



BMJ Open Prediction models for prostate cancer to be used in the primary care setting: a systematic review

Mohammad Aladwani,¹ Artitaya Lophatananon ,¹ William Ollier,^{1,2} Kenneth Muir ¹

To cite: Aladwani M, Lophatananon A, Ollier W, *et al.* Prediction models for prostate cancer to be used in the primary care setting: a systematic review. *BMJ Open* 2020;**10**:e034661. doi:10.1136/bmjopen-2019-034661

► Prepublication history and additional material for this paper are available online. To view these files, please visit the journal online (<http://dx.doi.org/10.1136/bmjopen-2019-034661>).

Received 02 October 2019

Revised 21 May 2020

Accepted 29 May 2020



© Author(s) (or their employer(s)) 2020. Re-use permitted under CC BY. Published by BMJ.

¹Division of Population Health, Health Services Research and Primary Care School of Health Sciences Faculty of Biology, Medicine and Health, The University of Manchester, Manchester, UK

²School of Healthcare Science, Manchester Metropolitan University Faculty of Science and Engineering, Manchester, UK

Correspondence to

Dr Kenneth Muir;
kenneth.muir@manchester.ac.uk

ABSTRACT

Objective To identify risk prediction models for prostate cancer (PCa) that can be used in the primary care and community health settings.

Design Systematic review.

Data sources MEDLINE and Embase databases combined from inception and up to the end of January 2019.

Eligibility Studies were included based on satisfying all the following criteria: (i) presenting an evaluation of PCa risk at initial biopsy in patients with no history of PCa, (ii) studies not incorporating an invasive clinical assessment or expensive biomarker/genetic tests, (iii) inclusion of at least two variables with prostate-specific antigen (PSA) being one of them, and (iv) studies reporting a measure of predictive performance. The quality of the studies and risk of bias was assessed by using the Prediction model Risk Of Bias ASsessment Tool (PROBAST).

Data extraction and synthesis Relevant information extracted for each model included: the year of publication, source of data, type of model, number of patients, country, age, PSA range, mean/median PSA, other variables included in the model, number of biopsy cores to assess outcomes, study endpoint(s), cancer detection, model validation and model performance.

Results An initial search yielded 109 potential studies, of which five met the set criteria. Four studies were cohort-based and one was a case-control study. PCa detection rate was between 20.6% and 55.8%. Area under the curve (AUC) was reported in four studies and ranged from 0.65 to 0.75. All models showed significant improvement in predicting PCa compared with being based on PSA alone. The difference in AUC between extended models and PSA alone was between 0.06 and 0.21.

Conclusion Only a few PCa risk prediction models have the potential to be readily used in the primary healthcare or community health setting. Further studies are needed to investigate other potential variables that could be integrated into models to improve their clinical utility for PCa testing in a community setting.

INTRODUCTION

Prostate cancer (PCa) is the second most common cancer and the fifth leading cause of cancer-attributed death in men worldwide with an estimated incidence of 1 276 106 and 358 989 deaths in 2018.¹ In the UK, around 47 200 new cases of PCa were reported in

Strengths and limitations of this study

- The review focussed on risk prediction models for PCa for use in primary care.
- The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) approach was followed in identifying relevant articles and reporting this study.
- We used the Prediction model Risk Of Bias ASsessment Tool (PROBAST) to assess the quality and risk of bias in the included models.
- The search strategy was restricted to two databases and a manual search, to retrieve original studies.

2015, accounting for 26% of all new cancer cases in males. PCa deaths in the UK of were around 11 600 in 2016.² The global projections of PCa incidence and mortality for 2030 are 1.7 and 0.5 million, respectively.³ The highest incidence of PCa is seen in western societies.⁴ The significant increase of PCa incidence and diagnosis over the last three decades can be attributed mainly to the widespread implementation of the prostate-specific antigen (PSA serum test after it had been introduced in the late 1980s.^{5 6}

The strong association of PSA with PCa,^{7 8} along with it being a relatively inexpensive test,⁹ has made PSA a key biomarker in the diagnostic process of PCa and for the recommendation of a confirmatory prostate biopsy.^{7 9} PSA is, however, not a cancer-specific marker.^{5 10} Conditions such as benign prostate hypertrophy (BPH), prostatitis and other non-malignant prostatic conditions can also elevate PSA level, thus introducing uncertainty to the application of the test.^{11–14} This highlights limitations of the PSA test regarding its specificity and sensitivity, and it being largely dependent on setting a ‘diagnostic’ cut-off point, which often leads to an unacceptable number of false-positive and false-negative results.^{5 10 15 16} Such issues are likely to be the part of the explanation for

the significant number of unnecessary biopsies currently being performed each year. Such procedures are associated with adverse side effects for patients and also increases healthcare costs.^{17 18}

To address such PSA test limitations, researchers have incorporated other measurable factors into approaches for the early detection of PCa; these 'risk assessment tools' are based on statistical models designed to improve the accuracy and performance of the PSA test.^{19–22} Logistic regression and artificial neural network (ANN) models are now considered to be the most common and effective statistical techniques in aiding the development of new models to enhance early PCa diagnosis.²³ These PCa risk prediction models can be used to aid the testing of men for further investigations.

Currently in the UK, there is no population-based screening programme for PCa. The ultimate goal of PCa screening is to find intermediate and high risk of PCa rather than low-risk PCa that would not require treatment but will give emotional burden to the patient once detected and unnecessary treatment in some patients. An important potential advantage of the extended risk models is their ability to provide a more accurate estimation of PCa risk. This may ultimately lead to their use in patient counselling and decision-making.^{24–27} Such models have already achieved better results in predicting probabilities of outcome compared with clinical judgement.^{28 29} Furthermore, it has been reported that using such predictive models may minimise the rate of unnecessary biopsies.³⁰

Recently, there has been a substantial increase in the development of predictive models to help clinicians assess risk and decide which man to send to clinical setting to further investigate for a possible diagnosis of PCa.^{22 26 30–35} The majority of these models are designed for use in clinical settings, where costs are less of an issue and most include the need for a clinical examination such as digital rectal examination (DRE) or trans-rectal ultrasonography (TRUS). One of the main limitations of DRE is its poor performance, especially at low PSA levels, and it is highly subjective to inter-observer variability.^{36–38} A meta-analysis study revealed that DRE has positive predictive value of only 18%.³⁹ Similarly, TRUS has been reported for having poor accuracy at low PSA levels^{40 41} and small PCa might not be palpable on DRE or visualisation on TRUS.⁴⁰ Furthermore, less than 40% of PCa detected by DRE are potentially curable, making it less beneficial for early diagnosis.⁴² Several studies showed that there is fear, anxiety and embarrassment among some men, in particular Black men, regarding the DRE test.^{43–46} Another disadvantage of the DRE is the fact that it is a potentially uncomfortable test.^{47–52} This may explain why the DRE is a barrier for some men to participate in PCa screening if it includes DRE test. Lee *et al* reported that 74% to 84% of Black men may not maintain annual DRE screening,⁵³ while another study found that it may prevent 22% of men from participating.⁵⁴ Since TRUS needs to be performed by a skilled urologist, this means men have to make an

appointment with a clinic in a different location, which makes the screening less convenient. As a result, men may feel reluctant to have such tests performed.

