EUROPEAN UROLOGY XXX (2017) XXX–XXX

available at www.sciencedirect.com
journal homepage: www.europeanurology.com

European Association of Urology

Prostate Cancer

The Stockholm-3 Model for Prostate Cancer Detection: Algorithm Update, Biomarker Contribution, and Reflex Test Potential

Peter Ström [a], *Tobias Nordström* [a,b], *Henrik Grönberg* [a], *Martin Eklund* [a,*]

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

Article info

Article history:
Accepted December 27, 2017

Associate Editor:
Matthew Cooperberg

Statistical Editor:
Andrew Vickers

Keywords:
Paired screen-positive design
Prostate cancer screening
Prostate-specific antigen
Reflex test
Stockholm-3 model

Abstract

Background: It has been shown that the Stockholm-3 model (S3M) outperforms prostate-specific antigen (PSA) as a screening tool for prostate cancer.
Objective: To update the S3M, to give a detailed account of the value of each predictor in the S3M, and to evaluate the S3M as a reflex test for men with PSA ≥3 ng/ml.
Design, setting, and participants: During 2012–2015, the Stockholm-3 study evaluated the S3M relative to PSA as tests for Gleason score ≥7 prostate cancers among men aged 50–69 yr. The participants (n = 59 159) underwent both tests, and biopsy was recommended if at least one was positive. A total of 5073 men had a biopsy because of elevated PSA (≥3 ng/ml).
Outcome measurements and statistical analysis: Logistic regression was used to update the S3M: intact PSA was removed, *HOXB13* was included, and the model was fitted to data from the Stockholm-3 training and validation cohorts. To compare S3M with PSA, we fixed the sensitivity for detection of high-grade cancer and evaluated the performance as the number of biopsies needed to achieve that sensitivity for each test.
Results and limitations: The updated S3M slightly improved the area under the receiver operating characteristic curve compared to previously published results (0.75 vs 0.74). When used as a reflex test for men with PSA ≥3 ng/ml, S3M reduced the number of biopsies needed by 34% compared to the use of PSA alone, with equal sensitivity. A limitation is the ethnically homogeneous population.
Conclusions: A major problem with PSA screening—too many unnecessary biopsies—can be mitigated if S3M is used as a reflex test.
Patient summary: To find aggressive prostate cancer with the minimum number of negative biopsies and detection of clinically insignificant cancers, we evaluated the use of a personalized diagnostic prediction model as a second test for men with a positive prostate-specific antigen (PSA) test. We found that this two-step approach could reduce prostate biopsies by a third compared to using PSA alone.

* Corresponding author. Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm SE-171 77, Sweden. Tel. +46 73 71211611; Fax: +46 8314975.
E-mail address: martin.eklund@ki.se (M. Eklund).

1. Introduction

Although prostate-specific antigen (PSA) measurement is suitable for prostate cancer screening because of its low cost and noninvasive nature, it has low specificity at acceptable sensitivity levels. This is because nonmalignant conditions such as inflammations and benign prostate hyperplasia may cause increases in PSA levels, and prostate cancer can exist without an increase in PSA. The low specificity leads to frequent prostate biopsies in men with benign conditions,

https://doi.org/10.1016/j.eururo.2017.12.028

and to overdiagnosis of indolent prostate cancer. Use of PSA for disease screening is therefore controversial [1]. One way to improve the specificity of prostate cancer screening is to use a second test—a reflex test—for men with increased PSA levels [2]. The reflex test needs to be more predictive of cancer than PSA alone, and since only men with a higher risk of prostate cancer undergo the test, a higher cost may be justified. There are several possible tests for this purpose, such as percentage free PSA, the 4K score [3], the Prostate Health Index (PHI) [4], PCA3 [5], and RC3 [6].

In 2015, Grönberg et al [7] published results for the Stockholm-3 (STHLM3) study, in which the individualized prediction model S3M was compared to PSA ≥3 ng/ml as a screening test for prostate cancer. The study was designed so that both tests would detect the same number of Gleason score (GS) ≥7 cancers, and the tests were evaluated in terms of the number of biopsies needed to achieve this. With maintained sensitivity for detecting GS ≥7 disease, use of S3M saved 32% of prostate biopsies compared to screening using PSA ≥3 ng/ml as an indication for a prostate biopsy.

Knowledge of the contributions of separate components in suggested tests is important to understand the performance of the underlying biomarkers and was not presented in detail by Grönberg et al [8]. Furthermore, it is not clear how the S3M will perform in a reflex test setting. Here, we describe an update of the S3M test and analyze the predictive contribution of the biomarkers included in the S3M. We also assess the usefulness of the S3M in a reflex setting whereby the test is used only for subjects with elevated PSA.

2. Patients and methods

2.1. Participants and study design

The STHLM3 study was a prospective and population-based diagnostic trial designed to compare S3M with PSA ≥3 ng/ml as indications for prostate biopsy [7]. It consisted of a training (n = 11 130) and a validation cohort (n = 47 688), both of randomly invited (no overlap) men aged 50–69 yr and without a previous prostate cancer diagnosis, from Stockholm County, Sweden (Table 1). Data for the training cohort were used to fit the S3M, which was subsequently evaluated in the validation study. STHLM3 used a paired screen-positive design in which the S3M test was analyzed for all participants with PSA ≥1 ng/ml [9]. Each participant was then recommended prostate biopsy if he had PSA ≥3 ng/ml or a S3M probability of GS ≥7 prostate cancer above a fixed threshold. The S3M threshold was set such that both tests (PSA and S3M) detected the same number of GS ≥7 cancers. The indication for biopsy referral was blinded to the study participants, the urologists, and the pathologist. The biopsy procedure followed a standardized protocol using 10–12-core systematic biopsies, with 12 cores if the prostate volume was >35 cm^3.

