

# BMJ Open

BMJ Open is committed to open peer review. As part of this commitment we make the peer review history of every article we publish publicly available.

When an article is published we post the peer reviewers' comments and the authors' responses online. We also post the versions of the paper that were used during peer review. These are the versions that the peer review comments apply to.

The versions of the paper that follow are the versions that were submitted during the peer review process. They are not the versions of record or the final published versions. They should not be cited or distributed as the published version of this manuscript.

BMJ Open is an open access journal and the full, final, typeset and author-corrected version of record of the manuscript is available on our site with no access controls, subscription charges or pay-per-view fees (<http://bmjopen.bmj.com>).

If you have any questions on BMJ Open's open peer review process please email [info.bmjopen@bmj.com](mailto:info.bmjopen@bmj.com)

# BMJ Open

## Non-clinical and non-genetic risk prediction models for prostate cancer: a systematic review

Journal:	<i>BMJ Open</i>
Manuscript ID	bmjopen-2019-034661
Article Type:	Original research
Date Submitted by the Author:	02-Oct-2019
Complete List of Authors:	Aladwani , Mohammad; The University of Manchester, Division of Population Health, Health Services Research and Primary Care School of Health Sciences Faculty of Biology, Medicine and Health Lophatananon, Artitaya ; The University of Manchester, Division of Population Health, Health Services Research and Primary Care School of Health Sciences Faculty of Biology, Medicine and Health Ollier, William; The University of Manchester, Division of Population Health, Health Services Research and Primary Care School of Health Sciences Faculty of Biology, Medicine and Health; Manchester Metropolitan University Faculty of Science and Engineering, School of Healthcare Science Muir, Kenneth; The University of Manchester, Division of Population Health, Health Services Research and Primary Care School of Health Sciences Faculty of Biology, Medicine and Health
Keywords:	EPIDEMIOLOGY, UROLOGY, Prostate disease < UROLOGY

SCHOLARONE™  
Manuscripts



I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our [licence](#).

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which [Creative Commons](#) licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

1  
2  
3  
4 **Full title: Non-clinical and non-genetic risk prediction models for**  
5 **prostate cancer: a systematic review**  
6  
7

8 **Institutional addresses and e-mail addresses of the contributing authors:**  
9

10 Dr Mohammad Aladwani

11  
12 Division of Population Health, Health Services Research and Primary Care, School of  
13 Health Sciences Faculty of Biology, Medicine and Health, The University of Manchester,  
14 Manchester, UK, M13 9PT  
15  
16

17  
18 E-mail: [mohammad.al-adwani@postgrad.manchester.ac.uk](mailto:mohammad.al-adwani@postgrad.manchester.ac.uk)  
19

20 Dr Artitaya Lophatananon

21  
22 Division of Population Health, Health Services Research and Primary Care, School of  
23 Health Sciences Faculty of Biology, Medicine and Health, The University of Manchester,  
24 Manchester, UK, M13 9PT  
25  
26

27  
28 E-mail: [Artitaya.lophananon@.manchester.ac.uk](mailto:Artitaya.lophananon@.manchester.ac.uk)  
29

30 Professor William Ollier,

31  
32 Division of Population Health, Health Services Research and Primary Care, School of  
33 Health Sciences Faculty of Biology, Medicine and Health, The University of Manchester,  
34 Manchester, UK, M13 9PT;  
35  
36

37  
38 School of Healthcare Science, Faculty of Science and Engineering, Manchester  
39 Metropolitan University, John Dalton Building Manchester M1 5GD UK  
40  
41

42  
43 E-mail: [Bill.Ollier@manchester.ac.uk](mailto:Bill.Ollier@manchester.ac.uk)  
44

45 **Address and contact details of the corresponding author:**  
46

47 Professor Kenneth Muir

48  
49 Division of Population Health, Health Services Research and Primary Care,  
50

51 School of Health Sciences Faculty of Biology, Medicine and Health,  
52

53 The University of Manchester, Manchester,  
54

55  
56 UK, M13 9PT  
57  
58  
59  
60

1  
2  
3 E-mail: [Kenneth.muir@manchester.ac.uk](mailto:Kenneth.muir@manchester.ac.uk)  
4

5 Telephone: +44-161-2785677  
6  
7

8  
9  
10 Key words: Prostate cancer, Prostate specific antigen, PSA, Prostate cancer screening,  
11 Risk Prediction, Risk tools, Prediction models  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

For peer review only

## Abstract

### Objective

To identify and compare existing risk prediction models for prostate cancer (PCa) that have a potential use in a primary care and community health hub settings.

### Design

Systematic review.

### Data sources

MEDLINE and EMBASE databases combined up to, and including, January 2019.

### Eligibility

Studies were considered relevant based on the following criteria; (i) presenting an evaluation of prospective risk of PCa at initial biopsy, (ii) patients having no history of PCa, (iii) studies not incorporating an invasive clinical assessment or a predictive genetic test, (iv) inclusion of at least two variables with PSA being one of them, and (v) studies reporting a measure of predictive performance.

### Results

An initial search yielded 109 studies, of which five met the set criteria. Four studies were cohort-based and one was a case-control study. One study used an artificial neural network (ANN) to build the prediction model, and the remaining four studies applied logistic regression methodology. The sample sizes analysed within these studies ranged from 151 to 3,773 subjects. After prostate specific antigen (PSA), the most common model inputs were free PSA (fPSA), free-to-total PSA ratio (%fPSA) and age. PCa detection rate was between 20.6% and 55.8%. 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 just 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 used in the primary health care / community health centres. The key elements for any model considered to

1  
2  
3 be of use at a community level are cost, feasibility and its performance for predicting  
4 PCa. Of all the models we investigated, issues around model development were  
5 identified. Despite increased sensitivity and specificity being achieved using models that  
6 incorporate other variables than just PSA alone, model calibration was poorly reported  
7 and external validation was carried out in only one study. Further studies are needed to  
8 investigate other potential variables that could be integrated into models to improve their  
9 clinical utility for PCa screening in a community-based setting.  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

## 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 <sup>1</sup>. In the UK, around 47,200 new cases of PCa were reported in 2015, accounting for 26% of all new cancer cases in males. PCa had a mortality within the UK of around 11,600 in 2016 <sup>2</sup>. The global projections of PCa incidence and mortality for 2030 are 1.7 and 0.5 million, respectively <sup>3</sup>. The highest incidence of PCa is seen in western societies <sup>4</sup>. The significant increase of PCa incidence and diagnosis over the last three decades can be attributed mainly to the widespread implementation of the PSA serum test after it had been introduced in the late 1980s <sup>5,6</sup>.

The strong association of PSA with PCa <sup>7,8</sup>, along with it being a relatively inexpensive test <sup>9</sup>, made PSA a key biomarker in the diagnostic process of PCa and the recommendation of a confirmatory prostate biopsy <sup>7,9</sup>. PSA is, however, not a cancer-specific marker <sup>5,10</sup>. 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 <sup>11-14</sup>. This highlights some 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 <sup>5,10,15,16</sup>. Such issues are likely to be the explanation for the significant number of unnecessary biopsies currently performed each year. Such procedures are associated with adverse side effects for patients and also increases health care costs <sup>17,18</sup>.

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 <sup>19-22</sup>. Logistic regression and artificial neural network (ANNs) models are now considered to be the most common and effective statistical techniques in aiding confident early PCa diagnosis <sup>23</sup>. An important advantage of these models is their ability to tailor accurate estimation of PCa risk individually. This may ultimately lead to their use in patient counselling and decision-making <sup>24-27</sup>. Such models have already



1  
2  
3 achieved better results in predicting probabilities of outcome than clinical judgment <sup>28 29</sup>.  
4  
5 Furthermore, it has been reported that using such predictive models may minimize the  
6  
7 rate of unnecessary biopsies <sup>30</sup>. Recently there has been a substantial increase in the  
8  
9 development of predictive models to help clinicians assess risk and make a diagnosis of  
10  
11 PCa <sup>22 26 30-35</sup>. The majority of these models are designed for use in clinical settings and  
12  
13 most include the need for a clinical examination. Some of these procedures, such as  
14  
15 digital rectal examination (DRE) or trans-rectal ultrasonography (TRUS), are considered  
16  
17 invasive and to potentially cause anxiety <sup>36</sup>. As a result, men may feel reluctant to have  
18  
19 such tests performed.

20  
21 This systematic review of the literature was undertaken to identify risk prediction models  
22  
23 that do not incorporate invasive clinical procedures or genetic variables but have  
24  
25 potential application for use in primary care and community settings. The performance  
26  
27 of the models reviewed for detecting PCa at initial biopsy have been compared using  
28  
29 'reported area under the curve' (AUC) and/or sensitivity - specificity testing.

## 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60

### 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)  
<sup>37</sup>.

#### Data sources and search strategy

A literature search was performed using Medline 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"). All articles defined (published 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, appropriate research articles  
were selected manually from the reference lists of any relevant review articles.

## Eligibility criteria

As this review focuses on PCa risk prediction based in community healthcare settings, all studies were selected on the following inclusion criteria (i) evaluating the prospective risk for PCa at initial biopsy (ii) study subjects used in the development of risk prediction models had no prior history of PCa (iii) studies that reported risk assessment tools not using genetic or “invasive” clinical tests/examinations such as DRE or TRUS (iv) studies that included a minimum of two variables of which PSA had to be one of them [an elevated PSA test in UK primary care is usually the first sign and rationale for suggesting a diagnosis of PCa within NICE guidelines] and (v) 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 (iii) articles that were not published in English. There were no time boundaries regarding the publication year.

Screening titles, abstract, and full-texts was carried out by one reviewer (MA). Any concerns about the eligibility of a study were resolved by discussion with the second reviewer (AL).

## 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 subjects, 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 validation, and model performance.

## 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<sup>27</sup>. Discrimination refers to how well the prediction model can differentiate subjects in different outcome classes according to their predicted risks. It is often assessed by measuring the area under the receiver operating characteristic (ROC) curve<sup>38</sup>. 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

1  
2  
3 discrimination. However, even with perfect discrimination, observed risk can differ from  
4 expected risk. Therefore, calibration has an important role in model evaluation <sup>39</sup>.  
5 Calibration represents the agreement between expected and observed outcomes <sup>40</sup>. A  
6 well-calibrated model is achieved when the calibration slope is close to 1. When the  
7 calibration slope is less than 1 it indicates that the model under-estimates low risks and  
8 over-estimate high risks <sup>41</sup>.  
9

10  
11 Due to the heterogeneity of the studies included, conducting a meta-analysis was not  
12 applicable.  
13

### 14 **Study quality assessment**

15  
16 The quality of the studies included in this review was assessed using the PROBAST  
17 tool <sup>42</sup>. This tool has been developed specifically to assess the risk of bias and  
18 applicability for prediction model studies. The tool consists of four domains and has 20  
19 signalling questions that facilitate to reach overall judgment of risk of bias, as well as  
20 issues relating to applicability.  
21  
22

### 23 **Patient and Public involvement**

24  
25 No patients were involved in setting the research question or the outcome measures,  
26 nor were they involved in the design and implementation of the study. There are no  
27 plans to involve patients in dissemination.  
28

## 29 **RESULTS**

30  
31 A total of 102 publications were identified using the search strategy as shown in Fig.1.  
32 An additional nine articles were identified through manual searches from a bibliography  
33 of reviewed articles. At the first filter step, a total of 109 titles and abstracts were  
34 screened for eligibility after removing two duplicates. In the second filter step, 60 papers  
35 failed to meet the inclusion criteria and were excluded, resulting in 49 articles. The final  
36 step of filtering yielded only five studies that were considered to be eligible (i.e. passed  
37 all set criteria) and were thus included in this systematic review.  
38

### 39 **Study characteristics**

40  
41 Four of the five studies included were cohort studies and one was a case-control study.  
42 The characteristics of each of these studies and populations are summarised in Table 1.  
43  
44  
45  
46  
47  
48  
49  
50

Table 1: Characteristics of the included studies

Author & year	Type of model	Type of study	Sample No.	Location	Population type	Age	PSA range	Media n PSA	No. of biopsy cores	Cancer detection
Carlson, 1998 <sup>43</sup>	Logistic regression	Cohort	Model dev = 3773 Validation = 525	Baltimore, USA	Referral	≥ 45	4-20 ng/ml	—	Sextant biopsy	32%
Babaian, 2000 <sup>44</sup>	Neural network = 3 ANNs	Cohort	151	Texas, USA	Screening program	40-75	2.5-4 ng/ml	—	11 cores	24.5%
Jansen, 2010 <sup>45</sup>	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 program	≥ 50	2-10 ng/ml	~ 4.4	≥6 cores	Site 1 = 55.8% Site 2 = 49.6%
Hill, 2013 <sup>46</sup>	Logistic regression	Case-control	1378	Florida, USA	Hospital referral	40-90	≥ 4 ng/ml	—	N/A	20.6%
Lazzeri, 2013 <sup>47</sup>	Logistic regression	Cohort	646	European multi-centre; Italy, Germany, France, Spain, and the UK	Referral	>45	2-10 ng/ml	~ 5.8	≥ 12 cores	40.1%

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

Subjects used to build the risk models varied across these studies. Of the five studies, three studies included men from referral populations <sup>43 46 47</sup> and two studies from screening programs <sup>44 45</sup>. The sample sizes ranged from 151 to 3,773 with three studies derived from US cohorts <sup>43 44 46</sup> and two from Europe <sup>45 47</sup>. Four studies used logistic regression methodology to build their model, whereas one study used an artificial neural network-based approach (ANN) <sup>44</sup>. The minimum age of participants was 40 years old <sup>44 46</sup> and the minimum PSA level was 2 ng/ml <sup>45 47</sup>.

### **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 prostate-specific antigen (p2PSA), percentage of p2PSA to free PSA (%p2PSA), prostate health index (PHI), levels of haemoglobin, albumin, creatinine and red blood cell count (RBC), haematuria, mean corpuscular volume (MCV), and prostatic acid phosphatase.

Table 2: Variables used in the prostate cancer risk prediction models\*

Author & year	Variables used in the model				
	Total PSA	Free PSA	Percent free PSA	Age	Other variables
Carlson, 1998 <sup>43</sup>	✓		✓	✓	
Babaian, 2000 <sup>44</sup>	✓	✓		✓	Creatinine Kinase, Prostatic acid phosphatase
Jansen, 2010 <sup>45</sup> (Site 1)	✓	✓			p2PSA
Jansen, 2010 <sup>45</sup> (Site 2)	✓	✓			p2PSA
Hill, 2013 <sup>46</sup> (Method 1)	✓			✓	HGB, RBC, Haematuria, Creatinine, MCV, and ethnicity "Black"
Hill, 2013 <sup>46</sup> (Method 2)	✓			✓	HGB, RBC, Creatinine, and MCV
Lazzeri, 2013 <sup>47</sup> (Model 1)	✓	✓	✓		
Lazzeri, 2013 <sup>47</sup> (Model 2)	✓	✓	✓		p2PSA
Lazzeri, 2013 <sup>47</sup> (Model 3)	✓	✓	✓		%p2PSA
Lazzeri, 2013 <sup>47</sup> (Model 4)	✓	✓	✓		PHI

\*HGB= Haemoglobin. RBC= Red blood cells. MCV= Mean corpuscular volume. PHI= Prostate Health Index. p2PSA= Precursor of prostate specific antigen, %p2PSA= p2PSA/fPSA

## Outcome

Study endpoint also varied among the studies selected. Two studies evaluated the accuracy of detecting any PCa<sup>43 44</sup> and three studies examined the pathologic Gleason score<sup>45-47</sup>. Although Jansen et al did not build a model to predict the aggressiveness of prostate cancer, they assessed the relationship of each variable individually with a Gleason score  $\geq 7$ . PCa was determined by taking a needle biopsy. All subjects in the five studies underwent prostate biopsy. The least number of biopsy cores used were 6<sup>43 45</sup> and the highest were  $\geq 12$ <sup>47</sup>. One study did not report the number of biopsy cores taken<sup>46</sup>. 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 to PSA alone (AUC of 0.755 vs. 0.585, respectively)<sup>45</sup>. The AUC values in other studies ranged from 0.648 to 0.74. Model performance is presented in Fig 2.

One study did not provide the AUC but instead reported an increase of 11% in specificity over percent free PSA alone with 95% sensitivity<sup>43</sup>. Lazzeri et al,<sup>47</sup> presented results from four separated models discriminating PCa with a Gleason score of  $\geq 7$ . Both model 2 (which includes the base model tPSA, fPSA and %fPSA in addition to p2PSA) and model 3 (which includes base model plus PHI) show the highest levels of discrimination with an AUC of 0.67. In the study of Hill<sup>46</sup> the authors classified PCa stages differently and built their two models accordingly. In their 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 to 0.59 for PSA alone. In the second model, the discrimination analysis was based on PCa stages II, III, IV versus PCa stage I, prostatic interstitial neoplasm (PIN), 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 to 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. The differences of AUC for PSA alone and the extended model are presented in Table 3.

Table 3: The difference of area under the curve (AUC) for PSA alone and extended model

Study	AUC for PSA	AUC for model	$\Delta$ AUC (Model – PSA)
Carlson <sup>43</sup>	NA	NA	NA
Babaian <sup>44</sup>	0.64 %fPSA	0.74	0.1
Jansen <sup>45</sup> (site 1)	0.58	0.75	0.17
Jansen <sup>45</sup> (site 2)	0.53	0.7	0.16
Hill <sup>46</sup> (method 1)	0.59	0.68	0.09
Hill <sup>46</sup> (method 2)	0.63	0.72	0.09
Lazzeri <sup>47</sup>	0.50 for any PC  0.54 for Gleason score $\geq 7$	Model 1 = 0.65 Model 1 (Gleason score $\geq 7$ ) = 0.60	0.15 0.06
		Model 2 = 0.71 Model 2 (Gleason score $\geq 7$ ) = 0.67	0.21 0.13
		Model 3 = 0.704 Model 3 (Gleason score $\geq 7$ ) = 0.67	0.2 0.13
		Model 4 = 0.71 Model 4 (Gleason score $\geq 7$ ) = 0.672	0.21 0.13

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 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 4: Sensitivity and specificity profile at different levels for each model

Study	Sensitivity	Specificity	Probability cut-off	Positive predictive value	Negative predictive value
Carlson <sup>43</sup>	99	18	>15	≤47	N/A
	95	34	18	51	NA
	89	43	20	42	NA
Babaian <sup>44</sup>	95	51	NA	39	97
	92	62	NA	44	96
	89	62	NA	43	95
Jansen <sup>45</sup> (Site 1)	95	23.9	NA	NA	NA
Jansen <sup>45</sup> (Site 1)	90	30.1	NA	NA	NA
Jansen <sup>45</sup> (Site 2)	95	23.2	NA	NA	NA
Jansen <sup>45</sup> (Site 2)	90	36.2	NA	NA	NA
Hill <sup>46</sup> (Method 1)	90.9	17.6	33	47.1	70.5
Hill <sup>46</sup> (Method 2)	89.8	28	13	20.6	91.3
Hill <sup>46</sup> (Method 1)	80.5	37.1	37	50.9	70.2
Hill <sup>46</sup> (Method 2)	78.2	45	15	28.7	88.8
Hill <sup>46</sup> (Method 1)	39.9	81.4	48	63.4	62.6
Hill <sup>46</sup> (Method 2)	45.8	79.5	23	36.7	85
* Lazzeri <sup>47</sup> model reported only sensitivity and specificity for predictive variables individually and at sensitivity of 90, %p2PSA and %fPSA achieved the highest specificity					

Table 5 summarises the validation and calibration results for the studies analysed. Model calibration was reported in two studies <sup>43 47</sup>. 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 <sup>45 46</sup>. Only one study reported an external validation using an additional data set consisting of 525 patients <sup>43</sup>. 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 <sup>44</sup>. Lazzeri used 200 bootstrap re-samples to minimise overfitting bias <sup>47</sup>.

Table 5: Validation and calibration for included models

Author & Year	Validation	Calibration
Carlson, 1998 <sup>43</sup>	External validation on additional data set consisting of 525 patients	Calibration plot
Babaian, 2000 <sup>44</sup>	Cross-validation and separate data set of 151	N/A
Jansen, 2010 <sup>45</sup>	N/A	N/A
Hill, 2013 <sup>46</sup>	N/A	N/A
Lazzeri, 2013 <sup>47</sup>	Internal validation using 200 bootstrap resamples	Internal calibration using the Hosmer-Lemeshow goodness-of-fit test

### 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 tool, particularly in the analysis domain. For instance, one study applied univariable analysis to select predictors <sup>43</sup>. Three studies did not measure calibration <sup>44-46</sup>. Furthermore, two studies did not count for optimism and overfitting by using internal validation methods <sup>45</sup>

1  
2  
3 46. Whereas one study did not use appropriate measures for model performance i.e.  
4 AUC, this study reported the calibration <sup>43</sup>.

5  
6  
7 The event per variable (EPV) was lower than recommended (< 10) in the Babaian <sup>44</sup>  
8 study indicating inadequate power. Four studies did not report missing data or how they  
9 handled it <sup>43-45</sup> <sup>47</sup>. The remaining study used complete-case analysis and excluded  
10 subjects with missing data on laboratory biomarkers (n=75) <sup>46</sup>. The PROBAST  
11 guidelines state that in a prediction model study where any risk of bias and applicability  
12 is low in all four domains, a regrading to high risk of bias should be considered when the  
13 study did not validate the model externally <sup>42</sup>. Thus, although the quality assessment for  
14 the Lazzeri study <sup>47</sup> was graded low risk in all the four domains, since the study did not  
15 report any external validation of the model, the assessment of the study has been  
16 regraded to high risk of bias. A full quality assessment for all studies is presented in  
17 Table 6.  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

Table 6: Quality assessment for risk of bias and applicability concern for included studies

Study	ROB				Applicability			Overall	
	Participants	Predictors	Outcome	Analysis	Participants	Predictors	Outcome	ROB	Applicability
Carlson <sup>43</sup>	+	+	-	-	+	+	-	-	-
Babaian <sup>44</sup>	+	+	+	-	-	+	+	-	-
Jansen <sup>45</sup>	-	-	-	-	+	-	-	-	-
Hill <sup>46</sup>	-	+	+	-	-	+	+	-	-
Lazzeri <sup>47</sup>	+	+	+	+	+	+	+	-	+

+ indicates a low risk of bias and applicability; - indicates a high risk of bias and applicability; ? indicates unclear bias and concern regarding applicability.

## DISCUSSION

Despite the large number of PCa risk prediction models, the majority still include clinical inputs and/or genetic variants; few models exist that do not include specialist clinical input or genetic testing. To our knowledge, this is the first study to examine risk prediction models for PCa that do not include clinical and genetic variables.

Our study identified five unique models that met the set criteria. The Carlson model<sup>43</sup> has the largest population (3,773 subjects) when compared to 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<sup>33</sup>. More importantly, by not reporting the AUC estimate, the model raises some doubts regarding the reliability of the model and its implementation<sup>33</sup>. It will also make comparison to other models not applicable<sup>48</sup>.