This systematic review of the literature was undertaken to identify risk prediction models that do not incorporate invasive or more costly clinical procedures or extensive biomarkers but have potential application for use in primary care and community settings. As low cost is a primary concern for community use, for this review, we set an indicative threshold of approximately three to five times the cost of a PSA test for inclusion. This excluded a number of models that contain new and emerging biomarker or single nucleotide polymorphism panels. As a number of persons are referred to the clinical setting, costs are less of an issue. The performance of the models reviewed for detecting PCa at initial biopsy have been compared using 'reported area under the curve' (AUC) and/or sensitivity-specificity testing.

METHODS

The approach used to identify and select relevant articles was based on the application of the 'Preferred Reporting Items for Systematic Reviews and Meta-Analysis' (PRISMA).⁵⁵

Data sources and search strategy

A literature search was performed using MEDLINE (via Ovid) and Embase databases. The 'medical subject heading' (MeSH) terms, combined with Boolean logic operators 'AND' and 'OR', were applied to retrieve relevant articles. The terms used for the search were 'Prostatic Neoplasms' AND ('Initial biopsy' OR 'first biopsy' OR 'early detection of cancer') AND ('nomograms' OR 'artificial neural networks' OR 'risk assessment' OR 'statistical model'). The full search strategy is provided in a online supplementary file 1. All articles defined (published since the inception of the databases and up to the end of January 2019) were subsequently further filtered as being those only published in English language and with an abstract. Further to using the above search databases, the research articles were selected manually from the reference lists of any relevant review articles. Google Scholar and MEDLINE searches were also carried out to identify independent study for external validation for each model included in this review. The results are presented in online supplementary file 2.

Eligibility criteria

As this review focusses on PCa risk prediction based in community healthcare settings, all studies were selected on the following inclusion criteria: (i) evaluating the risk for PCa at initial biopsy in patients who had no prior history of PCa, (ii) studies that reported 'low cost' risk assessment tools (ie, those not including more expensive genetic or biomarker test) or 'invasive' clinical tests/examinations (such as DRE or TRUS), (iii) studies that included a minimum of two variables of which PSA had

to be one of them (on the basis that an elevated PSA test in UK primary care is usually the first sign and rationale for suggesting a need for further investigation of PCa within NICE guidelines), and (iv) studies that reported AUC and/or sensitivity and specificity of the diagnostic/predictive tool. The exclusion criteria used were: (i) articles with models that were built and based on repeat or mixed biopsies, (ii) studies that only validate an existing model, and (iii) articles that were not published in English. There were no time boundaries regarding the publication year.

Screening of the titles, abstracts and full-texts was carried out by two reviewers (MA, AL). Any concerns about the eligibility of a study were resolved by discussion with a third reviewer (KM).

Data extraction

A data extraction form was developed to collect all relevant information. For each study used in this review, the items extracted included: year of publication, source of data, type of model, number of patients, country where it was performed, age, PSA range, mean/median PSA, number of biopsy cores, variables included in the model, study endpoint(s), cancer detection, model performance and model validation.

Evaluating the performance of the risk models

Prediction models can be evaluated against various criteria. The most critical measurements of model performance are discrimination and calibration.²⁷ Discrimination refers to how well the prediction model can differentiate patients in different outcome classes according to their predicted risks. It is often assessed by measuring the area under the receiver operating characteristic curve.⁵⁶ It also requires setting a series of thresholds to separate low and high ranges of predicted outcomes. A value of 0.5 indicates no discrimination, while a value of 1 indicates perfect discrimination. However, even with perfect discrimination, observed risk can differ from expected risk. Therefore, calibration has an important role in model evaluation.⁵⁷ Calibration represents the agreement between expected and observed outcomes.⁵⁸ A well-calibrated model is achieved when the calibration slope is close to 1. When the calibration slope is less than 1, it indicates that the model underestimates low risks and overestimates high risks.⁵⁹

Due to the heterogeneity of the studies included, conducting a meta-analysis was not applicable.

Study quality assessment

The quality of the studies included in this review was assessed using the Prediction model Risk Of Bias Assessment Tool (PROBAST).⁶⁰ This tool has been developed specifically to assess the risk of bias and applicability for prediction model studies. The tool consists of four domains and has 20 signalling questions that facilitate reaching overall judgement of risk of bias, as well as issues relating to applicability.



Figure 1 Flow diagram of studies included using the Preferred Reporting Items for Systematic Reviews and Meta-Analyses method. DRE, digital rectal examination; PCA3, prostate cancer antigen 3; PV, prostate volume; SNP, single nucleotide polymorphism.

Patient and public involvement

No patients were involved in setting the research question or the outcome measures, nor were they involved in the design and implementation of the study. There are no plans to involve patients in dissemination.

RESULTS

A total of 102 publications were identified using the search strategy as shown in figure 1. An additional nine articles were identified through manual searches from a bibliography of reviewed articles. At the first filter step, a total of 109 titles and abstracts were screened for eligibility after removing two duplicates. In the second filter step, 60 papers failed to meet the inclusion criteria and were excluded, resulting in 49 articles. The final step of filtering yielded only five studies that were considered to be eligible (ie, passed all set criteria) and were thus included in this systematic review. There was no independent study identified for external validation for included models.

Study characteristics

Four of the five studies included were based on cohort studies and one was a case-control study. The characteristics of each of these studies and populations are summarised in table 1. Details of PSA assays used in the models are presented in online supplementary file 3.

Patients used to build the risk models varied across these studies. Of the five studies, three studies included men from referral populations^{61–63} and two studies from screening programmes.^{64 65} The sample sizes ranged from 151 to 3773, with three studies derived from US

Table 1 Characteristics of the included studies

Author and year	Type of model	Type of study	Sample no.	Location	Population type	Age	Median age	PSA range	Median PSA	No. of biopsy cores	Cancer detection
Carlson <i>et al</i> , 1998 ⁶¹	Logistic regression	Cohort	Model dev=3773 Validation=525	Baltimore, USA	Referral	≥45	—	4 to 20ng/ mL	—	Sextant biopsy	32%
Babaian <i>et al</i> , 2000 ⁶⁴	Neural network=3 ANNs	Cohort	151	Texas, USA	Screening programme	40 to 75	62	2.5 to 4ng/ mL	—	11 cores	24.50%
Jansen <i>et al</i> , 2010 ⁶⁵	Logistic regression	Cohort	Site 1=405 Site 2=351	Site 1 from the Rotterdam arm of the European Study of screening for Prostate cancer Site 2 Innsbruck, Austria	Screening programme	≥50	Site 1 (66) Site 2 (60)	2 to 10ng/ mL	~4.4	≥6 cores	Site 1=55.8% Site 2=49.6%
Hill <i>et al</i> , 2013 ⁶²	Logistic regression	Case-control	1378	Florida, USA	Hospital referral	40 to 90	—	≥4ng/mL	—	N/A	20.60%
Lazzeri <i>et al</i> , 2013 ⁶³	Logistic regression	Cohort	646	European multicentre; Italy, Germany, France, Spain, and the UK	Referral	>45	—	2 to 10ng/ mL	~5.8	≥12 cores	40.10%

ANN, artificial neural network ; PSA, prostate-specific antigen.

