Published in The Lancet Oncology, http://dx.doi.org/10.1016/S1470-2045(15)00361-7

Prostate cancer screening in men aged 50–69 years (STHLM3): a prospective population-based diagnostic study

Henrik Grönberg, Jan Adolfsson, Markus Aly, Tobias Nordström, Peter Wiklund, Yvonne Brandberg, James Thompson, Fredrik Wiklund, Johan Lindberg, Mark Clements, Lars Egevad, Martin Eklund

Summary

Background

The prostate-specific antigen (PSA) test is used to screen for prostate cancer but has a high false-positive rate that translates into unnecessary prostate biopsies and overdiagnosis of low-risk prostate cancers. We aimed to develop and validate a model to identify high-risk prostate cancer (with a Gleason score of at least 7) with better test characteristics than that provided by PSA screening alone.

Methods

The Stockholm 3 (STHLM3) study is a prospective, population-based, paired, screen-positive, diagnostic study of men without prostate cancer aged 50–69 years randomly invited by date of birth from the Swedish Population Register kept by the Swedish Tax Agency. Men with prostate cancer at enrolment were excluded from the study. The predefined STHLM3 model (a combination of plasma protein biomarkers [PSA, free PSA, intact PSA, hK2, MSMB, MIC1], genetic polymorphisms [232 SNPs], and clinical variables [age, family, history, previous prostate biopsy, prostate exam]), and PSA concentration were both tested in all participants enrolled. The primary aim was to increase the specificity compared with PSA without decreasing the sensitivity to diagnose high-risk prostate cancer. The primary outcomes were number of detected high-risk cancers (sensitivity) and the number of performed prostate biopsies (specificity). The STHLM3 training cohort was used to train the STHLM3 model, which was prospectively tested in the STHLM3 validation cohort. Logistic regression was used to test for associations between biomarkers and clinical variables and prostate cancer with a Gleason score of at least 7. This study is registered with ISCRTN.com, number ISRCTN84445406.

Findings

The STHLM3 model performed significantly better than PSA alone for detection of cancers with a Gleason score of at least 7 ($p<0·0001$), the area under the curve was 0·56 (95% CI 0·55–0·60) with PSA alone and 0·74 (95% CI 0·72–0·75) with the STHLM3 model. All variables used in the STHLM3 model were significantly associated with prostate cancers with a Gleason score of at least 7 ($p<0·05$) in a multiple logistic regression model. At the same level of sensitivity as the PSA test using a cutoff of ≥3 ng/mL to diagnose high-risk prostate cancer, use of the STHLM3 model could reduce the number of biopsies by 32% (95% CI 24–39) and could avoid 44% (35–54) of benign biopsies.

Interpretation

The STHLM3 model could reduce unnecessary biopsies without compromising the ability to diagnose prostate cancer with a Gleason score of at least 7, and could be a step towards personalised risk-based prostate cancer diagnostic programmes.

Funding

Stockholm County Council (Stockholms Läns Landsting).

Research in context

Evidence before this study

We searched PubMed between Jan 1, 2000, and 31 Dec, 2009, using search terms related to prostate cancer, biomarkers, and screening for articles written in English.

Prostate-specific antigen (PSA) is widely used as an initial screening test for prostate cancer and is largely credited with the reduction in prostate cancer mortality reported during the past two decades. The European Randomised Study of Screening for Prostate Cancer (ERSPC) showed a 21% reduction in prostate cancer mortality with structured PSA screening after 13 years. However, the poor specificity of PSA translates into many unnecessary prostate biopsies and overdiagnosis of low-risk prostate cancers. Subsequently, no governmental body has recommended structured PSA screening because of the potential harms of overdiagnosis.

PSA is the only biomarker that has been assessed prospectively and in randomised controlled trials. Alternative plasma protein biomarkers other than PSA have been proposed to address these concerns. In addition to total PSA, the other plasma protein biomarkers used in the STHLM3 study were selected by a systematic scientific literature search during 2010 and two subsequent validation studies using the STHLM2 cohort (appendix). Five additional biomarkers were selected based on their associations with the presence of prostate cancer with a Gleason score of at least 7. Urine-based markers were excluded because they would require a prostate massage, which makes them unsuitable to use in a population screening setting. On the basis of the literature and genetic assessment of previous studies, 254 SNPs were selected based on their association with prostate cancer risk.

Added value of this study

We showed that the STHLM3 model, a combination of plasma protein biomarkers, genetic polymorphisms, and clinical variables, can significantly improve prostate cancer screening specificity with the same sensitivity compared with the PSA testing using a cutoff of at least 3 ng/mL. The STHLM3 model can identify cancers with a Gleason score of at least 7 in men aged 50–69 years and identify clinically significant prostate cancers in the PSA concentration range of 1–3 ng/mL. The STHLM3 model also includes two novel biomarkers and genetics markers that have not been previously included in a prospective diagnostic study. Additionally, STHLM3 is the first large-scale diagnostic study in prostate cancer where biopsy decision is prospectively based on the results from the predefined STHLM3 model. Finally, STHLM3 is population based, thus minimising selection bias and increasing the generalisability.

Implications of all the available evidence

Together with the results from the European Randomized Study of Prostate Cancer, the findings from the STHLM3 study indicate that prostate cancer mortality can be reduced but with substantially fewer biopsies and reduced overdiagnosis.

Introduction

Levels of prostate-specific antigen (PSA) are widely used as an initial screening test for prostate cancer and is largely credited with the reduction in prostate cancer mortality reported during the past two decades.[1,2] Even with evidence that PSA-based screening has reduced prostate cancer mortality, no governmental body has recommended structured PSA testing because of the potential harms of overdiagnosis.[3,4] Nonetheless, opportunistic screening is frequent, resulting in 1 000 000 prostate biopsies annually in the USA.[5] There is growing concern about the increasing incidence of serious infections caused by multidrug-resistant bacteria and related infectious complications (eg, 2% of patients develop septicaemia in Stockholm) after prostate biopsy.[6,7]

Alternative plasma protein biomarkers have been proposed to address these concerns; however, none has been prospectively assessed in screening studies.[8,9] At least 100 single nucleotide polymorphisms (SNPs) have been identified, accounting for about 30% of the inherited risk for prostate cancer.[10] Combination of PSA with a genetic score based on these SNPs has been suggested to increase the specificity of prostate cancer testing.[11,12]

We report a risk-based model for prostate cancer screening that combines PSA, SNPs, clinical variables, and established and novel plasma protein biomarkers (the STHLM3 model). We aimed to assess whether the STHLM3 model could increase the specificity of detecting men with high risk (Gleason score of at least 7) prostate cancer and thus substantially reduce the proportion of men undergoing prostate biopsy, while maintaining the same sensitivity to detect high risk prostate cancers as the PSA test alone (which uses a threshold of at least 3 ng/mL).