Since the sensitivity for detecting GS ≥7 cancer was the same for both tests by design, the evaluation of the usefulness of the tests could be based on the number of biopsies needed for each test (ie, the number of participants with S3M above the adjusted threshold compared to the number of participants with PSA ≥3 ng/ml), and specifically, how many of these biopsies were cancer-free and cancers graded as GS 6.

In this study, we included all biopsied participants from the pilot study and the validation study, and 331 (of whom 34 underwent biopsy) additional participants who had not had a blood test before the database of the STHLM3 study was locked (total n = 59 149). We excluded men with PSA ≥10 ng/ml (n = 630) and men taking α-reductase inhibitor medication at inclusion (n = 1385; Table 1).

Table 1 – Characteristics of the study cohort (pilot and validation cohorts for the Stockholm-3 study)

	Patients, n (%)	
	All participants (n = 59 149)	Biopsied participants [a] (n = 7417)
Age [b]		
<49 yr	1791 (3)	45 (1)
50–54 yr	12923 (22)	640 (9)
55–59 yr	13570 (23)	1222 (16)
60–64 yr	14072 (24)	2031 (27)
65–69 yr	15998 (27)	3299 (44)
≥70	795 (1)	180 (2)
First-degree relative with prostate cancer		
Yes	7262 (12)	1118 (15)
No	51887 (88)	6299 (85)
Previous negative biopsy		
Yes	1976 (3)	505 (7)
No	57173 (97)	6912 (93)
Prostate-specific antigen		
<1 ng/ml	26159 (44)	2 (0)
1–3 ng/ml	25350 (43)	1938 (26)
3–5 ng/ml	4655 (8)	3461 (47)
5–10 ng/ml	2355 (4)	1613 (22)
≥10 ng/ml	630 (1)	403 (5)
Medication (5α-reductase inhibitor)		
Yes	1385 (2)	180 (2)
No	57764 (98)	7237 (98)
Digital rectal examination		
Abnormal	–	681 (9)
Normal	–	6736 (91)
Prostate volume [c]		
<35 ml	–	2705 (36)
35–50 ml	–	2497 (34)
≥50 ml	–	2215 (30)
Biopsy result		
Benign	–	4618 (62)
Gleason 3 + 3	–	1558 (21)
Gleason 3 + 4	–	759 (10)
Gleason 4 + 3	–	253 (3)
Gleason ≥ 4 + 4	–	229 (3)

[a] Participants were recommended a biopsy on a double–blind basis if they were positive for prostate-specific antigen ≥3 ng/ml or the Stockholm-3 model test.
[b] The validation study only included men aged 50–69 yr.
[c] Measured via transrectal ultrasound.

2.2. Predictors in S3M

The predictors in S3M include clinical variables (age, first-degree family history of prostate cancer, and a previous biopsy), blood biomarkers (total PSA, free PSA, ratio of free/total PSA, hK2, MIC1, and MSMB), genetic markers (a genetic score based on 254 single-nucleotide polymorphisms [SNPs] and an explicit variable for the *HOXB13* SNP), and prostate examination (digital rectal examination [DRE], and prostate volume). Details of the genetic score have been described by Grönberg et al [7]. The original version of S3M also included intact PSA, but because of interference between the kallikreins in the immunosorbent allergen chip assay it has been removed from S3M. In addition, a new biomarker is included, the *HOXB13* SNP, a rare germline mutation of the *HOXB13* gene with a large effect on the risk of prostate cancer [10]. It is present in 1.3% of healthy Swedish men and is 3.5-fold more common among prostate cancer patients with otherwise similar characteristics. All continuous predictors are included as linear effects and the others (family history,

Please cite this article in press as: Ström P, et al. The Stockholm-3 Model for Prostate Cancer Detection: Algorithm Update, Biomarker Contribution, and Reflex Test Potential. Eur Urol (2018), https://doi.org/10.1016/j.eururo.2017.12.028

previous biopsy, *HOXB13*, and DRE) as indicator variables in a logistic regression model.

2.3. Evaluation of individual predictions

The S3M predictions were evaluated in terms of the area under the receiver operating characteristic curve (AUC) and impact on model calibration. The predictors were evaluated individually in a bivariate model together with PSA, as the last predictor to enter the full S3M, and cumulatively as added in a prespecified order.

Because of the novelty of MIC1, MSMB, and the genetic score for use in a diagnostic prediction model, and the low prevalence of men with a previous prostate biopsy or carrying the *HOXB13* SNP, we assessed the impact of inclusion of these variables on model calibration (predicted vs observed risk). This was done using both the S3M and a model using all S3M predictors except the one investigated.