Table 2 Variables used in the prostate cancer risk prediction models

Author and year	Variables used in the model				
	Total PSA	Free PSA	Per cent free PSA	Age	Other variables
Carlson <i>et al</i> , 1998 ⁶¹	✓		✓	✓	
Babaian <i>et al</i> , 2000 ⁶⁴	✓	✓		✓	Creatine kinase, prostatic acid phosphatase
Jansen <i>et al</i> , 2010 (Site 1) ⁶⁵	✓	✓			p2PSA
Jansen <i>et al</i> , 2010 (Site 2) ⁶⁵	✓	✓			p2PSA
Hill <i>et al</i> , 2013 (Method 1) ⁶²	✓			✓	HGB, RBC, haematuria, creatinine, MCV and ethnicity 'Black'
Hill <i>et al</i> , 2013 (Method 2) ⁶²	✓			✓	HGB, RBC, creatinine and MCV
Lazzeri <i>et al</i> , 2013 (Model 1) ⁶³	✓	✓	✓		
Lazzeri <i>et al</i> , 2013 (Model 2) ⁶⁶	✓	✓	✓		p2PSA
Lazzeri <i>et al</i> , 2013 (Model 3) ⁶³	✓	✓	✓		%p2PSA
Lazzeri <i>et al</i> , 2013 (Model 4) ⁶³	✓	✓	✓		PHI

HGB, haemoglobin; MCV, mean corpuscular volume; PHI, prostate health index; p2PSA, precursor of PSA; %p2PSA, percentage of p2PSA to free PSA; PSA, prostate-specific antigen; RBC, red blood cells.

cohorts^{61 62 64} and two from Europe.^{63 65} Four studies used logistic regression methodology to build their model, whereas one study used an ANN-based approach.⁶⁴ The minimum age of participants was 40 years^{62 64} and the minimum PSA level was 2 ng/mL.^{63 65}

Variables in the model

Table 2 presents details of the variables used in each model. PSA level was used in all models, followed by free PSA (fPSA), age and free-to-total PSA ratio (%fPSA). Other variables also reported in the models included: precursor of PSA (p2PSA), percentage of p2PSA to fPSA (%p2PSA), prostate health index (PHI), levels of haemoglobin (HGB), albumin, creatinine and red blood cell count (RBC), haematuria, mean corpuscular volume (MCV) and prostatic acid phosphatase.

Outcome

The study endpoint also varied among the studies selected. Two studies evaluated the accuracy of detecting any PCa^{61 64} and three studies examined the pathologic Gleason score.^{62 63 65} Although Jansen *et al*, did not build a model to predict the aggressiveness of PCa, they assessed the relationship of each variable individually with a Gleason score ≥ 7 . PCa was determined by taking a needle biopsy. All patients in the five studies underwent prostate biopsy. The least number of biopsy cores used were six^{61 65} and the highest were ≥ 12 .⁶³ One study did not report the number of biopsy cores taken.⁶² PCa rates ranged from 20.6% to 55.8%.

Evaluating the performance of the risk models

For predicting any PCa, the Jansen *et al*, model used data from the Rotterdam arm of the European Study of Screening for PCa (ESPRC). Their model achieved the highest discrimination value when compared with PSA

alone (AUC of 0.755 vs 0.585, respectively).⁶⁵ The AUC values in other studies ranged from 0.648 to 0.74.

One study did not provide the AUC but instead reported an increase of 11% in specificity over per cent fPSA alone with 95% sensitivity.⁶¹ Lazzeri *et al*,⁶³ presented results from four separated models discriminating PCa with a Gleason score of ≥ 7 . Lazzeri's model 2 (which includes the base model total PSA, fPSA and %fPSA in addition to p2PSA) and model 3 (which includes base model plus PHI) showed the highest levels of discrimination out of the four models with an AUC of 0.67. In the study of Hill,⁶² the authors classified PCa stages differently and built their two models accordingly. In Hill's first model, the difference in the discrimination was analysed and based on all PCa versus non-cancerous prostate conditions where the AUC for this model was 0.68 compared with 0.59 for PSA alone. In Hill's second model, the discrimination analysis was based on PCa stages II, III, IV versus PCa stage I, prostatic interstitial neoplasm, BPH and prostatitis where stages I, II, III and IV are parallel to T1, T2, T3/T4 and metastatic PCa, respectively. The AUC for the second model was 0.72 compared with 0.63 for PSA alone. In general, four studies examined the AUC with PSA alone and all reported a benefit from the use of logistic regression or the trained ANN. Model performance and the differences between the AUC's for PSA alone and for the extended models are presented in table 3.

Sensitivity and specificity data are presented in table 4. At 95% sensitivity, the Babaian *et al* model shows the highest specificity (51%) whereas the Jansen model for both sites had the lowest specificity (~23.5%). In the Hill study, with a sensitivity of ~90%, the specificity was lower than in other studies (~18% and 28%) for method 1 and 2, respectively. In the study reported by Lazzeri, the sensitivity and specificity were not reported for the overall

Table 3 The difference of AUC for PSA alone and extended model

Study	AUC for PSA	AUC for model	Δ AUC (Model – PSA)
Carlson <i>et al</i> ⁶¹	NA	NA	NA
Babaian <i>et al</i> ⁶⁴	0.64 %fPSA	0.74	0.1
Jansen <i>et al</i> (Site 1) ⁶⁵	0.58	0.75	0.17
Jansen <i>et al</i> (Site 2) ⁶⁵	0.53	0.7	0.16
Hill <i>et al</i> (Method 1) ⁶²	0.59	0.68	0.09
Hill <i>et al</i> (Method 2) ⁶²	0.63	0.72	0.09
Lazzeri <i>et al</i> ⁶³	0.50 for any PC 0.54 for Gleason score ≥ 7	Model 1=0.65	0.15
		Model 1 (Gleason score ≥ 7)=0.60	0.06
		Model 2=0.71	0.21
		Model 2 (Gleason score ≥ 7)=0.67	0.13
		Model 3=0.704	0.2
		Model 3 (Gleason score ≥ 7)=0.67	0.13
		Model 4=0.71	0.21
		Model 4 (Gleason score ≥ 7)=0.672	0.13

AUC, area under the curve; %fPSA, free-to-total PSA ratio ; PC, prostate cancer; PSA, prostate-specific antigen.

model; instead their study reports sensitivity and specificity for predictive variables individually. The highest sensitivity (90.5%) of %p2PSA and %fPSA achieved the

highest specificity in predicting PCa at 21.5% and 22.8%, respectively. Percentage p2PSA and PHI were more associated with Gleason scores.