Babaian<sup>44</sup> 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, PSAD-TZ= 0.75). They included a number of uncommon pre-biopsy inputs into their algorithm such as prostatic acid phosphatase and creatinine kinase<sup>49</sup>. Furthermore, they used a tight PSA range (2.5-4.0ng/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 colleagues<sup>45</sup> 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 prediction and detection of PCa have

1  
2  
3 been reported by several other studies <sup>16 50-53</sup>. Jansen <sup>45</sup> showed that p2PSA has no  
4 clear association with aggressive PCa. However, the base model that includes p2PSA  
5 had the highest clinical significance in correlation to pathologic Gleason score with a p-  
6 value of 0.008 compared to %fPSA and PHI (p-value 0.01 and 0.02) respectively.  
7 Although they used archived blood samples and retrospective analysis, the results were  
8 similar to a prospective study of 268 patients <sup>16</sup>.  
9

10  
11  
12  
13  
14 Hill <sup>46</sup> used a case-control study to evaluate several laboratory biomarkers. They found  
15 HGB, RBC, haematuria, creatinine, PSA, Age, MCV, and ethnicity (“being black”) were  
16 statistically significantly associated in the first method (p < 0.05). In the second method,  
17 HGB, RBC, creatinine, PSA, age, and MCV were found to be statistically significantly  
18 correlated (p < 0.001) with PCa. However, since this study was designed as a case-  
19 control study, it would have been prone to uncontrolled confounding and selection bias.  
20 Moreover, the type of screening protocols used in Veterans’ Administrations may vary to  
21 those conducted in other healthcare systems; therefore, other populations may not have  
22 been represented. Furthermore, subjects with a PSA level < 4.0 ng/ml have not been  
23 investigated, and thus, the performance of the models are unknown for individuals in  
24 this group.  
25

26  
27  
28  
29  
30  
31  
32  
33 Lazzeri <sup>47</sup> in a European multi-centre study have evaluated similar biomarkers as in  
34 Jansen study with the same PSA range 2-10 ng/ml prospectively. They found no  
35 difference in both %p2PSA and PHI as individual PCa predictors with AUC of 0.67 (95%  
36 CI 0.64-0.71). However, the base model (consisting of PSA, fPSA, and %fPSA) that  
37 also included either p2PSA or PHI out-performed the base model alone and the base  
38 model including %p2PSA. In the analysis, the additive value of both p2PSA and PHI is  
39 0.064 and 0.056 for %p2PSA for predicting the risk of PCa. These additive values  
40 increased to 0.076 for both p2PSA and PHI and 0.073 for %p2PSA in predicting  
41 Gleason scores  $\geq 7$  for the disease. The usefulness of PHI in improving the predictive  
42 accuracy of PCa over total and free PSA has been confirmed and reported by several  
43 studies <sup>16 53-56</sup>.  
44  
45  
46  
47  
48  
49  
50  
51

52  
53 In general, only one study has validated their model externally <sup>43</sup>, whereas the  
54 remaining studies were either validated internally <sup>44 47</sup> or not reported any validation  
55  
56  
57  
58  
59  
60

1  
2  
3 methods <sup>45 46</sup>. Prediction models may not be equally applicable to all data sets as  
4 patients' characteristics may vary <sup>20 57</sup>. As a result, the generalisability of a model might  
5 be poor when it used in populations other than that used in building the model.  
6 Therefore, external validation should be conducted before applying any new model into  
7 general practice <sup>58 59</sup>.

8  
9  
10  
11  
12 Another key performance measure of any model that needs careful evaluation is  
13 calibration. A calibration plot with a calibration slope is more preferable than the  
14 Hosmer-Lemeshow test; it has been acknowledged that evaluating a good and well  
15 calibrated model based on a large dataset can still fail the Hosmer-Lemeshow test. In  
16 contrast, when evaluating a poorly calibrated model with a small dataset it can still pass  
17 the Hosmer-Lemeshow test <sup>60</sup>. In our analysis, three studies fail to report the calibration  
18 of the model <sup>44-46</sup> whilst the Carlson study <sup>43</sup> used a calibration plot and Lazzeri <sup>47</sup> used  
19 the Hosmer-Lemeshow test. Excluding calibration from the majority of models may  
20 explain why some models are not currently used in practice <sup>60</sup>.

21  
22  
23  
24  
25  
26  
27  
28 With regard to biopsy cores, only two studies used extended biopsy cores. Babaian <sup>44</sup>  
29 used an 11-core multi-site biopsy, whereas Lazzeri <sup>47</sup> used at least 12 biopsy cores.  
30 Moreover, two studies used six cores biopsy in their model <sup>43 45</sup>. The use of six-core  
31 biopsy has been criticized as not being adequate in detecting PCa <sup>61</sup> and that models  
32 developed using sextant biopsy are less accurate than when a 10-core biopsy is used  
33 <sup>62</sup>. As a result, the European guideline for clinical PCa recommended an extended  
34 biopsy as standard practice for PCa detection <sup>63</sup>.

35  
36  
37  
38  
39  
40  
41 It is worth noting that all five reviewed models performed better than just PSA alone.  
42 However, none of them has both high specificity and sensitivity. The level of sensitivity  
43 has been increased, and despite enhancement in the specificity, it is still considered  
44 low. Specificity is crucial when it comes to being used in a population setting as men  
45 without PCa should be ruled out as much as possible from further invasive engagement  
46 with the health system.

47  
48  
49  
50  
51  
52 Our review suggests that none of the reviewed models provided an ideal performance in  
53 predicting PCa with high sensitivity and high specificity. It is particularly important when  
54 considering the application of PCa risk prediction at the population level that the tool  
55

1  
2  
3 used should be able to both detect the outcome and filter out people with no disease.  
4  
5 As there is robust evidence suggesting that the clinical relevance of PSA range to the  
6 detection of PCa differs across age groups <sup>64-66</sup>, any future model should consider PSA  
7 threshold in relation to a specific age range. Risk prediction models for PCa should  
8 therefore take account of age.  
9  
10

### 11 **Comparison with other studies**

12  
13 To our knowledge, three systematic reviews of PCa prediction tools have been  
14 published <sup>20 26 27</sup>. In the Louie et al. review, risk models were included that were  
15 externally validated in at least five study populations for the purpose of meta-analysis  
16 and only six studies were included in their analysis. Furthermore, all the studies  
17 included incorporated clinical tests such as DRE and/or TRUS-PV <sup>26</sup>. Schroder and  
18 Kattan <sup>20</sup> reviewed models that were built to predict the likelihood of having a positive  
19 prostate biopsy for cancer. However, it appears that they also included models where  
20 subjects had a previous negative biopsy. As such, some of the models included  
21 variables related to biopsy results and cores. The review by Shariat and colleagues  
22 examined different types of predictive tools <sup>27</sup>. They explored tools that predict PCa on  
23 initial and repeat biopsy, pathologic stages, biochemical recurrence after radical  
24 prostatectomy, metastasis, survival, and life expectancy. Similarly, virtually all of the  
25 prediction tools that were based on initial biopsy included variables based on invasive  
26 procedures.  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38

### 39 **Strengths and limitations of this study**

40  
41 This report is the first to review risk prediction tools for PCa that can be used in primary  
42 care and community settings. Any prediction model should therefore be simple to use,  
43 based on non-invasive tests, be feasible at a population level and at low cost. We  
44 carried out an extensive data extraction relating to important features and  
45 characteristics for each study included, such as modelling method, source of data,  
46 sample size, variables, discrimination, validation, and cancer detection rate. We have  
47 also followed PRISMA guidelines for identifying eligible articles as well as for reporting  
48 this study. In addition, the PROBAST tool was adopted to assess the quality and risk of  
49 bias for each prediction model.  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60



1  
2  
3 Our study has some limitations. Our aim was to identify prediction models that have the  
4 potential to be implemented in non-clinical settings, and consequently our search  
5 strategy was just focus to retrieve relevant studies for this dedicated purpose.  
6 Furthermore, we excluded articles that were not published in English or did not have an  
7 abstract. Moreover, studies that were only validating other existing models have not  
8 been included in this analysis. As a result of our strict inclusion criteria, our review  
9 includes a limited number of eligible studies.  
10  
11  
12  
13  
14

15  
16 A previous systematic review suggested that the majority of relevant studies could be  
17 identified through a manual search of articles reference lists instead of a database alone  
18 <sup>20</sup>. We identified four eligible studies using this approach. Given the small number of  
19 models identified by the approach we followed, that can be applied in primary care  
20 settings compared to the large number relating to wider existing models, it is unlikely  
21 that we have not included any study that would affect the results of our review.  
22  
23  
24  
25

### 26 **Implications and future research**

27  
28 It is now accepted that the PSA test and its derivatives have some limitations for  
29 detecting PCa as defined by biopsy <sup>67</sup>. As a consequence, a considerable number of  
30 PCa prediction models have been built to improve prediction accuracy. This has  
31 resulted in a plethora of PCa risk prediction tools, with to date more than 100 models  
32 described <sup>68 69</sup>. There is evidence that some of these models show benefit and have  
33 better performance over just PSA measurement alone <sup>20</sup>. It also has been demonstrated  
34 that some of these models out-performed clinical experts in predicting PCa <sup>28 29</sup>.  
35 Although such models are not designed to replace specialist clinical judgment or patient  
36 preferences <sup>57 67 70</sup>, they can help in patient counselling and aid clinicians to decide  
37 whether a prostate biopsy should be taken or not <sup>58 70 71</sup>.  
38  
39  
40  
41  
42  
43  
44  
45

46 Given the small number of risk prediction models for PCa that do not incorporate clinical  
47 or genetic tests, the discrimination of these reviewed models ranged between poor to  
48 moderate (AUC range ~0.65 to ~0.75); in addition some issues relating to their study  
49 design and analysis raises the risk of bias. Consequently none of these models could  
50 be currently recommended for use in a primary care and community health care setting.  
51 Several guidelines are against using PSA test based screening for PCa; the US  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 Preventive Services task force, the Canadian task force on preventive health and the  
4 American College of Preventive Medicine do not currently recommend PSA-based  
5 testing due to insufficient evidence <sup>72-74</sup>. This has made it harder to convince  
6 policymakers to adopt PCa screening programme.  
7  
8  
9

10 The first guideline of PROSTATE CANCER UK states *"In the future, health*  
11 *professionals should look at a man's PSA level alongside other known risk factors as*  
12 *part of a risk assessment tool, when one becomes available."* <sup>75</sup> However, the vast  
13 majority of the PCa risk prediction models are not suitable for routine use as they  
14 include clinical and genetic tests and are not validated externally in other cohorts.  
15 Therefore, the main challenge in the UK, is to develop a risk prediction tool that is  
16 reliable, cheap, is applicable for as wide an ethnicity as possible, and most importantly  
17 is easy to use and can be implemented at a primary care level <sup>76</sup>.  
18  
19  
20  
21  
22  
23  
24

25 The value of such risk tools is that they will help to stratify men at high risk of developing  
26 PCa earlier so that they have appropriate management and/or surveillance programme  
27 as early as possible. Such tools should help physicians have a better understanding of  
28 the risk for this disease and simplify the procedures and discussions with patients when  
29 recommending further specialist-led investigations. Furthermore, using the appropriate  
30 risk prediction tool will avoid men from undergoing inappropriate further and frequent  
31 testing <sup>76</sup>. This will reduce any associated costs of inappropriate tests and decrease the  
32 burden on health care delivery systems.  
33  
34  
35  
36  
37  
38

39 It is crucial to address these issues by identify all possible risk factors for PCa that are  
40 non-clinical, non-genetic, and easy to use and interpret. There is a pressing need to  
41 develop a risk prediction tool using all appropriate factors into a robust multivariable  
42 analysis and validate the model externally to eliminate applicability and generalisability  
43 concerns. Only when this is achieved will it be possible to introduce a prostate cancer  
44 screening programme fit for purpose.  
45  
46  
47  
48  
49

## 50 **CONCLUSION**

51  
52 There is a paucity of suitable models that incorporate non-clinical, non-genetic inputs  
53 and which can be used at a primary care level and in other community health services.  
54 Existing models have several limitations in both study design and reporting performance  
55  
56  
57  
58  
59  
60

measures. Future research should take into account these key issues and explore other risk factors.

## ARTICLE SUMMARY

- There is abundance of risk models for prostate cancer that aimed to improve the performance of PSA test.
- The majority of these models are designed for use in clinical settings and some have issue with reliability and/or model validation.
- This is the first systematic review that focused on prostate cancer risk prediction that has the potential to be used in primary care and any other community health hubs.
- Few risk models that do not incorporate clinical and/or genetic tests and have potential to be applied in primary care and community settings.
- Future research must address all issues related to risk prediction model development and its potential application.

## ACKNOWLEDGEMENT

MA is supported by Kuwait's Ministry of Health. 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]).

## AUTHOR CONTRIBUTIONS

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

No funding is required for conducting this study.

## COMPETING INTERESTS

All authors have completed the ICMJE uniform disclosure form at [www.icmje.org/coi\\_disclosure.pdf](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 three years; no other relationships or activities that could appear to have influenced the submitted work.

## ETHICAL APPROVAL

1  
2  
3 Not applicable as this study did not involve any personal data.  
4

#### 5 **DATA SHARING**

6  
7 Data extraction sheet and the PROBAST assessment form for risk of bias is available upon  
8 request.  
9  
10

#### 11 **Figure legend**

12 Figure 1: Flow diagram of studies included using PRISMA method

13 Figure 2: Model performance for each study using area under the curve (AUC)  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

## REFERENCES

1. 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(6):394-424.
2. Cancer Research UK. Prostate cancer statistics 2017 [Available from: <https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/prostate-cancer#heading-One>].
3. Center MM, Jemal A, Lortet-Tieulent J, et al. International variation in prostate cancer incidence and mortality rates. *Eur Urol* 2012;61(6):1079-92.
4. 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(6):1388-400.
5. Dijkstra S, Mulders P, Schalken J. Clinical use of novel urine and blood based prostate cancer biomarkers: a review. *Clin Biochem* 2014;47(10-11):889-96.
6. Fontanella P, Benecchi L, Grasso A, et al. Decision-making tools in prostate cancer: from risk grouping to nomograms. *Minerva Urol Nefrol* 2017;69(6):556-66.
7. Stephan C, Ralla B, Jung K. Prostate-specific antigen and other serum and urine markers in prostate cancer. *Biochim Biophys Acta* 2014;1846(1):99-112.
8. Thompson IM, Pauler DK, Goodman PJ, et al. Prevalence of prostate cancer among men with a prostate-specific antigen level  $\leq$  4.0 ng per milliliter. *N Engl J Med* 2004;350(22):2239-46.
9. 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(1):18-26.
10. 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(1):18-29.
11. 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(2):293-306.
12. 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(4 Pt 1):1297-301.
13. Bunting PS. A guide to the interpretation of serum prostate specific antigen levels. *Clin Biochem* 1995;28(3):221-41.
14. 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(5):1650-55.
15. 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(2):95-101.

16. 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(2):214-22.
17. Heijnsdijk EA, der Kinderen A, Wever EM, et al. Overdetection, overtreatment and costs in prostate-specific antigen screening for prostate cancer. *Br J Cancer* 2009;101(11):1833-8.
18. 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(3):359-62.
19. 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(4):618-29.
20. Schroder 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(2):274-90.
21. Ross PL, Scardino PT, Kattan MW. A catalog of prostate cancer nomograms. *J Urol* 2001;165(5):1562-8.
22. 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(14):4400-7.
23. Hu X, Cammann H, Meyer HA, et al. Artificial neural networks and prostate cancer--tools for diagnosis and management. *Nat Rev Urol* 2013;10(3):174-82.
24. Caras RJ, Sterbis JR. Prostate cancer nomograms: a review of their use in cancer detection and treatment. *Curr Urol Rep* 2014;15(3):391.
25. Iremashvili V, Soloway MS, Pelaez L, et al. Comparative validation of nomograms predicting clinically insignificant prostate cancer. *Urology* 2013;81(6):1202-8.
26. 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(5):848-64.
27. Shariat SF, Kattan MW, Vickers AJ, et al. Critical review of prostate cancer predictive tools. *Future Oncol* 2009;5(10):1555-84.
28. Ross PL, Gerigk C, Gonen M, et al. Comparisons of nomograms and urologists' predictions in prostate cancer. *Semin Urol Oncol* 2002;20(2):82-8.
29. 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(6):1254-8.
30. 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(11):1200-6.
31. 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

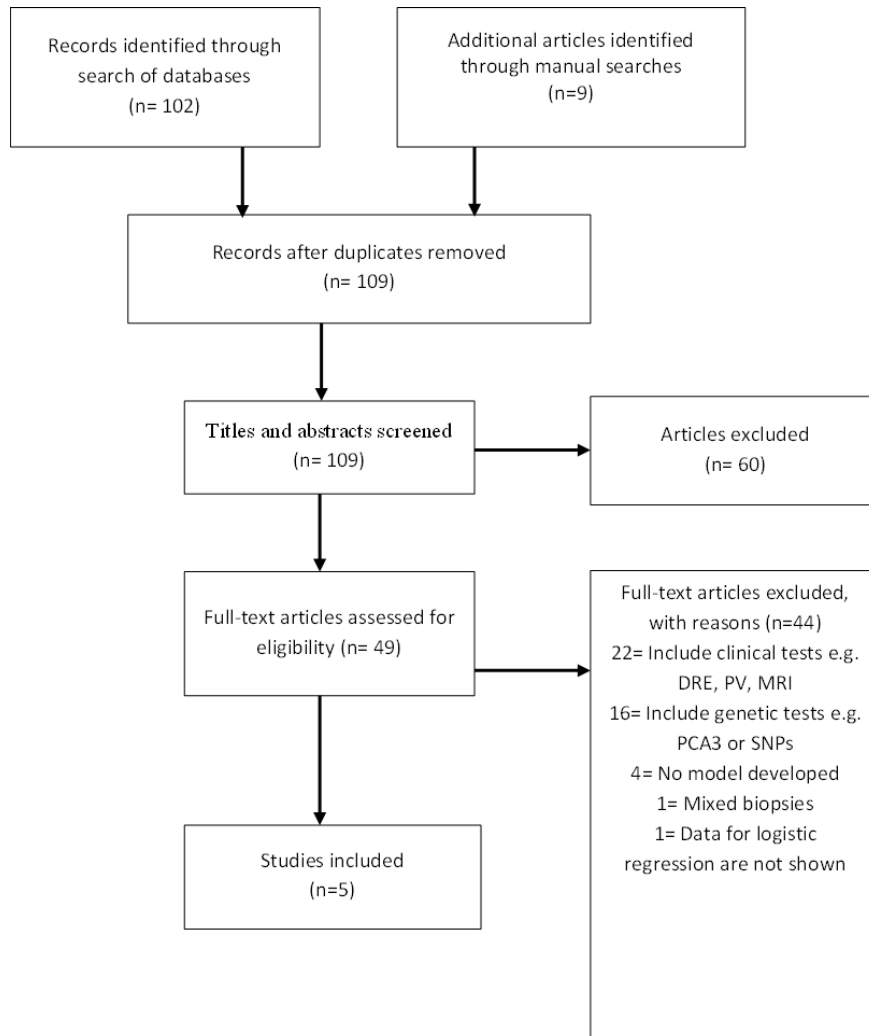


- 1  
2  
3 examination and a prostate-specific antigen between 0 and 4 ng/mL. *Urology*  
4 1999;54(4):709-13.  
5  
6 32. Garzotto M, Hudson RG, Peters L, et al. Predictive modeling for the presence of prostate  
7 carcinoma using clinical, laboratory, and ultrasound parameters in patients with prostate  
8 specific antigen levels  $\leq 10$  ng/mL. *Cancer* 2003;98(7):1417-22.  
9  
10 33. Karakiewicz PI, Benayoun S, Kattan MW, et al. Development and validation of a nomogram  
11 predicting the outcome of prostate biopsy based on patient age, digital rectal examination  
12 and serum prostate specific antigen. *J Urol* 2005;173(6):1930-4.  
13  
14 34. Kranse R, Roobol M, Schroder FH. A graphical device to represent the outcomes of a logistic  
15 regression analysis. *Prostate* 2008;68(15):1674-80.  
16  
17 35. Thompson IM, Ankerst DP, Chi C, et al. Assessing prostate cancer risk: results from the  
18 Prostate Cancer Prevention Trial. *J Natl Cancer Inst* 2006;98(8):529-34.  
19  
20 36. Brindle LA, Oliver SE, Dedman D, et al. Measuring the psychosocial impact of population-  
21 based prostate-specific antigen testing for prostate cancer in the UK. *BJU Int*  
22 2006;98(4):777-82.  
23  
24 37. Moher D, Liberati A, Tetzlaff J, et al. Preferred reporting items for systematic reviews and  
25 meta-analyses: the PRISMA statement. *PLoS Med* 2009;6(7):e1000097.  
26  
27 38. Park SH, Goo JM, Jo CH. Receiver operating characteristic (ROC) curve: practical review  
28 for radiologists. *Korean J Radiol* 2004;5(1):11-8.  
29  
30 39. Han K, Song K, Choi BW. How to Develop, Validate, and Compare Clinical Prediction  
31 Models Involving Radiological Parameters: Study Design and Statistical Methods.  
32 *Korean J Radiol* 2016;17(3):339-50.  
33  
34 40. Steyerberg EW. Clinical prediction models: a practical approach to development, validation,  
35 and updating: Springer Science & Business Media 2008.  
36  
37 41. Pavlou M, Ambler G, Seaman SR, et al. How to develop a more accurate risk prediction  
38 model when there are few events. *BMJ* 2015;351:h3868.  
39  
40 42. Wolff RF, Moons KGM, Riley RD, et al. PROBAST: A Tool to Assess the Risk of Bias and  
41 Applicability of Prediction Model Studies. *Ann Intern Med* 2019;170(1):51-58.  
42  
43 43. Carlson GD, Calvanese CB, Partin AW. An algorithm combining age, total prostate-specific  
44 antigen (PSA), and percent free PSA to predict prostate cancer: results on 4298 cases.  
45 *Urology* 1998;52(3):455-61.  
46  
47 44. Babaian RJ, Fritsche H, Ayala A, et al. Performance of a neural network in detecting prostate  
48 cancer in the prostate-specific antigen reflex range of 2.5 to 4.0 ng/mL. *Urology*  
49 2000;56(6):1000-6.  
50  
51 45. Jansen FH, van Schaik RH, Kurstjens J, et al. Prostate-specific antigen (PSA) isoform p2PSA  
52 in combination with total PSA and free PSA improves diagnostic accuracy in prostate  
53 cancer detection. *Eur Urol* 2010;57(6):921-7.  
54  
55 46. Hill OT, Mason TJ, Schwartz SW, et al. Improving prostate cancer detection in veterans  
56 through the development of a clinical decision rule for prostate biopsy. *BMC Urol*  
57 2013;13:6.  
58  
59