2.4. Reflex test

Men in STHLM3 with PSA ≥3 ng/ml were biopsied regardless of S3M results (n = 5073). For these men, we can evaluate the S3M test as a reflex test. By accepting a fraction of missed GS ≥7 cancers compared to PSA ≥3 ng/ml (here 10% and 20%), we compared the S3M to the corresponding fractions of missed GS ≥7 cancers by further increasing the PSA threshold for biopsy indication. Since the sensitivity is then the same for the two tests (S3M and PSA), the evaluation used the reduction in the number of biopsies compared to PSA ≥3 ng/ml for biopsy referral as the performance measure. The biopsies that could be avoided were also assessed by stratification as benign biopsies and GS 6 cancers. We also evaluated three additional models: a model with age and four kallikreins (total PSA, free PSA, intact PSA, and hK2) as predictors; a model with only clinical variables (total PSA, age, DRE, and prostate volume); and PSA density (total PSA/prostate volume). The two first models involved logistic regression with the predictors included as in S3M, while PSA density was used as is (ie, positivity defined as being above a certain threshold).

Furthermore, cancer length (in mm) and the percentage of positive cores for GS ≥7 cancers missed by the S3M and PSA tests were compared to all GS ≥7 cancers in the cohort.

2.5. Statistical analyses

All risk predictions for the participants (S3M and models comprising a subset of predictors) were based on tenfold cross-validation, and all confidence intervals (CIs) presented are 95% two-sided bootstrap intervals based on 1000 bootstrap samples [11]. *HOXB13* inclusion in the model was decided a priori, and the cross-validated model was prespecified (ie, no model selection). The software used was R version 3.2.5 [12].

3. Results

3.1. Predictors in S3M

The AUC for PSA alone was 0.58 (95% CI 0.57–0.60) and the AUC for the updated S3M was 0.75 (95% CI 0.73–0.77; Table 2). When used together with PSA in a bivariate model, all biomarkers showed a higher AUC than PSA alone, indicating independent value to discriminate for GS ≥7 prostate cancer. The AUC rapidly increased when clinical variables (DRE, previous biopsies, and prostate volume) were included (Cumulative), but each additional predictor increased the AUC by one unit at most. However, together these additional predictors increased the AUC from 0.71 to 0.76. The AUC was almost unaffected by taking out a single predictor from the model (Remove). Prostate volume resulted in the greatest increase in AUC when the predictors were individually considered together with PSA, and the largest loss in AUC when removed from the full set of predictors in S3M.

Calibration of the predictions improved among participants with the 10% highest or 10% lowest values for the genetic score, MIC1, or MSMB when compared to a model with all S3M predictors except each of these three. The calibration also improved for *HOXB13* carriers and men with

Table 2 – Performance in predicting Gleason score ≥7 prostate cancer for S3M, PSA, and submodels comprising S3M components

Test	AUC (95% CI)		
S3M	0.75 (0.73–0.77)		
PSA	0.58 (0.57–0.60)		
Individual predictors	**Bivariate [a]**	**Cumulative [b]**	**Remove [c]**
Age	0.59 (0.57–0.61)	0.59 (0.57–0.61)	0.75 (0.74–0.77)
Digital rectal examination	0.63 (0.61–0.64)	0.63 (0.61–0.65)	0.75 (0.73–0.76)
Previous biopsies	0.61 (0.59–0.63)	0.65 (0.63–0.66)	0.75 (0.74–0.77)
Prostate volume	0.67 (0.66–0.69)	0.71 (0.69–0.73)	0.74 (0.73–0.76)
Family history	0.59 (0.57–0.61)	0.71 (0.70–0.73)	0.76 (0.74–0.77)
Free PSA	0.65 (0.63–0.67)	0.72 (0.71–0.74)	0.76 (0.74–0.78)
Free/total PSA ratio	0.65 (0.63–0.67)	0.73 (0.71–0.74)	0.76 (0.74–0.77)
Intact PSA [d]	0.58 (0.56–0.60)	0.74 (0.72–0.75)	0.75 (0.73–0.77)
hK2	0.59 (0.57–0.61)	0.75 (0.74–0.77)	0.75 (0.73–0.76)
MIC1	0.59 (0.57–0.61)	0.75 (0.74–0.77)	0.76 (0.74–0.77)
MSMB	0.60 (0.58–0.62)	0.76 (0.74–0.77)	0.76 (0.74–0.77)
HOXB13	0.59 (0.56–0.60)	0.76 (0.74–0.77)	0.76 (0.74–0.77)
Genetic score	0.61 (0.59–0.63)	0.76 (0.74–0.77)	0.76 (0.74–0.77)

S3M = Stockhom-3 model; PSA = prostate-specific antigen; AUC = area under the receiver operating characteristic curve; CI = confidence interval.
[a] Individual S3M biomarkers in combination with PSA (including intact PSA).
[b] The cumulative performance for inclusion of each biomarker in the order presented (including intact PSA).
[c] The remaining value after removing the biomarker from the full set of predictors (including intact PSA).
[d] Intact PSA is no longer part of S3M but was evaluated among the set of individual predictors.