Table 5 summarises the validation and calibration results for the studies included. Model calibration was reported in two studies.^{61 63} Carlson plotted the observed and expected risks using calibration plots, whereas Lazzeri used the Hosmer-Lemeshow goodness-of-fit test. In terms of validation, two studies did not report model validation.^{62 65} Only one study reported an external validation using an additional data set consisting of 525 patients.⁶¹ Cross-validation using multiple re-sampling schemes was used in the Babaian study; however, they did not report the number of time this was performed.⁶⁴ Lazzeri used 200 bootstrap re-samples to minimise overfitting bias.⁶³

Study quality assessment

Quality assessment was carried out by two reviewers (MA and AL) with any discordance resolved by a third reviewer (KM). The assessment of results suggested some issue of study quality according to the criteria as set in the PROBAST, particularly in the analysis domain. For instance, one study applied univariable analysis to select predictors.⁶¹ Three studies did not measure calibration.^{62 64 65} Furthermore, two studies did not account for optimism and overfitting by using internal validation methods.^{62 65} Whereas one study did not use appropriate measures for model performance that is, AUC, this study reported the calibration.⁶¹

The event per variable was lower than recommended (<10)^{59 66} in the Babaian⁶⁴ study, indicating inadequate

Table 4 Sensitivity and specificity profile at different levels for each model*

Study	Sensitivity	Specificity	Probability cut-off	Positive predictive value	Negative predictive value
Carlson <i>et al</i> ⁶¹	99	18	>15	≤ 47	NA
	95	34	18	51	NA
	89	43	20	42	NA
Babaian <i>et al</i> ⁶⁴	95	51	NA	39	97
	92	62	NA	44	96
	89	62	NA	43	95
Jansen <i>et al</i> (Site 1) ⁶⁵	95	23.9	NA	NA	NA
Jansen <i>et al</i> (Site 1) ⁶⁵	90	30.1	NA	NA	NA
Jansen <i>et al</i> (Site 2) ⁶⁵	95	23.2	NA	NA	NA
Jansen <i>et al</i> (Site 2) ⁶⁵	90	36.2	NA	NA	NA
Hill <i>et al</i> (Method 1) ⁶²	90.9	17.6	33	47.1	70.5
Hill <i>et al</i> (Method 2) ⁶²	89.8	28	13	20.6	91.3
Hil <i>et al</i> (Method 1) ⁶²	80.5	37.1	37	50.9	70.2
Hill <i>et al</i> (Method 2) ⁶²	78.2	45	15	28.7	88.8
Hill <i>et al</i> (Method 1) ⁶²	39.9	81.4	48	63.4	62.6
Hill <i>et al</i> (Method 2) ⁶²	45.8	79.5	23	36.7	85

*Lazzeri⁶³ model reported only sensitivity and specificity for predictive variables individually and at sensitivity of 90, %p2PSA and %fPSA achieved the highest specificity
%fPSA, free-to-total PSA ratio ; NA, not applicable; %p2PSA, percentage of p2PSA to free PSA .

Table 5 Validation and calibration for included models

Author and year	Validation	Calibration
Carlson <i>et al</i> , 1998 ⁶¹	External validation on additional data set consisting of 525 patients	Calibration plot
Babaian <i>et al</i> , 2000 ⁶⁴	Cross-validation and separate data set of 151	NA
Jansen <i>et al</i> , 2010 ⁶⁵	NA	NA
Hill <i>et al</i> , 2013 ⁶²	NA	NA
Lazzeri <i>et al</i> , 2013 ⁶³	Internal validation using 200 bootstrap resamples	Internal calibration using the Hosmer-Lemeshow goodness-of-fit test

power. Four studies did not report missing data or how they handled it.^{61 63–65} The remaining study used complete case analysis and excluded patients with missing data on laboratory biomarkers (n=75).⁶² The PROBAST guidelines state that in a prediction model study where any risk of bias and applicability is low in all four domains, a regrading to high risk of bias should be considered when the study did not validate the model externally.⁶⁰ Thus, although the quality assessment for the Lazzeri study⁶³ was graded low risk in all the four domains, since the study did not report any external validation of the model, the assessment of the study has been regraded to high risk of bias according to the PROBAST criteria. A full quality assessment for all studies is presented in [table 6](#).

DISCUSSION

Despite the large number of PCa risk prediction models, the majority still include clinical inputs and/or more costly biomarker or genetic panels; few low cost models exist that do not include specialist clinical input or more expensive further testing that limits their use for population wide assessments. To our knowledge, this is the first study to examine risk prediction models for PCa that are low cost and do not include clinical and genetic variables, and are based on single time-point assessment.

Our study identified five unique models that met the set criteria. The Carlson model⁶¹ has the largest population (3773 patients) when compared with the other four studies. Although they reported an 11% increase in specificity, they did not report AUC predictive estimates. It has been acknowledged that sensitivity and specificity results

are dependent on the prevalence of the disease. Hence, the comparison between populations where the PCa prevalence may vary (especially in early detection) will be difficult.³³ More importantly, by not reporting the AUC estimate, the model raises some doubts regarding the reliability of the model and its implementation.³³ It will also make comparison to other models not applicable.⁶⁷

Babaian⁶⁴ developed an algorithm and compared the performance of the ANN to PSA density (total PSA divided by prostate volume) (PSAD), %fPSA and transition zone density (PSAD-TZ). Their ANN demonstrated a significant increase of model specificity that reached 51% when sensitivity was held at 95%. This was better than the specificity value of each individual variable such as %fPSA (10%), PSAD (39%), and PSAD-TZ (22%). In terms of AUC, the ANN achieved a moderate accuracy (0.74), being the second highest among all studies included. However, the ANN model did not show significant improvement when compared with a model fitted with only individual variables (AUC for %fPSA=0.64, PSAD=0.74 and PSAD-TZ=0.75). They included a number of uncommon pre-biopsy inputs into their algorithm such as prostatic acid phosphatase and creatine kinase.⁶⁸ Furthermore, they used a tight PSA range (2.5 to 4.0 ng/mL) which meant that their model may be less suitable for patients with PSA level below or above that range, thus limiting its generalisability.

The study by Jansen and colleagues⁶⁵ demonstrated that adding p2PSA to the base model of PSA and fPSA significantly enhanced the PCa predictive value and specificity. The association and added value of p2PSA in the

Table 6 Quality assessment for ROB and applicability concern for included studies

Study	ROB*				Applicability			Overall	
	Participants	Predictors	Outcome	Analysis	Participants	Predictors	Outcome	ROB	Applicability
Carlson <i>et al</i> ⁶¹	+	+	-	-	+	+	-	-	-
Babaian <i>et al</i> ⁶⁴	+	+	+	-	-	+	+	-	-
Jansen <i>et al</i> ⁶⁵	-	-	-	-	+	-	-	-	-
Hill <i>et al</i> ⁶²	-	+	+	-	-	+	+	-	-
Lazzeri <i>et al</i> ⁶³	+	+	+	+	+	+	+	-	+

+ indicates a ROB or applicability; - indicates a high ROB or applicability.

*ROB, risk of bias.

prediction and detection of PCa have been reported by several other studies.^{16 69–72} Jansen⁶⁵ showed that p2PSA has no clear association with aggressive PCa. However, the base model that includes p2PSA had the highest clinical significance in correlation to pathologic Gleason score with a p value of 0.008 compared with %fPSA and PHI (p value 0.01 and 0.02, respectively). Although they used archived blood samples and retrospective analysis, the results were similar to a prospective study of 268 patients.¹⁶

Hill⁶² used a case-control study to evaluate several laboratory biomarkers. They found HGB, RBC, haematuria, creatinine, PSA, age, MCV and ethnicity ('being Black') were statistically significantly associated in the first method ($p < 0.05$). In the second method, HGB, RBC, creatinine, PSA, age and MCV were found to be statistically significantly correlated ($p < 0.001$) with PCa. However, since this study was designed as a case-control study, it would have been more prone to uncontrolled confounding and selection bias. Moreover, the type of screening protocols used in Veterans' Administrations may vary to those conducted in other healthcare systems; therefore, the results may not be applicable to other populations. Furthermore, patients with a PSA level < 4.0 ng/mL have not been investigated, and thus, the performance of the models are unknown for individuals in this group.

