study protocol:

Randomisation arm	**MRI**	**No biopsy**	**SBx only**	**TBx only**	**SBx and TBx**
Standard arm	MRI or no MRI	Include	Include	Include	Include
Experimental arm	MRI or no MRI	Include	Include	Include	Include

Randomised allocation and analysis group for ITT analyses:

Randomisation arm	**Test received**	**Analysis group**
Standard arm	Systematic biopsy (with or without MRI)	Standard arm
Standard arm	Targeted biopsy	Standard arm
Standard arm	Systematic biopsy and Targeted biopsy	Standard arm
Standard arm	No biopsy / Other	Standard arm
Experimental arm	Systematic biopsy (with or without MRI)	Experimental arm
Experimental arm	Targeted biopsy	Experimental arm

Experimental arm	Systematic biopsy and Targeted biopsy	Experimental arm
Experimental arm	No biopsy / Other	Experimental arm

2. *PP population* includes men who:
    a. are included in the ITT population,
    b. have a valid PSA value and Stockholm3 score,
    c. have a complete systematic pathology report if randomised to standard arm (but no targeted pathology report),
    d. have a complete MRI report if randomised to experimental arm and (i), (ii), or (iii), as appropriate,
        i. have a complete systematic and a complete targeted pathology report if PIRADS≥3,
        ii. have a complete systematic pathology report if PIRADS<3 and Stockholm3≥25% (but no targeted pathology report),
        iii. did not undergo any biopsy if PIRADS<3 and Stockholm3<25%.

Conditions for excluding patients from the PP population, based on deviations from the study protocol:

Randomisation arm	**Further specifications**	**No biopsy**	**SBx only**	**TBx only**	**SBx and TBx**
Standard arm	No MRI	Exclude	Include	Exclude	Exclude
Standard arm	MRI	Exclude	Exclude	Exclude	Exclude
Experimental arm	No MRI	Exclude	Exclude	Exclude	Exclude
Experimental arm	MRI PIRADS≥3	Exclude	Exclude	Exclude	Include
Experimental arm	MRI PIRADS<3 and Stockholm3≥25%	Exclude	Include	Exclude	Exclude
Experimental arm	MRI PIRADS<3 and Stockholm3<25%	Include	Exclude	Exclude	Exclude

The analyses will be performed and reported on both the ITT and PP population.

Patients' characteristics

Patients' characteristics will be presented with descriptive statistics, overall, by study arm, 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.

Analyses

Data structure

The table below lays out the general data structure for the STHLM3-MRI trial.

Standard arm

	Endpoint = positive (yes)		Endpoint = negative (no)	
	PSA test positive	PSA test negative	PSA test positive	PSA test negative
Stockholm3 test positive	a	b	e	f
Stockholm3 test negative	c	[d]	g	[h]
Experimental arm				
	Endpoint = positive (yes)		Endpoint = negative (no)	
	PSA test positive	PSA test negative	PSA test positive	PSA test negative
Stockholm3 test positive	a’	b’	e’	f’
Stockholm3 test negative	c’	[d’]	g’	[h’]

Note: since the study protocol dictates that only those patients who screen positive on either screening test are referred for further work-up, the number of patients reported between brackets are unknown. In a standard screen-positive study, the total number of patients [d]+[h] (and [d’]+[h’]) is known. However, this study combines a paired and a randomised design and only those men who screen positive on either test are randomised. Therefore, the quantities [d]+[h] and [d’]+[h’] are unknown in the present study.

Contrasts between study arms (unpaired design)

Analyses will compare the difference in detection probabilities (DPs) between study arms. The DP is the probability of being endpoint-positive given the study arm and, possibly, either or both screening tests being positive. For example the DP of ISUP1 in men randomised to the experimental arm and Stockholm3-positive is equal to Pr(ISUP1 | Stockholm3≥11%, Experimental arm).