Please cite this article in press as: Ström P, et al. The Stockholm-3 Model for Prostate Cancer Detection: Algorithm Update, Biomarker Contribution, and Reflex Test Potential. Eur Urol (2018), https://doi.org/10.1016/j.eururo.2017.12.028

Table 3 – Calibration of S3M in subgroups according to values for selected biomarkers, comparing S3M including all but the specified biomarker, the full S3M, and the observed risk in the corresponding group

	Subjects	Predicted risk (%)		Observed
	(*n*)	S3M – SB	Full S3M	risk (%)
Genetic score [a]				
10% highest	656	16%	19%	19
80% middle	5247	15%	15%	15
10% lowest	656	12%	10%	9
MIC1 [a]				
10% highest	657	17%	19%	21
80% middle	5251	15%	14%	14
10% lowest	651	14%	13%	14
MSMB [a]				
10% highest	655	16%	19%	20
80% middle	5237	15%	15%	15
10% lowest	667	14%	12%	11
HOXB13 [b]				
Yes	87	28%	34%	34
No	4986	15%	15%	15
Previous biopsy				
No	6169	15%	15%	15
Yes	390	16%	6%	6

S3M = Stockholm-3 model; S3M – SB = S3M without the specified biomarker.
[a] These variables occur as continuous variables in S3M but are grouped here in quantiles (10%, 80%, 10%) to asses calibration among subjects with unusually high or low values for these biomarkers.
[b] *HOXB13* calibration was only performed for participants with PSA ≥3 ng/ml, since everyone with this SNP mutation (and PSA ≥1 ng/ml) was recommended biopsy.

a previous negative biopsy when these predictors were included in the model (Table 3).

3.2. *Reflex tests: results for men with PSA ≥3 ng/ml*

All 5073 men with PSA ≥3 ng/ml were biopsied. Of these biopsies, 791 (16%) were GS ≥7 cancer, 1078 (21%) were GS 6 cancer, and 3204 (63%) were benign. Table 4 evaluates alternative strategies to reduce the number of biopsies by reflex tests among men with PSA ≥3 ng/ml. The cost for reducing the number of biopsies is a lower detection rate of GS ≥7 cancers. Here we chose to accept 10% (n = 79) or 20% (n = 158) missed GS ≥7 cancers. Missing 10% and 20% of the GS ≥7 cancers corresponds to increasing the PSA cutoff for biopsy referral to 3.2 ng/ml and 3.4 ng/ml, respectively. This would reduce the total number of biopsies by 14% (n = 710) and 27% (n = 1370). The corresponding reductions would be 15% (n = 480) and 29% (n = 929) for benign biopsies, and 15% (n = 162) and 28% (n = 302) for GS 6 cancer biopsies. By instead using percentage free PSA as a reflex test, the reduction in total biopsies would increase to 18% and 32% for 10% and 20% missed GS ≥7 cancers, respectively. When S3M is used, 33% (n = 1674) and 52% (n = 2638) fewer biopsies would be performed. Among benign biopsies 42% (n = 1346) and 62% (n = 1986) would be avoided, while GS 6 cancer biopsies would be reduced by 26% (n = 280) and 43% (n = 464). The three additional models evaluated (PSA density, clinical model, and age + four kallikreins) all showed similar performance, with AUCs of 0.69–0.71 and reduced biopsies of 23–26% and 40–43% for 10% and 20% missed GS ≥7 cancers, respectively.

As a reflex test, S3M would reduce the number of necessary biopsies by an estimated 34% compared to simply raising the PSA threshold. This number was calculated for a reflex test missing 20% of GS ≥7 cancers, whereby 2435 biopsies would be performed after an S3M reflex test and 3703 biopsies by raising the PSA threshold.

In addition to the evaluation of the individual predictors in Table 4, we also considered the remaining performance after removing each predictor separately from the model, with only volume resulting in a meaningful decrease in biopsies, as shown in Supplementary Table 1. This table also presents an alternative cumulative order that more closely represents the clinical workflow, where DRE and volume are added last.

Figure 1 shows the implication of using S3M as a reflex test in contrast to raising the PSA threshold for biopsy referral among men with PSA in the range 3–10 ng/ml. The figure complements the first and second rows of Table 4 by showing the total number of biopsies needed by the fraction of missed GS ≥7 cancers we are willing to accept.

Table 5 compares the characteristics of the GS ≥7 cancers that would have been missed in the reflex test to all the GS ≥7 cancers among men with PSA ≥3 ng/ml. A cancer length of <10mm in the biopsy specimen was more common ($p < 0.01$) among the missed cancers from a S3M reflex test (69%) compared to all GS ≥7 cancers (46%). There was no significant difference ($p = 0.28$) between those missed by S3M (69%) and by increasing the PSA threshold (63%). The fraction of the 10–12 cores that were positive for cancer among the cancers missed by S3M is lower compared to all GS ≥7 cancers ($p < 0.01$) and cancers missed by PSA ($p < 0.01$). For GS ≥7 cancers missed by S3M, 77% had <34% positive cores. The corresponding numbers for all cancers and cancers missed by PSA are 56% and 64%, respectively. The Gleason grade distributions are similar among all cancers and the missed cancers.

4. Discussion

After removing intact PSA and including the *HOXB13* SNP in the set of predictors, and fitting the S3M to more data, we estimated an AUC of 0.75 (95% CI 0.73–0.77) for predicting GS ≥7 prostate cancers, which is a slight improvement compared to the result reported by Grönberg et al (AUC 0.74, 95% CI 0.72–0.75). From Tables 2 and 4 it is evident that the collection of predictors together is beneficial for reducing unnecessary biopsies. However, the loss in performance when biomarkers are left out from the full S3M model is small, and for some of the weakest predictors there are no (or only negligible) benefits for the population. However, they may still be important for a minority of men with a rare exposure (eg, the *HOXB13* risk SNP). The most substantial contributions are from prostate volume and free PSA.