- 1  
2  
3 47. Lazzeri M, Haese A, de la Taille A, et al. Serum isoform [-2]proPSA derivatives significantly  
4 improve prediction of prostate cancer at initial biopsy in a total PSA range of 2-10 ng/ml:  
5 a multicentric European study. *Eur Urol* 2013;63(6):986-94.  
6
- 7 48. Porter CR, O'Donnell C, Crawford ED, et al. Predicting the outcome of prostate biopsy in a  
8 racially diverse population: a prospective study. *Urology* 2002;60(5):831-35.  
9
- 10 49. Porter CR, Gamito EJ, Crawford ED, et al. Model to predict prostate biopsy outcome in large  
11 screening population with independent validation in referral setting. *Urology*  
12 2005;65(5):937-41.  
13
- 14 50. Catalona WJ, Bartsch G, Rittenhouse HG, et al. Serum pro-prostate specific antigen  
15 preferentially detects aggressive prostate cancers in men with 2 to 4 ng/ml prostate  
16 specific antigen. *J Urol* 2004;171(6 Pt 1):2239-44.  
17
- 18 51. Sokoll LJ, Sanda MG, Feng Z, et al. A prospective, multicenter, National Cancer Institute  
19 Early Detection Research Network study of [-2]proPSA: improving prostate cancer  
20 detection and correlating with cancer aggressiveness. *Cancer Epidemiol Biomarkers Prev*  
21 2010;19(5):1193-200.  
22
- 23 52. Sokoll LJ, Wang Y, Feng Z, et al. [-2]proenzyme prostate specific antigen for prostate cancer  
24 detection: a national cancer institute early detection research network validation study. *J*  
25 *Urol* 2008;180(2):539-43; discussion 43.  
26
- 27 53. Stephan C, Vincendeau S, Houlgatte A, et al. Multicenter evaluation of [-2]proprostate-  
28 specific antigen and the prostate health index for detecting prostate cancer. *Clin Chem*  
29 2013;59(1):306-14.  
30
- 31 54. Le BV, Griffin CR, Loeb S, et al. [-2]Proenzyme prostate specific antigen is more accurate  
32 than total and free prostate specific antigen in differentiating prostate cancer from benign  
33 disease in a prospective prostate cancer screening study. *J Urol* 2010;183(4):1355-9.  
34
- 35 55. Loeb S, Sokoll LJ, Broyles DL, et al. Prospective multicenter evaluation of the Beckman  
36 Coulter Prostate Health Index using WHO calibration. *J Urol* 2013;189(5):1702-06.  
37
- 38 56. Lughezzani G, Lazzeri M, Haese A, et al. Multicenter European external validation of a  
39 Prostate Health Index-based nomogram for predicting prostate cancer at extended  
40 biopsy. *Eur Urol* 2014;66(5):906-12.  
41
- 42 57. Chun FK, Karakiewicz PI, Briganti A, et al. A critical appraisal of logistic regression-based  
43 nomograms, artificial neural networks, classification and regression-tree models, look-up  
44 tables and risk-group stratification models for prostate cancer. *BJU Int* 2007;99(4):794-  
45 800.  
46
- 47 58. Herman MP, Dorsey P, John M, et al. Techniques and predictive models to improve prostate  
48 cancer detection. *Cancer* 2009;115(S13):3085-99.  
49
- 50 59. Trottier G, Roobol MJ, Lawrentschuk N, et al. Comparison of risk calculators from the  
51 Prostate Cancer Prevention Trial and the European Randomized Study of Screening for  
52 Prostate Cancer in a contemporary Canadian cohort. *BJU Int* 2011;108(8 Pt 2):E237-44.  
53
- 54 60. Collins GS, de Groot JA, Dutton S, et al. External validation of multivariable prediction  
55 models: a systematic review of methodological conduct and reporting. *BMC Med Res*  
56 *Methodol* 2014;14(1):40.  
57  
58  
59



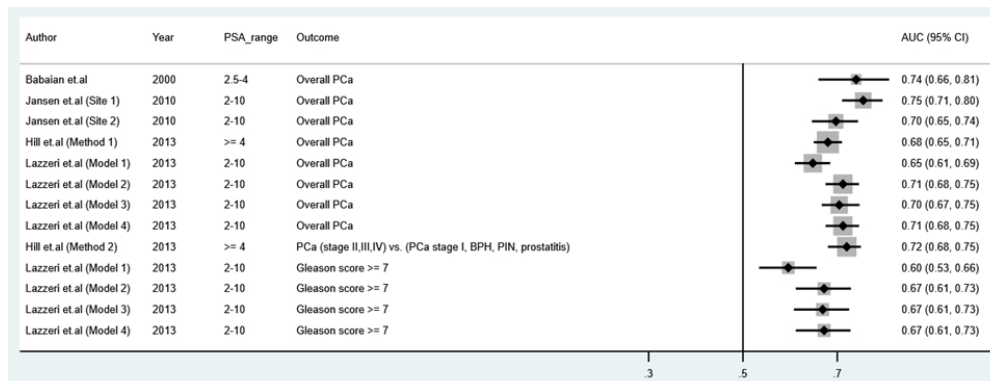
- 1  
2  
3 61. Miyake H, Kurahashi T, Takenaka A, et al. Improved accuracy for predicting the Gleason  
4 score of prostate cancer by increasing the number of transrectal biopsy cores. *Urol Int*  
5 2007;79(4):302-6.  
6
- 7 62. Chun FK, Briganti A, Graefen M, et al. Development and external validation of an extended  
8 10-core biopsy nomogram. *Eur Urol* 2007;52(2):436-44.  
9
- 10 63. Aus G, Abbou CC, Bolla M, et al. EAU guidelines on prostate cancer. *Eur Urol*  
11 2005;48(4):546-51.  
12
- 13 64. Vickers AJ, Ulmert D, Serio AM, et al. The predictive value of prostate cancer biomarkers  
14 depends on age and time to diagnosis: towards a biologically-based screening strategy.  
15 *Int J Cancer* 2007;121(10):2212-7.  
16
- 17 65. Steuber T, O'Brien MF, Lilja H. Serum markers for prostate cancer: a rational approach to  
18 the literature. *Eur Urol* 2008;54(1):31-40.  
19
- 20 66. Carlsson S, Assel M, Sjoberg D, et al. Influence of blood prostate specific antigen levels at  
21 age 60 on benefits and harms of prostate cancer screening: population based cohort study.  
22 *BMJ* 2014;348:g2296.  
23
- 24 67. Cavadas V, Osorio L, Sabell F, et al. Prostate cancer prevention trial and European  
25 randomized study of screening for prostate cancer risk calculators: a performance  
26 comparison in a contemporary screened cohort. *Eur Urol* 2010;58(4):551-8.  
27
- 28 68. Vickers AJ. Prediction models: revolutionary in principle, but do they do more good than  
29 harm. *J Clin Oncol* 2011;29(22):2951-52.  
30
- 31 69. Shariat SF, Karakiewicz PI, Margulis V, et al. Inventory of prostate cancer predictive tools.  
32 *Curr Opin Urol* 2008;18(3):279-96.  
33
- 34 70. Kawakami S, Numao N, Okubo Y, et al. Development, validation, and head-to-head  
35 comparison of logistic regression-based nomograms and artificial neural network models  
36 predicting prostate cancer on initial extended biopsy. *Eur Urol* 2008;54(3):601-11.  
37
- 38 71. Hernandez DJ, Han M, Humphreys EB, et al. Predicting the outcome of prostate biopsy:  
39 comparison of a novel logistic regression-based model, the prostate cancer risk  
40 calculator, and prostate-specific antigen level alone. *BJU Int* 2009;103(5):609-14.  
41
- 42 72. Moyer VA. Screening for prostate cancer: US Preventive Services Task Force  
43 recommendation statement. *Ann Intern Med* 2012;157(2):120-34.  
44
- 45 73. Bell N, Connor Gorber S, Shane A, et al. Recommendations on screening for prostate cancer  
46 with the prostate-specific antigen test. *CMAJ* 2014;186(16):1225-34.  
47
- 48 74. Lim LS, Sherin K, Committee APP. Screening for prostate cancer in US men: ACPM  
49 position statement on preventive practice. *Am J Prev Med* 2008;34(2):164-70.  
50
- 51 75. Prostate Cancer UK. Consensus statements on PSA testing in asymptomatic men in the UK -  
52 information for health professionals 2016 [Available from:  
53 <https://prostatecanceruk.org/about-us/projects-and-policies/consensus-on-psa-testing>  
54 accessed 20-8 2019.  
55
- 56 76. Frame I, Cant S. Current challenges in prostate cancer: an interview with Prostate Cancer  
57 UK. *BMC Med* 2015;13(1):166.  
58  
59



Flow diagram of studies included using PRISMA method

78x91mm (300 x 300 DPI)

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60



Model performance for each study using area under the curve (AUC)

81x33mm (300 x 300 DPI)

# Reporting checklist for systematic review and meta-analysis.

Based on the PRISMA guidelines.

## Instructions to authors

Complete this checklist by entering the page numbers from your manuscript where readers will find each of the items listed below.

Your article may not currently address all the items on the checklist. Please modify your text to include the missing information. If you are certain that an item does not apply, please write "n/a" and provide a short explanation.

Upload your completed checklist as an extra file when you submit to a journal.

In your methods section, say that you used the PRISMA reporting guidelines, and cite them as:

Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement

	Reporting Item	Page Number
<b>Title</b>		
	<a href="#">#1</a> Identify the report as a systematic review, meta-analysis, or both.	3
<b>Abstract</b>		
Structured summary	<a href="#">#2</a> Provide a structured summary including, as applicable: background; objectives; data sources; study eligibility criteria, participants, and interventions; study appraisal and synthesis methods; results; limitations; conclusions and implications of key findings; systematic review registration number	3-4
<b>Introduction</b>		
Rationale	<a href="#">#3</a> Describe the rationale for the review in the context of what is already known.	5-6
Objectives	<a href="#">#4</a> Provide an explicit statement of questions being addressed with	6

reference to participants, interventions, comparisons, outcomes, and study design (PICOS).

## Methods

Protocol and registration	<a href="#">#5</a>	Indicate if a review protocol exists, if and where it can be accessed (e.g., Web address) and, if available, provide registration information including the registration number.	n/a
Eligibility criteria	<a href="#">#6</a>	Specify study characteristics (e.g., PICOS, length of follow-up) and report characteristics (e.g., years considered, language, publication status) used as criteria for eligibility, giving rational	6
Information sources	<a href="#">#7</a>	Describe all information sources in the search (e.g., databases with dates of coverage, contact with study authors to identify additional studies) and date last searched.	6
Search	<a href="#">#8</a>	Present full electronic search strategy for at least one database, including any limits used, such that it could be repeated.	6
Study selection	<a href="#">#9</a>	State the process for selecting studies (i.e., for screening, for determining eligibility, for inclusion in the systematic review, and, if applicable, for inclusion in the meta-analysis).	7
Data collection process	<a href="#">#10</a>	Describe the method of data extraction from reports (e.g., piloted forms, independently by two reviewers) and any processes for obtaining and confirming data from investigators.	7
Data items	<a href="#">#11</a>	List and define all variables for which data were sought (e.g., PICOS, funding sources), and any assumptions and simplifications made.	8
Risk of bias in individual studies	<a href="#">#12</a>	Describe methods used for assessing risk of bias in individual studies (including specification of whether this was done at the study or outcome level, or both), and how this information is to be used in any data synthesis.	8
Summary measures	<a href="#">#13</a>	State the principal summary measures (e.g., risk ratio, difference in means).	7
Planned methods of analysis	<a href="#">#14</a>	Describe the methods of handling data and combining results of studies, if done, including measures of consistency (e.g., I <sup>2</sup> ) for each meta-analysis.	6-8
Risk of bias	<a href="#">#15</a>	Specify any assessment of risk of bias that may affect the	15-16

1	across studies		cumulative evidence (e.g., publication bias, selective reporting within studies).	
2				
3				
4	Additional	<a href="#">#16</a>	Describe methods of additional analyses (e.g., sensitivity or	n/a
5	analyses		subgroup analyses, meta-regression), if done, indicating which	
6			were pre-specified.	
7				
8				
9	<b>Results</b>			
10				
11	Study selection	<a href="#">#17</a>	Give numbers of studies screened, assessed for eligibility, and	8-9
12			included in the review, with reasons for exclusions at each stage,	
13			ideally with a <a href="#">flow diagram</a> .	
14				
15				
16				
17	Study	<a href="#">#18</a>	For each study, present characteristics for which data were	8-9
18	characteristics		extracted (e.g., study size, PICOS, follow-up period) and provide	
19			the citation.	
20				
21				
22	Risk of bias	<a href="#">#19</a>	Present data on risk of bias of each study and, if available, any	15-17
23	within studies		outcome-level assessment (see Item 12).	
24				
25				
26	Results of	<a href="#">#20</a>	For all outcomes considered (benefits and harms), present, for	13-15
27	individual studies		each study: (a) simple summary data for each intervention group	
28			and (b) effect estimates and confidence intervals, ideally with a	
29			forest plot.	
30				
31				
32				
33	Synthesis of	<a href="#">#21</a>	Present the main results of the review. If meta-analyses are	12-15
34	results		done, include for each, confidence intervals and measures of	
35			consistency.	
36				
37				
38	Risk of bias	<a href="#">#22</a>	Present results of any assessment of risk of bias across studies	15-17
39	across studies		(see Item 15).	
40				
41				
42	Additional	<a href="#">#23</a>	Give results of additional analyses, if done (e.g., sensitivity or	n/a
43	analysis		subgroup analyses, meta-regression [see Item 16]).	
44				
45				
46	<b>Discussion</b>			
47				
48	Summary of	<a href="#">#24</a>	Summarize the main findings, including the strength of evidence	18-21
49	Evidence		for each main outcome; consider their relevance to key groups	
50			(e.g., health care providers, users, and policy makers	
51				
52				
53	Limitations	<a href="#">#25</a>	Discuss limitations at study and outcome level (e.g., risk of bias),	21-22
54			and at review level (e.g., incomplete retrieval of identified	
55			research, reporting bias).	
56				
57				
58				
59				
60				

1	Conclusions	<a href="#">#26</a>	Provide a general interpretation of the results in the context of	23
2			other evidence, and implications for future research.	
3				
4				
5	<b>Funding</b>			
6				
7	Funding	<a href="#">#27</a>	Describe sources of funding or other support (e.g., supply of	24
8			data) for the systematic review; role of funders for the systematic	
9			review.	
10				
11				

12 The PRISMA checklist is distributed under the terms of the Creative Commons Attribution License  
13 CC-BY. This checklist was completed on 30. September 2019 using <https://www.goodreports.org/>, a  
14 tool made by the [EQUATOR Network](#) in collaboration with [Penelope.ai](#)  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

# BMJ Open

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

Journal:	<i>BMJ Open</i>
Manuscript ID	bmjopen-2019-034661.R1
Article Type:	Original research
Date Submitted by the Author:	03-Mar-2020
Complete List of Authors:	Aladwani , Mohammad; The University of Manchester, Division of Population Health, Health Services Research and Primary Care School of Health Sciences Faculty of Biology, Medicine and Health Lophatananon, Artitaya ; The University of Manchester, Division of Population Health, Health Services Research and Primary Care School of Health Sciences Faculty of Biology, Medicine and Health Ollier, William; The University of Manchester, Division of Population Health, Health Services Research and Primary Care School of Health Sciences Faculty of Biology, Medicine and Health; Manchester Metropolitan University Faculty of Science and Engineering, School of Healthcare Science Muir, Kenneth; The University of Manchester, Division of Population Health, Health Services Research and Primary Care School of Health Sciences Faculty of Biology, Medicine and Health
<b>Primary Subject Heading</b>:	Public health
Secondary Subject Heading:	Urology, Epidemiology
Keywords:	EPIDEMIOLOGY, UROLOGY, Prostate disease < UROLOGY

SCHOLARONE™  
Manuscripts





I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our [licence](#).

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which [Creative Commons](#) licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

1  
2  
3  
4 **Full title: Prediction models for prostate cancer to be used in the**  
5 **primary care setting: a systematic review.**  
6  
7

8 **Institutional addresses and e-mail addresses of the contributing authors:**  
9

10 Dr Mohammad Aladwani

11  
12 Division of Population Health, Health Services Research and Primary Care, School of  
13 Health Sciences Faculty of Biology, Medicine and Health, The University of Manchester,  
14 Manchester, M13 9PT, UK  
15  
16

17  
18 E-mail: [mohammad.al-adwani@postgrad.manchester.ac.uk](mailto:mohammad.al-adwani@postgrad.manchester.ac.uk)  
19

20 Dr Artitaya Lophatananon

21  
22 Division of Population Health, Health Services Research and Primary Care, School of  
23 Health Sciences Faculty of Biology, Medicine and Health, The University of Manchester,  
24 Manchester, M13 9PT,UK  
25  
26

27  
28 E-mail: [artitaya.lophananon@.manchester.ac.uk](mailto:artitaya.lophananon@.manchester.ac.uk)  
29

30 Professor William Ollier,

31  
32 Division of Population Health, Health Services Research and Primary Care, School of  
33 Health Sciences Faculty of Biology, Medicine and Health, The University of Manchester,  
34 Manchester, M13 9PT, UK  
35  
36

37  
38 School of Healthcare Science, Faculty of Science and Engineering, Manchester  
39 Metropolitan University, John Dalton Building, Manchester, M1 5GD, UK  
40  
41

42  
43 E-mail: [Bill.Ollier@manchester.ac.uk](mailto:Bill.Ollier@manchester.ac.uk)  
44

45 **Address and contact details of the corresponding author:**  
46

47 Professor Kenneth R. Muir

48  
49 Division of Population Health, Health Services Research and Primary Care,  
50 School of Health Sciences, Faculty of Biology, Medicine and Health,  
51 The University of Manchester, Manchester, M13 9PT, UK  
52  
53

54  
55 E-mail: [kenneth.muir@manchester.ac.uk](mailto:kenneth.muir@manchester.ac.uk)  
56  
57  
58  
59  
60

1  
2  
3 Telephone: +44-161-275-5518  
4  
5  
6  
7

8 Key words: Prostate cancer, prostate specific antigen, PSA, prostate cancer screening,  
9  
10 risk Prediction, risk tools, prediction models, primary care, community.  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

For peer review only

## 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 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 PROBAST tool.

### Data extraction and synthesis

Relevant information extracted for each model included; the year of publication, source of data, type of model, number of subjects, 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%. AUC was reported in four studies and ranged from 0.65 to 0.75. All models showed significant improvement in predicting PCa compared with

1  
2  
3 being based on PSA alone. The difference in AUC between extended models and PSA  
4 alone was between 0.06 and 0.21.  
5

## 6 7 **Conclusions**

8  
9 Only a few PCa risk prediction models have the potential to be readily used in the  
10 primary health care or community health setting. Further studies are needed to  
11 investigate other potential variables that could be integrated into models to improve their  
12 clinical utility for PCa testing in a community setting.  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

## Strengths and limitations

- The review focussed on risk prediction models for PCa for use in primary care.
- The PRISMA approach was followed in identifying relevant articles and reporting this study.
- We used the PROBAST tool 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.

## 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<sup>1</sup>. In the UK, around 47,200 new cases of PCa were reported in 2015, accounting for 26% of all new cancer cases in males. PCa deaths in the UK were around 11,600 in 2016<sup>2</sup>. The global projections of PCa incidence and mortality for 2030 are 1.7 and 0.5 million, respectively<sup>3</sup>. The highest incidence of PCa is seen in western societies<sup>4</sup>. The significant increase of PCa incidence and diagnosis over the last three decades can be attributed mainly to the widespread implementation of the PSA serum test after it had been introduced in the late 1980s<sup>5 6</sup>.

The strong association of PSA with PCa<sup>7 8</sup>, along with it being a relatively inexpensive test<sup>9</sup>, has made PSA a key biomarker in the diagnostic process of PCa and for the recommendation of a confirmatory prostate biopsy<sup>7 9</sup>. PSA is, however, not a cancer-specific marker<sup>5 10</sup>. 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<sup>11-14</sup>. 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<sup>5 10 15 16</sup>. Such issues are likely to be the part of the explanation for the significant number of unnecessary biopsies currently performed each year. Such procedures are associated with adverse side effects for patients and also increases health care costs<sup>17 18</sup>.

1  
2  
3 To address such PSA test limitations, researchers have incorporated other measurable  
4 factors into approaches for the early detection of PCa; these “risk assessment tools” are  
5 based on statistical models designed to improve the accuracy and performance of the  
6 PSA test <sup>19-22</sup>. Logistic regression and artificial neural network (ANNs) models are now  
7 considered to be the most common and effective statistical techniques in aiding  
8 development of new models to enhance early PCa diagnosis <sup>23</sup>. These PCa risk  
9 prediction models can be used to aid the testing of men for further investigations.

15  
16 Currently in the UK, there is no population based screening programme for PCa. The  
17 ultimate goal of PCa screening is to find intermediate and high risk of PCa rather than  
18 low risk PCa that would not require treatment but will give emotional burden to the  
19 patient once detected and unnecessary treatment in some patients. An important  
20 potential advantage of the extended risk models is their ability to provide a more  
21 accurate estimation of PCa risk. This may ultimately lead to their use in patient  
22 counselling and decision-making <sup>24-27</sup>. Such models have already achieved better  
23 results in predicting probabilities of outcome compare to clinical judgment <sup>28 29</sup>.  
24 Furthermore, it has been reported that using such predictive models may minimize the  
25 rate of unnecessary biopsies <sup>30</sup>.

33  
34 Recently there has been a substantial increase in the development of predictive models  
35 to help clinicians assess risk and decide which man to send to clinical setting to further  
36 investigate for a possible diagnosis of PCa <sup>22 26 30-35</sup>. The majority of these models are  
37 designed for use in clinical settings, where costs are less of an issue and most include  
38 the need for a clinical examination such as digital rectal examination (DRE) or trans-  
39 rectal ultrasonography (TRUS). One of the main limitations of DRE is that it has poor  
40 performance, especially at low PSA levels, and it is highly subjective to inter-observer  
41 variability <sup>36-38</sup>. A meta-analysis study revealed that DRE has positive predictive value of  
42 only 18% <sup>39</sup>. Similarly, TRUS has been reported for having poor accuracy at low PSA  
43 levels <sup>40 41</sup> and small PCa might not be palpable on DRE or visualisation on TRUS <sup>40</sup>.  
44 Furthermore, less than 40% of PCa detected by DRE are potentially curable, making it  
45 less beneficial for early diagnosis <sup>42</sup>. Several studies showed that there are fear, anxiety  
46 and embarrassment among some men, in particular black men, regarding the DRE test  
47 <sup>43-46</sup>. Another disadvantage of the DRE is the fact it is a potentially uncomfortable test <sup>47-</sup>

1  
2  
3 5<sup>2</sup>. This may explain why the DRE is a barrier for some men to participate in PCa  
4 screening if it is including DRE test. Lee et al reported that 74%-84% of black men may  
5 not maintain annual DRE screening <sup>53</sup>, while another study found that it may prevent  
6 22% of men to participate <sup>54</sup>. Since TRUS need to be performed by a skilled urologist,  
7 this means men have to make an appointment with a clinic in different location, which  
8 makes the screening less convenience. As a result, men may feel reluctant to have  
9 such tests performed.

10  
11  
12  
13  
14  
15  
16 This systematic review of the literature was undertaken to identify risk prediction models  
17 that do not incorporate invasive or more costly clinical procedures or extensive  
18 biomarkers but have potential application for use in primary care and community  
19 settings. As low cost is a primary concern for community use, for this review we set an  
20 indicative threshold of approximately 3-5 times the cost of a PSA test for inclusion. This  
21 excluded a number of models that contain new and emerging biomarker or SNP panels.  
22 As the numbers of person referred to clinical setting costs are less of an issue. The  
23 performance of the models reviewed for detecting PCa at initial biopsy have been  
24 compared using 'reported area under the curve' (AUC) and/or sensitivity - specificity  
25 testing.

## 33 **METHODS**

34  
35  
36 The approach used to identify and select relevant articles was based on the application  
37 of the '*Preferred Reporting Items for Systematic Reviews and Meta-Analysis*' (PRISMA)  
38 <sup>55</sup>.