Absolute scale. The absolute difference in DPs is defined as the DP in the experimental arm minus the DP in the standard arm ($\Delta DP = DP_{Exp} - DP_{Std}$) or vice versa, as appropriate. It is estimated by plugging into the formula the observed proportions. An approximate 100(1-α)% two-sided Wald confidence interval for ΔDP is calculated as

$$\widehat{\Delta DP} \pm z_{\alpha/2}\sqrt{\frac{\widehat{DP}_{Exp}*(1-\widehat{DP}_{Exp})}{n_{Exp}} + \frac{\widehat{DP}_{Std}*(1-\widehat{DP}_{Std})}{n_{Std}}}.$$

Relative scale. The relative difference in DPs is defined as the DP in the experimental arm divided by the DP in the standard arm ($rDP = DP_{Exp}/DP_{Std}$) or vice versa, as appropriate. It is estimated by plugging into the formula the observed proportions. An approximate 100(1-α)% two-sided Wald confidence interval for rDP is calculated as

$$exp\left(log(\widehat{rDP}) \pm z_{\alpha/2}\sqrt{\frac{1}{\widehat{DP}_{Exp}*n_{Exp}} - \frac{1}{n_{Exp}} + \frac{1}{\widehat{DP}_{Std}*n_{Std}} - \frac{1}{n_{Std}}}\right).$$

Contrasts 3 vs 1, 2 vs 1, and 9 vs 7 will be analysed using absolute differences (Study 1). Contrasts 6 vs 1, and 5 vs 1 will be analysed using relative differences (Study 2).

Interpretation as relative sensitivity

It should be noted that rDPs can (under the assumption of no false positive biopsies, see SAP Appendix 1) be interpreted as relative true positive fractions (rTPF) (ie, relative sensitivities).

Contrasts within study arm (paired design)

Analyses will compare the true positive fraction between Stockholm3 and PSA, within either study arm. Comparisons will be made on a relative scale. rTPF is defined as $TPF_{Stockholm3}/TPF_{PSA}$ or vice versa, as appropriate.

The rTPF (standard arm) is estimated as (a+b)/(a+c) (or (a+c)/(a+b), as appropriate) and an approximate 100(1-α)% two-sided confidence interval for rTPF is calculated as

$$exp\left(log(\widehat{rTPF}) \pm z_{\alpha/2}\sqrt{\frac{b+c}{(a+b)(a+c)}}\right).$$

Analogous formulas are used for comparisons within the experimental arm and for the relative False Positive Fraction.

Note: the quantity (a+b)/(a+c) estimates the rTPF in both the enrolled and randomised population. It also estimates the ratio of detection probabilities like Pr(ISUP≥2, Stockholm≥11% | Experimental arm)/Pr(ISUP≥2, PSA≥3 ng/ml | Experimental arm).

Non-inferiority and superiority tests

The null and the alternative hypothesis for a non-inferiority test for the ΔDP are:

$$H_0: \Delta DP \leq -\delta$$
$$H_a: \Delta DP > -\delta$$

where $\delta > 0$ is the non-inferiority margin. This means that non-inferiority for a specific endpoint will be claimed if the lower boundary of the two-sided $(2\alpha \times 100)\%$ confidence interval for the ΔDP does not cover $-\delta$.

The null and the alternative hypothesis for a superiority test for the ΔDP are:

$$H_0: \Delta DP \leq \theta$$
$$H_a: \Delta DP > \theta$$

where $\theta \geq 0$ is the superiority margin. This means that superiority for a specific endpoint will be claimed if the lower boundary of the two-sided $(2\alpha \times 100)\%$ confidence interval for the ΔDP does not cover θ.

For comparisons on a relative scale (rTPF), the null and the alternative hypothesis for non-inferiority and superiority tests are

$$H_0: rDP \leq exp(-\delta)$$

$$H_a: rDP > exp(-\delta)$$

and

$$H_0: rDP \leq exp(\theta)$$
$$H_a: rDP > exp(\theta)$$

respectively, with non-inferiority and superiority margins equal to $\delta > 0$ and $\theta \geq 0$.

One-sided p-values will be calculated based on the test considered (non-inferiority or superiority).

Switching from non-inferiority to superiority: if the two-sided $(2(1-\alpha) \times 100)\%$ confidence interval for ΔDP (rTPF) not only lies entirely above the non-inferiority margin, but also above the superiority margin, superiority will be claimed at the same alpha-level set for the non-inferiority test. In this case, we will also calculate the p-value associated with a test for superiority.