Lazzeri *et al.*⁶³ in a European multicentre study have evaluated similar biomarkers as in Jansen study with the same PSA range 2 to 10 ng/mL prospectively. They found no difference in both %p2PSA and PHI as individual PCa predictors with AUC of 0.67 (95% CI 0.64 to 0.71). However, the base model (consisting of PSA, fPSA and %fPSA) that also included either p2PSA or PHI outperformed the base model alone and the base model including %p2PSA. In the analysis, the additive value of both p2PSA and PHI is 0.064 and 0.056 for %p2PSA for predicting the risk of PCa. These additive values increased to 0.076 for both p2PSA and PHI, and 0.073 for %p2PSA in predicting Gleason scores ≥ 7 for the disease. The usefulness of PHI in improving the predictive accuracy of PCa over total and free PSA has been confirmed and reported by several studies.^{16 72–75}

In general, only one study has validated their model externally,⁶¹ whereas the remaining studies were either validated internally^{63 64} or did not report any validation methods.^{62 65} Prediction models may not be equally applicable to all data sets as patients' characteristics may vary.^{20 76} As a result, the generalisability of a model might be poor when it used in populations other than that used in building the model. Therefore, external validation should be conducted before applying any new model into general practice.^{77 78}

Another key performance measure of any model that needs careful evaluation is calibration. A calibration plot with a calibration slope is more preferable than the Hosmer-Lemeshow test; it has been acknowledged that evaluating a good and well calibrated model based on a large data set can still fail the Hosmer-Lemeshow test. In contrast, when evaluating a poorly calibrated model with

a small data set it can still pass the Hosmer-Lemeshow test.⁷⁹ In our analysis, three studies fail to report the calibration of the model^{62 64 65} while the Carlson study⁶¹ used a calibration plot and Lazzeri⁶³ used the Hosmer-Lemeshow test. Excluding calibration from the majority of models may explain why some models are not currently used in practice.⁷⁹

With regard to biopsy cores, only two studies used extended biopsy cores. Babaian⁶⁴ used an 11-core multisite biopsy, whereas Lazzeri⁶³ used at least 12 biopsy cores. Moreover, two studies used six cores biopsy in their model.^{61 65} The use of six-core biopsy has been criticised as not being adequate in detecting PCa⁸⁰ and that models developed using sextant biopsy are less accurate than when a 10-core biopsy is used.⁷⁶ As a result, the European guideline for clinical PCa recommended an extended biopsy as standard practice for PCa detection.⁸¹

It is worth noting that all five reviewed models performed better than just PSA alone. However, none of them has both high specificity and sensitivity. The level of sensitivity has been increased, and despite enhancement in the specificity, it is still considered low. Specificity is crucial when it comes to being used in a population setting as men without PCa should be ruled out as much as possible from further invasive engagement with the health system.

Our review therefore suggests that none of the reviewed models provide an ideal performance in predicting PCa with high sensitivity and high specificity. It is particularly important when considering the application of PCa risk prediction at the population level that the tool used should be able to both detect the outcome and filter out people with no disease. As there is robust evidence suggesting the clinical relevance of PSA range to the detection of PCa differs across age groups,^{82–84} any future model should consider PSA threshold in relation to a specific age range. Risk prediction models for PCa should therefore take account of age.

Out of the five reviewed models, the Lazzeri model 2, has the greatest potential to be implemented in primary care. It achieved the least risk of bias and had fair discrimination for both any and aggressive PCa. It also had the largest improvement in discrimination performance compared with PSA alone. Moreover, except for the p2PSA that requires a specific assay, the included variables are common and easy to measure. However, before it could be used, the model requires to be validated externally.

Comparison with other studies

To our knowledge, three systematic reviews of PCa prediction tools have been published.^{20 26 27} In the Louie *et al* review, risk models were included that were externally validated in at least five study populations for the purpose of meta-analysis and only six studies were included in their analysis. Furthermore, all the studies included incorporated clinical tests such as DRE and/or TRUS-PV.²⁶ Schroder and Kattan²⁰ reviewed models that were

built to predict the likelihood of having a positive prostate biopsy for cancer. However, it appears that they also included models where patients had a previous negative biopsy. As such, some of the models included variables related to biopsy results and cores. The review by Shariat and colleagues examined different types of predictive tools.²⁷ They explored tools that predict PCa on initial and repeat biopsy, pathologic stages, biochemical recurrence after radical prostatectomy, metastasis, survival and life expectancy. Similarly, virtually all the prediction tools that were based on initial biopsy included variables based on invasive procedures.

Strengths and limitations of this study

This report is the first to review risk prediction tools for PCa that can be used in primary care and community settings. Any prediction model should therefore be simple to use, based on non-invasive tests, be feasible at a population level and at low cost. We carried out an extensive data extraction relating to important features and characteristics for each study included, such as modelling method, source of data, sample size, variables, discrimination, validation and cancer detection rate. We have also followed PRISMA guidelines for identifying eligible articles as well as for reporting this study. In addition, the PROBAST was adopted to assess the quality and risk of bias for each prediction model.

Our study has some limitations. Our aim was to identify prediction models that have the potential to be implemented in a primary care or community setting, and consequently our search strategy was to retrieve relevant studies for this specific purpose. Furthermore, we excluded articles that were not published in English or did not have an abstract. Moreover, only two databases were searched, besides manual search, to retrieve original studies.

A previous systematic review suggested that the majority of relevant studies could be identified through a manual search of articles reference lists instead of a database alone.²⁰ We identified four eligible studies using this approach. Given the small number of models identified by the approach we followed, that can be applied in primary care settings compared with the large number relating to wider existing models, it is unlikely that we have not included any study that would affect the results of our review.

Implications and future research

It is now accepted that the PSA test and its derivatives have some limitations for detecting PCa as defined by subsequent biopsy.⁸⁵ As a consequence, a considerable number of PCa prediction models have been built to improve prediction accuracy. This has resulted in a plethora of PCa risk prediction tools, with to date more than 100 models described.^{86 87} There is evidence that some of these models show benefit and have better performance over just PSA measurement alone.²⁰ It also has been demonstrated that some of these models out-performed clinical

experts in predicting PCa.^{28 29} Although such models are not designed to replace specialist clinical judgement or patient preferences,^{76 85 88} they can help in patient counselling and aid clinicians to decide whether a prostate biopsy should be taken or not.^{77 88 89}

Given the small number of risk prediction models for PCa that do not incorporate clinical or genetic tests, the discrimination of these reviewed models ranged between poor to moderate (AUC range ~0.65 to ~0.75); in addition there were some issues relating to their study design and analysis raises the risk of bias. Consequently, none of these models could be currently recommended for use in a primary care and community healthcare setting. Several guidelines are against using PSA test based screening for PCa; the US Preventive Services task force, the Canadian task force on preventive health and the American College of Preventive Medicine do not currently recommend PSA-based testing due to insufficient evidence.^{90–92} This has made it difficult, so far, to convince policymakers to adopt PCa screening programme.