### 41 **Data sources and search strategy**

42  
43 A literature search was performed using Medline (via OVID) and Embase databases.  
44 The 'medical subject heading' (MeSH) terms, combined with Boolean logic operators  
45 'AND' and 'OR', were applied to retrieve relevant articles. The terms used for the search  
46 were "Prostatic Neoplasms" AND ("Initial biopsy" OR "first biopsy" OR "early detection of  
47 cancer") AND ("nomograms" OR "artificial neural networks" OR "risk assessment" OR  
48 "statistical model"). The full search strategy is provided in a supplementary file (S1). All  
49 articles defined (published since the inception of the databases and up to the end of  
50 January 2019) were subsequently further filtered as being those only published in  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60



1  
2  
3 English language and with an abstract. Further to using the above search databases,  
4 the research articles were selected manually from the reference lists of any relevant  
5 review articles. Google Scholar and Medline searches were also carried out to identify  
6 independent study for external validation for each model included in this review. The  
7 results are presented in supplementary table (S2).  
8  
9

### 11 **Eligibility criteria**

12  
13  
14 As this review focuses on PCa risk prediction based in community healthcare settings,  
15 all studies were selected on the following inclusion criteria (i) evaluating the risk for PCa  
16 at initial biopsy in subjects who had no prior history of PCa (ii) studies that reported “low  
17 cost” risk assessment tools (i.e. those not including more expensive genetic, biomarker  
18 test) or “invasive” clinical tests/examinations (such as DRE or TRUS) (iii) studies that  
19 included a minimum of two variables of which PSA had to be one of them [on the basis  
20 that an elevated PSA test in UK primary care is usually the first sign and rationale for  
21 suggesting a need for further investigation of PCa within NICE guidelines] and (iv)  
22 studies that reported AUC and/or sensitivity and specificity of the diagnostic/predictive  
23 tool. The exclusion criteria used were (i) articles with models that were built and based  
24 on repeat or mixed biopsies (ii) studies that only validate an existing model (iii) articles  
25 that were not published in English. There were no time boundaries regarding the  
26 publication year.  
27  
28  
29

30  
31 Screening of the titles, abstracts, and full-texts was carried out by two reviewers (MA,  
32 AL). Any concerns about the eligibility of a study were resolved by discussion with a  
33 third reviewer (KM).  
34  
35

### 36 **Data extraction**

37  
38 A data extraction form was developed to collect all relevant information. For each study  
39 used in this review, the items extracted included; year of publication, source of data,  
40 type of model, number of subjects, country where it was performed, age, PSA range,  
41 mean/median PSA, number of biopsy cores, variables included in the model, study  
42 endpoint(s), cancer detection, model performance, and model validation.  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

## 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 <sup>27</sup>. Discrimination refers to how well the prediction model can differentiate subjects in different outcome classes according to their predicted risks. It is often assessed by measuring the area under the receiver operating characteristic (ROC) curve <sup>56</sup>. 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 <sup>57</sup>. Calibration represents the agreement between expected and observed outcomes <sup>58</sup>. 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 under-estimates low risks and over-estimate high risks <sup>59</sup>.

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 PROBAST tool <sup>60</sup>. 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 to reach overall judgment of risk of bias, as well as issues relating to applicability.

## 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 Fig.1. An additional nine articles were identified through manual searches from a bibliography

1  
2  
3 of reviewed articles. At the first filter step, a total of 109 titles and abstracts were  
4 screened for eligibility after removing two duplicates. In the second filter step, 60 papers  
5 failed to meet the inclusion criteria and were excluded, resulting in 49 articles. The final  
6 step of filtering yielded only five studies that were considered to be eligible (i.e. passed  
7 all set criteria) and were thus included in this systematic review. There is no  
8 independent study identified for external validation for included models.  
9

### 14 **Study characteristics**

15  
16 Four of the five studies included were based on cohort studies and one was a case-  
17 control study. The characteristics of each of these studies and populations are  
18 summarised in Table 1. Details of PSA assays used in the models are presented in  
19 supplementary table (S3).  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

Table 1: Characteristics of the included studies.

Author & 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, 1998 <sup>61</sup>	Logistic regression	Cohort	Model dev = 3773 Validation = 525	Baltimore, USA	Referral	≥ 45	—	4-20 ng/ml	—	Sextant biopsy	32%
Babaian, 2000 <sup>62</sup>	Neural network = 3 ANNs	Cohort	151	Texas, USA	Screening program	40-75	62	2.5-4 ng/ml	—	11 cores	24.5%
Jansen, 2010 <sup>63</sup>	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 program	≥ 50	Site 1 (66) Site 2 (60)	2-10 ng/ml	~ 4.4	≥6 cores	Site 1 = 55.8% Site 2 = 49.6%
Hill, 2013 <sup>64</sup>	Logistic regression	Case-control	1378	Florida, USA	Hospital referral	40-90	—	≥ 4 ng/ml	—	N/A	20.6%
Lazzeri, 2013 <sup>65</sup>	Logistic regression	Cohort	646	European multi-centre; Italy, Germany, France, Spain, and the UK	Referral	>45	—	2-10 ng/ml	~ 5.8	≥ 12 cores	40.1%

1  
2  
3 Subjects used to build the risk models varied across these studies. Of the five studies,  
4 three studies included men from referral populations <sup>61 64 65</sup> and two studies from  
5 screening programs <sup>62 63</sup>. The sample sizes ranged from 151 to 3,773 with three studies  
6 derived from US cohorts <sup>61 62 64</sup> and two from Europe <sup>63 65</sup>. Four studies used logistic  
7 regression methodology to build their model, whereas one study used an artificial neural  
8 network-based approach (ANN) <sup>62</sup>. The minimum age of participants was 40 years old <sup>62</sup>  
9 <sup>64</sup> and the minimum PSA level was 2 ng/ml <sup>63 65</sup>.

### 16 17 **Variables in the model**

18  
19 Table 2 presents details of the variables used in each model. PSA level was used in all  
20 models, followed by free PSA (fPSA), age, and free-to-total PSA ratio (%fPSA). Other  
21 variables also reported in the models included; precursor of prostate-specific antigen  
22 (p2PSA), percentage of p2PSA to free PSA (%p2PSA), prostate health index (PHI),  
23 levels of haemoglobin, albumin, creatinine and red blood cell count (RBC), haematuria,  
24 mean corpuscular volume (MCV), and prostatic acid phosphatase.  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

Table 2: Variables used in the prostate cancer risk prediction models\*.

Author & year	Variables used in the model				
	Total PSA	Free PSA	Percent free PSA	Age	Other variables
Carlson, 1998 <sup>61</sup>	✓		✓	✓	
Babaian, 2000 <sup>62</sup>	✓	✓		✓	Creatinine Kinase, Prostatic acid phosphatase
Jansen, 2010 <sup>63</sup> (Site 1)	✓	✓			p2PSA
Jansen, 2010 <sup>63</sup> (Site 2)	✓	✓			p2PSA
Hill, 2013 <sup>64</sup> (Method 1)	✓			✓	HGB, RBC, Haematuria, Creatinine, MCV, and ethnicity "Black"
Hill, 2013 <sup>64</sup> (Method 2)	✓			✓	HGB, RBC, Creatinine, and MCV
Lazzeri, 2013 <sup>65</sup> (Model 1)	✓	✓	✓		
Lazzeri, 2013 <sup>65</sup> (Model 2)	✓	✓	✓		p2PSA
Lazzeri, 2013 <sup>65</sup> (Model 3)	✓	✓	✓		%p2PSA
Lazzeri, 2013 <sup>65</sup> (Model 4)	✓	✓	✓		PHI

\*HGB= Haemoglobin. RBC= Red blood cells. MCV= Mean corpuscular volume. PHI= Prostate Health Index. p2PSA= Precursor of prostate specific antigen, %p2PSA= p2PSA/fPSA

## Outcome

The study endpoint also varied among the studies selected. Two studies evaluated the accuracy of detecting any PCa<sup>61 62</sup> and three studies examined the pathologic Gleason score<sup>63-65</sup>. Although Jansen et al, did not build a model to predict the aggressiveness of prostate cancer, they assessed the relationship of each variable individually with a Gleason score  $\geq 7$ . PCa was determined by taking a needle biopsy. All subjects in the five studies underwent prostate biopsy. The least number of biopsy cores used were 6<sup>61 63</sup> and the highest were  $\geq 12$ <sup>65</sup>. One study did not report the number of biopsy cores taken<sup>64</sup>. 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 to PSA alone (AUC of 0.755 vs. 0.585, respectively)<sup>63</sup>. 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 percent free PSA alone with 95% sensitivity<sup>61</sup>. Lazzeri et al,<sup>65</sup> presented results from four separated models discriminating PCa with a Gleason score of  $\geq 7$ . Lazzeri's model 2 (which includes the base model tPSA, 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 4 models with an AUC of 0.67. In the study of Hill<sup>64</sup> 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 to 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 (PIN), 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 to 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.

Table 3: The difference of area under the curve (AUC) for PSA alone and extended model.

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

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 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 4: Sensitivity and specificity profile at different levels for each model.

Study	Sensitivity	Specificity	Probability cut-off	Positive predictive value	Negative predictive value
Carlson <sup>61</sup>	99	18	>15	≤47	N/A
	95	34	18	51	NA
	89	43	20	42	NA
Babaian <sup>62</sup>	95	51	NA	39	97
	92	62	NA	44	96
	89	62	NA	43	95
Jansen <sup>63</sup> (Site 1)	95	23.9	NA	NA	NA
Jansen <sup>63</sup> (Site 1)	90	30.1	NA	NA	NA
Jansen <sup>63</sup> (Site 2)	95	23.2	NA	NA	NA
Jansen <sup>63</sup> (Site 2)	90	36.2	NA	NA	NA
Hill <sup>64</sup> (Method 1)	90.9	17.6	33	47.1	70.5
Hill <sup>64</sup> (Method 2)	89.8	28	13	20.6	91.3
Hill <sup>64</sup> (Method 1)	80.5	37.1	37	50.9	70.2
Hill <sup>64</sup> (Method 2)	78.2	45	15	28.7	88.8
Hill <sup>64</sup> (Method 1)	39.9	81.4	48	63.4	62.6
Hill <sup>64</sup> (Method 2)	45.8	79.5	23	36.7	85
* Lazzeri <sup>65</sup> model reported only sensitivity and specificity for predictive variables individually and at sensitivity of 90, %p2PSA and %fPSA achieved the highest specificity					

Table 5 summarises the validation and calibration results for the studies included. Model calibration was reported in two studies <sup>61</sup> <sup>65</sup>. 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 <sup>63</sup> <sup>64</sup>. Only one study reported an external validation using an additional data set consisting of 525 patients <sup>61</sup>. 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 <sup>62</sup>. Lazzeri used 200 bootstrap re-samples to minimise overfitting bias <sup>65</sup>.

Table 5: Validation and calibration for included models.

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

### 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 tool, particularly in the analysis domain. For instance, one study applied univariable analysis to select predictors <sup>61</sup>. Three studies did not measure calibration <sup>62-64</sup>. Furthermore, two studies did not account for optimism and overfitting by using internal validation methods <sup>63 64</sup>. Whereas one study did not use appropriate measures for model performance i.e. AUC, this study reported the calibration <sup>61</sup>.

The event per variable (EPV) was lower than recommended ( $< 10$ ) <sup>59 66</sup> in the Babaian <sup>62</sup> study indicating inadequate power. Four studies did not report missing data or how they handled it <sup>61-63 65</sup>. The remaining study used complete-case analysis and excluded subjects with missing data on laboratory biomarkers ( $n=75$ ) <sup>64</sup>. The PROBAST

1  
2  
3 guidelines state that in a prediction model study where any risk of bias and applicability  
4 is low in all four domains, a regrading to high risk of bias should be considered when the  
5 study did not validate the model externally <sup>60</sup>. Thus, although the quality assessment for  
6 the Lazzeri study <sup>65</sup> was graded low risk in all the four domains, since the study did not  
7 report any external validation of the model, the assessment of the study has been  
8 regraded to high risk of bias according to the PROBAST criteria. A full quality  
9 assessment for all studies is presented in Table 6.  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

Table 6: Quality assessment for risk of bias and applicability concern for included studies.

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

\* ROB- Risk of bias

+ indicates a low risk of bias or applicability; - indicates a high risk of bias or applicability.

## 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 model<sup>61</sup> has the largest population (3,773 subjects) when compared to 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<sup>33</sup>. More importantly, by not reporting the AUC estimate, the model raises some doubts regarding the reliability of the model and its implementation<sup>33</sup>. It will also make comparison to other models not applicable<sup>67</sup>.

Babaian<sup>62</sup> 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, PSAD-TZ= 0.75). They included a number of uncommon pre-biopsy inputs into their algorithm such as prostatic acid phosphatase and creatinine kinase<sup>68</sup>. Furthermore, they used a tight PSA range (2.5-4.0ng/ml) which meant that their model may be less suitable for patients with PSA level below or above that range, thus limiting its generalisability.

1  
2  
3 The study by Jansen and colleagues<sup>63</sup> demonstrated that adding p2PSA to the base  
4 model of PSA and fPSA significantly enhanced the PCa predictive value and specificity.  
5 The association and added value of p2PSA in the prediction and detection of PCa have  
6 been reported by several other studies<sup>16 69-72</sup>. Jansen<sup>63</sup> showed that p2PSA has no  
7 clear association with aggressive PCa. However, the base model that includes p2PSA  
8 had the highest clinical significance in correlation to pathologic Gleason score with a p-  
9 value of 0.008 compared to %fPSA and PHI (p-value 0.01 and 0.02) respectively.  
10 Although they used archived blood samples and retrospective analysis, the results were  
11 similar to a prospective study of 268 patients<sup>16</sup>.

12  
13 Hill<sup>64</sup> used a case-control study to evaluate several laboratory biomarkers. They found  
14 HGB, RBC, haematuria, creatinine, PSA, Age, MCV, and ethnicity ("being black") were  
15 statistically significantly associated in the first method ( $p < 0.05$ ). In the second method,  
16 HGB, RBC, creatinine, PSA, age, and MCV were found to be statistically significantly  
17 correlated ( $p < 0.001$ ) with PCa. However, since this study was designed as a case-  
18 control study, it would have been more prone to uncontrolled confounding and selection  
19 bias. Moreover, the type of screening protocols used in Veterans' Administrations may  
20 vary to those conducted in other healthcare systems; therefore, the results may not be  
21 applicable to other populations. Furthermore, subjects with a PSA level  $< 4.0$  ng/ml  
22 have not been investigated, and thus, the performance of the models are unknown for  
23 individuals in this group.

24  
25 Lazzeri et al,<sup>65</sup> in a European multi-centre study have evaluated similar biomarkers as  
26 in Jansen study with the same PSA range 2-10 ng/ml prospectively. They found no  
27 difference in both %p2PSA and PHI as individual PCa predictors with AUC of 0.67 (95%  
28 CI 0.64-0.71). However, the base model (consisting of PSA, fPSA, and %fPSA) that  
29 also included either p2PSA or PHI out-performed the base model alone and the base  
30 model including %p2PSA. In the analysis, the additive value of both p2PSA and PHI is  
31 0.064 and 0.056 for %p2PSA for predicting the risk of PCa. These additive values  
32 increased to 0.076 for both p2PSA and PHI and 0.073 for %p2PSA in predicting  
33 Gleason scores  $\geq 7$  for the disease. The usefulness of PHI in improving the predictive  
34 accuracy of PCa over total and free PSA has been confirmed and reported by several  
35 studies<sup>16 72-75</sup>.

1  
2  
3 In general, only one study has validated their model externally <sup>61</sup>, whereas the  
4 remaining studies were either validated internally <sup>62 65</sup> or not reported any validation  
5 methods <sup>63 64</sup>. Prediction models may not be equally applicable to all data sets as  
6 patients' characteristics may vary <sup>20 76</sup>. As a result, the generalisability of a model might  
7 be poor when it used in populations other than that used in building the model.  
8 Therefore, external validation should be conducted before applying any new model into  
9 general practice <sup>77 78</sup>.

10  
11  
12 Another key performance measure of any model that needs careful evaluation is  
13 calibration. A calibration plot with a calibration slope is more preferable than the  
14 Hosmer-Lemeshow test; it has been acknowledged that evaluating a good and well  
15 calibrated model based on a large dataset can still fail the Hosmer-Lemeshow test. In  
16 contrast, when evaluating a poorly calibrated model with a small dataset it can still pass  
17 the Hosmer-Lemeshow test <sup>79</sup>. In our analysis, three studies fail to report the calibration  
18 of the model <sup>62-64</sup> whilst the Carlson study <sup>61</sup> used a calibration plot and Lazzeri <sup>65</sup> used  
19 the Hosmer-Lemeshow test. Excluding calibration from the majority of models may  
20 explain why some models are not currently used in practice <sup>79</sup>.

21  
22  
23 With regard to biopsy cores, only two studies used extended biopsy cores. Babaian <sup>62</sup>  
24 used an 11-core multi-site biopsy, whereas Lazzeri <sup>65</sup> used at least 12 biopsy cores.  
25 Moreover, two studies used six cores biopsy in their model <sup>61 63</sup>. The use of six-core  
26 biopsy has been criticized as not being adequate in detecting PCa <sup>80</sup> and that models  
27 developed using sextant biopsy are less accurate than when a 10-core biopsy is used  
28 <sup>81</sup>. As a result, the European guideline for clinical PCa recommended an extended  
29 biopsy as standard practice for PCa detection <sup>82</sup>.

30  
31  
32 It is worth noting that all five reviewed models performed better than just PSA alone.  
33 However, none of them has both high specificity and sensitivity. The level of sensitivity  
34 has been increased, and despite enhancement in the specificity, it is still considered  
35 low. Specificity is crucial when it comes to being used in a population setting as men  
36 without PCa should be ruled out as much as possible from further invasive engagement  
37 with the health system.  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 Our review therefore suggests that none of the reviewed models provide an ideal  
4 performance in predicting PCa with high sensitivity and high specificity. It is particularly  
5 important when considering the application of PCa risk prediction at the population level  
6 that the tool used should be able to both detect the outcome and filter out people with  
7 no disease. As there is robust evidence suggesting that the clinical relevance of PSA  
8 range to the detection of PCa differs across age groups<sup>83-85</sup>, any future model should  
9 consider PSA threshold in relation to a specific age range. Risk prediction models for  
10 PCa should therefore take account of age.  
11  
12

### 13 **Comparison with other studies**

14  
15 To our knowledge, three systematic reviews of PCa prediction tools have been  
16 published<sup>20 26 27</sup>. In the Louie et al. review, risk models were included that were  
17 externally validated in at least five study populations for the purpose of meta-analysis  
18 and only six studies were included in their analysis. Furthermore, all the studies  
19 included incorporated clinical tests such as DRE and/or TRUS-PV<sup>26</sup>. Schroder and  
20 Kattan<sup>20</sup> reviewed models that were built to predict the likelihood of having a positive  
21 prostate biopsy for cancer. However, it appears that they also included models where  
22 subjects had a previous negative biopsy. As such, some of the models included  
23 variables related to biopsy results and cores. The review by Shariat and colleagues  
24 examined different types of predictive tools<sup>27</sup>. They explored tools that predict PCa on  
25 initial and repeat biopsy, pathologic stages, biochemical recurrence after radical  
26 proctectomy, metastasis, survival, and life expectancy. Similarly, virtually all of the  
27 prediction tools that were based on initial biopsy included variables based on invasive  
28 procedures.  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43

### 44 **Strengths and limitations of this study**

45  
46 This report is the first to review risk prediction tools for PCa that can be used in primary  
47 care and community settings. Any prediction model should therefore be simple to use,  
48 based on non-invasive tests, be feasible at a population level and at low cost. We  
49 carried out an extensive data extraction relating to important features and  
50 characteristics for each study included, such as modelling method, source of data,  
51 sample size, variables, discrimination, validation, and cancer detection rate. We have  
52  
53  
54  
55  
56  
57  
58  
59  
60



1  
2  
3 also followed PRISMA guidelines for identifying eligible articles as well as for reporting  
4 this study. In addition, the PROBAST tool was adopted to assess the quality and risk of  
5 bias for each prediction model.  
6  
7

8  
9 Our study has some limitations. Our aim was to identify prediction models that have the  
10 potential to be implemented in a primary care or community settings, and consequently  
11 our search strategy was to retrieve relevant studies for this specific purpose.  
12 Furthermore, we excluded articles that were not published in English or did not have an  
13 abstract. Moreover, only two databases were searched, besides manual search, to  
14 retrieve original studies.  
15  
16  
17  
18

19 A previous systematic review suggested that the majority of relevant studies could be  
20 identified through a manual search of articles reference lists instead of a database alone  
21 <sup>20</sup>. We identified four eligible studies using this approach. Given the small number of  
22 models identified by the approach we followed, that can be applied in primary care  
23 settings compared to the large number relating to wider existing models, it is unlikely  
24 that we have not included any study that would affect the results of our review.  
25  
26  
27  
28  
29

### 30 **Implications and future research**

31  
32 It is now accepted that the PSA test and its derivatives have some limitations for  
33 detecting PCa as defined by subsequent biopsy <sup>86</sup>. As a consequence, a considerable  
34 number of PCa prediction models have been built to improve prediction accuracy. This  
35 has resulted in a plethora of PCa risk prediction tools, with to date more than 100  
36 models described <sup>87 88</sup>. There is evidence that some of these models show benefit and  
37 have better performance over just PSA measurement alone <sup>20</sup>. It also has been  
38 demonstrated that some of these models out-performed clinical experts in predicting  
39 PCa <sup>28 29</sup>. Although such models are not designed to replace specialist clinical judgment  
40 or patient preferences <sup>76 86 89</sup>, they can help in patient counselling and aid clinicians to  
41 decide whether a prostate biopsy should be taken or not <sup>77 89 90</sup>.  
42  
43  
44  
45  
46  
47  
48  
49

50 Given the small number of risk prediction models for PCa that do not incorporate clinical  
51 or genetic tests, the discrimination of these reviewed models ranged between poor to  
52 moderate (AUC range ~0.65 to ~0.75); in addition there were some issues relating to  
53 their study design and analysis raises the risk of bias. Consequently none of these  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 models could be currently recommended for use in a primary care and community  
4 health care setting. Several guidelines are against using PSA test based screening for  
5 PCa; the US Preventive Services task force, the Canadian task force on preventive  
6 health and the American College of Preventive Medicine do not currently recommend  
7 PSA-based testing due to insufficient evidence<sup>91-93</sup>. This has made it difficult, so far, to  
8 convince policymakers to adopt PCa screening programme.  
9

10  
11  
12  
13  
14 The first guideline of PROSTATE CANCER UK states *"In the future, health*  
15 *professionals should look at a man's PSA level alongside other known risk factors as*  
16 *part of a risk assessment tool, when one becomes available."*<sup>94</sup> However, the vast  
17 majority of the current PCa risk prediction models are not suitable for routine use as  
18 they include clinical and genetic tests and are not validated externally in other cohorts.  
19 Therefore, the main challenge in the UK, remains to develop a risk prediction tool that is  
20 reliable, cheap, is applicable for as wide an ethnicity as possible, and most importantly  
21 is easy to use and can be implemented at a primary care level<sup>95</sup>.  
22  
23  
24  
25  
26  
27