Primary endpoints

Contrasts between study arms: we will assess the non-inferiority of the experimental arm versus the standard arm in detecting ISUP≥2 cancers ($\delta = 0.04$ or $exp(-\delta) = 0.78$, as appropriate). The α level is set to 0.025. Two-sided 95% confidence intervals will be reported.

Contrasts within study arms: we will assess the non-inferiority of the Stockholm3 test versus PSA in detecting ISUP≥2 cancers ($exp(-\delta) = 0.78$) and the superiority of PSA versus the Stockholm3 test in detecting ISUP1 cancers (i.e. a lower proportion of ISUP1 cancer detected according to the Stockholm3 test) ($exp(-\theta) = 1$). The α level is set to 0.025 for both tests. Two-sided 95% confidence intervals will be reported.

Key secondary endpoints

Contrasts between study arms: we will assess the superiority of the standard arm versus the experimental arm in detecting ISUP1 cancers (i.e. a lower proportion of ISUP1 cancer detected in the experimental arm) ($\theta = 0$ or $exp(-\theta) = 1$, as appropriate). The α level is set to 0.025. Two-sided 95% confidence intervals will be reported.

Contrasts within study arms: we will assess the superiority of PSA versus the Stockholm3 test in detecting ISUP1 cancers (i.e. a lower proportion of ISUP1 cancer detected according to the Stockholm3 test) ($exp(-\theta) = 1$). The α level is set to 0.025. Two-sided 95% confidence intervals will be reported.

Secondary endpoints

We will report the proportion of men with post-biopsy SAEs (see table “Secondary endpoints”) by study arm, where applicable.

Sample size calculations

Original sample size calculation (performed in March 2017)

Basic data and assumptions used in the sample size calculations

We used data from the STHLM3 trial[10] for sample size calculations. In this data, 18% of men with PSA ≥ 3 had a clinically significant prostate cancer when biopsied with SBx. We further

noted that rTPF=1.25 for clinically significant prostate cancer comparing MRI+TBx with SBx based on the results from a meta-analysis (Schoots et al. 2015[3]). We set the noninferiority delta to 4 percentage points for demonstrating noninferiority with respect to sensitivity of clinically significant prostate cancer. We set the alpha to 0.025.

Primary contrast

This study was originally powered for the contrast of Strategy 6 vs 1. Simulating 1000 trials (by bootstrapping from the STHLM3 data) under the assumptions outlined in the preceding section 303 men need to biopsied in the SBx arm based on PSA ≥ 3 to have 80% power to demonstrate non-inferior sensitivity of S3M+MRI+TBx compared with PSA+SBx. This means that at least 415 men need to be biopsied in the SBx arm (since some men are not randomized based on PSA ≥ 3 but on S3M ≥ 11%) and, consequently, 623 to the MRI arm (because of the 2:3 randomization). Total number of men undergoing workup according to protocol (SBx in the no MRI arm and MRI and TBx if PI-RADS ≥ 3 in the MRI arm) is thus 1038. Assuming 20% dropout, 1300 men need to be randomized.

Updated sample size calculations (performed during Spring and Summer 2019)

We revised the sample size calculations above in order to have sufficient statistical power to answer comparisons of Stockholm3 vs. PSA within the experimental arm.

The updated, final sample size is the result of a balance between time and financial constraints on one hand and the need to maximise the power for contrasting Strategies 6 vs. 3 on the other hand. In fact, different assumptions about the joint probability of ISUP≥2 and screening positive on the PSA test [Pr(ISUP≥2, PSA≥3 ng/ml)] and about the TPF of the PSA test [Pr(PSA≥3 ng/ml | ISUP≥2)] lead to different required sample sizes.

Based on the sample size calculation, we decided to increase the number of men invited into the study to ~50000. Based on an updated estimate of the expected participation rate (25%), this will lead to ~12500 men included in the study and to ~2100 randomised men based on PSA or Stockholm3 (assuming ~16.5% of the enrolled men will test positive on either screening test; 13% passed on PSA alone, leading to ~1600 randomized men with PSA ≥ 3 ng/ml), ~1700 of whom will complete the diagnostic chain, assuming a 20% drop-out rate.