Methods

Study design and participants

STHLM3 was a prospective, population-based, diagnostic study following a paired, screen-positive design, in which we compared the STHLM3 model with PSA in men aged 50–69 years from Stockholm, Sweden. Men, irrespective of any comorbidity except prostate cancer, were randomly selected by date of birth from the Swedish Population Register kept by the Swedish Tax Agency and invitations were posted to them. The two screening methods, PSA and the STHLM3 model, were both tested in each study participant. PSA is the only biomarker prospectively assessed in population-based screening trials with a positive effect on prostate cancer mortality[2] and the clinical usefulness of other biomarkers in this context is limited. Subsequently, we chose to use a PSA concentration of at least 3 ng/mL as the reference to infer the same mortality effect as seen in these trials. The STHLM3 model is a test consisting of a combination of plasma protein biomarkers (PSA, free PSA, intact PSA, hK2, MSMB, and MIC1), genetic markers, clinical variables (age, family history, previous prostate biopsy), and a prostate exam (digital rectal exam and prostate volume). Plasma protein biomarkers used in STHLM3 were selected from a scientific literature search and two subsequent validation studies (appendix). For the genetic markers, we tested the 254 SNPs shown to be associated with prostate cancer in our previous studies.[11,13] These SNPs were combined in a genetic score using odds ratios estimated from cohorts in these previous studies (appendix).[11,13] We subsequently ranked the SNPs according to their p value and included SNPs in the genetic score in the order of the ranked list. SNPs that could not be genotyped reliably were excluded from the score, leaving 232 SNPs in the STHLM3 model.

STHLM3 was done in two separate phases. The STHLM3 training cohort recruited men from May, 2012, to May, 2013, followed by the STHLM3 validation cohort, which recruited men from August, 2013, to December, 2014. The training cohort (appendix) was used to train and predefine the STHLM3 model algorithm. The validation cohort was used to prospectively test the STHLM3 algorithm. Local and governmental ethics committees reviewed and approved the study protocol. All participants provided written informed consent.

Procedures

The study participants were enrolled and blood was drawn at 67 clinical laboratories in Stockholm collaborating with STHLM3. Previous prostate biopsy sample records (within 10 years before inclusion in STHLM3) were taken from highly accurate health-care registers,[14] together with self-reported family history of prostate cancer (first-degree relatives). PSA levels were analysed in all patients and in those with a PSA concentration of at least 1 ng/mL, additional biomarkers were analysed (genetic and plasma protein markers).

Men with PSA concentrations of less than 1 ng/mL have very low prevalence of high-risk cancers. The PCPT trial[15] showed that almost half the study population had a PSA concentration less than 1 ng/mL and only 7·2% of high-risk cancers were identified in these. Moreover, men with a PSA concentration of less than 1 ng/mL have a very low risk of dying with prostate cancer within 25 years of their diagnosis.[16] This low-risk group was therefore excluded from being tested with the additional biomarkers. Men with a PSA concentration of at least 3 ng/mL or an STHLM3 model indicating high risk were considered to be at an increased risk of high-risk prostate cancer and were referred to a urologist. The urologist performed digital rectal exams, prostate volume measurements, and transrectal prostate biopsy. According to a standardised biopsy protocol, 10 core biopsies were taken if the prostate volume was less than 35 cm. and 12 core biopsies were taken if the volume was greater or equal to 35 cm.. A single pathologist assessed all biopsies to reduce interobserver variance (appendix). Participating urologists and the pathologist were blinded to biomarker results and PSA concentration. Plasma protein biomarkers were analysed using Thermo Fisher Scientific's ISAC multiplex platform. Genotyping was done using the QuantStudio 12K Flex Real-Time PCR System (appendix).

Outcomes

The primary aim of STHLM3 was to increase the specificity of a combined prostate cancer test compared with the PSA test without decreasing the sensitivity of high-risk prostate cancer. The primary endpoints were the number of detected high-risk prostate cancers (sensitivity) and the number of performed prostate biopsies (specificity).

Statistical analysis

Logistic regression was used to test for associations between predictors (biomarkers and clinical variables measured in the STHLM3 training cohort) and cancers with a Gleason score of at least 7 (cancers with Gleason scores of 6 and benign biopsies were treated as controls). Non-linear models and semi-supervised models were also examined but did not improve predictive performance. We let τ denote the STHLM3 model cutoff yielding the same sensitivity as a PSA concentration of at least 3 ng/mL to detect cancers with a Gleason score of at least 7. On the basis of the STHLM3 training cohort data, a cutoff of τ^* was chosen using five-fold cross-validation, such that the sensitivity of the STHLM3 model with cutoff τ^* was estimated to be 7% higher than with a PSA concentration of more than 3 ng/mL. This oversampling was done to ensure that τ was estimable (ie, that as many or more high-risk cancers would be diagnosed with τ^* as an STHLM3 model cutoff for biopsy compared with a PSA concentration of at least 3 ng/mL as a cutoff for biopsy). Assuming the relative false-positive fraction of the STHLM3 model compared with PSA to be 0·81 (for a relative true positive fraction of 1·07), we calculated that a sample size of 48 000 men would be needed to show, with 92% power, a non-inferior sensitivity and superior specificity of the STHLM3 model (appendix).

The trained model was used prospectively in the STHLM3 validation cohort. Study participants were referred to have a biopsy sample taken if they had a predicted STHLM3-model risk of having cancer with a Gleason score of at least 7 exceeding τ^* or a PSA concentration of at least 3 ng/mL.