28  
29 The value of such risk tools is that they will help to stratify men at high risk of developing  
30 PCa earlier so that they have appropriate management and/or surveillance programme  
31 as early as possible and therefore, may fit into the clinical pathway. Such tools should  
32 help physicians have a better understanding of the risk for this disease and simplify the  
33 procedures and discussions with patients when recommending further specialist-led  
34 investigations such as DRE and/or MRI where a decision on whether a biopsy should or  
35 not perform is concluded. Furthermore, using the appropriate risk prediction tool will  
36 avoid men from undergoing inappropriate further and frequent testing<sup>95</sup>. This will  
37 reduce any associated costs of inappropriate tests and decrease the burden on health  
38 care delivery systems.  
39  
40  
41  
42  
43  
44  
45

46  
47 It is crucial to address these issues by identify all possible risk factors for PCa that are  
48 non-clinical, non-genetic, and easy to use and interpret. There remains a pressing need  
49 to develop a risk prediction tool in the future using all appropriate factors (potentially  
50 also including genetics once there is infrastructure in place for genetic testing in the  
51 primary care and the cost comes down) into a robust multivariable analysis and validate  
52 the model externally to eliminate applicability and generalisability concerns. Only when  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 this is achieved will it be possible to introduce a prostate cancer screening programme  
4 fit for purpose.  
5

## 6 7 **CONCLUSION**

8  
9 There is a paucity of suitable low cost risk models that incorporate non-clinical, non-  
10 genetic inputs and which can be used at a primary care level and in other community  
11 health services. Existing models have limitations reflecting both study design and  
12 reporting performance measures. Future research should take into account these key  
13 issues and explore other risk factors for incorporation into further models.  
14  
15  
16  
17

## 18 **ACKNOWLEDGEMENT**

19  
20 MA is supported by Kuwait's Ministry of Health. KM and AL are supported by the NIHR  
21 Manchester Biomedical Research Centre and by the ICEP ("This work was also supported by  
22 CRUK [grant number C18281/A19169]). KM is also supported by the Allan Turing Institute.  
23  
24  
25

## 26 **AUTHOR CONTRIBUTIONS**

27  
28 MA, AL and KM were involved in study conception, idea and design. Data acquisition and  
29 extraction was obtained by MA. Data synthesis and interpretation was carried out by MA, AL  
30 and WO. MA and AL drafted the manuscript. All authors approved the final version of  
31 manuscript. KM is the study guarantor.  
32  
33

## 34 **FUNDING**

35  
36 No funding is required for conducting this study.  
37  
38

## 39 **COMPETING INTERESTS**

40  
41 All authors have completed the ICMJE uniform disclosure form at  
42 [www.icmje.org/coi\\_disclosure.pdf](http://www.icmje.org/coi_disclosure.pdf) and declare: no support from any organisation for the  
43 submitted work; no financial relationships with any organisations that might have an interest in  
44 the submitted work in the previous three years; no other relationships or activities that could  
45 appear to have influenced the submitted work.  
46  
47  
48

## 49 **ETHICAL APPROVAL**

50  
51 Not applicable as this study did not involve any personal data.  
52  
53  
54  
55  
56  
57  
58  
59  
60

## DATA SHARING

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

## Figure legend

Figure 1: Flow diagram of studies included using PRISMA method

## REFERENCES

1. 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: a cancer journal for clinicians* 2018;68(6):394-424.
2. Cancer Research UK. Prostate cancer statistics 2017 [Available from: <https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/prostate-cancer#heading-One>].
3. Center MM, Jemal A, Lortet-Tieulent J, et al. International variation in prostate cancer incidence and mortality rates. *European urology* 2012;61(6):1079-92.
4. 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(6):1388-400.
5. Dijkstra S, Mulders P, Schalken J. Clinical use of novel urine and blood based prostate cancer biomarkers: a review. *Clin Biochem* 2014;47(10-11):889-96.
6. Fontanella P, Benecchi L, Grasso A, et al. Decision-making tools in prostate cancer: from risk grouping to nomograms. *Minerva Urol Nefrol* 2017;69(6):556-66.
7. Stephan C, Ralla B, Jung K. Prostate-specific antigen and other serum and urine markers in prostate cancer. *Biochim Biophys Acta* 2014;1846(1):99-112.
8. Thompson IM, Pauler DK, Goodman PJ, et al. Prevalence of prostate cancer among men with a prostate-specific antigen level  $\leq$  4.0 ng per milliliter. *N Engl J Med* 2004;350(22):2239-46.
9. 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(1):18-26.
10. 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(1):18-29.
11. 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(2):293-306.
12. 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(4 Pt 1):1297-301.

13. Bunting PS. A guide to the interpretation of serum prostate specific antigen levels. *Clin Biochem* 1995;28(3):221-41.
14. 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(5):1650-55.
15. Mistry K, Cable G. Meta-analysis of prostate-specific antigen and digital rectal examination as screening tests for prostate carcinoma. *The Journal of the American Board of Family Practice* 2003;16(2):95-101. [published Online First: 2003/04/01]
16. 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(2):214-22.
17. Heijnsdijk EA, der Kinderen A, Wever EM, et al. Overdetection, overtreatment and costs in prostate-specific antigen screening for prostate cancer. *British journal of cancer* 2009;101(11):1833-8. doi: 10.1038/sj.bjc.6605422 [published Online First: 2009/11/12]
18. 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(3):359-62. doi: 10.1016/j.eururo.2010.10.036 [published Online First: 2010/11/03]
19. 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(4):618-29. doi: 10.1016/j.eururo.2016.08.003 [published Online First: 2016/08/30]
20. Schroder 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(2):274-90.
21. Ross PL, Scardino PT, Kattan MW. A catalog of prostate cancer nomograms. *J Urol* 2001;165(5):1562-8. [published Online First: 2001/05/09]
22. 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. *Clinical cancer research : an official journal of the American Association for Cancer Research* 2008;14(14):4400-7. doi: 10.1158/1078-0432.Ccr-07-4713 [published Online First: 2008/07/17]
23. Hu X, Cammann H, Meyer HA, et al. Artificial neural networks and prostate cancer--tools for diagnosis and management. *Nature reviews Urology* 2013;10(3):174-82. doi: 10.1038/nrurol.2013.9 [published Online First: 2013/02/13]
24. Caras RJ, Sterbis JR. Prostate cancer nomograms: a review of their use in cancer detection and treatment. *Current urology reports* 2014;15(3):391. doi: 10.1007/s11934-013-0391-0 [published Online First: 2014/01/24]
25. Iremashvili V, Soloway MS, Pelaez L, et al. Comparative validation of nomograms predicting clinically insignificant prostate cancer. *Urology* 2013;81(6):1202-8. doi: 10.1016/j.urology.2013.01.062 [published Online First: 2013/04/09]

- 1
  - 2
  - 3
  - 4
  - 5
  - 6
  - 7
  - 8
  - 9
  - 10
  - 11
  - 12
  - 13
  - 14
  - 15
  - 16
  - 17
  - 18
  - 19
  - 20
  - 21
  - 22
  - 23
  - 24
  - 25
  - 26
  - 27
  - 28
  - 29
  - 30
  - 31
  - 32
  - 33
  - 34
  - 35
  - 36
  - 37
  - 38
  - 39
  - 40
  - 41
  - 42
  - 43
  - 44
  - 45
  - 46
  - 47
  - 48
  - 49
  - 50
  - 51
  - 52
  - 53
  - 54
  - 55
  - 56
  - 57
  - 58
  - 59
  - 60
26. Louie KS, Seigneurin A, Cathcart P, et al. Do prostate cancer risk models improve the predictive accuracy of PSA screening? A meta-analysis. *Annals of Oncology* 2015;26:848-64. doi: 10.1093/annonc/mdu525
27. Shariat SF, Kattan MW, Vickers AJ, et al. Critical review of prostate cancer predictive tools. *Future oncology (London, England)* 2009;5(10):1555-84. doi: 10.2217/fon.09.121 [published Online First: 2009/12/17]
28. Ross PL, Gerigk C, Gonen M, et al. Comparisons of nomograms and urologists' predictions in prostate cancer. *Seminars in urologic oncology* 2002;20(2):82-8. [published Online First: 2002/05/16]
29. 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(6):1254-8.
30. 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. *The Prostate* 2012;72(11):1200-6. doi: 10.1002/pros.22470 [published Online First: 2012/01/04]
31. 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(4):709-13.
32. 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  $\leq$  10 ng/mL. *Cancer* 2003;98(7):1417-22. doi: 10.1002/cncr.11668 [published Online First: 2003/09/26]
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(6):1930-4.
34. Kranse R, Roobol M, Schroder FH. A graphical device to represent the outcomes of a logistic regression analysis. *The Prostate* 2008;68(15):1674-80. doi: 10.1002/pros.20840 [published Online First: 2008/08/21]
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(8):529-34.
36. Schröder FH, Kruger AB, Rietbergen J, et al. Evaluation of the digital rectal examination as a screening test for prostate cancer. *Journal of the National Cancer Institute* 1998;90(23):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). *The Prostate* 2001;47(4):252-61.
38. Smith DS, Catalona WJ. Interexaminer variability of digital rectal examination in detecting prostate cancer. *Urology* 1995;45(1):70-74.



- 1
  - 2
  - 3
  - 4
  - 5
  - 6
  - 7
  - 8
  - 9
  - 10
  - 11
  - 12
  - 13
  - 14
  - 15
  - 16
  - 17
  - 18
  - 19
  - 20
  - 21
  - 22
  - 23
  - 24
  - 25
  - 26
  - 27
  - 28
  - 29
  - 30
  - 31
  - 32
  - 33
  - 34
  - 35
  - 36
  - 37
  - 38
  - 39
  - 40
  - 41
  - 42
  - 43
  - 44
  - 45
  - 46
  - 47
  - 48
  - 49
  - 50
  - 51
  - 52
  - 53
  - 54
  - 55
  - 56
  - 57
  - 58
  - 59
  - 60
39. Barocas DA, Cowan JE, Smith JA, Jr., et al. What percentage of patients with newly diagnosed carcinoma of the prostate are candidates for surveillance? An analysis of the CaPSURE database. 2008
40. Vis A, Kranse R, Roobol M, et al. Serendipity in detecting disease in low prostate-specific antigen ranges. *BJU international* 2002;89(4):384-89.
41. Littrup PJ, Bailey SE. Prostate cancer: the role of transrectal ultrasound and its impact on cancer detection and management. *Radiologic Clinics of North America* 2000;38(1):87-113.
42. Gerber GS, Chodak GW. Routine screening for cancer of the prostate. *JNCI: Journal of the National Cancer Institute* 1991;83(5):329-35.
43. Gelfand DE, Parzuchowski J, Cort M, et al. Digital rectal examinations and prostate cancer screening: attitudes of African American men. *Oncology Nursing Forum-Oncology Nursing Society* 1995;22(8):1253-58.
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: Interdisciplinary International Journal of the American Cancer Society* 1996;78(3):471-79.
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. *American journal of men's health* 2007;1(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 Epidemiology and Prevention Biomarkers* 2006;15(2):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. *Journal of health care for the poor and underserved* 2013;24(1):311-31.
48. Ogunsanya ME, Brown CM, Odedina FT, et al. Beliefs regarding prostate cancer screening among black males aged 18 to 40 Years. *American journal of men's health* 2017;11(1):41-53.
49. Clarke-Tasker VA. What we thought we knew: African American males' perceptions of prostate cancer and screening methods. *ABNF Journal* 2002;13(3):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. *Archivos espanoles de urologia* 2008;61(7):850-54.
51. Essink-Bot M-L, de Koning HJ, Nijs HG, et al. Short-term effects of population-based screening for prostate cancer on health-related quality of life. *Journal of National Cancer Institute* 1998;90(12):925-31.
52. Appleton L, Wyatt D, Perkins E, et al. The impact of prostate cancer on men's everyday life. *European journal of cancer care* 2015;24(1):71-84.

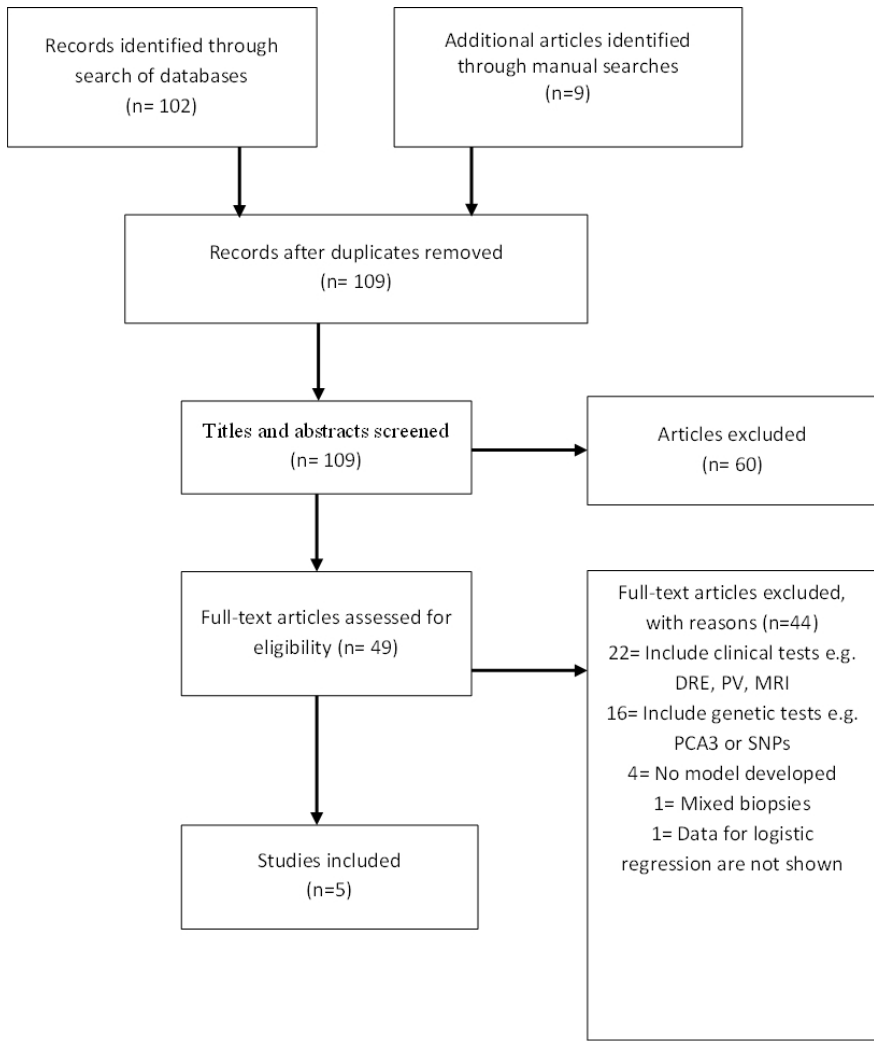
- 1  
2  
3 53. Lee DJ, Consedine NS, Spencer BA. Barriers and facilitators to digital rectal examination  
4 screening among African-American and African-Caribbean men. *Urology*  
5 2011;77(4):891-98.  
6
- 7 54. Nagler HM, Gerber EW, Homel P, et al. Digital rectal examination is barrier to population-  
8 based prostate cancer screening. *Urology* 2005;65(6):1137-40.  
9
- 10 55. Moher D, Liberati A, Tetzlaff J, et al. Preferred reporting items for systematic reviews and  
11 meta-analyses: the PRISMA statement. *PLoS medicine* 2009;6(7):e1000097.  
12
- 13 56. Park SH, Goo JM, Jo CH. Receiver operating characteristic (ROC) curve: practical review  
14 for radiologists. *Korean journal of radiology* 2004;5(1):11-8. doi:  
15 10.3348/kjr.2004.5.1.11 [published Online First: 2004/04/06]  
16
- 17 57. Han K, Song K, Choi BW. How to Develop, Validate, and Compare Clinical Prediction  
18 Models Involving Radiological Parameters: Study Design and Statistical Methods.  
19 *Korean journal of radiology* 2016;17(3):339-50. doi: 10.3348/kjr.2016.17.3.339  
20 [published Online First: 2016/05/03]  
21
- 22 58. Steyerberg EW. Clinical prediction models: a practical approach to development, validation,  
23 and updating: Springer Science & Business Media 2008.  
24
- 25 59. Pavlou M, Ambler G, Seaman SR, et al. How to develop a more accurate risk prediction  
26 model when there are few events. *Bmj* 2015;351:h3868.  
27
- 28 60. Wolff RF, Moons KGM, Riley RD, et al. PROBAST: A Tool to Assess the Risk of Bias and  
29 Applicability of Prediction Model Studies. *Ann Intern Med* 2019;170(1):51-58.  
30
- 31 61. Carlson GD, Calvanese CB, Partin AW. An algorithm combining age, total prostate-specific  
32 antigen (PSA), and percent free PSA to predict prostate cancer: results on 4298 cases.  
33 *Urology* 1998;52(3):455-61.  
34
- 35 62. Babaian RJ, Fritsche H, Ayala A, et al. Performance of a neural network in detecting prostate  
36 cancer in the prostate-specific antigen reflex range of 2.5 to 4.0 ng/mL. *Urology*  
37 2000;56(6):1000-6.  
38
- 39 63. Jansen FH, van Schaik RH, Kurstjens J, et al. Prostate-specific antigen (PSA) isoform p2PSA  
40 in combination with total PSA and free PSA improves diagnostic accuracy in prostate  
41 cancer detection. *European urology* 2010;57(6):921-27.  
42
- 43 64. Hill OT, Mason TJ, Schwartz SW, et al. Improving prostate cancer detection in veterans  
44 through the development of a clinical decision rule for prostate biopsy. *BMC Urol*  
45 2013;13:6.  
46
- 47 65. Lazzeri M, Haese A, de la Taille A, et al. Serum isoform [-2]proPSA derivatives significantly  
48 improve prediction of prostate cancer at initial biopsy in a total PSA range of 2-10 ng/ml:  
49 a multicentric European study. *Eur Urol* 2013;63(6):986-94.  
50
- 51 66. Moons KG, de Groot JA, Bouwmeester W, et al. Critical appraisal and data extraction for  
52 systematic reviews of prediction modelling studies: the CHARMS checklist. *PLoS*  
53 *medicine* 2014;11(10)  
54
- 55 67. Porter CR, O'Donnell C, Crawford ED, et al. Predicting the outcome of prostate biopsy in a  
56 racially diverse population: a prospective study. *Urology* 2002;60(5):831-35.  
57  
58  
59



- 1  
2  
3 68. Porter CR, Gamito EJ, Crawford ED, et al. Model to predict prostate biopsy outcome in large  
4 screening population with independent validation in referral setting. *Urology*  
5 2005;65(5):937-41.  
6
- 7 69. Catalona WJ, Bartsch G, Rittenhouse HG, et al. Serum pro-prostate specific antigen  
8 preferentially detects aggressive prostate cancers in men with 2 to 4 ng/ml prostate  
9 specific antigen. *J Urol* 2004;171(6 Pt 1):2239-44.  
10
- 11 70. Sokoll LJ, Sanda MG, Feng Z, et al. A prospective, multicenter, National Cancer Institute  
12 Early Detection Research Network study of [-2]proPSA: improving prostate cancer  
13 detection and correlating with cancer aggressiveness. *Cancer Epidemiol Biomarkers Prev*  
14 2010;19(5):1193-200.  
15
- 16 71. Sokoll LJ, Wang Y, Feng Z, et al. [-2]proenzyme prostate specific antigen for prostate cancer  
17 detection: a national cancer institute early detection research network validation study. *J*  
18 *Urol* 2008;180(2):539-43; discussion 43.  
19
- 20 72. Stephan C, Vincendeau S, Houlgatte A, et al. Multicenter evaluation of [-2]proprostate-  
21 specific antigen and the prostate health index for detecting prostate cancer. *Clin Chem*  
22 2013;59(1):306-14.  
23
- 24 73. Le BV, Griffin CR, Loeb S, et al. [-2]Proenzyme prostate specific antigen is more accurate  
25 than total and free prostate specific antigen in differentiating prostate cancer from benign  
26 disease in a prospective prostate cancer screening study. *J Urol* 2010;183(4):1355-9.  
27
- 28 74. Loeb S, Sokoll LJ, Broyles DL, et al. Prospective multicenter evaluation of the Beckman  
29 Coulter Prostate Health Index using WHO calibration. *J Urol* 2013;189(5):1702-06.  
30
- 31 75. Lughezzani G, Lazzeri M, Haese A, et al. Multicenter European external validation of a  
32 Prostate Health Index-based nomogram for predicting prostate cancer at extended  
33 biopsy. *Eur Urol* 2014;66(5):906-12.  
34
- 35 76. Chun FK-H, Briganti A, Graefen M, et al. Development and external validation of an  
36 extended 10-core biopsy nomogram. *European urology* 2007;52:436-44.  
37
- 38 77. Herman MP, Dorsey P, John M, et al. Techniques and predictive models to improve prostate  
39 cancer detection. *Cancer* 2009;115(S13):3085-99.  
40
- 41 78. Trottier G, Roobol MJ, Lawrentschuk N, et al. Comparison of risk calculators from the  
42 Prostate Cancer Prevention Trial and the European Randomized Study of Screening for  
43 Prostate Cancer in a contemporary Canadian cohort. *BJU Int* 2011;108(8 Pt 2):E237-44.  
44
- 45 79. Collins GS, de Groot JA, Dutton S, et al. External validation of multivariable prediction  
46 models: a systematic review of methodological conduct and reporting. *BMC Med Res*  
47 *Methodol* 2014;14(1):40.  
48
- 49 80. Miyake H, Kurahashi T, Takenaka A, et al. Improved accuracy for predicting the Gleason  
50 score of prostate cancer by increasing the number of transrectal biopsy cores. *Urol Int*  
51 2007;79(4):302-6.  
52
- 53 81. Chun FK, Briganti A, Graefen M, et al. Development and external validation of an extended  
54 10-core biopsy nomogram. *Eur Urol* 2007;52(2):436-44.  
55  
56  
57  
58  
59

- 1  
2  
3 82. Aus G, Abbou CC, Bolla M, et al. EAU guidelines on prostate cancer. *Eur Urol*  
4 2005;48(4):546-51.  
5
- 6 83. Vickers AJ, Ulmert D, Serio AM, et al. The predictive value of prostate cancer biomarkers  
7 depends on age and time to diagnosis: towards a biologically-based screening strategy.  
8 *Int J Cancer* 2007;121(10):2212-7.  
9
- 10 84. Steuber T, O'Brien MF, Lilja H. Serum markers for prostate cancer: a rational approach to  
11 the literature. *Eur Urol* 2008;54(1):31-40.  
12
- 13 85. Carlsson S, Assel M, Sjoberg D, et al. Influence of blood prostate specific antigen levels at  
14 age 60 on benefits and harms of prostate cancer screening: population based cohort study.  
15 *Bmj* 2014;348:g2296.  
16
- 17 86. Cavadas V, Osorio L, Sabell F, et al. Prostate cancer prevention trial and European  
18 randomized study of screening for prostate cancer risk calculators: a performance  
19 comparison in a contemporary screened cohort. *Eur Urol* 2010;58(4):551-8.  
20
- 21 87. Vickers AJ. Prediction models: revolutionary in principle, but do they do more good than  
22 harm. *J Clin Oncol* 2011;29(22):2951-52.  
23
- 24 88. Shariat SF, Karakiewicz PI, Margulis V, et al. Inventory of prostate cancer predictive tools.  
25 *Curr Opin Urol* 2008;18(3):279-96.  
26
- 27 89. Kawakami S, Numao N, Okubo Y, et al. Development, validation, and head-to-head  
28 comparison of logistic regression-based nomograms and artificial neural network models  
29 predicting prostate cancer on initial extended biopsy. *Eur Urol* 2008;54(3):601-11.  
30
- 31 90. Hernandez DJ, Han M, Humphreys EB, et al. Predicting the outcome of prostate biopsy:  
32 comparison of a novel logistic regression-based model, the prostate cancer risk  
33 calculator, and prostate-specific antigen level alone. *BJU Int* 2009;103(5):609-14.  
34
- 35 91. Moyer VA. Screening for prostate cancer: US Preventive Services Task Force  
36 recommendation statement. *Ann Intern Med* 2012;157(2):120-34.  
37
- 38 92. Bell N, Connor Gorber S, Shane A, et al. Recommendations on screening for prostate cancer  
39 with the prostate-specific antigen test. *CMAJ* 2014;186(16):1225-34.  
40
- 41 93. Lim LS, Sherin K, Committee APP. Screening for prostate cancer in US men: ACPM  
42 position statement on preventive practice. *Am J Prev Med* 2008;34(2):164-70.  
43
- 44 94. Prostate Cancer UK. Consensus statements on PSA testing in asymptomatic men in the UK -  
45 information for health professionals 2016 [Available from:  
46 <https://prostatecanceruk.org/about-us/projects-and-policies/consensus-on-psa-testing>  
47 accessed 20-8 2019.  
48
- 49 95. Frame I, Cant S. Current challenges in prostate cancer: an interview with Prostate Cancer  
50 UK. *BMC Med* 2015;13(1):166.  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60



Flow diagram of studies included using PRISMA method

78x91mm (300 x 300 DPI)

## Search Strategy

- 1) \* Prostatic Neoplasms/
- 2) Initial biopsy.mp
- 3) First biopsy.mp
- 4) \* “Early Detection of Cancer”/mt [Methods]
- 5) 2 or 3 or 4
- 6) 1 and 5
- 7) Nomograms/
- 8) “Neural Networks (Computer)”/
- 9) Risk Assessment/
- 10) Models, Statistical/
- 11) 7 or 8 or 9 or 10
- 12) 6 and 11
- 13) Limit 12 to English language
- 14) Limit 13 to abstracts

Table S2: Search results for independent studies that did external validation for included models.