The first guideline of PROSTATE CANCER UK states, “In the future, health professionals should look at a man’s PSA level alongside other known risk factors as part of a risk assessment tool, when one becomes available.”⁹³ However, the vast majority of the current PCa risk prediction models are not suitable for routine use as they include clinical and genetic tests and are not validated externally in other cohorts. Therefore, the main challenge in the UK, remains to develop a risk prediction tool that is reliable, cheap, is applicable for as wide an ethnicity as possible, and, most importantly, is easy to use and can be implemented at a primary care level.⁹⁴

The value of such risk tools is that they will help to stratify men at high risk of developing PCa earlier so that they have appropriate management and/or surveillance programme as early as possible and, therefore, may fit into the clinical pathway. Such tools should help physicians have a better understanding of the risk for this disease and simplify the procedures and discussions with patients when recommending further specialist-led investigations such as DRE and/or MRI where a decision on whether a biopsy should or not perform is concluded. Furthermore, using the appropriate risk prediction tool will avoid men from undergoing inappropriate further and frequent testing.⁹⁴ This will reduce any associated costs of inappropriate tests and decrease the burden on healthcare delivery systems.

It is crucial to address these issues by identifying all possible risk factors for PCa that are non-clinical, non-genetic, and easy to use and interpret. There remains a pressing need to develop a risk prediction tool in the future using all appropriate factors (potentially also including genetics once there is infrastructure in place for genetic testing in the primary care and the cost comes down) into a robust multivariable analysis and validate the model externally to eliminate applicability and generalisability concerns. Only when this is achieved will it be

possible to introduce a PCa screening programme fit for purpose.

CONCLUSION

There is a paucity of suitable low-cost risk models that incorporate non-clinical, non-genetic inputs and which can be used at a primary care level and in other community health services. Existing models have limitations reflecting both study design and reporting performance measures. Future research should take into account these key issues and explore other risk factors for incorporation into further models.

Acknowledgements MA is supported by Kuwait's Ministry of Health.

Contributors MA, AL and KM were involved in study conception, idea and design. Data acquisition and extraction was obtained by MA. Data synthesis and interpretation was carried out by MA, AL and WO. MA and AL drafted the manuscript. All authors approved the final version of manuscript. KM is the study guarantor.

Funding KM and AL are supported by the NIHR Manchester Biomedical Research Centre and by the ICEP (This work was also supported by CRUK (grant number C18281/A19169)). KM is also supported by the Allan Turing Institute.

Competing interests All authors have completed the ICMJE uniform disclosure form at http://www.icmje.org/coi_disclosure.pdf and declare: no support from any organisation for the submitted work; no financial relationships with any organisations that might have an interest in the submitted work in the previous 3 years; no other relationships or activities that could appear to have influenced the submitted work.

Patient and public involvement Patients and/or the public were not involved in the design, or conduct, or reporting, or dissemination plans of this research.

Patient consent for publication Not required.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available upon reasonable request. Data extraction sheet and the PROBAST assessment form for risk of bias is available upon request.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution 4.0 Unported (CC BY 4.0) license, which permits others to copy, redistribute, remix, transform and build upon this work for any purpose, provided the original work is properly cited, a link to the licence is given, and indication of whether changes were made. See: <https://creativecommons.org/licenses/by/4.0/>.

ORCID iDs

Aritaya Lophatananon <http://orcid.org/0000-0003-0550-4657>

Kenneth Muir <http://orcid.org/0000-0001-6429-988X>

REFERENCES

- Bray F, Ferlay J, Soerjomataram I, et al. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* 2018;68:394–424.
- Cancer Research UK. Prostate cancer statistics, 2017. Available: <https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/prostate-cancer#heading-One>
- Center MM, Jemal A, Lortet-Tieulent J, et al. International variation in prostate cancer incidence and mortality rates. *Eur Urol* 2012;61:1079–92.
- Zhou CK, Check DP, Lortet-Tieulent J, et al. Prostate cancer incidence in 43 populations worldwide: an analysis of time trends overall and by age group. *Int J Cancer* 2016;138:1388–400.
- Dijkstra S, Mulders PFA, Schalken JA. Clinical use of novel urine and blood based prostate cancer biomarkers: a review. *Clin Biochem* 2014;47:889–96.
- Fontanella P, Benecchi L, Grasso A, et al. Decision-Making tools in prostate cancer: from risk grouping to nomograms. *Minerva Urol Nefrol* 2017;69:556–66.
- Stephan C, Ralla B, Jung K. Prostate-Specific antigen and other serum and urine markers in prostate cancer. *Biochim Biophys Acta* 2014;1846:99–112.
- Thompson IM, Pauler DK, Goodman PJ, et al. Prevalence of prostate cancer among men with a prostate-specific antigen level < or =4.0 ng per milliliter. *N Engl J Med* 2004;350:2239–46.
- Pereira-Azevedo N, Verbeek JFM, Nieboer D, et al. Head-To-Head comparison of prostate cancer risk calculators predicting biopsy outcome. *Transl Androl Urol* 2018;7:18–26.
- Stephan C, Cammann H, Meyer H-A, et al. Psa and new biomarkers within multivariate models to improve early detection of prostate cancer. *Cancer Lett* 2007;249:18–29.
- Polascik TJ, Oesterling JE, Partin AW. Prostate specific antigen: a decade of discovery--what we have learned and where we are going. *J Urol* 1999;162:293–306.
- Stamey TA, Caldwell M, McNeal JE, et al. The prostate specific antigen era in the United States is over for prostate cancer: what happened in the last 20 years? *J Urol* 2004;172:1297–301.
- Bunting PS. A guide to the interpretation of serum prostate specific antigen levels. *Clin Biochem* 1995;28:221–41.
- Catalona WJ, Partin AW, Sanda MG, et al. A multi-center study of [–2] pro-prostate-specific antigen (PSA) in combination with PSA and free PSA for prostate cancer detection in the 2.0 to 10.0 ng/mL PSA range. *J Urol* 2011;185:1650–5.
- Mistry K, Cable G. Meta-Analysis of prostate-specific antigen and digital rectal examination as screening tests for prostate carcinoma. *J Am Board Fam Pract* 2003;16:95–101.
- Guazzoni G, Nava L, Lazzeri M, et al. Prostate-Specific antigen (PSA) isoform p2PSA significantly improves the prediction of prostate cancer at initial extended prostate biopsies in patients with total PSA between 2.0 and 10 ng/ml: results of a prospective study in a clinical setting. *Eur Urol* 2011;60:214–22.
- Heijnsdijk EAM, de Konderen A, Wever EM, et al. Overdetection, overtreatment and costs in prostate-specific antigen screening for prostate cancer. *Br J Cancer* 2009;101:1833–8.
- Roobol MJ. Is prostate cancer screening bad or good? summary of a debate at the innovation in urology meeting, September 17–19, 2010, Milan, Italy. *Eur Urol* 2011;59:359–62.
- Mottet N, Bellmunt J, Bolla M, et al. EAU-ESTRO-SIOG guidelines on prostate cancer. Part 1: screening, diagnosis, and local treatment with curative intent. *Eur Urol* 2017;71:618–29.
- Schröder F, Kattan MW. The comparability of models for predicting the risk of a positive prostate biopsy with prostate-specific antigen alone: a systematic review. *Eur Urol* 2008;54:274–90.
- Ross PL, Scardino PT, Kattan MW. A catalog of prostate cancer nomograms. *J Urol* 2001;165:1562–8.
- Shariat SF, Karakiewicz PI, Suardi N, et al. Comparison of nomograms with other methods for predicting outcomes in prostate cancer: a critical analysis of the literature. *Clin Cancer Res* 2008;14:4400–7.
- Hu X, Cammann H, Meyer H-A, et al. Artificial neural networks and prostate cancer--tools for diagnosis and management. *Nat Rev Urol* 2013;10:174–82.
- Caras RJ, Sterbis JR. Prostate cancer nomograms: a review of their use in cancer detection and treatment. *Curr Urol Rep* 2014;15:391.
- Iremashvili V, Soloway MS, Pelaez L, et al. Comparative validation of nomograms predicting clinically insignificant prostate cancer. *Urology* 2013;81:1202–8.
- Louie KS, Seigneurin A, Cathcart P, et al. Do prostate cancer risk models improve the predictive accuracy of PSA screening? A meta-analysis. *Ann Oncol* 2015;26:848–64.
- Shariat SF, Kattan MW, Vickers AJ, et al. Critical review of prostate cancer predictive tools. *Future Oncol* 2009;5:1555–84.
- Ross PL, Gerigk C, Gonen M, et al. Comparisons of nomograms and urologists' predictions in prostate cancer. *Semin Urol Oncol* 2002;20:82–8.
- Walz J, Gallina A, Perrotte P, et al. Clinicians are poor raters of life-expectancy before radical prostatectomy or definitive radiotherapy for localized prostate cancer. *BJU Int* 2007;100:1254–8.
- Ouzaid I, Yates DR, Hupertan V, et al. A direct comparison of the diagnostic accuracy of three prostate cancer nomograms designed to predict the likelihood of a positive initial transrectal biopsy. *Prostate* 2012;72:1200–6.
- Eastham JA, May R, Robertson JL, et al. Development of a nomogram that predicts the probability of a positive prostate biopsy in men with an abnormal digital rectal examination and a prostate-specific antigen between 0 and 4 ng/mL. *Urology* 1999;54:709–13.
- Garzotto M, Hudson RG, Peters L, et al. Predictive modeling for the presence of prostate carcinoma using clinical, laboratory, and ultrasound parameters in patients with prostate specific antigen levels < or = 10 ng/mL. *Cancer* 2003;98:1417–22.