The paired, screen-positive design is an efficient design to estimate relative changes in test characteristics.[17,18] A paired, screen-positive design was used in the STHLM3 study to estimate ratios between the STHLM3 model and PSA tests for the sensitivity for high risk prostate cancers, for the false-positive fraction of men with benign biopsies and cancers with a Gleason score of 6, and for the number of total biopsies. After the STHLM3 validation cohort closed, we computed an estimate of τ, $\tau^{\wedge}$, by setting the relative sensitivity for high-risk cancers equal to 1 and solving for τ (that is, the same number of cancers with a Gleason score of at least 7 for a PSA concentration of at least 3 ng/mL and for the STHLM3 model cutoff being at least τ). CIs were computed with the non-parametric bootstrap method with 1000 bootstrapped datasets. Briefly, for each bootstrapped dataset, we solved for the STHLM3 model cutoff yielding a relative true positive fraction equal to 1 and then calculated the proportion of patients that had a biopsy taken, the percentage of patients with a tumour of Gleason score 6 that were spared biopsy, the percentage reduction in benign biopsies, and the numbers of detected cancers stratified by Gleason score and total cancer length above or below 10 mm. CIs were then computed with methods described by Efron and colleagues.[19] Age-stratified (5-year strata) results and results using other endpoints[20] (number of all prostate cancers, Gleason score of at least 4 + 3, and CAPRA score[20] of at least 3 as dependent variables) were computed as additional analyses.

For model comparisons, we calculated the area under the curve (AUC) with 95% CI calculated using the bootstrap method. All p values are two-sided and a p value of less than 0·05 was considered significant. We used R statistical software version 3.1 for all analyses (appendix). This study is registered with ISRCTN.com, number ISRCTN84445406 Subsequent to registration, but before enrolment, the study design was changed from a randomised design to that reported here. Continuous quality controls were implemented for all processes and data integrity was monitored regularly throughout the study (appendix).

Role of the funding source

The funder of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report. The corresponding author had full access to all the data in the study and they had final responsibility for the decision to submit for publication.

Results

Of 260 000 men aged 50–69 years with a residential address in Stockholm, 145 905 were randomly selected and invited to the study. 1633 men with prevalent prostate cancer at recruitment were excluded (figure 1). Between May 28, 2012, and May 31, 2013, 11 130 men were recruited to the STHLM3 training cohort. The results from the training cohort are summarised in the appendix.

113 082 men aged 50–69 years were invited to participate in the STHLM3 validation cohort and 1263 men were excluded because of previous prostate cancer diagnosis. Between Aug 5, 2013, and Dec 30, 2014, 47 688 (42%) men chose to enrol in the STHLM3 validation cohort (figure 1). On the basis of the results of the PSA test and the STHLM3 model, or both, 7606 men in the validation cohort were recommended to urological consultation and 5426 (71%) had a prostate biopsy sample taken. Baseline characteristics of the validation cohort are shown (table 1).

All variables used in the STHLM3 model were significantly associated with high-risk prostate cancers ($p<0·05$) in a multiple logistic regression model. STHLM3 did significantly better ($p<0·0001$) than PSA for detection of high-risk prostate cancers; the AUC was 0·56 (95% CI 0·55–0·60) with PSA alone and 0·74 (95% CI 0·72–0·75; table 2). In a stepwise AUC analysis, the established risk factors (age, family history, and biopsy history), the combined biomarker

score (genetic score and plasma protein biomarkers), and the prostate exam (digital rectal exam and prostate volume) all independently improved the test characteristics of the STHLM3 model (table 2).

Using the same sensitivity for the STHLM3 model as for the PSA test to detect high-risk prostate cancers (corresponding to $\tau^{\wedge}$ =0·10; 95% CI 0·09–0·12; ie, 10% predicted risk), the STHLM3 model would reduce the number of biopsies by 32% (95% CI 24–39). Moreover, using the STHLM3 model would have reduced the number of benign biopsies by 44% (95% CI 35–54) (figure 2). Of the 603 high-risk cancers identified by the STHLM3 model, 124 (21%) were identified in the PSA range 1–3 ng/mL. Furthermore, there was no significant difference in the total cancer length of the high-risk prostate cancers detected by the STHLM3 compared with those detected with a PSA of at least 3 ng/mL (p=0·82; figure 2). The number of cancers with a Gleason score of 6 identified by the STHLM3 model was 722, whereas 867 were identified using PSA alone; a reduction of 17% (95% CI 7–26). The 104 cancers with a Gleason score of 6 not identified by the STHLM3 model all had a total cancer length of less than 10 mm in the biopsy samples (figure 3).

Subanalysis of 5-year age strata with a fixed sensitivity (10% risk) for high risk prostate cancers in each age class showed similar performance of the STHLM3 model in all age classes (appendix). Sensitivity analysis showed that the STHLM3 model performance is robust to endpoints other than a Gleason score of at least 7 (eg, a Gleason score of at least 4 + 3 prostate cancers or CAPRA score 0–2; table 3). Comparing the STHLM3 model to PSA alone excluding the prostate exam (digital rectal exam and prostate volume) maintained the robustness of the STHLM3 model (appendix).

Discussion

We have shown that a combination of plasma protein biomarkers, genetic polymorphisms, and clinical variables can improve the specificity of prostate cancer screening significantly compared with PSA in men aged 50–69 years. Use of the STHLM3 model in structured screening could reduce the number of prostate biopsy samples taken by about a third compared with the use of PSA screening. Importantly, this can be achieved without compromising the number of high-risk cancers diagnosed. We identified an equal Gleason score distribution for cancers detected by PSA versus the STHLM3 model, apart from small cancers with a Gleason score of 6 (figure 3). The clinical usefulness of the STHLM3 model is further displayed in the appendix where the STHLM3 model is compared with PSA in a hypothetical example of 10 000 tested men.

We aimed to develop a model for prostate cancer screening with better test characteristics than PSA. PSA was chosen as a comparison because it is the most widely used screening biomarker and the only biomarker that has been assessed prospectively in a randomised controlled trial with prostate cancer mortality as an endpoint. The European Randomized Study of Prostate Cancer (ERSPC)[2] showed a 21% decrease in prostate cancer mortality using PSA screening after 13 years of follow-up. To infer the same mortality reduction as reported in ERSPC, we fixed the sensitivity of the STHLM3 model for detection of high-risk cancers to be equal to that of a PSA concentration of at least 3 ng/mL.

Use of the STHLM3 model could result in a 17% reduction in the number of Gleason score 6 cancers that were biopsied. Although some debate remains as to whether Gleason score 6 cancers can progress to higher-grade disease and although postoperative upgrading of Gleason score 6 cancers is not uncommon, longitudinal cohort studies[21,22] suggest that most of these lesions are indolent and therefore need no treatment. The STHLM3 model would permit a substantial reduction in the number of men needlessly being diagnosed and treated for small and insignificant prostate cancers. Gleason score 6 cancers not diagnosed by the STHLM3 model were all less than 10 mm in total cancer length, indicating that most of these cancers were clinically insignificant.