Although intact PSA is no longer part of S3M, we included it in the list of individual predictors for

Please cite this article in press as: Ström P, et al. The Stockholm-3 Model for Prostate Cancer Detection: Algorithm Update, Biomarker Contribution, and Reflex Test Potential. Eur Urol (2018), https://doi.org/10.1016/j.eururo.2017.12.028

Table 4 – Reduction in biopsies for a reflex test missing 10% and 20% GS ≥7 cancers among subjects positive for the first test (PSA ≥3 ng/ml) in total and for benign samples and GS 6 cancers

Test	AUC (95% CI)	Missing 10% GS ≥7 cancers				Missing 20% GS ≥7 cancers			
		Biopsy referral	Reduction in biopsies, % (95% CI)			Biopsy referral	Reduction in biopsies, % (95% CI)		
			Benign	GS 6	Total		Benign	GS 6	Total
S3M	0.76 (0.74–0.77)	0.08	42 (36–48)	26 (22–31)	33 (29–38)	0.11	62 (58–67)	43 (38–47)	52 (48–55)
PSA	0.58 (0.56–0.60)	3.2 ng/ml	15 (13–18)	15 (13–18)	14 (12–16)	3.4 ng/ml	29 (25–33)	28 (23–32)	27 (24–30)
Free/total PSA ratio	0.64 (0.62–0.67)	28%	20 (17–24)	17 (13–21)	18 (15–21)	31%	36 (32–40)	31 (27–36)	32 (30–36)
PSA density [a]	0.69 (0.67–0.70)	0.07 ng/(ml × cm^3)	27 (24–33)	20 (16–26)	23 (20–28)	0.09 ng/(ml × cm^3)	45 (41–49)	37 (33–41)	40 (36–42)
Clinical model [b]	0.71 (0.69–0.73)	0.09	31 (26–37)	24 (20–30)	26 (22–31)	0.12	50 (45–55)	40 (35–46)	43 (39–47)
Age + four kallikreins [c]	0.70 (0.68–0.72)	0.08	29 (23–33)	20 (16–24)	24 (20–27)	0.11	47 (42–51)	37 (31–41)	40 (36–44)
Bivariate [d]									
Age	0.59 (0.56–0.61)	–	15 (13–18)	14 (12–17)	14 (12–17)	–	28 (24–33)	26 (21–31)	26 (23–30)
DRE	0.63 (0.60–0.65)	–	18 (15–21)	18 (14–21)	17 (14–19)	–	33 (28–37)	31 (25–35)	30 (26–34)
Previous biopsies	0.61 (0.59–0.63)	–	20 (17–24)	16 (13–20)	18 (15–21)	–	33 (30–37)	27 (24–32)	29 (27–33)
Prostate volume	0.69 (0.67–0.70)	–	28 (25–33)	20 (16–24)	24 (21–27)	–	45 (41–49)	36 (31–40)	39 (36–42)
Family history	0.59 (0.57–0.61)	–	16 (12–19)	15 (11–18)	15 (12–17)	–	29 (25–34)	25 (21–30)	27 (24–31)
Free PSA	0.65 (0.63–0.67)	–	20 (17–26)	17 (13–23)	18 (15–23)	–	36 (32–42)	32 (27–37)	33 (29–38)
Free/total PSA ratio	0.65 (0.63–0.67)	–	20 (18–25)	18 (15–23)	18 (16–22)	–	36 (33–43)	33 (29–39)	33 (30–38)
Intact PSA [e]	0.58 (0.56–0.61)	–	14 (11–17)	15 (12–18)	14 (11–16)	–	28 (24–32)	27 (23–31)	27 (23–30)
hK2	0.59 (0.56–0.61)	–	16 (13–19)	13 (10–16)	14 (12–17)	–	30 (26–33)	26 (22–30)	28 (24–30)
MIC1	0.59 (0.57–0.61)	–	17 (12–19)	15 (12–18)	15 (12–17)	–	30 (26–35)	29 (25–35)	28 (25–32)
MSMB	0.60 (0.58–0.62)	–	17 (14–19)	13 (10–16)	15 (12–17)	–	31 (27–35)	27 (23–32)	28 (25–32)
HOXB13	0.59 (0.57–0.61)	–	16 (12–20)	16 (12–19)	15 (12–18)	–	28 (24–33)	26 (22–31)	26 (23–30)
Genetic score	0.61 (0.59–0.63)	–	22 (18–25)	14 (10–17)	19 (15–20)	–	36 (31–39)	25 (21–30)	31 (27–34)
Cumulative [f]									
Age	0.59 (0.56–0.61)	–	15 (13–18)	14 (12–17)	14 (12–17)	–	28 (24–33)	26 (21–31)	26 (23–30)
DRE	0.62 (0.60–0.65)	–	16 (13–20)	14 (11–18)	15 (13–18)	–	31 (27–35)	29 (25–34)	29 (26–32)
Previous biopsies	0.65 (0.62–0.67)	–	21 (18–25)	16 (12–19)	18 (16–21)	–	36 (31–41)	29 (24–35)	32 (28–36)
Prostate volume	0.72 (0.70–0.74)	–	36 (31–41)	25 (20–30)	29 (25–34)	–	52 (46–57)	39 (34–45)	44 (40–48)
Family history	0.72 (0.70–0.74)	–	36 (31–41)	25 (20–29)	30 (25–34)	–	53 (47–58)	39 (34–45)	45 (40–49)
Free PSA	0.73 (0.71–0.75)	–	37 (31–42)	25 (20–29)	30 (25–34)	–	55 (51–59)	42 (37–47)	47 (43–50)
Free/total PSA ratio	0.73 (0.71–0.75)	–	38 (31–42)	25 (20–29)	31 (26–34)	–	55 (51–59)	43 (38–47)	47 (43–50)
Intact PSA [e]	0.75 (0.73–0.76)	–	39 (33–45)	25 (20–32)	32 (27–37)	–	58 (53–62)	44 (39–48)	49 (45–52)
hK2	0.76 (0.74–0.78)	–	43 (38–48)	27 (22–31)	34 (30–38)	–	61 (56–66)	43 (38–48)	51 (47–55)
MIC1	0.76 (0.74–0.78)	–	42 (38–47)	26 (22–32)	33 (30–38)	–	62 (57–66)	43 (39–49)	51 (48–55)
MSMB	0.77 (0.75–0.78)	–	42 (36–48)	26 (21–31)	34 (29–39)	–	63 (59–67)	44 (40–48)	52 (49–56)
HOXB13	0.77 (0.75–0.79)	–	42 (37–48)	26 (21–31)	34 (30–39)	–	63 (59–67)	44 (40–49)	52 (49–56)
Genetic score	0.77 (0.75–0.79)	–	44 (40–49)	26 (22–30)	35 (31–39)	–	64 (59–67)	44 (39–48)	53 (49–56)