- 33 Karakiewicz PI, Benayoun S, Kattan MW, *et al.* Development and validation of a nomogram predicting the outcome of prostate biopsy based on patient age, digital rectal examination and serum prostate specific antigen. *J Urol* 2005;173:1930–4.
- 34 Kranske R, Roobol M, Schröder FH. A graphical device to represent the outcomes of a logistic regression analysis. *Prostate* 2008;68:1674–80.
- 35 Thompson IM, Ankerst DP, Chi C, *et al.* Assessing prostate cancer risk: results from the prostate cancer prevention trial. *J Natl Cancer Inst* 2006;98:529–34.
- 36 Schroder FH, Kruger AB, Rietbergen J, *et al.* Evaluation of the digital rectal examination as a screening test for prostate cancer. *J Natl Cancer Inst* 1998;90:1817–23.
- 37 Vis AN, Hoedemaeker RF, Roobol M, *et al.* Tumor characteristics in screening for prostate cancer with and without rectal examination as an initial screening test at low PSA (0.0–3.9 ng/ml). *Prostate* 2001;47:252–61.
- 38 Smith DS, Catalona WJ. Interexaminer variability of digital rectal examination in detecting prostate cancer. *Urology* 1995;45:70–4.
- 39 Barocas DA, Cowan JE, Smith JA, *et al.* What percentage of patients with newly diagnosed carcinoma of the prostate are candidates for surveillance? an analysis of the CaPSURE database. , 2008: 180, 1330–5.
- 40 Vis AN, Kranske R, Roobol M, *et al.* Serendipity in detecting disease in low prostate-specific antigen ranges. *BJU Int* 2002;89:384–9.
- 41 Littrup PJ, Bailey SE. Prostate cancer: the role of transrectal ultrasound and its impact on cancer detection and management. *Radiol Clin North Am* 2000;38:87–113.
- 42 Gerber GS, Chodak GW. Routine screening for cancer of the prostate. *JNCI: Journal of the National Cancer Institute* 1991;83:329–35.
- 43 Gelfand DE, Parzuchowski J, Cort M, *et al.* Digital rectal examinations and prostate cancer screening: attitudes of African American men. *Oncol Nurs Forum* 1995;22:1253–8.
- 44 Myers RE, Wolf TA, McKee L, *et al.* Factors associated with intention to undergo annual prostate cancer screening among African American men in Philadelphia. *Cancer* 1996;78:471–9.
- 45 Consedine NS, Horton D, Ungar T, *et al.* Fear, knowledge, and efficacy beliefs differentially predict the frequency of digital rectal examination versus prostate specific antigen screening in ethnically diverse samples of older men. *Am J Mens Health* 2007;1:29–43.
- 46 Consedine NS, Morgenstern AH, Kudadjie-Gyamfi E, *et al.* Prostate cancer screening behavior in men from seven ethnic groups: the fear factor. *Cancer Epidemiol Biomarkers Prev* 2006;15:228–37.
- 47 Ukoli FA, Patel K, Hargreaves M, *et al.* A tailored prostate cancer education intervention for low-income African Americans: impact on knowledge and screening. *J Health Care Poor Underserved* 2013;24:311–31.
- 48 Ogunsanya ME, Brown CM, Odedina FT, *et al.* Beliefs regarding prostate cancer screening among black males aged 18 to 40 years. *Am J Mens Health* 2017;11:41–53.
- 49 Clarke-Tasker VA, Wade R. What we thought we knew: African American males' perceptions of prostate cancer and screening methods. *Abnf J* 2002;13:56.
- 50 Romero FR, Romero AW, Brenny Filho T, *et al.* Patients' perceptions of pain and discomfort during digital rectal exam for prostate cancer screening. *Arch Esp Urol* 2008;61:850–4.
- 51 Essink-Bot ML, de Koning HJ, Nijis HG, *et al.* Short-Term effects of population-based screening for prostate cancer on health-related quality of life. *J Natl Cancer Inst* 1998;90:925–31.
- 52 Appleton L, Wyatt D, Perkins E, *et al.* The impact of prostate cancer on men's everyday life. *Eur J Cancer Care* 2015;24:71–84.
- 53 Lee DJ, Consedine NS, Spencer BA. Barriers and facilitators to digital rectal examination screening among African-American and African-Caribbean men. *Urology* 2011;77:891–8.
- 54 Nagler HM, Gerber EW, Homel P, *et al.* Digital rectal examination is barrier to population-based prostate cancer screening. *Urology* 2005;65:1137–40.
- 55 Moher D, Liberati A, Tetzlaff J, *et al.* Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med* 2009;6:e1000097.
- 56 Park SH, Goo JM, Jo C-H. Receiver operating characteristic (ROC) curve: practical review for radiologists. *Korean J Radiol* 2004;5:11–18.
- 57 Han K, Song K, Choi BW, *et al.* How to develop, validate, and compare clinical prediction models involving radiological parameters: study design and statistical methods. *Korean J Radiol* 2016;17:339–50.
- 58 Steyerberg EW. *Clinical prediction models: a practical approach to development, validation, and updating.* Springer Science & Business Media, 2008.
- 59 Pavlou M, Ambler G, Seaman SR, *et al.* How to develop a more accurate risk prediction model when there are few events. *BMJ* 2015;351:h3868.
- 60 Wolff RF, Moons KGM, Riley RD, *et al.* PROBAST: a tool to assess the risk of bias and applicability of prediction model studies. *Ann Intern Med* 2019;170:51–8.
- 61 Carlson GD, Calvanese CB, Partin AW. An algorithm combining age, total prostate-specific antigen (PSA), and percent free PSA to predict prostate cancer: results on 4298 cases. *Urology* 1998;52:455–61.
- 62 Hill OT, Mason TJ, Schwartz SW, *et al.* Improving prostate cancer detection in veterans through the development of a clinical decision rule for prostate biopsy. *BMC Urol* 2013;13:6.
- 63 Lazzeri M, Haese A, de la Taille A, *et al.* Serum isoform [-2]proPSA derivatives significantly improve prediction of prostate cancer at initial biopsy in a total PSA range of 2–10 ng/ml: a multicentric European study. *Eur Urol* 2013;63:986–94.
- 64 Babaian RJ, Fritsche H, Ayala A, *et al.* Performance of a neural network in detecting prostate cancer in the prostate-specific antigen reflex range of 2.5 to 4.0 ng/mL. *Urology* 2000;56:1000–6.
- 65 Jansen FH, van Schaik RHN, Kurstjens J, *et al.* Prostate-Specific antigen (PSA) isoform p2PSA in combination with total PSA and free PSA improves diagnostic accuracy in prostate cancer detection. *Eur Urol* 2010;57:921–7.
- 66 Moons KGM, de Groot JAH, Bouwmeester W, *et al.* Critical appraisal and data extraction for systematic reviews of prediction modelling studies: the charms checklist. *PLoS Med* 2014;11:e1001744.
- 67 Porter CR, O'Donnell C, Crawford ED, *et al.* Predicting the outcome of prostate biopsy in a racially diverse population: a prospective study. *Urology* 2002;60:831–5.
- 68 Porter CR, Gamito EJ, Crawford ED, *et al.* Model to predict prostate biopsy outcome in large screening population with independent validation in referral setting. *Urology* 2005;65:937–41.
- 69 Catalona WJ, Bartsch G, Rittenhouse HG, *et al.* Serum pro-prostate specific antigen preferentially detects aggressive prostate cancers in men with 2 to 4 ng/ml prostate specific antigen. *J Urol* 2004;171:2239–44.
- 70 Sokoll LJ, Sanda MG, Feng Z, *et al.* A prospective, multicenter, National Cancer Institute Early Detection Research Network study of [-2]proPSA: improving prostate cancer detection and correlating with cancer aggressiveness. *Cancer Epidemiol Biomarkers Prev* 2010;19:1193–200.
- 71 Sokoll LJ, Wang Y, Feng Z, *et al.* [-2]proenzyme prostate specific antigen for prostate cancer detection: a national cancer institute early detection research network validation study. *J Urol* 2008;180:539–43.
- 72 Stephan C, Vincendeau S, Houlgatte A, *et al.* Multicenter evaluation of [-2]proprostate-specific antigen and the prostate health index for detecting prostate cancer. *Clin Chem* 2013;59:306–14.
- 73 Le BV, Griffin CR, Loeb S, *et al.* [-2]Proenzyme prostate specific antigen is more accurate than total and free prostate specific antigen in differentiating prostate cancer from benign disease in a prospective prostate cancer screening study. *J Urol* 2010;183:1355–9.
- 74 Loeb S, Sokoll LJ, Broyles DL, *et al.* Prospective multicenter evaluation of the Beckman Coulter prostate health index using who calibration. *J Urol* 2013;189:1702–6.
- 75 Lughezzani G, Lazzeri M, Haese A, *et al.* Multicenter European external validation of a prostate health index-based nomogram for predicting prostate cancer at extended biopsy. *Eur Urol* 2014;66:906–12.
- 76 Chun FK-H, Briganti A, Graefen M, *et al.* Development and external validation of an extended 10-core biopsy nomogram. *Eur Urol* 2007;52:436–45.
- 77 Herman MP, Dorsey P, John M, *et al.* Techniques and predictive models to improve prostate cancer detection. *Cancer* 2009;115:3085–99.
- 78 Trottier G, Roobol MJ, Lawrentschuk N, *et al.* Comparison of risk calculators from the prostate cancer prevention trial and the European randomized study of screening for prostate cancer in a contemporary Canadian cohort. *BJU Int* 2011;108:E237–44.
- 79 Collins GS, de Groot JA, Dutton S, *et al.* External validation of multivariable prediction models: a systematic review of methodological conduct and reporting. *BMC Med Res Methodol* 2014;14:40.
- 80 Miyake H, Kurahashi T, Takenaka A, *et al.* Improved accuracy for predicting the Gleason score of prostate cancer by increasing the number of transrectal biopsy cores. *Urol Int* 2007;79:302–6.
- 81 Aus G, Abbou CC, Bolla M, *et al.* EAU guidelines on prostate cancer. *Eur Urol* 2005;48:546–51.
- 82 Vickers AJ, Ulmert D, Serio AM, *et al.* The predictive value of prostate cancer biomarkers depends on age and time to diagnosis: towards a biologically-based screening strategy. *Int J Cancer* 2007;121:2212–7.