Information from the prostate exam (prostate volume and digital rectal exam) has been shown to be important for predicting cancers with a Gleason score of at least 7 and improves the STHLM3 model.[12] From a practical standpoint, to avoid having to do a digital rectal exam and prostate volume measurements on all men, we foresee a situation where an individual prostate volume threshold for a biopsy sample recommendation is written explicitly in the urology referral based on the result from the STHLM3 model (eg, on the basis of the STHLM3 model, a biopsy sample is recommended if the prostate volume is <50 cm. or if the digital rectal exam is positive). The genetic score did not improve the overall predictive performance of the STHLM3 model as much as shown in the training cohort (appendix); however, it did improve predictive performance (table 2), it is inexpensive to measure (and the price is constantly dropping), it only needs to be measured once in a man's lifetime, and it is important for men with a very high genetic risk (men in the top decile of the genetic score have a 25% risk of cancer with a Gleason score of at least 7).

To put these numbers in perspective, a 30% reduction in prostate biopsies translates to 300 000 fewer procedures annually in the USA.[5] Having a prostate biopsy sample taken can cause pain and rectal bleeding and increases the risk of a serious infection with multiresistant bacteria.[6] Thus, the improved specificity of the STHLM3 model could result in savings in terms of reduced treatment morbidity, costs to the individual patient, and to the health-care system. A full health economic assessment will be published in a separate report. Future reports will also include an assessment of increasing the PSA concentration level cutoffs for the STHLM3 model testing (set to 1 ng/mL in the STHLM3 model) because the group of men with a PSA concentration between 1 ng/mL and 2 ng/mL showed a low prevalence of cancers with a Gleason score of at least 7.

Two commercial tests based on the kallikreins are available, the Prostate Health Index (PHI) and the 4KScore. Used as a reflex test to PSA, both have been found to increase specificity compared with PSA, but at the risk of missing 5–15% of high-risk cancers that would normally be detected in a clinical setting.[9,23–25] STHLM3 avoids this problem with the

design—ie, prospectively biopsying based on the STHLM3 model results, and the fixed sensitivity between the STHLM3 model and PSA. The STHLM3 model includes all four kallikreins included in 4KScore. A comparative study between 4KScore and PHI has shown that they have similar performance.[26] The 4KScore was recently assessed in the large UK ProtecT study.[27] Although it is difficult to compare the results between different studies, we believe that the results from ProtecT and the results presented here both show the value of using a structured prediction algorithm for biopsy recommendations.[27] Several other prostate cancer risk calculators are also available, combining PSA, free PSA, and clinical variables.[28] To our knowledge, the STHLM3 model is the only model that has been prospectively assessed in a large diagnostic study.

Population-based data from the Stockholm PSA and Biopsy Registry[14] for more than 420 000 men being tested for PSA during the past 10 years suggests that the use of clinical risk calculators is very limited, at least in Stockholm. In Stockholm, the following variables are available in routine clinical care: age, total PSA, repeated PSA testing (PSA velocity), free PSA, family history, prostate volume, PSA density, digital rectal exam, and previous biopsy history. However, the detection rate of high-risk cancers in clinical practice and using PSA alone in this study are similar. The conclusion is that PSA alone, as used in STHLM3, is at least as good as diagnostic practice for detecting high-risk cancers. These results show the difficulty of interpreting several biomarkers and clinical variables at the same time. Some clinicians and health-care systems use age-adjusted PSA cutoffs for biopsy recommendation. In STHLM3, the AUC improvement of using age as a predictor together with PSA was very small (table 2). This indicates that age-specific PSA cutoffs do not improve much on using PSA alone in this study.

The STHLM3 model includes two novel biomarkers for prostate cancer. β-microseminoprotein (MSMB) is one of the most common proteins in human semen and is highly expressed in normal prostate. Several studies[29,30] have indicated that MSMB is downregulated in prostate cancer, particularly in high-grade tumours, and tissue expression MSMB is associated with biochemical recurrences after radical prostatectomy. Macrophage inhibitory cytokine 1 (MIC1), also known as GDF15, is involved in inflammation regulation and apoptotic pathways in injured tissues. In-vitro studies[31] suggest that it plays an important part in the progression of prostate cancer. High serum concentrations of MIC1 were highly associated with metastatic prostate cancer in a Swedish study of 1442 men with prostate cancer.[32]

Comparing AUC statistics from other published biomarker studies are not informative as the receiver operating characteristic (ROC) curves and AUC values in this study are relative. Absolute true and false-positive rates cannot be assessed in a screen-positive design because the true disease rate is not known. Subsequently, comparisons of relative ROC curves are only valid within the STHLM3 dataset and cannot be compared with other studies. For example, because of the marked age-related prevalence of prostate cancer, a difference of 5–10 years in mean age of the study population will substantially change the ROC curve. Similarly, the PSA range of the study population will greatly affect the ROC curve. Instead of using AUC statistics as a main outcome, we used a clinically interpretable endpoint, the number of prostate biopsies.

Our study has some limitations. We cannot assess the performance of the STHLM3 model in a retesting situation because the testing was done only once in this study. Furthermore, STHLM3 is a diagnostic trial, which is not designed to address long-term outcomes—eg, prostate cancer mortality (although we believe the results shown in figure 2 indicate that any effect on mortality by using the STHLM3 model instead of PSA will be small). We are planning follow-up studies to address these research questions. At the time of start of this study in 2011, we chose to use a Gleason score of at least 7 as the main outcome in STHLM3. The view of what is a clinically significant prostate cancer is controversial and varies within the professional community. The STHLM3 model is flexible and when using other outcomes the performance is similar (table 3).

STHLM3 was done in Stockholm, Sweden, and most participants were of northern European descent. However, robust evidence exists[9] that suggests that protein biomarkers in prostate cancer predict outcome similarly in other white populations in different parts of the world. Thus, we predict that the STHLM3 model would be generalisable among these populations. Although most of the SNPs used in this study are also significantly associated with prostate cancer in other ethnicities,[33,34] differences exist between SNP profiles from groups of different ethnic origin. The use of the STHLM3 model in these ethnic groups needs to be validated.

The strengths of the STHLM3 study are that it is population based with a prospective design and uses a predefined algorithm tested in the independent STHLM3 validation cohort. Moreover, the STHLM3 study consists of a large random subsample of the male population aged 50–69 years in Stockholm, where opportunistic PSA testing is common.[14] No significant differences in age and previous PSA testing were reported between participants in STHLM3 and men undergoing PSA testing in Stockholm[14] and the education level of participants in STHLM3 was similar to that in the general male population in Stockholm.[35] Importantly, a subanalysis including 35% of PSA-naïve participants showed that the STHLM3 model worked equally well in these men as in men with a previous history of PSA testing (33% reduced biopsies for PS naive men and 31% for men with previous PSA).