GS = Gleason score; PSA = prostate-specific antigen; S3M = Stockholm-3 model; DRE = digital rectal examination.
[a] PSA/prostate volume; the referral value is the actual value for the density, not a probability of GS ≥7 cancer.
[b] PSA, age, DRE, and prostate volume.
[c] Age, PSA, free PSA, free/total PSA ratio, hK2, and intact PSA.
[d] Individual S3M biomarkers in combination with PSA (including intact PSA).
[e] Intact PSA is no longer part of S3M but was evaluated among the set of individual predictors.
[f] The cumulative performance by including each biomarker in the order presented (including Intact PSA).

comparison. This is the reason for the potentially better performance compared to the original S3M (ie, AUC >0.75) when a predictor is removed from the rest. Table 3 shows that the weaker predictors (MSMB, MIC1, and genetic score) also contribute to improving the risk prediction for subjects with high or low values for these markers. Similarly, *HOXB13* also improves the model calibration, although the prevalence of *HOXB13* mutations in the population is too low to meaningfully affect the AUC.

The cost for advanced biomarker-based prostate cancer tests ranges from $500 upwards [2]. S3M is still not a commercial test and it is difficult to judge the price of the test. Estimates indicate that it will fall in the lower range despite the inclusion of genetic factors rather than just protein biomarkers. Combined biomarker tests such as S3M and others are superior to PSA alone, but their higher cost motivates initial triaging based on PSA alone. Here we analyzed S3M as a reflex test for men with PSA ≥3 ng/ml, meaning that to save biopsies compared to PSA ≥3 ng/ml there must be a sacrifice in terms of some missed cancers, as already considered in similar studies [3,4,13].

If accepting a decrease of 10% in detection of GS ≥7 cancers, S3M would reduce biopsies by 33%, while the three additional models evaluated (PSA density, clinical model, and age + four kallikreins) would achieve biopsy reductions in the range 23–26%, and simply increasing the PSA threshold would reduce the number of biopsies by 14%. This highlights the value of utilizing all S3M predictors, including those with only a modest incremental increase in model accuracy. This is further illustrated by

Please cite this article in press as: Ström P, et al. The Stockholm-3 Model for Prostate Cancer Detection: Algorithm Update, Biomarker Contribution, and Reflex Test Potential. Eur Urol (2018), https://doi.org/10.1016/j.eururo.2017.12.028