<b>Study</b>	<b>No. of citations in Google Scholar</b>	<b>No. of citations in Medline</b>	<b>No. of external validation found in Google Scholar</b>	<b>No. of external validation found in Medline</b>
Carlson, 1998	117	20	Null	Null
Babaian, 2000	128	13	Null	Null
Jansen, 2010	231	44	Null	Null
Hill, 2010	5	0	Null	Null
Lazzeri, 2013	195	0	Null	Null

Supp Table 1: Type of assays used to measure PSA and free PSA.

Study	PSA Assay	Free PSA Assay
Carlson, 1998	PA immunoassay (Tosoh)	Investigational double-antibody radioimmunoassay (PSA II, DIANON Systems)
Babaian, 2000	PSA immunometric assay (Tosoh)	Tandem R assay (Beckman Coulter, San Diego, Calif)
Jansen, 2010	Access 2 Immunoassay System (Beckman Coulter, Brea, CA, USA)	Access 2 Immunoassay System (Beckman Coulter, Brea, CA, USA)
Hill, 2013	N/A	N/A
Lazzeri, 2013	Access 2 Immunoassay System (Beckman Coulter, Brea, CA, USA)	Access 2 Immunoassay System (Beckman Coulter, Brea, CA, USA)

# Reporting checklist for systematic review and meta-analysis.

Based on the PRISMA guidelines.

## Instructions to authors

Complete this checklist by entering the page numbers from your manuscript where readers will find each of the items listed below.

Your article may not currently address all the items on the checklist. Please modify your text to include the missing information. If you are certain that an item does not apply, please write "n/a" and provide a short explanation.

Upload your completed checklist as an extra file when you submit to a journal.

In your methods section, say that you used the PRISMA reporting guidelines, and cite them as:

Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement

	Reporting Item	Page Number
<b>Title</b>		
	<a href="#">#1</a> Identify the report as a systematic review, meta-analysis, or both.	3
<b>Abstract</b>		
Structured summary	<a href="#">#2</a> Provide a structured summary including, as applicable: background; objectives; data sources; study eligibility criteria, participants, and interventions; study appraisal and synthesis methods; results; limitations; conclusions and implications of key findings; systematic review registration number	3-4
<b>Introduction</b>		
Rationale	<a href="#">#3</a> Describe the rationale for the review in the context of what is already known.	5-6
Objectives	<a href="#">#4</a> Provide an explicit statement of questions being addressed with	6

reference to participants, interventions, comparisons, outcomes, and study design (PICOS).

## Methods

Protocol and registration	<a href="#">#5</a>	Indicate if a review protocol exists, if and where it can be accessed (e.g., Web address) and, if available, provide registration information including the registration number.	n/a
Eligibility criteria	<a href="#">#6</a>	Specify study characteristics (e.g., PICOS, length of follow-up) and report characteristics (e.g., years considered, language, publication status) used as criteria for eligibility, giving rational	6
Information sources	<a href="#">#7</a>	Describe all information sources in the search (e.g., databases with dates of coverage, contact with study authors to identify additional studies) and date last searched.	6
Search	<a href="#">#8</a>	Present full electronic search strategy for at least one database, including any limits used, such that it could be repeated.	6
Study selection	<a href="#">#9</a>	State the process for selecting studies (i.e., for screening, for determining eligibility, for inclusion in the systematic review, and, if applicable, for inclusion in the meta-analysis).	7
Data collection process	<a href="#">#10</a>	Describe the method of data extraction from reports (e.g., piloted forms, independently by two reviewers) and any processes for obtaining and confirming data from investigators.	7
Data items	<a href="#">#11</a>	List and define all variables for which data were sought (e.g., PICOS, funding sources), and any assumptions and simplifications made.	8
Risk of bias in individual studies	<a href="#">#12</a>	Describe methods used for assessing risk of bias in individual studies (including specification of whether this was done at the study or outcome level, or both), and how this information is to be used in any data synthesis.	8
Summary measures	<a href="#">#13</a>	State the principal summary measures (e.g., risk ratio, difference in means).	7
Planned methods of analysis	<a href="#">#14</a>	Describe the methods of handling data and combining results of studies, if done, including measures of consistency (e.g., I <sup>2</sup> ) for each meta-analysis.	6-8
Risk of bias	<a href="#">#15</a>	Specify any assessment of risk of bias that may affect the	15-16



1	across studies		cumulative evidence (e.g., publication bias, selective reporting within studies).	
2				
3				
4	Additional	<a href="#">#16</a>	Describe methods of additional analyses (e.g., sensitivity or	n/a
5	analyses		subgroup analyses, meta-regression), if done, indicating which	
6			were pre-specified.	
7				
8				
9	<b>Results</b>			
10				
11	Study selection	<a href="#">#17</a>	Give numbers of studies screened, assessed for eligibility, and	8-9
12			included in the review, with reasons for exclusions at each stage,	
13			ideally with a <a href="#">flow diagram</a> .	
14				
15				
16				
17	Study	<a href="#">#18</a>	For each study, present characteristics for which data were	8-9
18	characteristics		extracted (e.g., study size, PICOS, follow-up period) and provide	
19			the citation.	
20				
21				
22	Risk of bias	<a href="#">#19</a>	Present data on risk of bias of each study and, if available, any	15-17
23	within studies		outcome-level assessment (see Item 12).	
24				
25				
26	Results of	<a href="#">#20</a>	For all outcomes considered (benefits and harms), present, for	13-15
27	individual studies		each study: (a) simple summary data for each intervention group	
28			and (b) effect estimates and confidence intervals, ideally with a	
29			forest plot.	
30				
31				
32				
33	Synthesis of	<a href="#">#21</a>	Present the main results of the review. If meta-analyses are	12-15
34	results		done, include for each, confidence intervals and measures of	
35			consistency.	
36				
37				
38	Risk of bias	<a href="#">#22</a>	Present results of any assessment of risk of bias across studies	15-17
39	across studies		(see Item 15).	
40				
41				
42	Additional	<a href="#">#23</a>	Give results of additional analyses, if done (e.g., sensitivity or	n/a
43	analysis		subgroup analyses, meta-regression [see Item 16]).	
44				
45				
46	<b>Discussion</b>			
47				
48	Summary of	<a href="#">#24</a>	Summarize the main findings, including the strength of evidence	18-21
49	Evidence		for each main outcome; consider their relevance to key groups	
50			(e.g., health care providers, users, and policy makers	
51				
52				
53	Limitations	<a href="#">#25</a>	Discuss limitations at study and outcome level (e.g., risk of bias),	21-22
54			and at review level (e.g., incomplete retrieval of identified	
55			research, reporting bias).	
56				
57				
58				
59				
60				

1	Conclusions	<a href="#">#26</a>	Provide a general interpretation of the results in the context of	23
2			other evidence, and implications for future research.	
3				
4				
5	<b>Funding</b>			
6				
7	Funding	<a href="#">#27</a>	Describe sources of funding or other support (e.g., supply of	24
8			data) for the systematic review; role of funders for the systematic	
9			review.	
10				
11				

12 The PRISMA checklist is distributed under the terms of the Creative Commons Attribution License  
13 CC-BY. This checklist was completed on 30. September 2019 using <https://www.goodreports.org/>, a  
14 tool made by the [EQUATOR Network](#) in collaboration with [Penelope.ai](#)  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

# BMJ Open

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

Journal:	<i>BMJ Open</i>
Manuscript ID	bmjopen-2019-034661.R2
Article Type:	Original research
Date Submitted by the Author:	21-May-2020
Complete List of Authors:	Aladwani , Mohammad; The University of Manchester, Division of Population Health, Health Services Research and Primary Care School of Health Sciences Faculty of Biology, Medicine and Health Lophatananon, Artitaya ; The University of Manchester, Division of Population Health, Health Services Research and Primary Care School of Health Sciences Faculty of Biology, Medicine and Health Ollier, William; The University of Manchester, Division of Population Health, Health Services Research and Primary Care School of Health Sciences Faculty of Biology, Medicine and Health; Manchester Metropolitan University Faculty of Science and Engineering, School of Healthcare Science Muir, Kenneth; The University of Manchester, Division of Population Health, Health Services Research and Primary Care School of Health Sciences Faculty of Biology, Medicine and Health
<b>Primary Subject Heading</b>:	Public health
Secondary Subject Heading:	Urology, Epidemiology
Keywords:	EPIDEMIOLOGY, UROLOGY, Prostate disease < UROLOGY

SCHOLARONE™  
Manuscripts



I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our [licence](#).

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which [Creative Commons](#) licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

1  
2  
3  
4 **Full title: Prediction models for prostate cancer to be used in the**  
5 **primary care setting: a systematic review.**  
6  
7

8 **Institutional addresses and e-mail addresses of the contributing authors:**  
9

10 Dr Mohammad Aladwani

11  
12 Division of Population Health, Health Services Research and Primary Care, School of  
13 Health Sciences Faculty of Biology, Medicine and Health, The University of Manchester,  
14 Manchester, M13 9PT, UK  
15  
16

17  
18 E-mail: [mohammad.al-adwani@postgrad.manchester.ac.uk](mailto:mohammad.al-adwani@postgrad.manchester.ac.uk)  
19

20 Dr Artitaya Lophatananon

21  
22 Division of Population Health, Health Services Research and Primary Care, School of  
23 Health Sciences Faculty of Biology, Medicine and Health, The University of Manchester,  
24 Manchester, M13 9PT,UK  
25  
26

27  
28 E-mail: [artitaya.lophananon@.manchester.ac.uk](mailto:artitaya.lophananon@.manchester.ac.uk)  
29

30 Professor William Ollier,

31  
32 Division of Population Health, Health Services Research and Primary Care, School of  
33 Health Sciences Faculty of Biology, Medicine and Health, The University of Manchester,  
34 Manchester, M13 9PT, UK  
35  
36

37  
38 School of Healthcare Science, Faculty of Science and Engineering, Manchester  
39 Metropolitan University, John Dalton Building, Manchester, M1 5GD, UK  
40  
41

42  
43 E-mail: [Bill.Ollier@manchester.ac.uk](mailto:Bill.Ollier@manchester.ac.uk)  
44

45 **Address and contact details of the corresponding author:**  
46

47 Professor Kenneth R. Muir

48  
49 Division of Population Health, Health Services Research and Primary Care,  
50 School of Health Sciences, Faculty of Biology, Medicine and Health,  
51 The University of Manchester, Manchester, M13 9PT, UK  
52  
53

54  
55 E-mail: [kenneth.muir@manchester.ac.uk](mailto:kenneth.muir@manchester.ac.uk)  
56  
57  
58  
59  
60

1  
2  
3 Telephone: +44-161-275-5518  
4  
5  
6  
7

8 Key words: Prostate cancer, prostate specific antigen, PSA, prostate cancer screening,  
9  
10 risk Prediction, risk tools, prediction models, primary care, community.  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

For peer review only

## 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 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 PROBAST tool.

### Data extraction and synthesis

Relevant information extracted for each model included; the year of publication, source of data, type of model, number of subjects, 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%. 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.

## Conclusions

Only a few PCa risk prediction models have the potential to be readily used in the primary health care 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.

For peer review only



## Strengths and limitations

- The review focussed on risk prediction models for PCa for use in primary care.
- The PRISMA approach was followed in identifying relevant articles and reporting this study.
- We used the PROBAST tool 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.

## 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 <sup>1</sup>. In the UK, around 47,200 new cases of PCa were reported in 2015, accounting for 26% of all new cancer cases in males. PCa deaths in the UK of were around 11,600 in 2016 <sup>2</sup>. The global projections of PCa incidence and mortality for 2030 are 1.7 and 0.5 million, respectively <sup>3</sup>. The highest incidence of PCa is seen in western societies <sup>4</sup>. The significant increase of PCa incidence and diagnosis over the last three decades can be attributed mainly to the widespread implementation of the PSA serum test after it had been introduced in the late 1980s <sup>5 6</sup>.

The strong association of PSA with PCa <sup>7 8</sup>, along with it being a relatively inexpensive test <sup>9</sup>, has made PSA a key biomarker in the diagnostic processs of PCa and for the recommendation of a confirmatory prostate biopsy <sup>7 9</sup>. PSA is, however, not a cancer-specific marker <sup>5 10</sup>. 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 <sup>11-14</sup>. 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 <sup>5 10 15 16</sup>. Such issues are likely to be the part of the explanation for the significant number of unnecessary biopsies currently performed each year. Such procedures are associated with adverse side effects for patients and also increases health care costs <sup>17 18</sup>.

1  
2  
3 To address such PSA test limitations, researchers have incorporated other measurable  
4 factors into approaches for the early detection of PCa; these “risk assessment tools” are  
5 based on statistical models designed to improve the accuracy and performance of the  
6 PSA test <sup>19-22</sup>. Logistic regression and artificial neural network (ANNs) models are now  
7 considered to be the most common and effective statistical techniques in aiding  
8 development of new models to enhance early PCa diagnosis <sup>23</sup>. These PCa risk prediction  
9 models can be used to aid the testing of men for further investigations.

15  
16 Currently in the UK, there is no population based screening programme for PCa. The  
17 ultimate goal of PCa screening is to find intermediate and high risk of PCa rather than low  
18 risk PCa that would not require treatment but will give emotional burden to the patient  
19 once detected and unnecessary treatment in some patients. An important potential  
20 advantage of the extended risk models is their ability to provide a more accurate  
21 estimation of PCa risk. This may ultimately lead to their use in patient counselling and  
22 decision-making <sup>24-27</sup>. Such models have already achieved better results in predicting  
23 probabilities of outcome compare to clinical judgment <sup>28 29</sup>. Furthermore, it has been  
24 reported that using such predictive models may minimize the rate of unnecessary  
25 biopsies <sup>30</sup>.

33  
34 Recently there has been a substantial increase in the development of predictive models  
35 to help clinicians assess risk and decide which man to send to clinical setting to further  
36 investigate for a possible diagnosis of PCa <sup>22 26 30-35</sup>. The majority of these models are  
37 designed for use in clinical settings, where costs are less of an issue and most include  
38 the need for a clinical examination such as digital rectal examination (DRE) or trans-rectal  
39 ultrasonography (TRUS). One of the main limitations of DRE is that it has poor  
40 performance, especially at low PSA levels, and it is highly subjective to inter-observer  
41 variability <sup>36-38</sup>. A meta-analysis study revealed that DRE has positive predictive value of  
42 only 18% <sup>39</sup>. Similarly, TRUS has been reported for having poor accuracy at low PSA  
43 levels <sup>40 41</sup> and small PCa might not be palpable on DRE or visualisation on TRUS <sup>40</sup>.  
44 Furthermore, less than 40% of PCa detected by DRE are potentially curable, making it  
45 less beneficial for early diagnosis <sup>42</sup>. Several studies showed that there are fear, anxiety  
46 and embarrassment among some men, in particular black men, regarding the DRE test  
47 <sup>43-46</sup>. Another disadvantage of the DRE is the fact it is a potentially uncomfortable test <sup>47-</sup>

1  
2  
3 5<sup>2</sup>. This may explain why the DRE is a barrier for some men to participate in PCa  
4 screening if it is including DRE test. Lee et al reported that 74%-84% of black men may  
5 not maintain annual DRE screening <sup>53</sup>, while another study found that it may prevent 22%  
6 of men to participate <sup>54</sup>. Since TRUS need to be performed by a skilled urologist, this  
7 means men have to make an appointment with a clinic in different location, which makes  
8 the screening less convenience. As a result, men may feel reluctant to have such tests  
9 performed.

10  
11  
12  
13  
14  
15  
16 This systematic review of the literature was undertaken to identify risk prediction models  
17 that do not incorporate invasive or more costly clinical procedures or extensive  
18 biomarkers but have potential application for use in primary care and community settings.  
19 As low cost is a primary concern for community use, for this review we set an indicative  
20 threshold of approximately 3-5 times the cost of a PSA test for inclusion. This excluded a  
21 number of models that contain new and emerging biomarker or SNP panels. As the  
22 numbers of person referred to clinical setting costs are less of an issue. The performance  
23 of the models reviewed for detecting PCa at initial biopsy have been compared using  
24 'reported area under the curve' (AUC) and/or sensitivity - specificity testing.  
25  
26  
27  
28  
29  
30

## 31 **METHODS**

32  
33  
34 The approach used to identify and select relevant articles was based on the application  
35 of the '*Preferred Reporting Items for Systematic Reviews and Meta-Analysis*' (PRISMA)  
36 <sup>55</sup>.  
37  
38

### 39 **Data sources and search strategy**

40  
41  
42 A literature search was performed using Medline (via OVID) and Embase databases. The  
43 'medical subject heading' (MeSH) terms, combined with Boolean logic operators 'AND'  
44 and 'OR', were applied to retrieve relevant articles. The terms used for the search were  
45 "Prostatic Neoplasms" AND ("Initial biopsy" OR "first biopsy" OR "early detection of  
46 cancer") AND ("nomograms" OR "artificial neural networks" OR "risk assessment" OR  
47 "statistical model"). The full search strategy is provided in a Supplementary File 1. All  
48 articles defined (published since the inception of the databases and up to the end of  
49 January 2019) were subsequently further filtered as being those only published in English  
50 language and with an abstract. Further to using the above search databases, the research  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 articles were selected manually from the reference lists of any relevant review articles.  
4 Google Scholar and Medline searches were also carried out to identify independent study  
5 for external validation for each model included in this review. The results are presented  
6 in Supplementary File 2.  
7  
8  
9

### 10 **Eligibility criteria**

11  
12 As this review focuses on PCa risk prediction based in community healthcare settings, all  
13 studies were selected on the following inclusion criteria (i) evaluating the risk for PCa at  
14 initial biopsy in subjects who had no prior history of PCa (ii) studies that reported “low  
15 cost” risk assessment tools (i.e. those not including more expensive genetic, biomarker  
16 test) or “invasive” clinical tests/examinations (such as DRE or TRUS) (iii) studies that  
17 included a minimum of two variables of which PSA had to be one of them [on the basis  
18 that an elevated PSA test in UK primary care is usually the first sign and rationale for  
19 suggesting a need for further investigation of PCa within NICE guidelines] and (iv) studies  
20 that reported AUC and/or sensitivity and specificity of the diagnostic/predictive tool. The  
21 exclusion criteria used were (i) articles with models that were built and based on repeat  
22 or mixed biopsies (ii) studies that only validate an existing model (iii) articles that were not  
23 published in English. There were no time boundaries regarding the publication year.  
24  
25  
26  
27  
28  
29  
30  
31  
32

33 Screening of the titles, abstracts, and full-texts was carried out by two reviewers (MA, AL).  
34 Any concerns about the eligibility of a study were resolved by discussion with a third  
35 reviewer (KM).  
36  
37  
38

### 39 **Data extraction**

40  
41 A data extraction form was developed to collect all relevant information. For each study  
42 used in this review, the items extracted included; year of publication, source of data, type  
43 of model, number of subjects, country where it was performed, age, PSA range,  
44 mean/median PSA, number of biopsy cores, variables included in the model, study  
45 endpoint(s), cancer detection, model performance, and model validation.  
46  
47  
48  
49

### 50 **Evaluating the performance of the risk models**

51  
52 Prediction models can be evaluated against various criteria. The most critical  
53 measurements of model performance are discrimination and calibration<sup>27</sup>. Discrimination  
54 refers to how well the prediction model can differentiate subjects in different outcome  
55  
56  
57  
58  
59

1  
2  
3 classes according to their predicted risks. It is often assessed by measuring the area  
4 under the receiver operating characteristic (ROC) curve<sup>56</sup>. It also requires setting a series  
5 of thresholds to separate low and high ranges of predicted outcomes. A value of 0.5  
6 indicates no discrimination, while a value of 1 indicates perfect discrimination. However,  
7 even with perfect discrimination, observed risk can differ from expected risk. Therefore,  
8 calibration has an important role in model evaluation<sup>57</sup>. Calibration represents the  
9 agreement between expected and observed outcomes<sup>58</sup>. A well-calibrated model is  
10 achieved when the calibration slope is close to 1. When the calibration slope is less than  
11 1 it indicates that the model under-estimates low risks and over-estimate high risks<sup>59</sup>.

12  
13  
14  
15  
16  
17  
18  
19 Due to the heterogeneity of the studies included, conducting a meta-analysis was not  
20 applicable.  
21

### 22 23 **Study quality assessment**

24  
25 The quality of the studies included in this review was assessed using the PROBAST tool  
26<sup>60</sup>. This tool has been developed specifically to assess the risk of bias and applicability  
27 for prediction model studies. The tool consists of four domains and has 20 signalling  
28 questions that facilitate to reach overall judgment of risk of bias, as well as issues relating  
29 to applicability.  
30  
31  
32

### 33 34 **Patient and Public involvement**

35  
36 No patients were involved in setting the research question or the outcome measures, nor  
37 were they involved in the design and implementation of the study. There are no plans to  
38 involve patients in dissemination.  
39  
40

## 41 42 **RESULTS**

43  
44 A total of 102 publications were identified using the search strategy as shown in Fig.1. An  
45 additional nine articles were identified through manual searches from a bibliography of  
46 reviewed articles. At the first filter step, a total of 109 titles and abstracts were screened  
47 for eligibility after removing two duplicates. In the second filter step, 60 papers failed to  
48 meet the inclusion criteria and were excluded, resulting in 49 articles. The final step of  
49 filtering yielded only five studies that were considered to be eligible (i.e. passed all set  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 criteria) and were thus included in this systematic review. There is no independent study  
4 identified for external validation for included models.  
5

### 6 7 **Study characteristics**

8  
9 Four of the five studies included were based on cohort studies and one was a case-control  
10 study. The characteristics of each of these studies and populations are summarised in  
11 Table 1. Details of PSA assays used in the models are presented in Supplementary File  
12  
13  
14 3.  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

Table 1: Characteristics of the included studies.