The updated sample size will give:

1. 50–90% power for the contrast S3M+TBx+SBx versus PSA+TBx+SBx. This is based on the following assumptions: 0.015–0.03 probability of detecting ISUP≥2 and screening positive on the PSA test (Grönberg et al. 2015[9]), 0.57–0.63 TPF for the PSA test (Thompson et al. 2005), rTPF (Stockholm3≥11% vs PSA≥3) equal to 1 (Grönberg et al. 2015[9]), and a conservative DDR estimate (Alonzo, Pepe, and Moskowitz 2002[13]). The non-inferiority margin was set to $\delta = -log(0.78)$ and alpha to 0.025.

   Note: the contrast S3M+TBx+SBx versus PSA+TBx+SBx is nested within the experimental study arm, where about 7500 men will be included (due to the 2:3 randomisation). We further assumed a 20% drop-out rate, leading to 6000 men available for the analyses.

DR_psa	TPR_psa	power
0.015	0.57	0.48
0.020	0.57	0.60
0.025	0.57	0.70
0.030	0.57	0.77
0.015	0.59	0.52
0.020	0.59	0.64
0.025	0.59	0.73
0.030	0.59	0.81

0.015	0.61	0.55
0.020	0.61	0.67
0.025	0.61	0.77
0.030	0.61	0.84
0.015	0.63	0.59
0.020	0.63	0.71
0.025	0.63	0.80
0.030	0.63	0.87

2. More than 90% power for the contrast PSA+MRI+TBx+SBx vs PSA+SBx (rDP = 1.3, noninferiority margin for ΔDP $\delta = 0.04$, alpha = 0.025).

Thus, contrast 1 drives the required samples size of the study. We therefore powered the study (with respect to sample size for sending out invitations and enrolling participants) to have sufficient statistical power to answer comparisons of Stockholm3 vs. PSA within the experimental arm.

Note: no correction for multiple comparisons was made. This means that each of the three tests (with respect to the ISUP 2, ISUP 1, and biopsy endpoints) has an approximate type I error rate of 2.5% if the corresponding null hypothesis is true. If all three null hypotheses are true and we assume the tests to be independent, the overall type I error rate is approximately 7%. In reality however, these hypotheses are strongly correlated. Thus, the overall type 1 error rate is bounded below by 2.5% and above by 7%.

R code for the power calculations is available at:
https://gist.github.com/anddis/fc1a265d102b509b0eacd59ab065661a

Subgroup analyses

Subgroup analyses will be performed for the following subpopulations:

- Age: [50, 60), [60, 70), [70, 75) years
- PSA: [1.5, 3)[3], [3, 4), [4,10), [10, +inf) ng/mL
- Screen-naive vs not screen-naive patients
- Biopsy-naive vs not biopsy-naive patients

Statistical tests for effect heterogeneity across subpopulations will be performed by jointly testing the interaction (product) terms in generalised linear models or marginal models (Pepe and Alonzo 2001), as appropriate. No correction for multiple comparisons will be made.

Additional analyses

Due to updates in the Stockholm3 assay system to reduce measurement errors in the biomarkers included in the Stockholm3 risk prediction model, Stockholm3 ≥ 11% as a selection criterion for randomization may be more sensitive than PSA ≥ 3 (i.e. more men with clinically significant prostate cancer will be randomized based on the criterion Stockholm3 ≥ 11% compared with PSA ≥ 3). If this turns out to be true, we will perform analyses where we "count backwards" (increase the S3M cutoff) and compare biopsy rates at identical sensitivity for clinically significant prostate cancer when comparing diagnostic strategies involving S3M compared with PSA ≥ 3 (as described in Grönberg et al.[9]).