Modern imaging techniques, such as MRI combined with targeted biopsies, will most likely further reduce overdiagnosis and unnecessary prostate biopsies.[36] With use of the STHLM3 model, MRI could be used more cost-effectively by identifying men at increased risk of clinically significant prostate cancer for subsequent MRI referral. Additionally, the identification of additional, new biomarkers can be incorporated within the STHLM3 model as they arise, ultimately improving its performance. We have established a systematic assessment programme of potential new markers using the 58 000 samples obtained in STHLM3 and 26 000 samples from earlier studies.[26]

We argue that the STHLM3 study has several important novel aspects, which increases its potential to change clinical practice. The STHLM3 model identified clinically significant cancers (with a Gleason score of at least 7) in men with low PSA concentration ranges (1–3 ng/mL), representing 40% of the population aged 50–69 years. STHLM3 is a large-

scale population-based study, thus minimising selection bias and increasing generalisability. The STHLM3 model also includes two novel plasma biomarkers and genetics, which have never before been included in a prospective diagnostic study.

Overall, the STHLM3 model can be used as an aid to identify high risk prostate cancers in men aged 50–69 years, with a PSA concentration of at least 1 ng/mL, reducing the number of prostate biopsies and the detection of clinically insignificant disease, while maintaining the sensitivity to clinically significant prostate cancer.

Acknowledgments

We thank all study participants, the STHLM3 core management group for taking care of all contact with participants and organising the databases, KI Biobank at Karolinska Institutet for taking care of all blood samples, aliquoting plasma, extracting DNA, and doing all genotyping, Karolinska University Hospital Laboratory for organising sample handling in more than 60 outpatient laboratories throughout Stockholm, the STHLM3 outpatient urologists taking care of and doing prostate biopsies in more than 7000 STHLM3 participants, Unilabs AB in Stockholm and Histocenter AB in Göteborg for their very high quality handling of more than 80 000 prostate biopsy cores for further pathology assessment, Laboratoriemedicin in Falun for great help with the training cohort, the Thermo Fisher Scientific team in Uppsala and the QuantStudio team in Europe and California for being true partners in the STHLM3 project, and the STHLM3 Scientific Advisory Board who have reviewed and advised on everything from the study protocol to the final results. The main funder of the STHLM3 study is the Stockholm County Council (Stockholms Läns Landsting) who is the main provider of health care in Stockholm. Auxiliary funding for pilots and infrastructure was provided by the Swedish Cancer Society (Cancerfonden), Restaurants Against Cancer (RAC), and the Swedish Research Council (Vetenskapsrådet), Odd Fellow in Västerås, Swedish Research Council for Health Working Life and Welfare (FORTE), and Swedish e-Science Research Center (SeRC). The STHLM3 study is a part of the Linnaeus Center CRISP "Predication and prevention of breast and prostate cancer" funded by the Swedish Research Council.

References

1 Etzioni R, Gulati R, Tsodikov A, et al. The prostate cancer conundrum revisited: treatment changes and prostate cancer mortality declines. Cancer 2012; 118: 5955–63.

2 Schroder 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.

3 Loeb S, Bjurlin MA, Nicholson J, et al. Overdiagnosis and overtreatment of prostate cancer. Eur Urol 2014; 65: 1046–55.

4 Moyer VA, Force USPST. Screening for prostate cancer: US Preventive Services Task Force recommendation statement. Annals Intern Med 2012; 157: 120–34.

5 Loeb S, Vellekoop A, Ahmed HU, et al. Systematic review of complications of prostate biopsy. Eur Urol 2013; 64: 876–92.

6 Aly M, Dyrdak R, Nordstrom T, et al. Rapid increase in multidrugresistant enteric bacilli blood stream infection after prostate biopsy—a 10-year population-based cohort study. Prostate 2015; 75: 947–56.

7 Carignan A, Roussy JF, Lapointe V, Valiquette L, Sabbagh R, Pepin J. Increasing risk of infectious complications after transrectal ultrasound-guided prostate biopsies: time to reassess antimicrobial prophylaxis? Eur Urol 2012; 62: 453–59.

8 Schroder FH, Roobol MJ. Defining the optimal prostate-specific antigen threshold for the diagnosis of prostate cancer. Curr Opin Urol 2009; 19: 227–31.

9 Bruzzese D, Mazzarella C, Ferro M, et al. Prostate health index vs percent free prostate-specific antigen for prostate cancer detection in men with "gray" prostate-specific antigen levels at first biopsy: systematic review and meta-analysis. Transl Res 2014; 164: 444–51.

10 Al Olama AA, Kote-Jarai Z, Berndt SI, et al. A meta-analysis of 87 040 individuals identifies 23 new susceptibility loci for prostate cancer. Nat Genet 2014; 46: 1103–09.

11 Aly M, Wiklund F, Xu J, et al. Polygenic risk score improves prostate cancer risk prediction: results from the Stockholm-1 cohort study. Eur Urol 2011; 60: 21–28.

12 Kader AK, Sun J, Reck BH, et al. Potential impact of adding genetic markers to clinical parameters in predicting prostate biopsy outcomes in men following an initial negative biopsy: findings from the REDUCE trial. Eur Urol 2012; 62: 953–61.

13 Lindmark F, Zheng SL, Wiklund F, et al. H6D polymorphism in macrophage-inhibitory cytokine-1 gene associated with prostate cancer. J Natl Cancer Inst 2004; 96: 1248–54.

14 Nordstrom T, Aly M, Clements MS, Weibull CE, Adolfsson J, Gronberg H. Prostate-specific antigen (PSA) testing is prevalent and increasing in Stockholm County, Sweden, despite no recommendations for PSA screening: results from a population-based study, 2003–2011. Eur Urol 2013; 63: 419–25.

15 Thompson IM, Goodman PJ, Tangen CM, et al. The influence of finasteride on the development of prostate cancer. N Engl J Med 2003; 349: 215–24.