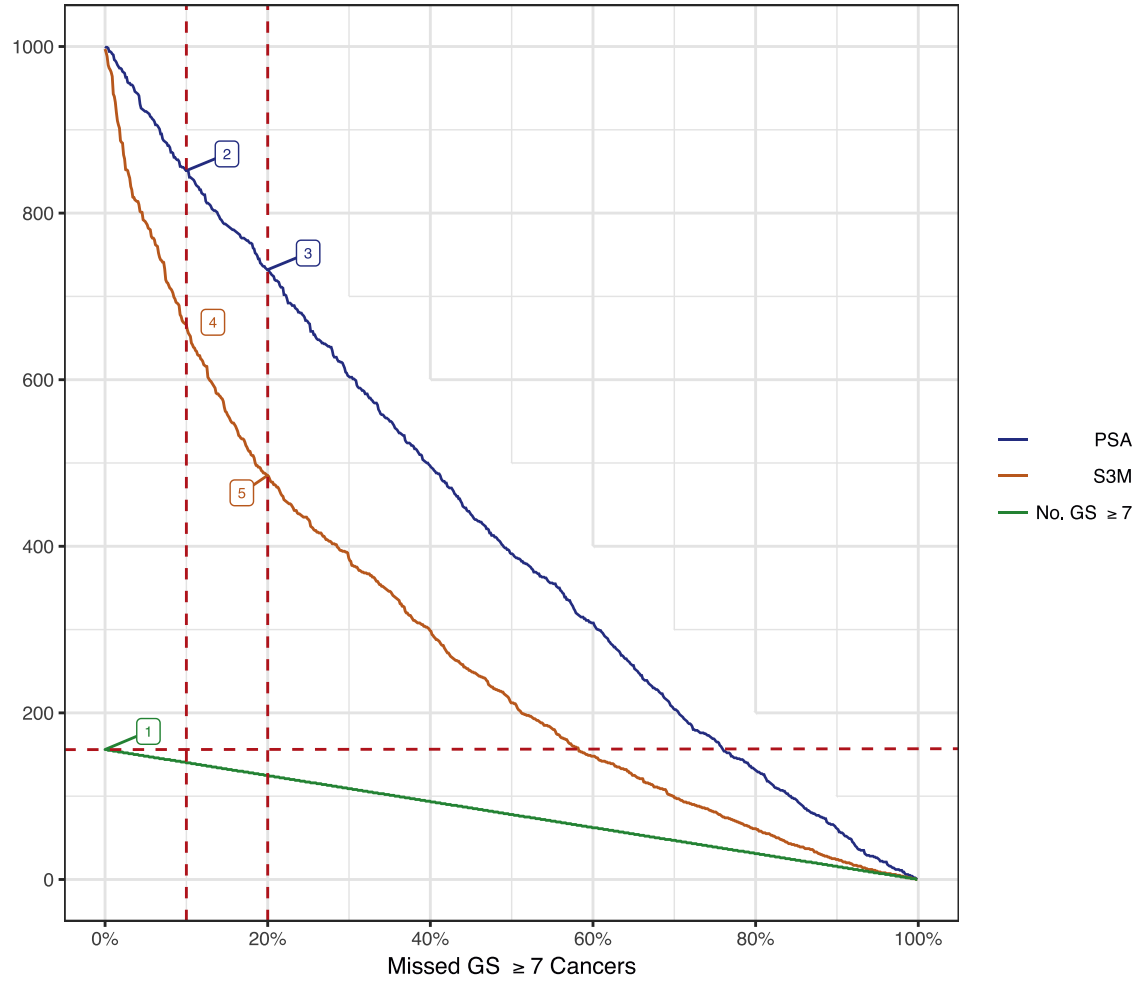

Fig. 1 – Clinical implications of using the Stockholm-3 model (S3M) as a reflex test for 1000 men aged 50–69 yr with prostate-specific antigen (PSA) in the range 3–10 ng/ml. GS = Gleason score.

the higher AUC for S3M (0.76) compared to the three additional models evaluated (PSA density, clinical model, and age + four kallikreins with AUC of 0.69, 0.71, and 0.70, respectively) for men with PSA ≥3 ng/ml. If we accept 20% missed GS ≥7 cancers compared to PSA ≥3 ng/ml for biopsy referral, S3M would save more than half of the biopsies. The reduction in benign biopsies is particularly large (62% would be avoided). Biopsies in men with cancer that would have been missed by a S3M reflex test contained less cancer than the overall positive biopsies in the study, in terms of both total cancer length and number of positive cores. This was also the case for cancers missed by S3M compared to cancers missed by simply raising the PSA threshold.

This missed cancers compared to PSA ≥3 ng/ml can be compensated for by lowering the PSA threshold for a reflex test, for example to 1.5 ng/ml as suggested by Crawford et al [2] or 1 ng/ml as used by Grönberg et al [7]. The full range of outcomes when using different PSA thresholds for S3M testing and different S3M thresholds for biopsy recommendation was described by Nordström et al [14].

Table 5 – Tumor characteristics of the GS ≥7 cancers included in the study cohort (*n* = 6559) and of the 20% GS ≥7 cancers missed by a reflex test for these men based on S3M and PSA

	Patients, *n* (%)		
	All GS ≥7 cancers (*n* = 969)	Missed GS ≥7 cancers (*n* = 194)	
		PSA	S3M
Cancer length			
<10 mm	448 (46)	123 (63)	134 (69)
≥10 mm	521 (54)	71 (37)	60 (31)
Positive cores			
<34%	540 (56)	125 (64)	150 (77)
≥34%	429 (44)	69 (36)	44 (23)
Grade			
GS 3 + 4	645 (67)	140 (72)	136 (70)
GS 4 + 3	188 (19)	36 (19)	39 (20)
GS ≥ 4 + 4	136 (14)	18 (9)	19 (10)

GS = Gleason score; S3M = Stockholm-3 model; PSA = prostate-specific antigen.

Please cite this article in press as: Ström P, et al. The Stockholm-3 Model for Prostate Cancer Detection: Algorithm Update, Biomarker Contribution, and Reflex Test Potential. Eur Urol (2018), https://doi.org/10.1016/j.eururo.2017.12.028

The strength of this study is the sample size of nearly 60 000 participants, with more than 7000 of these undergoing biopsy. Each of these biopsies was examined by the same pathologist (Professor Lars Egevad), which decreases the risk of systematic differences in the evaluation of prostate tissues. Furthermore, the STHLM3 study was blinded, whereby the urologist, patient, or pathologist did not know the result of the PSA test or the prediction from the S3M test, only whether the patient was recommended biopsy or not. The main weakness of STHLM3 is that the population is ethnically homogeneous, with most participants of northern European descent.