We will artificially randomise (2:3) those men who screened negative on both screening tests. By doing this, the totals [d]+[h] and [d']+[h'] will become known, which in turn will allow the estimation of quantities like Pr(ISUP≥2, Stockholm3≥11% | Standard arm) (i.e., the probability of ISUP≥2 and Stockholm3≥11% in enrolled men randomised to the standard arm). Contrasts between the two study arms with respect to these quantities will be

[3] Where applicable.

performed using the same methodology described in the section “Contrasts between study arms (unpaired design)”.

We may in additional analyses use regression models to model the DP, TPF, and FPF given covariates. We will employ standard generalised linear models or marginal models (Pepe and Alonzo 2001[14]), as appropriate.

Data Safety Management Board (DSMB)

See protocol.

Handling of missing data

Missing data with respect to outcome data (most importantly, participants who are recommended biopsy but never undergo the procedure) will primarily be handled by performing analyses on both the ITT and the PP populations. The analysis on the PP population inherently makes a missing-completely-at-random (MCAR) assumption. However, there is a chance that there is uneven dropout levels in the two arms. For example, men randomized to undergo MRI may to a higher degree choose to undergo biopsy since there is visual feedback of a lesion. If deemed necessary to understand and interpret study results, we may therefore perform multiple imputation based on the Stockholm3 score to impute outcomes for men who drop out of the study before the biopsy is performed. Briefly, if imputation is performed, it will follow the following protocol:

- *Systematic biopsy arm.* The Stockholm3 test, which is calibrated to systematic biopsy outcomes, will be used to impute biopsy outcome on men who did not undergo biopsy (despite a study recommendation to do so) by performing a Bernoulli experiment using the predicted Stockholm3 risk score for ISUP 2 cancer as a parameter. The analysis will be performed on 1000 multiple imputation datasets and summarized.
- *Experimental biopsy arm.* We will, by using data with the STHLM3MRI trial, fit a model to associate Stockholm3 test and PI-RADS score result to TBx outcome (Stockholm3-TBx). Using this model, we will proceed in a similar way as for the control arm. I.e., we will repeatedly impute outcome using Bernoulli experiments with the predictions from the Stockholm3-TBx model as parameter. The analysis will be performed on 1000 multiple imputation datasets and summarized.

Covid-19 addendum 200223

The Covid-19 pandemic puts a tremendous strain on the entire healthcare system, meaning that the STHLM3-MRI phase II study will be impacted and lead times for patients in the trial will become potentially very long. It is not unlikely likely (at the time of writing 200223) that the study will be prolonged by many months and even years. Therefore, we have decided to open up for the possibility to report on endpoints as they mature. I.e., if STHLM3MRI is unable to continue recruit participants and perform tests according to the study protocol, we open for the possibility to not have one finalized database lock that will be used for all analyses. Rather, we may then lock a database for a specific analysis when there is enough data in the study to test the hypothesis corresponding to the analysis. In particular, we already have enough data collected in the trial for Publication 1. From an ethical point of view, we believe this is the least bad possible approach under the current circumstances. We have a large dataset already collected in the study and we believe that it makes sense to use these data to benefit of patients as soon as possible for the endpoints and analyses that are possible to analyse, rather than waiting for a limited set of men who are left in the study and -- due to Covid-19 -- may not be able to complete the study protocol for a very long time. This plan has been communicated to and approved by the trial’s DSMB.

Additional comment written 200905: The trial could be completed despite the covid-19 pandemic, and the addendum above will not have to be activated. This means that we will have one single database lock for performing all analyses, as originally planned.

Post-hoc analyses

Ignoring biopsy results on men with negative MRI and Stockholm3 ≥ 25%

In order to estimate results in the counterfactual scenario where participants with negative MRI and Stockholm3 ≥25% would not have been referred to undergo systematic biopsy, we will perform analyses where these biopsies are ignored. To be clear, we will not exclude these men from the analyses, but ignore their biopsy and biopsy outcome (i.e., they would enter the analyses as not having had a biopsy or any potential cancer diagnosis). The participants with a negative MRI and Stockholm3 ≥25% will in these analyses thus contribute to the counts of performed MRI scans, but not to the count of biopsies or the cancer count. Apart from this, the analyses will be performed identically as detailed above.