16 Vickers AJ, Cronin AM, Bjork T, et al. Prostate specific antigen concentration at age 60 and death or metastasis from prostate cancer: case-control study. BMJ 2010; 341: c4521.
17 Schatzkin A, Connor RJ, Taylor PR, Bunnag B. Comparing new and old screening tests when a reference procedure cannot be performed on all screenees. Example of automated cytometry for early detection of cervical cancer. Am J Epidemiol 1987; 125: 672–78.
18 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: 1432–38.
19 Efron B. Better bootstrap confidence intervals. J Am Stat Assoc 1987; 82: 171–185.
20 Punnen S, Freedland SJ, Presti JC Jr, et al. Multi-institutional validation of the CAPRA-S score to predict disease recurrence and mortality after radical prostatectomy. Eur Urol 2014; 65: 1171–77.
21 Rider JR, Sandin F, Andren O, Wiklund P, Hugosson J, Stattin P. Long-term outcomes among noncuratively treated men according to prostate cancer risk category in a nationwide, population-based study. Eur Urol 2013; 63: 88–96.
22 Eggener SE, Badani K, Barocas DA, et al. Gleason 6 prostate cancer: translating biology into population health. J Urol 2015; 194: 626–34.
23 Loeb S, Sanda MG, Broyles DL, et al. The prostate health index selectively identifies clinically significant prostate cancer. J Urol 2015; 193: 1163–69.
24 Vedder MM, de Bekker-Grob EW, Lilja HG, et al. The added value of percentage of free to total prostate-specific antigen, PCA3, and a kallikrein panel to the ERSPC risk calculator for prostate cancer in prescreened men. Eur Urol 2014; 66: 1109–15.
25 Vickers AJ, Gupta A, Savage CJ, et al. A panel of kallikrein marker predicts prostate cancer in a large, population-based cohort followed for 15 years without screening. Cancer Epidemiol Biomarkers Prev 2011; 20: 255–61.
26 Nordstrom T, Vickers A, Assel M, Lilja H, Gronberg H, Eklund M. Comparison Between the four-kallikrein panel and prostate health index for predicting prostate cancer. Eur Urol 2015; 68: 139–46.
27 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; published online Apr 11. DOI:10.1093/jnci/djv095.
28 van Vugt HA, Kranse R, Steyerberg EW, et al. Prospective validation of a risk calculator which calculates the probability of a positive prostate biopsy in a contemporary clinical cohort. Eur J Cancer 2012; 48: 1809–15.
29 Whitaker HC, Kote-Jarai Z, Ross-Adams H, et al. The rs10993994 risk allele for prostate cancer results in clinically relevant changes in microseminoprotein-beta expression in tissue and urine. PLoS One 2010; 5: e13363.
30 Dahlman A, Rexhepaj E, Brennan DJ, et al. Evaluation of the prognostic significance of MSMB and CRISP3 in prostate cancer using automated image analysis. Mod Pathol 2011; 24: 708–19.
31 Bruzzese F, Hagglof C, Leone A, et al. Local and systemic protumorigenic effects of cancer-associated fibroblast-derived GDF15. Cancer Res 2014; 74: 3408–17.
32 Brown DA, Lindmark F, Stattin P, et al. Macrophage inhibitory cytokine 1: a new prognostic marker in prostate cancer. Clin Cancer Res 2009; 15: 6658–64.
33 Xu J, Mo Z, Ye D, et al. Genome-wide association study in Chinese men identifies two new prostate cancer risk loci at 9q31.2 and 19q13.4. Nat Genet 2012; 44: 1231–35.
34 Han Y, Signorello LB, Strom SS, et al. Generalizability of established prostate cancer risk variants in men of African ancestry. Int J Cancer 2015; 136: 1210–17.
35 Statistics Sweden (SCB). Statistikdatabasen. http://www. statistikdatabasen.scb.se (accessed March 1, 2015).
36 Futterer JJ, Briganti A, De Visschere P, et al. Can clinically significant prostate cancer be detected with multiparametric magnetic resonance imaging? A systematic review of the literature. Eur Urol 2015; 68: 738.

Figures and Tables

	Participants enrolled to the study (n=47 688)	Participants that had a biopsy sample taken* (n=5426)
Age (years)		
50–54	11723 (25%)	492 (9%)
55–59	10924 (23%)	897 (17%)
60–64	11159 (23%)	1460 (27%)
65–69	13882 (29%)	2577 (47%)
First-degree relative with prostate cancer (self-reported)		
Yes	5872 (12%)	834 (15%)
No	41816 (88%)	4592 (85%)
Participants that had previously (within 10 years of inclusion) had a negative biopsy		
Yes	1739 (4%)	431 (8%)
No	45949 (96%)	4995 (92%)
Had a PSA test within 10 years of inclusion		
Yes	31435 (66%)	3709 (68%)
No	16253 (34%)	1717 (32%)
PSA at inclusion (ng/mL)		
Median	1·1 (0·6–1·6)	3·8 (3·0–4·8)
<1	21175 (44%)	0
1 to <3	20136 (42%)	1036 (19%)
3 to <5	4042 (8%)	2878 (53%)
5 to <10	1834 (4%)	1196 (22%)
10 or more	510 (1%)	316 (6%)
5-α-reductase inhibitors at inclusion		
Yes†	1179 (2%)	131 (2%)
No	46509 (98%)	5295 (98%)
Digital rectal exam		
Abnormal	–	524 (10%)
Normal	–	4902 (90%)
Prostate volume (mL)‡		
<35	–	1829 (34%)
35–50	–	1889 (35%)
>50	–	1708 (31%)
Biopsy results		
Benign	–	3320 (61%)
Gleason score of 3+3	–	1189 (22%)
Gleason score of 3+4	–	566 (10%)
Gleason score of 4+3	–	182 (3%)
Gleason score of 4+4 or more	–	169 (3%)

Data are n (%) or median (IQR). *Biopsied up to Feb 28, 2015. About 300 biopsy samples planned for March and April, 2015, are not included in the analysis. †Prostate volume was assessed by transrectal ultrasound. ‡Men with a prostate specific antigen concentration of at least 10 ng/mL and men receiving 5-α-reductase inhibitors are excluded from analysis. Five men had both a PSA concentration of at least 10 ng/mL and were receiving 5-α-reductase inhibitors.