5. Conclusions

We have shown that we can reduce biopsies by using a refined test in the diagnostic pathway before deciding on prostate biopsy. The STHLM3 study prospectively demonstrated a 32% reduction in biopsies using S3M compared to PSA $\geq$3 ng/ml for biopsy referral, without any loss in sensitivity by also detecting GS $\geq$7 cancers for the PSA range 1–3 ng/ml. Here we demonstrate an equally large reduction (34%) when S3M is used as a reflex test for men with PSA $\geq$3 ng/ml. This study adds to the evidence that improved risk stratification using biomarker models can improve prostate cancer diagnosis.

Author contributions: Martin Eklund had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Study concept and design: Ström; Nordström; Eklund; Grönberg.
Acquisition of data: Nordström; Eklund; Grönberg.
Analysis and interpretation of data: Ström; Eklund.
Drafting of the manuscript: Ström.
Critical revision of the manuscript for important intellectual content: Eklund; Nordström; Grönberg.
Statistical analysis: Ström.
Obtaining funding: Grönberg; Eklund.
Administrative, technical, or material support: None.
Supervision: Eklund; Grönberg; Nordström.
Other: None.

Financial disclosures: Martin Eklund certifies that all conflicts of interest, including specific financial interests and relationships and affiliations relevant to the subject matter or materials discussed in the manuscript (eg, employment/affiliation, grants or funding, consultancies, honoraria, stock ownership or options, expert testimony, royalties, or patents filed, received, or pending), are the following: Henrik Grönberg has five prostate cancer diagnostic–related patents pending, has patent applications licensed to Thermo Fisher Scientific, and might receive royalties from sales related to these patents. Martin Eklund is named on four of these five patent applications. The Karolinska Institutet collaborates with Thermo Fisher Scientific in developing the technology for STHLM3.

Funding/Support and role of the sponsor: The main funder of the STHLM3 study was Stockholm County Council (Stockholms Läns Landsting), the main provider of health care in Stockholm. Funding was provided by the Swedish Cancer Society (Cancerfonden), the Swedish Research Council (Vetenskapsrådet), and the Swedish Research Council for Health, Working Life and Welfare (FORTE). The STHLM3 study is part of the Linnaeus Center CRISP "Prediction and Prevention of Breast and Prostate Cancer" funded by the Swedish Research Council. The sponsors played no direct role in the study.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.eururo.2017.12.028.

References

[1] Fleshner K, Carlsson SV, Roobol MJ. The effect of the USPSTF PSA screening recommendation on prostate cancer incidence patterns in the USA. Nat Rev Urol 2017;14:26–37.
[2] Crawford ED, Rosenberg MT, Partin AW, et al. An approach using PSA levels of 1.5 ng/ml as the cutoff for prostate cancer screening in primary care. Urology 2016;96:116–20.
[3] Parekh DJ, Punnen S, Sjoberg DD, et al. A multi-institutional prospective trial in the USA confirms that the 4Kscore accurately identifies men with high-grade prostate cancer. Eur Urol 2015;68:464–70.
[4] de la Calle C, Patil D, Wei JT, et al. Multicenter evaluation of the Prostate Health Index to detect aggressive prostate cancer in biopsy naive men. J Urol 2015;194:65–72.
[5] Deras IL, Aubin SM, Blase A, et al. PCA3: a molecular urine assay for predicting prostate biopsy outcome. J Urol 2008;179:1587–92.
[6] Roobol MJ, Verbeek JF, van der Kwast T, Kummerlin IP, Kweldam CF, van Leenders GJ. Improving the Rotterdam European Randomized Study of Screening for Prostate Cancer risk calculator for initial prostate biopsy by incorporating the 2014 International Society of Urological Pathology Gleason grading and cribriform growth. Eur Urol 2017;72:45–51.
[7] Gronberg 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.
[8] Carlsson SV, Kattan MW. Prostate cancer: personalized risk—stratified screening or abandoning it altogether? Nat Rev Clin Oncol 2016;13:140–2.
[9] Pepe MS, Feng Z, Janes H, Bossuyt PM, Potter JD. Pivotal evaluation of the accuracy of a biomarker used for classification or prediction: standards for study design. J Natl Cancer Inst 2008;100:1432–8.
[10] Karlsson R, Aly M, Clements M, et al. A population-based assessment of germline HOXB13 G84E mutation and prostate cancer risk. Eur Urol 2014;65:169–76.
[11] Efron B. The 1977 Rietz Lecture. Bootstrap methods—another look at the jackknife. Ann Stat 1979;7:1–26.
[12] R Foundation for Statistical Computing. R: a language and environment for statistical computing. www.r-project.org/foundation/.
[13] Bryant RJ, Sjoberg DD, Vickers AJ, et al. Predicting high-grade cancer at ten-core prostate biopsy using four kallikrein markers measured in blood in the ProtecT study. J Natl Cancer Inst 2015;107:djv095.
[14] Nordström T, Grönberg H, Adolfsson J, et al. Balancing overdiagnosis and early detection of prostate cancer using the Stockholm-3 model. Eur Urol Focus. In press. https://doi.org/10.1016/j.euf.2016.11.016.