Author & 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, 1998 <sup>61</sup>	Logistic regression	Cohort	Model dev = 3773 Validation = 525	Baltimore, USA	Referral	≥ 45	—	4-20 ng/ml	—	Sextant biopsy	32%
Babaian, 2000 <sup>62</sup>	Neural network = 3 ANNs	Cohort	151	Texas, USA	Screening program	40-75	62	2.5-4 ng/ml	—	11 cores	24.5%
Jansen, 2010 <sup>63</sup>	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 program	≥ 50	Site 1 (66) Site 2 (60)	2-10 ng/ml	~ 4.4	≥ 6 cores	Site 1 = 55.8% Site 2 = 49.6%
Hill, 2013 <sup>64</sup>	Logistic regression	Case-control	1378	Florida, USA	Hospital referral	40-90	—	≥ 4 ng/ml	—	N/A	20.6%
Lazzeri, 2013 <sup>65</sup>	Logistic regression	Cohort	646	European multi-centre; Italy, Germany, France, Spain, and the UK	Referral	>45	—	2-10 ng/ml	~ 5.8	≥ 12 cores	40.1%

1  
2  
3 Subjects used to build the risk models varied across these studies. Of the five studies,  
4 three studies included men from referral populations <sup>61 64 65</sup> and two studies from  
5 screening programs <sup>62 63</sup>. The sample sizes ranged from 151 to 3,773 with three studies  
6 derived from US cohorts <sup>61 62 64</sup> and two from Europe <sup>63 65</sup>. Four studies used logistic  
7 regression methodology to build their model, whereas one study used an artificial neural  
8 network-based approach (ANN) <sup>62</sup>. The minimum age of participants was 40 years old <sup>62</sup>  
9 <sup>64</sup> and the minimum PSA level was 2 ng/ml <sup>63 65</sup>.

### 17 **Variables in the model**

19 Table 2 presents details of the variables used in each model. PSA level was used in all  
20 models, followed by free PSA (fPSA), age, and free-to-total PSA ratio (%fPSA). Other  
21 variables also reported in the models included; precursor of prostate-specific antigen  
22 (p2PSA), percentage of p2PSA to free PSA (%p2PSA), prostate health index (PHI), levels  
23 of haemoglobin, albumin, creatinine and red blood cell count (RBC), haematuria, mean  
24 corpuscular volume (MCV), and prostatic acid phosphatase.  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60



Table 2: Variables used in the prostate cancer risk prediction models\*.

Author & year	Variables used in the model				
	Total PSA	Free PSA	Percent free PSA	Age	Other variables
Carlson, 1998 <sup>61</sup>	✓		✓	✓	
Babaian, 2000 <sup>62</sup>	✓	✓		✓	Creatinine Kinase, Prostatic acid phosphatase
Jansen, 2010 <sup>63</sup> ( <i>Site 1</i> )	✓	✓			p2PSA
Jansen, 2010 <sup>63</sup> ( <i>Site 2</i> )	✓	✓			p2PSA
Hill, 2013 <sup>64</sup> ( <i>Method 1</i> )	✓			✓	HGB, RBC, Haematuria, Creatinine, MCV, and ethnicity "Black"
Hill, 2013 <sup>64</sup> ( <i>Method 2</i> )	✓			✓	HGB, RBC, Creatinine, and MCV
Lazzeri, 2013 <sup>65</sup> ( <i>Model 1</i> )	✓	✓	✓		
Lazzeri, 2013 <sup>65</sup> ( <i>Model 2</i> )	✓	✓	✓		p2PSA
Lazzeri, 2013 <sup>65</sup> ( <i>Model 3</i> )	✓	✓	✓		%p2PSA
Lazzeri, 2013 <sup>65</sup> ( <i>Model 4</i> )	✓	✓	✓		PHI

\*HGB= Haemoglobin. RBC= Red blood cells. MCV= Mean corpuscular volume. PHI= Prostate Health Index. p2PSA= Precursor of prostate specific antigen, %p2PSA= p2PSA/fPSA

## Outcome

The study endpoint also varied among the studies selected. Two studies evaluated the accuracy of detecting any PCa<sup>61 62</sup> and three studies examined the pathologic Gleason score<sup>63-65</sup>. Although Jansen et al, did not build a model to predict the aggressiveness of prostate cancer, they assessed the relationship of each variable individually with a Gleason score  $\geq 7$ . PCa was determined by taking a needle biopsy. All subjects in the five studies underwent prostate biopsy. The least number of biopsy cores used were 6<sup>61 63</sup> and the highest were  $\geq 12$ <sup>65</sup>. One study did not report the number of biopsy cores taken<sup>64</sup>. 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 to PSA alone (AUC of 0.755 vs. 0.585, respectively)<sup>63</sup>. 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 percent free PSA alone with 95% sensitivity<sup>61</sup>. Lazzeri et al,<sup>65</sup> presented results from four separated models discriminating PCa with a Gleason score of  $\geq 7$ . Lazzeri's model 2 (which includes the base model tPSA, 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 4 models with an AUC of 0.67. In the study of Hill<sup>64</sup> 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 to 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 (PIN), 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 to 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.

Table 3: The difference of area under the curve (AUC) for PSA alone and extended model.

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

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 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 4: Sensitivity and specificity profile at different levels for each model.

Study	Sensitivity	Specificity	Probability	Positive	Negative
-------	-------------	-------------	-------------	----------	----------

			cut-off	predictive value	predictive value
	99	18	>15	≤47	N/A
Carlson <sup>61</sup>	95	34	18	51	NA
	89	43	20	42	NA
	95	51	NA	39	97
Babaian <sup>62</sup>	92	62	NA	44	96
	89	62	NA	43	95
Jansen <sup>63</sup> (Site 1)	95	23.9	NA	NA	NA
Jansen <sup>63</sup> (Site 1)	90	30.1	NA	NA	NA
Jansen <sup>63</sup> (Site 2)	95	23.2	NA	NA	NA
Jansen <sup>63</sup> (Site 2)	90	36.2	NA	NA	NA
Hill <sup>64</sup> (Method 1)	90.9	17.6	33	47.1	70.5
Hill <sup>64</sup> (Method 2)	89.8	28	13	20.6	91.3
Hill <sup>64</sup> (Method 1)	80.5	37.1	37	50.9	70.2
Hill <sup>64</sup> (Method 2)	78.2	45	15	28.7	88.8
Hill <sup>64</sup> (Method 1)	39.9	81.4	48	63.4	62.6
Hill <sup>64</sup> (Method 2)	45.8	79.5	23	36.7	85
* Lazzeri <sup>65</sup> model reported only sensitivity and specificity for predictive variables individually and at sensitivity of 90, %p2PSA and %fPSA achieved the highest specificity					

Table 5 summarises the validation and calibration results for the studies included. Model calibration was reported in two studies <sup>61 65</sup>. 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 <sup>63 64</sup>. Only one study reported an external validation using an additional data set consisting of 525 patients <sup>61</sup>. 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 <sup>62</sup>. Lazzeri used 200 bootstrap re-samples to minimise overfitting bias <sup>65</sup>.

Table 5: Validation and calibration for included models.

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

### 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 tool, particularly in the analysis domain. For instance, one study applied univariable analysis to select predictors <sup>61</sup>. Three studies did not measure calibration <sup>62-64</sup>. Furthermore, two studies did not account for optimism and overfitting by using internal validation methods <sup>63 64</sup>. Whereas one study did not use appropriate measures for model performance i.e. AUC, this study reported the calibration <sup>61</sup>.

The event per variable (EPV) was lower than recommended ( $< 10$ ) <sup>59 66</sup> in the Babaian <sup>62</sup> study indicating inadequate power. Four studies did not report missing data or how they handled it <sup>61-63 65</sup>. The remaining study used complete-case analysis and excluded subjects with missing data on laboratory biomarkers ( $n=75$ ) <sup>64</sup>. 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 <sup>60</sup>. Thus, although the quality assessment for the Lazzeri

1  
2  
3 study <sup>65</sup> was graded low risk in all the four domains, since the study did not report any  
4 external validation of the model, the assessment of the study has been regraded to high  
5 risk of bias according to the PROBAST criteria. A full quality assessment for all studies is  
6 presented in Table 6.  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

Table 6: Quality assessment for risk of bias and applicability concern for included studies.

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

\* ROB- Risk of bias

+ indicates a low risk of bias or applicability; - indicates a high risk of bias or applicability.

## 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 model<sup>61</sup> has the largest population (3,773 subjects) when compared to 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<sup>33</sup>. More importantly, by not reporting the AUC estimate, the model raises some doubts regarding the reliability of the model and its implementation<sup>33</sup>. It will also make comparison to other models not applicable<sup>67</sup>.

Babaian<sup>62</sup> 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, PSAD-TZ= 0.75). They included a number of uncommon pre-biopsy inputs into their algorithm such as prostatic acid phosphatase and creatinine kinase<sup>68</sup>. Furthermore, they used a tight PSA range (2.5-4.0ng/ml) which meant that their model may be less suitable for patients with PSA level below or above that range, thus limiting its generalisability.



1  
2  
3 The study by Jansen and colleagues<sup>63</sup> demonstrated that adding p2PSA to the base  
4 model of PSA and fPSA significantly enhanced the PCa predictive value and specificity.  
5 The association and added value of p2PSA in the prediction and detection of PCa have  
6 been reported by several other studies<sup>16 69-72</sup>. Jansen<sup>63</sup> showed that p2PSA has no clear  
7 association with aggressive PCa. However, the base model that includes p2PSA had the  
8 highest clinical significance in correlation to pathologic Gleason score with a p-value of  
9 0.008 compared to %fPSA and PHI (p-value 0.01 and 0.02) respectively. Although they  
10 used archived blood samples and retrospective analysis, the results were similar to a  
11 prospective study of 268 patients<sup>16</sup>.

12  
13 Hill<sup>64</sup> used a case-control study to evaluate several laboratory biomarkers. They found  
14 HGB, RBC, haematuria, creatinine, PSA, Age, MCV, and ethnicity (“being black”) were  
15 statistically significantly associated in the first method ( $p < 0.05$ ). In the second method,  
16 HGB, RBC, creatinine, PSA, age, and MCV were found to be statistically significantly  
17 correlated ( $p < 0.001$ ) with PCa. However, since this study was designed as a case-control  
18 study, it would have been more prone to uncontrolled confounding and selection bias.  
19 Moreover, the type of screening protocols used in Veterans’ Administrations may vary to  
20 those conducted in other healthcare systems; therefore, the results may not be applicable  
21 to other populations. Furthermore, subjects with a PSA level  $< 4.0$  ng/ml have not been  
22 investigated, and thus, the performance of the models are unknown for individuals in this  
23 group.

24  
25 Lazzeri et al,<sup>65</sup> in a European multi-centre study have evaluated similar biomarkers as in  
26 Jansen study with the same PSA range 2-10 ng/ml prospectively. They found no  
27 difference in both %p2PSA and PHI as individual PCa predictors with AUC of 0.67 (95%  
28 CI 0.64-0.71). However, the base model (consisting of PSA, fPSA, and %fPSA) that also  
29 included either p2PSA or PHI out-performed the base model alone and the base model  
30 including %p2PSA. In the analysis, the additive value of both p2PSA and PHI is 0.064  
31 and 0.056 for %p2PSA for predicting the risk of PCa. These additive values increased to  
32 0.076 for both p2PSA and PHI and 0.073 for %p2PSA in predicting Gleason scores  $\geq 7$   
33 for the disease. The usefulness of PHI in improving the predictive accuracy of PCa over  
34 total and free PSA has been confirmed and reported by several studies<sup>16 72-75</sup>.

1  
2  
3 In general, only one study has validated their model externally <sup>61</sup>, whereas the remaining  
4 studies were either validated internally <sup>62 65</sup> or not reported any validation methods <sup>63 64</sup>.  
5 Prediction models may not be equally applicable to all data sets as patients'  
6 characteristics may vary <sup>20 76</sup>. As a result, the generalisability of a model might be poor  
7 when it used in populations other than that used in building the model. Therefore, external  
8 validation should be conducted before applying any new model into general practice <sup>77 78</sup>.  
9

10  
11  
12  
13  
14 Another key performance measure of any model that needs careful evaluation is  
15 calibration. A calibration plot with a calibration slope is more preferable than the Hosmer-  
16 Lemeshow test; it has been acknowledged that evaluating a good and well calibrated  
17 model based on a large dataset can still fail the Hosmer-Lemeshow test. In contrast, when  
18 evaluating a poorly calibrated model with a small dataset it can still pass the Hosmer-  
19 Lemeshow test <sup>79</sup>. In our analysis, three studies fail to report the calibration of the model  
20 <sup>62-64</sup> whilst the Carlson study <sup>61</sup> used a calibration plot and Lazzeri <sup>65</sup> used the Hosmer-  
21 Lemeshow test. Excluding calibration from the majority of models may explain why some  
22 models are not currently used in practice <sup>79</sup>.  
23  
24  
25  
26  
27  
28  
29

30 With regard to biopsy cores, only two studies used extended biopsy cores. Babaian <sup>62</sup>  
31 used an 11-core multi-site biopsy, whereas Lazzeri <sup>65</sup> used at least 12 biopsy cores.  
32 Moreover, two studies used six cores biopsy in their model <sup>61 63</sup>. The use of six-core biopsy  
33 has been criticized as not being adequate in detecting PCa <sup>80</sup> and that models developed  
34 using sextant biopsy are less accurate than when a 10-core biopsy is used <sup>81</sup>. As a result,  
35 the European guideline for clinical PCa recommended an extended biopsy as standard  
36 practice for PCa detection <sup>82</sup>.  
37  
38  
39  
40  
41

42 It is worth noting that all five reviewed models performed better than just PSA alone.  
43 However, none of them has both high specificity and sensitivity. The level of sensitivity  
44 has been increased, and despite enhancement in the specificity, it is still considered low.  
45 Specificity is crucial when it comes to being used in a population setting as men without  
46 PCa should be ruled out as much as possible from further invasive engagement with the  
47 health system.  
48  
49  
50  
51  
52

53 Our review therefore suggests that none of the reviewed models provide an ideal  
54 performance in predicting PCa with high sensitivity and high specificity. It is particularly  
55  
56  
57  
58  
59  
60

1  
2  
3 important when considering the application of PCa risk prediction at the population level  
4 that the tool used should be able to both detect the outcome and filter out people with no  
5 disease. As there is robust evidence suggesting that the clinical relevance of PSA range  
6 to the detection of PCa differs across age groups<sup>83-85</sup>, any future model should consider  
7 PSA threshold in relation to a specific age range. Risk prediction models for PCa should  
8 therefore take account of age.  
9

10  
11  
12  
13  
14 Out of the five reviewed models, the Lazzeri model 2, has the greatest potential to be  
15 implemented in primary care. It achieved the least risk of bias and had fair discrimination  
16 for both any and aggressive PCa. It also had the largest improvement in discrimination  
17 performance compared to PSA alone. Moreover, except for the p2PSA that requires a  
18 specific assay, the included variables are common and easy to measure. However, before  
19 it could be used, the model requires to be validated externally.  
20  
21  
22  
23

### 24 **Comparison with other studies**

25  
26 To our knowledge, three systematic reviews of PCa prediction tools have been published  
27<sup>20 26 27</sup>. In the Louie et al. review, risk models were included that were externally validated  
28 in at least five study populations for the purpose of meta-analysis and only six studies  
29 were included in their analysis. Furthermore, all the studies included incorporated clinical  
30 tests such as DRE and/or TRUS-PV<sup>26</sup>. Schroder and Kattan<sup>20</sup> reviewed models that  
31 were built to predict the likelihood of having a positive prostate biopsy for cancer.  
32 However, it appears that they also included models where subjects had a previous  
33 negative biopsy. As such, some of the models included variables related to biopsy results  
34 and cores. The review by Shariat and colleagues examined different types of predictive  
35 tools<sup>27</sup>. They explored tools that predict PCa on initial and repeat biopsy, pathologic  
36 stages, biochemical recurrence after radical prostatectomy, metastasis, survival, and life  
37 expectancy. Similarly, virtually all of the prediction tools that were based on initial biopsy  
38 included variables based on invasive procedures.  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49

### 50 **Strengths and limitations of this study**

51  
52 This report is the first to review risk prediction tools for PCa that can be used in primary  
53 care and community settings. Any prediction model should therefore be simple to use,  
54 based on non-invasive tests, be feasible at a population level and at low cost. We carried  
55  
56  
57  
58  
59  
60

1  
2  
3 out an extensive data extraction relating to important features and characteristics for each  
4 study included, such as modelling method, source of data, sample size, variables,  
5 discrimination, validation, and cancer detection rate. We have also followed PRISMA  
6 guidelines for identifying eligible articles as well as for reporting this study. In addition, the  
7 PROBAST tool was adopted to assess the quality and risk of bias for each prediction  
8 model.  
9

10  
11  
12  
13  
14 Our study has some limitations. Our aim was to identify prediction models that have the  
15 potential to be implemented in a primary care or community settings, and consequently  
16 our search strategy was to retrieve relevant studies for this specific purpose. Furthermore,  
17 we excluded articles that were not published in English or did not have an abstract.  
18 Moreover, only two databases were searched, besides manual search, to retrieve original  
19 studies.  
20

21 A previous systematic review suggested that the majority of relevant studies could be  
22 identified through a manual search of articles reference lists instead of a database alone  
23 <sup>20</sup>. We identified four eligible studies using this approach. Given the small number of  
24 models identified by the approach we followed, that can be applied in primary care  
25 settings compared to the large number relating to wider existing models, it is unlikely that  
26 we have not included any study that would affect the results of our review.  
27  
28  
29

### 30 **Implications and future research**

31  
32 It is now accepted that the PSA test and its derivatives have some limitations for detecting  
33 PCa as defined by subsequent biopsy <sup>86</sup>. As a consequence, a considerable number of  
34 PCa prediction models have been built to improve prediction accuracy. This has resulted  
35 in a plethora of PCa risk prediction tools, with to date more than 100 models described <sup>87</sup>  
36 <sup>88</sup>. There is evidence that some of these models show benefit and have better  
37 performance over just PSA measurement alone <sup>20</sup>. It also has been demonstrated that  
38 some of these models out-performed clinical experts in predicting PCa <sup>28 29</sup>. Although  
39 such models are not designed to replace specialist clinical judgment or patient  
40 preferences <sup>76 86 89</sup>, they can help in patient counselling and aid clinicians to decide  
41 whether a prostate biopsy should be taken or not <sup>77 89 90</sup>.  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 Given the small number of risk prediction models for PCa that do not incorporate clinical  
4 or genetic tests, the discrimination of these reviewed models ranged between poor to  
5 moderate (AUC range ~0.65 to ~0.75); in addition there were some issues relating to their  
6 study design and analysis raises the risk of bias. Consequently none of these models  
7 could be currently recommended for use in a primary care and community health care  
8 setting. Several guidelines are against using PSA test based screening for PCa; the US  
9 Preventive Services task force, the Canadian task force on preventive health and the  
10 American College of Preventive Medicine do not currently recommend PSA-based testing  
11 due to insufficient evidence <sup>91-93</sup>. This has made it difficult, so far, to convince  
12 policymakers to adopt PCa screening programme.

13  
14 The first guideline of PROSTATE CANCER UK states *"In the future, health professionals  
15 should look at a man's PSA level alongside other known risk factors as part of a risk  
16 assessment tool, when one becomes available."* <sup>94</sup> However, the vast majority of the  
17 current PCa risk prediction models are not suitable for routine use as they include clinical  
18 and genetic tests and are not validated externally in other cohorts. Therefore, the main  
19 challenge in the UK, remains to develop a risk prediction tool that is reliable, cheap, is  
20 applicable for as wide an ethnicity as possible, and most importantly is easy to use and  
21 can be implemented at a primary care level <sup>95</sup>.

22  
23 The value of such risk tools is that they will help to stratify men at high risk of developing  
24 PCa earlier so that they have appropriate management and/or surveillance programme  
25 as early as possible and therefore, may fit into the clinical pathway. Such tools should  
26 help physicians have a better understanding of the risk for this disease and simplify the  
27 procedures and discussions with patients when recommending further specialist-led  
28 investigations such as DRE and/or MRI where a decision on whether a biopsy should or  
29 not perform is concluded. Furthermore, using the appropriate risk prediction tool will avoid  
30 men from undergoing inappropriate further and frequent testing <sup>95</sup>. This will reduce any  
31 associated costs of inappropriate tests and decrease the burden on health care delivery  
32 systems.

33  
34 It is crucial to address these issues by identify all possible risk factors for PCa that are  
35 non-clinical, non-genetic, and easy to use and interpret. There remains a pressing need  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 to develop a risk prediction tool in the future using all appropriate factors (potentially also  
4 including genetics once there is infrastructure in place for genetic testing in the primary  
5 care and the cost comes down) into a robust multivariable analysis and validate the model  
6 externally to eliminate applicability and generalisability concerns. Only when this is  
7 achieved will it be possible to introduce a prostate cancer screening programme fit for  
8 purpose.  
9

## 14 **CONCLUSION**

16 There is a paucity of suitable low cost risk models that incorporate non-clinical, non-  
17 genetic inputs and which can be used at a primary care level and in other community  
18 health services. Existing models have limitations reflecting both study design and  
19 reporting performance measures. Future research should take into account these key  
20 issues and explore other risk factors for incorporation into further models.  
21  
22

## 25 **ACKNOWLEDGEMENT**

27 MA is supported by Kuwait's Ministry of Health.  
28

## 29 **AUTHOR CONTRIBUTIONS**

31 MA, AL and KM were involved in study conception, idea and design. Data acquisition and  
32 extraction was obtained by MA. Data synthesis and interpretation was carried out by MA, AL and  
33 WO. MA and AL drafted the manuscript. All authors approved the final version of manuscript. KM  
34 is the study guarantor.  
35  
36

## 38 **FUNDING**

40 KM and AL are supported by the NIHR Manchester Biomedical Research Centre and by the ICEP  
41 ("This work was also supported by CRUK [grant number C18281/A19169]). KM is also supported  
42 by the Allan Turing Institute.  
43  
44

## 46 **COMPETING INTERESTS**

47 All authors have completed the ICMJE uniform disclosure form at  
48 [www.icmje.org/coi\\_disclosure.pdf](http://www.icmje.org/coi_disclosure.pdf) and declare: no support from any organisation for the submitted  
49 work; no financial relationships with any organisations that might have an interest in the submitted  
50 work in the previous three years; no other relationships or activities that could appear to have  
51 influenced the submitted work.  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 **ETHICAL APPROVAL**  
4

5 Not applicable as this study did not involve any personal data.  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

For peer review only



## DATA SHARING

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

## Figure legend

Figure 1: Flow diagram of studies included using PRISMA method

## REFERENCES

1. 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: a cancer journal for clinicians* 2018;68(6):394-424.
2. Cancer Research UK. Prostate cancer statistics 2017 [Available from: <https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/prostate-cancer#heading-One>].
3. Center MM, Jemal A, Lortet-Tieulent J, et al. International variation in prostate cancer incidence and mortality rates. *European urology* 2012;61(6):1079-92.
4. 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(6):1388-400.
5. Dijkstra S, Mulders P, Schalken J. Clinical use of novel urine and blood based prostate cancer biomarkers: a review. *Clin Biochem* 2014;47(10-11):889-96.
6. Fontanella P, Benecchi L, Grasso A, et al. Decision-making tools in prostate cancer: from risk grouping to nomograms. *Minerva Urol Nefrol* 2017;69(6):556-66.
7. Stephan C, Ralla B, Jung K. Prostate-specific antigen and other serum and urine markers in prostate cancer. *Biochim Biophys Acta* 2014;1846(1):99-112.
8. Thompson IM, Pauler DK, Goodman PJ, et al. Prevalence of prostate cancer among men with a prostate-specific antigen level  $\leq$  4.0 ng per milliliter. *N Engl J Med* 2004;350(22):2239-46.
9. 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(1):18-26.
10. 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(1):18-29.
11. 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(2):293-306.
12. 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(4 Pt 1):1297-301.