Table 1: Baseline characteristics of the validation cohort

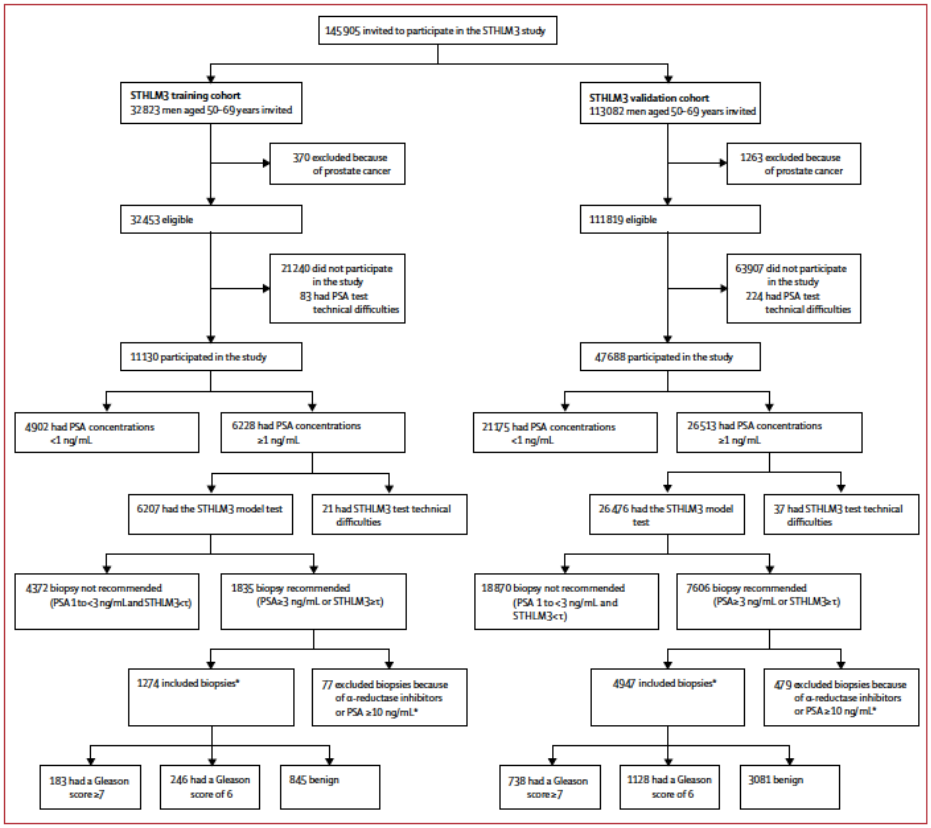

Figure 1: CONSORT diagram of the STHLM3 study
We let τ denote the STHLM3 model cutoff yielding the same sensitivity as a PSA concentration of at least 3 ng/mL to detect cancers with a Gleason score of at least 7. *Some participants did not have the recommended biopsies.

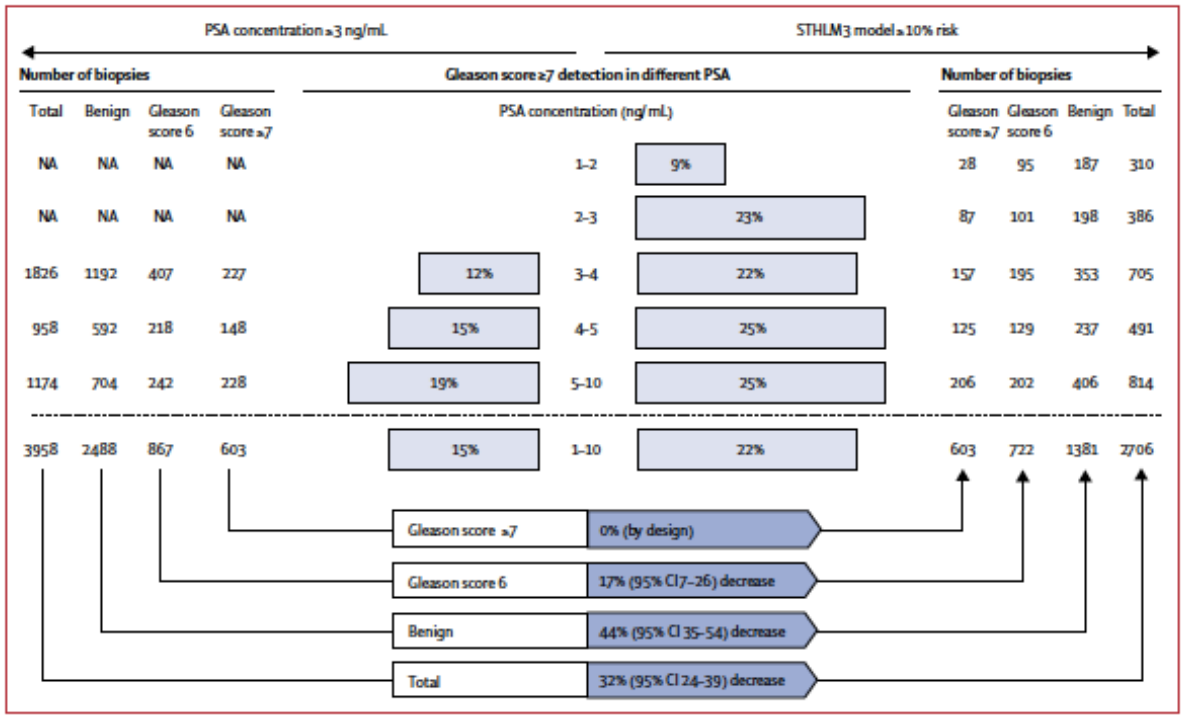

Figure 2: Comparison of biopsy results in the STHLM3 validation cohort (n=4947 biopsy samples) using PSA concentrations of at least 3 ng/mL versus the STHLM3 model with at least 10% risk of detecting cancers
PSA=prostate-specific antigen. NA=not applicable.

	Univariate p value	Multivariate p value	AUC (univariate)	AUC (stepwise multivariate)	Cumulative AUC	p value*	Saved biopsies compared with PSA≥3 ng/mL (%)
Total PSA	0·001	0·008	0·56 (0·54–0·59)	0·56 (0·54–0·59)	0·56 (0·55–0·60)	Reference	Reference
Risk factors	..	..	..	..	0·58 (0·56–0·60)	<0·0001	5% (3–8)
Age	<0·0001	<0·0001	0·54 (0·52–0·56)	0·57 (0·55–0·59)	..	..	..
Family history	0·005	0·004	0·52 (0·51–0·54)	0·58 (0·55–0·60)	..	..	..
Previous biopsies	<0·0001	<0·0001	0·51 (0·50–0·52)	0·58 (0·56–0·60)	..	..	..
Biomarkers†	..	..	..	..	0·70 (0·68–0·72)	<0·0001	27% (19–35)
Genetic score‡	<0·0001	0·006	0·54 (0·52–0·56)	0·60 (0·58–0·62)	..	..	..
MSMB	0·0002	0·0002	0·54 (0·52–0·56)	0·61 (0·59–0·63)	..	..	..
MIC1	<0·0001	0·047	0·53 (0·51–0·56)	0·62 (0·60–0·64)	..	..	..
Free PSA §	<0·0001	<0·0001	0·55 (0·53–0·57)	0·66 (0·64–0·68)	..	..	..
Intact PSA	0·194	<0·0001	0·52 (0·50–0·55)	0·69 (0·67–0·71)	..	..	..
hK2	<0·0001	<0·0001	0·55 (0·53–0·57)	0·70 (0·68–0·72)	..	..	..
Prostate exam¶	..	..	..	..	0·74 (0·72–0·75)	<0·0001	32% (24–39)
Digital rectal exam‖	<0·0001	<0·0001	0·57 (0·56–0·59)	0·72 (0·70–0·74)	..	..	..
Prostate volume‖	<0·0001	<0·0001	0·62 (0·60–0·64	0·74 (0·72–0·75)	..	..	..