Appendix 1

Let T and S be the events “SBx+TBx positive for a specific ISUP grade” (eg, ISUP≥2) and “SBx positive for a specific ISUP grade”, respectively. Let P be the event “PSA screening test above 3 ng/ml”. Let D be the event “the subject is positive for a specific ISUP grade (true, unobservable status)”.

The main between-arm contrast of Study 1, expressed in relative terms, is given by $\frac{\Pr(T=1|P=1)}{\Pr(S=1|P=1)}$.

This can be rewritten as:

$$\frac{\Pr(T=1|P=1)}{\Pr(S=1|P=1)} = \frac{\Pr(T=1,D=1|P=1)+\Pr(T=1,D=0|P=1)}{\Pr(S=1,D=1|P=1)+\Pr(S=1,D=0|P=1)}$$
$$= \frac{\Pr(T=1,D=1|P=1)+\Pr(T=1|D=0,P=1)\Pr(D=0|P=1)}{\Pr(S=1,D=1|P=1)+\Pr(S=1|D=0,P=1)\Pr(D=0|P=1)} \text{ [assumption]}$$
$$= \frac{\Pr(T=1|D=1,P=1)\Pr(D=1|P=1)}{\Pr(S=1|D=1,P=1)\Pr(D=1|P=1)} = \frac{\Pr(T=1|D=1,P=1)}{\Pr(S=1|D=1,P=1)} = rTPF$$

The third equality holds under the assumption that the FPFs $\Pr(T=1|D=0,P=1)$ and $\Pr(S=1|D=0,P=1)$ are equal to zero, while the fourth equality hinges on the fact that — because of randomisation— the probabilities $Pr(D=1|P=1)$ in the two study arms are the same in expectation.

The equation above can be extended to the other between-arm contrasts.

References

1. Pepe MS. *The Statistical Evaluation of Medical Tests for Classification and Prediction*. 1. publ. in paperback. Oxford Univ. Press; 2004.

2. Ahdoot M, Wilbur AR, Reese SE, et al. MRI-Targeted, Systematic, and Combined Biopsy for Prostate Cancer Diagnosis. *N Engl J Med*. 2020;382(10):917-928. doi:10.1056/NEJMoa1910038

3. 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(1):100-109. doi:10.1016/S1470-2045(18)30569-2

4. 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(6):722-728. doi:10.1016/j.eururo.2018.06.022

5. 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(2):204-210. doi:10.1016/j.eururo.2017.12.028

6. 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 confi rmatory study. *www.thelancet.com*. 2017;389. doi:10.1016/S0140-6736(16)32401-1

7. Schröder FH, Hugosson J, Roobol MJ, et al. Screening and Prostate-Cancer Mortality in a Randomized European Study. *N Engl J Med*. 2009;360(13):1320-1328. doi:10.1056/NEJMoa0810084

8. 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 Lond Engl*. 2014;384(9959):2027-2035. doi:10.1016/S0140-6736(14)60525-0

9. Grönberg H, Adolfsson J, Aly M, et al. Prostate cancer screening in men aged 50 to 69 years (STHLM3): a prospective population-based diagnostic study. *www.thelancet.com/oncology*. 2015;16. doi:10.1016/S1470-2045(15)00361-7

10. Kasivisvanathan V, Rannikko AS, Borghi M, et al. MRI-Targeted or Standard Biopsy for Prostate-Cancer Diagnosis. *N Engl J Med*. 2018;378(19):1767-1777. doi:10.1056/NEJMoa1801993

11. 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(3):438-450. doi:10.1016/j.eururo.2014.11.037

12. Thompson IM. Operating Characteristics of Prostate-Specific Antigen in Men With an Initial PSA Level of 3.0 ng/mL or Lower. *JAMA*. 2005;294(1):66. doi:10.1001/jama.294.1.66

13. Alonzo TA, Pepe MS, Moskowitz CS. Sample size calculations for comparative studies of medical tests for detecting presence of disease. *Stat Med*. 2002;21(6):835-852. doi:10.1002/sim.1058

14. Pepe MS. Comparing disease screening tests when true disease status is ascertained only for screen positives. *Biostatistics*. 2001;2(3):249-260. doi:10.1093/biostatistics/2.3.249