- 83 Steuber T, O'Brien MF, Lilja H. Serum markers for prostate cancer: a rational approach to the literature. *Eur Urol* 2008;54:31–40.
- 84 Carlsson S, Assel M, Sjoberg D, *et al.* Influence of blood prostate specific antigen levels at age 60 on benefits and harms of prostate cancer screening: population based cohort study. *BMJ* 2014;348:g2296.
- 85 Cavadas V, Osório L, Sabell F, *et al.* Prostate cancer prevention trial and European randomized study of screening for prostate cancer risk calculators: a performance comparison in a contemporary screened cohort. *Eur Urol* 2010;58:551–8.
- 86 Vickers AJ. Prediction models: revolutionary in principle, but do they do more good than harm? *J Clin Oncol* 2011;29:2951–2.
- 87 Shariat SF, Karakiewicz PI, Margulis V, *et al.* Inventory of prostate cancer predictive tools. *Curr Opin Urol* 2008;18:279–96.
- 88 Kawakami S, Numao N, Okubo Y, *et al.* Development, validation, and head-to-head comparison of logistic regression-based nomograms and artificial neural network models predicting prostate cancer on initial extended biopsy. *Eur Urol* 2008;54:601–11.
- 89 Hernandez DJ, Han M, Humphreys EB, *et al.* Predicting the outcome of prostate biopsy: comparison of a novel logistic regression-based model, the prostate cancer risk calculator, and prostate-specific antigen level alone. *BJU Int* 2009;103:609–14.
- 90 Moyer VA, U.S. Preventive Services Task Force. Screening for prostate cancer: U.S. preventive services Task force recommendation statement. *Ann Intern Med* 2012;157:120–34.
- 91 Bell N, Connor Gorber S, Shane A, *et al.* Recommendations on screening for prostate cancer with the prostate-specific antigen test. *CMAJ* 2014;186:1225–34.
- 92 Lim LS, Sherin K, ACPM Prevention Practice Committee. Screening for prostate cancer in U.S. men AcpM position statement on preventive practice. *Am J Prev Med* 2008;34:164–70.
- 93 Prostate Cancer UK. Consensus statements on PSA testing in asymptomatic men in the UK - information for health professionals, 2016. Available: <https://prostatecanceruk.org/about-us/projects-and-policies/consensus-on-psa-testing> [Accessed 20 Aug 2019].
- 94 Frame I, Cant S. Current challenges in prostate cancer: an interview with prostate cancer UK. *BMC Med* 2015;13:166.