Data are p value or AUC (95% CI), unless otherwise indicated. 442 men with PSA≥10 or 5-α-reductase inhibitor users were excluded. Additionally, 37 men with incomplete data were excluded from the analysis. PSA=prostate-specific antigen. AUC=area under the curve. MSMB=β-microseminoprotein. MIC1=macrophage inhibitory cytokine 1. hK2=human kallikrein 2. *p value from DeLong's test for differences in AUC. †The biomarker score was computed for each participant by combining the genetic score and five plasma biomarkers (MSMB, MIC1, free PSA, intact PSA, and hK2) using logistic regression. ‡The genetic score was computed for each participant by summing the number of risk alleles at each of the 232 SNPs multiplied by the logarithm of each SNP's odds ratio estimated from Swedish CAPS and STHLM1. §Both free PSA and ratio-free PSA or total PSA have been included in STHLM3 model. Univariate AUC for ratio-free PSA or total PSA is 0·63 (95% CI 0·61–0·65). ¶Because all blood-based markers will be used to refer men to a urological assessment (digital rectal exams and transrectal ultrasound), they are added to the model before adding digital rectal exams and prostate volume as predictors. ‖Prostate volume and digital rectal exams were only assessed in men who had biopsy samples taken.

Table 2: Test characteristics of the clinical variables and biomarkers included in the STHLM3 model for prediction of prostate cancers with a Gleason score of at least 7 in the STHLM3 validation cohort based on 4947 biopsy samples taken in men aged 50–69 years

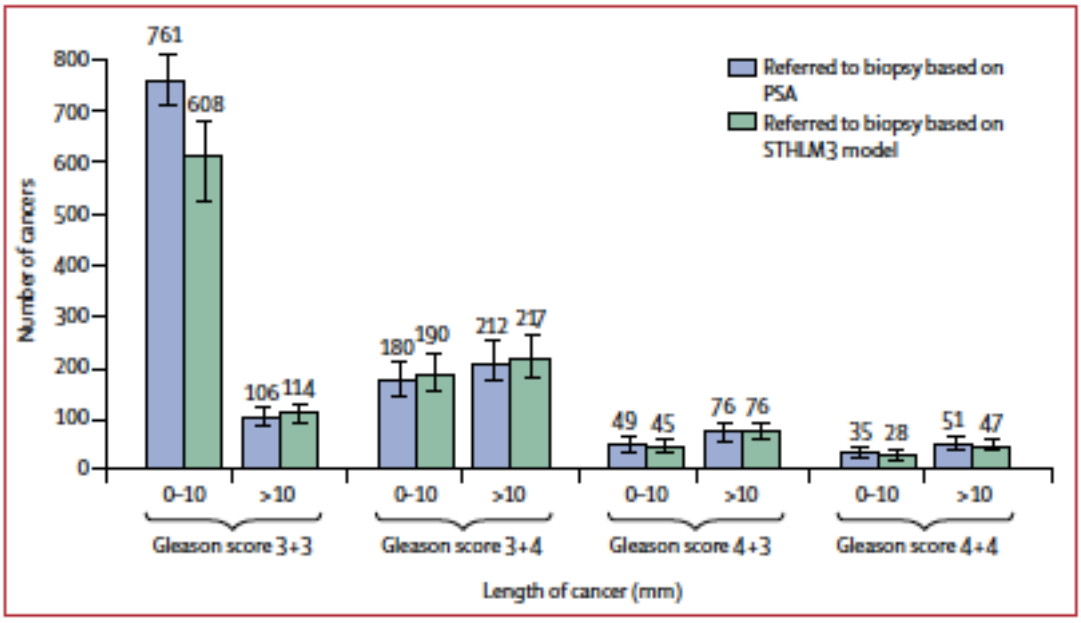

***Figure 3*: Comparison of the tumour characteristics of biopsy samples from the validation cohort with a Gleason score of at least 6 in the biopsy samples obtained from men referred using either the STHLM3 model with a 10% risk cutoff or PSA concentrations of at least 3 ng/mL**

The total number of cancers with a Gleason score of at least 7 are fixed and by design the same (n=603) in both groups. No significant differences in cancer length in cancers with a Gleason score of at least 7 were detected by the STHLM3 model versus PSA testing (p=0·82). The reduction in Gleason score 6 cancers detected by the STHLM3 model was for cancers <10 mm length (p<0·001). The error bars on the graph represent 95% CI. NA=not applicable. PSA=prostate-specific antigen.

	Area under the curve		Proportion of biopsies saved compared with PSA ≥3 ng/mL (%)		
	PSA test (95%CI)	STHLM3 model (95%CI)	All biopsies (95%CI)	Cancers with a Gleason score of 6 (95%CI)	Benign (95%CI)
All prostate cancers	0·52 (0·50–0·53)	0·69 (0·68–0·71)	20% (16–25)	5% (3–9)	32% (26–38)
Cancers with a Gleason score ≥7	0·56 (0·54–0·59)	0·74 (0·72–0·75)	32% (24–39)	17% (7–29)	44% (35–54)
Excluding very low-risk cancer*	0·64 (0·62–0·67)	0·78 (0·76–0·80)	30% (21–37)	15% (5–24)	43% (36–53)
Cancers with a Gleason score ≥(4+3)	0·60 (0·56–0·64)	0·74 (0·71–0·77)	23% (15–34)	8% (1–19)	35% (26–48)

The results are based on the STHLM3 validation cohort including 4947 biopsies done in men aged 50–69 years. PSA=prostate-specific antigen. *Very low-risk cancer is defined as a CAPRA score of 0–2.

Table 3: Test characteristics of the STHLM3 model using the different endpoints compared with the PSA test