13. Bunting PS. A guide to the interpretation of serum prostate specific antigen levels. *Clin Biochem* 1995;28(3):221-41.
14. 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(5):1650-55.
15. Mistry K, Cable G. Meta-analysis of prostate-specific antigen and digital rectal examination as screening tests for prostate carcinoma. *The Journal of the American Board of Family Practice* 2003;16(2):95-101. [published Online First: 2003/04/01]
16. 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(2):214-22.
17. Heijnsdijk EA, der Kinderen A, Wever EM, et al. Overdetection, overtreatment and costs in prostate-specific antigen screening for prostate cancer. *British journal of cancer* 2009;101(11):1833-8. doi: 10.1038/sj.bjc.6605422 [published Online First: 2009/11/12]
18. 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(3):359-62. doi: 10.1016/j.eururo.2010.10.036 [published Online First: 2010/11/03]
19. 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(4):618-29. doi: 10.1016/j.eururo.2016.08.003 [published Online First: 2016/08/30]
20. Schroder 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(2):274-90.
21. Ross PL, Scardino PT, Kattan MW. A catalog of prostate cancer nomograms. *J Urol* 2001;165(5):1562-8. [published Online First: 2001/05/09]
22. 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. *Clinical cancer research : an official journal of the American Association for Cancer Research* 2008;14(14):4400-7. doi: 10.1158/1078-0432.Ccr-07-4713 [published Online First: 2008/07/17]
23. Hu X, Cammann H, Meyer HA, et al. Artificial neural networks and prostate cancer--tools for diagnosis and management. *Nature reviews Urology* 2013;10(3):174-82. doi: 10.1038/nrurol.2013.9 [published Online First: 2013/02/13]
24. Caras RJ, Sterbis JR. Prostate cancer nomograms: a review of their use in cancer detection and treatment. *Current urology reports* 2014;15(3):391. doi: 10.1007/s11934-013-0391-0 [published Online First: 2014/01/24]
25. Iremashvili V, Soloway MS, Pelaez L, et al. Comparative validation of nomograms predicting clinically insignificant prostate cancer. *Urology* 2013;81(6):1202-8. doi: 10.1016/j.urology.2013.01.062 [published Online First: 2013/04/09]

- 1
  - 2
  - 3
  - 4
  - 5
  - 6
  - 7
  - 8
  - 9
  - 10
  - 11
  - 12
  - 13
  - 14
  - 15
  - 16
  - 17
  - 18
  - 19
  - 20
  - 21
  - 22
  - 23
  - 24
  - 25
  - 26
  - 27
  - 28
  - 29
  - 30
  - 31
  - 32
  - 33
  - 34
  - 35
  - 36
  - 37
  - 38
  - 39
  - 40
  - 41
  - 42
  - 43
  - 44
  - 45
  - 46
  - 47
  - 48
  - 49
  - 50
  - 51
  - 52
  - 53
  - 54
  - 55
  - 56
  - 57
  - 58
  - 59
  - 60
26. Louie KS, Seigneurin A, Cathcart P, et al. Do prostate cancer risk models improve the predictive accuracy of PSA screening? A meta-analysis. *Annals of Oncology* 2015;26:848-64. doi: 10.1093/annonc/mdu525
27. Shariat SF, Kattan MW, Vickers AJ, et al. Critical review of prostate cancer predictive tools. *Future oncology (London, England)* 2009;5(10):1555-84. doi: 10.2217/fon.09.121 [published Online First: 2009/12/17]
28. Ross PL, Gerigk C, Gonen M, et al. Comparisons of nomograms and urologists' predictions in prostate cancer. *Seminars in urologic oncology* 2002;20(2):82-8. [published Online First: 2002/05/16]
29. 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(6):1254-8.
30. 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. *The Prostate* 2012;72(11):1200-6. doi: 10.1002/pros.22470 [published Online First: 2012/01/04]
31. 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(4):709-13.
32. 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  $\leq$  10 ng/mL. *Cancer* 2003;98(7):1417-22. doi: 10.1002/cncr.11668 [published Online First: 2003/09/26]
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(6):1930-4.
34. Kranse R, Roobol M, Schroder FH. A graphical device to represent the outcomes of a logistic regression analysis. *The Prostate* 2008;68(15):1674-80. doi: 10.1002/pros.20840 [published Online First: 2008/08/21]
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(8):529-34.
36. Schröder FH, Kruger AB, Rietbergen J, et al. Evaluation of the digital rectal examination as a screening test for prostate cancer. *Journal of the National Cancer Institute* 1998;90(23):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). *The Prostate* 2001;47(4):252-61.
38. Smith DS, Catalona WJ. Interexaminer variability of digital rectal examination in detecting prostate cancer. *Urology* 1995;45(1):70-74.

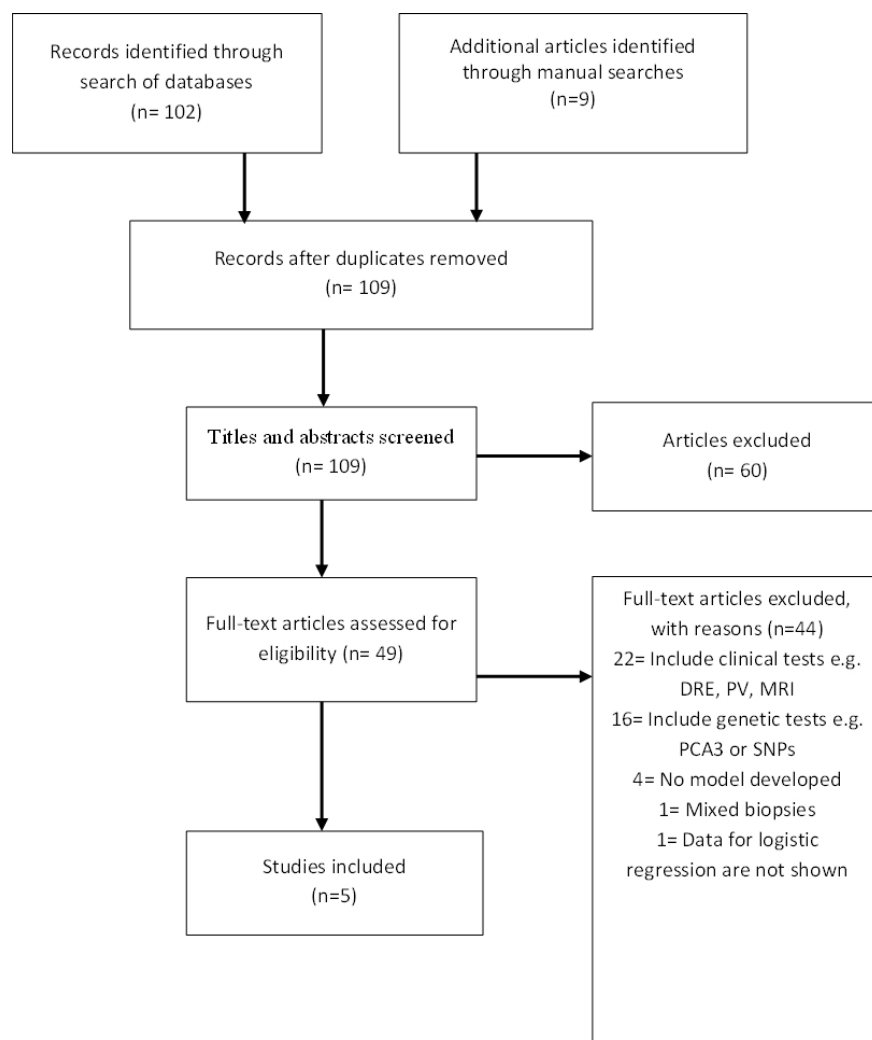
- 1
- 2
- 3
- 4 39. Barocas DA, Cowan JE, Smith JA, Jr., et al. What percentage of patients with newly diagnosed
- 5 carcinoma of the prostate are candidates for surveillance? An analysis of the CaPSURE
- 6 database. 2008
- 7
- 8 40. Vis A, Kranse R, Roobol M, et al. Serendipity in detecting disease in low prostate-specific
- 9 antigen ranges. *BJU international* 2002;89(4):384-89.
- 10
- 11 41. Littrup PJ, Bailey SE. Prostate cancer: the role of transrectal ultrasound and its impact on
- 12 cancer detection and management. *Radiologic Clinics of North America* 2000;38(1):87-
- 13 113.
- 14
- 15 42. Gerber GS, Chodak GW. Routine screening for cancer of the prostate. *JNCI: Journal of the*
- 16 *National Cancer Institute* 1991;83(5):329-35.
- 17
- 18 43. Gelfand DE, Parzuchowski J, Cort M, et al. Digital rectal examinations and prostate cancer
- 19 screening: attitudes of African American men. *Oncology Nursing Forum-Oncology*
- 20 *Nursing Society* 1995;22(8):1253-58.
- 21
- 22 44. Myers RE, Wolf TA, McKee L, et al. Factors associated with intention to undergo annual
- 23 prostate cancer screening among African American men in Philadelphia. *Cancer:*
- 24 *Interdisciplinary International Journal of the American Cancer Society* 1996;78(3):471-
- 25 79.
- 26
- 27 45. Consedine NS, Horton D, Ungar T, et al. Fear, knowledge, and efficacy beliefs differentially
- 28 predict the frequency of digital rectal examination versus prostate specific antigen
- 29 screening in ethnically diverse samples of older men. *American journal of men's health*
- 30 2007;1(1):29-43.
- 31
- 32 46. Consedine NS, Morgenstern AH, Kudadjie-Gyamfi E, et al. Prostate cancer screening behavior
- 33 in men from seven ethnic groups: the fear factor. *Cancer Epidemiology and Prevention*
- 34 *Biomarkers* 2006;15(2):228-37.
- 35
- 36 47. Ukoli FA, Patel K, Hargreaves M, et al. A tailored prostate cancer education intervention for
- 37 low-income African Americans: Impact on knowledge and screening. *Journal of health*
- 38 *care for the poor and underserved* 2013;24(1):311-31.
- 39
- 40 48. Ogunsanya ME, Brown CM, Odedina FT, et al. Beliefs regarding prostate cancer screening
- 41 among black males aged 18 to 40 Years. *American journal of men's health* 2017;11(1):41-
- 42 53.
- 43
- 44 49. Clarke-Tasker VA. What we thought we knew: African American males' perceptions of
- 45 prostate cancer and screening methods. *ABNF Journal* 2002;13(3):56.
- 46
- 47 50. Romero FR, Romero AW, Brenny Filho T, et al. Patients' perceptions of pain and discomfort
- 48 during digital rectal exam for prostate cancer screening. *Archivos espanoles de urologia*
- 49 2008;61(7):850-54.
- 50
- 51 51. Essink-Bot M-L, de Koning HJ, Nijs HG, et al. Short-term effects of population-based
- 52 screening for prostate cancer on health-related quality of life. *Journal of National Cancer*
- 53 *Institute* 1998;90(12):925-31.
- 54
- 55 52. Appleton L, Wyatt D, Perkins E, et al. The impact of prostate cancer on men's everyday life.
- 56 *European journal of cancer care* 2015;24(1):71-84.
- 57
- 58
- 59
- 60

- 1  
2  
3 53. Lee DJ, Consedine NS, Spencer BA. Barriers and facilitators to digital rectal examination  
4 screening among African-American and African-Caribbean men. *Urology* 2011;77(4):891-  
5 98.  
6
- 7 54. Nagler HM, Gerber EW, Homel P, et al. Digital rectal examination is barrier to population-  
8 based prostate cancer screening. *Urology* 2005;65(6):1137-40.  
9
- 10 55. Moher D, Liberati A, Tetzlaff J, et al. Preferred reporting items for systematic reviews and  
11 meta-analyses: the PRISMA statement. *PLoS medicine* 2009;6(7):e1000097.  
12
- 13 56. Park SH, Goo JM, Jo CH. Receiver operating characteristic (ROC) curve: practical review for  
14 radiologists. *Korean journal of radiology* 2004;5(1):11-8. doi: 10.3348/kjr.2004.5.1.11  
15 [published Online First: 2004/04/06]  
16
- 17 57. Han K, Song K, Choi BW. How to Develop, Validate, and Compare Clinical Prediction Models  
18 Involving Radiological Parameters: Study Design and Statistical Methods. *Korean journal*  
19 *of radiology* 2016;17(3):339-50. doi: 10.3348/kjr.2016.17.3.339 [published Online First:  
20 2016/05/03]  
21
- 22 58. Steyerberg EW. Clinical prediction models: a practical approach to development, validation,  
23 and updating: Springer Science & Business Media 2008.  
24
- 25 59. Pavlou M, Ambler G, Seaman SR, et al. How to develop a more accurate risk prediction model  
26 when there are few events. *Bmj* 2015;351:h3868.  
27
- 28 60. Wolff RF, Moons KGM, Riley RD, et al. PROBAST: A Tool to Assess the Risk of Bias and  
29 Applicability of Prediction Model Studies. *Ann Intern Med* 2019;170(1):51-58.  
30
- 31 61. Carlson GD, Calvanese CB, Partin AW. An algorithm combining age, total prostate-specific  
32 antigen (PSA), and percent free PSA to predict prostate cancer: results on 4298 cases.  
33 *Urology* 1998;52(3):455-61.  
34
- 35 62. Babaian RJ, Fritsche H, Ayala A, et al. Performance of a neural network in detecting prostate  
36 cancer in the prostate-specific antigen reflex range of 2.5 to 4.0 ng/mL. *Urology*  
37 2000;56(6):1000-6.  
38
- 39 63. Jansen FH, van Schaik RH, Kurstjens J, et al. Prostate-specific antigen (PSA) isoform p2PSA  
40 in combination with total PSA and free PSA improves diagnostic accuracy in prostate  
41 cancer detection. *European urology* 2010;57(6):921-27.  
42
- 43 64. Hill OT, Mason TJ, Schwartz SW, et al. Improving prostate cancer detection in veterans  
44 through the development of a clinical decision rule for prostate biopsy. *BMC Urol*  
45 2013;13:6.  
46
- 47 65. Lazzeri M, Haese A, de la Taille A, et al. Serum isoform [-2]proPSA derivatives significantly  
48 improve prediction of prostate cancer at initial biopsy in a total PSA range of 2-10 ng/ml:  
49 a multicentric European study. *Eur Urol* 2013;63(6):986-94.  
50
- 51 66. Moons KG, de Groot JA, Bouwmeester W, et al. Critical appraisal and data extraction for  
52 systematic reviews of prediction modelling studies: the CHARMS checklist. *PLoS*  
53 *medicine* 2014;11(10)  
54
- 55 67. Porter CR, O'Donnell C, Crawford ED, et al. Predicting the outcome of prostate biopsy in a  
56 racially diverse population: a prospective study. *Urology* 2002;60(5):831-35.  
57  
58  
59

- 1  
2  
3 68. Porter CR, Gamito EJ, Crawford ED, et al. Model to predict prostate biopsy outcome in large  
4 screening population with independent validation in referral setting. *Urology*  
5 2005;65(5):937-41.  
6
- 7 69. Catalona WJ, Bartsch G, Rittenhouse HG, et al. Serum pro-prostate specific antigen  
8 preferentially detects aggressive prostate cancers in men with 2 to 4 ng/ml prostate specific  
9 antigen. *J Urol* 2004;171(6 Pt 1):2239-44.  
10
- 11 70. Sokoll LJ, Sanda MG, Feng Z, et al. A prospective, multicenter, National Cancer Institute Early  
12 Detection Research Network study of [-2]proPSA: improving prostate cancer detection and  
13 correlating with cancer aggressiveness. *Cancer Epidemiol Biomarkers Prev*  
14 2010;19(5):1193-200.  
15
- 16 71. Sokoll LJ, Wang Y, Feng Z, et al. [-2]proenzyme prostate specific antigen for prostate cancer  
17 detection: a national cancer institute early detection research network validation study. *J*  
18 *Urol* 2008;180(2):539-43; discussion 43.  
19
- 20 72. Stephan C, Vincendeau S, Houlgatte A, et al. Multicenter evaluation of [-2]proprostate-specific  
21 antigen and the prostate health index for detecting prostate cancer. *Clin Chem*  
22 2013;59(1):306-14.  
23
- 24 73. Le BV, Griffin CR, Loeb S, et al. [-2]Proenzyme prostate specific antigen is more accurate  
25 than total and free prostate specific antigen in differentiating prostate cancer from benign  
26 disease in a prospective prostate cancer screening study. *J Urol* 2010;183(4):1355-9.  
27
- 28 74. Loeb S, Sokoll LJ, Broyles DL, et al. Prospective multicenter evaluation of the Beckman  
29 Coulter Prostate Health Index using WHO calibration. *J Urol* 2013;189(5):1702-06.  
30
- 31 75. Lughezzani G, Lazzeri M, Haese A, et al. Multicenter European external validation of a  
32 Prostate Health Index–based nomogram for predicting prostate cancer at extended biopsy.  
33 *Eur Urol* 2014;66(5):906-12.  
34
- 35 76. Chun FK-H, Briganti A, Graefen M, et al. Development and external validation of an extended  
36 10-core biopsy nomogram. *European urology* 2007;52:436-44.  
37
- 38 77. Herman MP, Dorsey P, John M, et al. Techniques and predictive models to improve prostate  
39 cancer detection. *Cancer* 2009;115(S13):3085-99.  
40
- 41 78. Trottier G, Roobol MJ, Lawrentschuk N, et al. Comparison of risk calculators from the Prostate  
42 Cancer Prevention Trial and the European Randomized Study of Screening for Prostate  
43 Cancer in a contemporary Canadian cohort. *BJU Int* 2011;108(8 Pt 2):E237-44.  
44
- 45 79. Collins GS, de Groot JA, Dutton S, et al. External validation of multivariable prediction  
46 models: a systematic review of methodological conduct and reporting. *BMC Med Res*  
47 *Methodol* 2014;14(1):40.  
48
- 49 80. Miyake H, Kurahashi T, Takenaka A, et al. Improved accuracy for predicting the Gleason score  
50 of prostate cancer by increasing the number of transrectal biopsy cores. *Urol Int*  
51 2007;79(4):302-6.  
52
- 53 81. Chun FK, Briganti A, Graefen M, et al. Development and external validation of an extended  
54 10-core biopsy nomogram. *Eur Urol* 2007;52(2):436-44.  
55  
56  
57  
58  
59



- 1
- 2
- 3
- 4 82. Aus G, Abbou CC, Bolla M, et al. EAU guidelines on prostate cancer. *Eur Urol*
- 5 2005;48(4):546-51.
- 6
- 7 83. Vickers AJ, Ulmert D, Serio AM, et al. The predictive value of prostate cancer biomarkers
- 8 depends on age and time to diagnosis: towards a biologically-based screening strategy. *Int*
- 9 *J Cancer* 2007;121(10):2212-7.
- 10
- 11 84. Steuber T, O'Brien MF, Lilja H. Serum markers for prostate cancer: a rational approach to the
- 12 literature. *Eur Urol* 2008;54(1):31-40.
- 13
- 14 85. Carlsson S, Assel M, Sjoberg D, et al. Influence of blood prostate specific antigen levels at age
- 15 60 on benefits and harms of prostate cancer screening: population based cohort study. *Bmj*
- 16 2014;348:g2296.
- 17
- 18 86. Cavadas V, Osorio L, Sabell F, et al. Prostate cancer prevention trial and European randomized
- 19 study of screening for prostate cancer risk calculators: a performance comparison in a
- 20 contemporary screened cohort. *Eur Urol* 2010;58(4):551-8.
- 21
- 22 87. Vickers AJ. Prediction models: revolutionary in principle, but do they do more good than harm.
- 23 *J Clin Oncol* 2011;29(22):2951-52.
- 24
- 25 88. Shariat SF, Karakiewicz PI, Margulis V, et al. Inventory of prostate cancer predictive tools.
- 26 *Curr Opin Urol* 2008;18(3):279-96.
- 27
- 28 89. Kawakami S, Numao N, Okubo Y, et al. Development, validation, and head-to-head
- 29 comparison of logistic regression-based nomograms and artificial neural network models
- 30 predicting prostate cancer on initial extended biopsy. *Eur Urol* 2008;54(3):601-11.
- 31
- 32 90. Hernandez DJ, Han M, Humphreys EB, et al. Predicting the outcome of prostate biopsy:
- 33 comparison of a novel logistic regression-based model, the prostate cancer risk calculator,
- 34 and prostate-specific antigen level alone. *BJU Int* 2009;103(5):609-14.
- 35
- 36 91. Moyer VA. Screening for prostate cancer: US Preventive Services Task Force recommendation
- 37 statement. *Ann Intern Med* 2012;157(2):120-34.
- 38
- 39 92. Bell N, Connor Gorber S, Shane A, et al. Recommendations on screening for prostate cancer
- 40 with the prostate-specific antigen test. *CMAJ* 2014;186(16):1225-34.
- 41
- 42 93. Lim LS, Sherin K, Committee APP. Screening for prostate cancer in US men: ACPM position
- 43 statement on preventive practice. *Am J Prev Med* 2008;34(2):164-70.
- 44
- 45 94. Prostate Cancer UK. Consensus statements on PSA testing in asymptomatic men in the UK -
- 46 information for health professionals 2016 [Available from:
- 47 <https://prostatecanceruk.org/about-us/projects-and-policies/consensus-on-psa-testing>
- 48 accessed 20-8 2019.
- 49
- 50 95. Frame I, Cant S. Current challenges in prostate cancer: an interview with Prostate Cancer UK.
- 51 *BMC Med* 2015;13(1):166.
- 52
- 53
- 54
- 55
- 56
- 57
- 58
- 59
- 60



Flow diagram of studies included using PRISMA method

78x91mm (300 x 300 DPI)

### Search Strategy

- 1) \* Prostatic Neoplasms/
- 2) Initial biopsy.mp
- 3) First biopsy.mp
- 4) \* “Early Detection of Cancer”/mt [Methods]
- 5) 2 or 3 or 4
- 6) 1 and 5
- 7) Nomograms/
- 8) “Neural Networks (Computer)”/
- 9) Risk Assessment/
- 10) Models, Statistical/
- 11) 7 or 8 or 9 or 10
- 12) 6 and 11
- 13) Limit 12 to English language
- 14) Limit 13 to abstracts



*Supplementary File 2: Search results for independent studies that did external validation for included models.*

<b>Study</b>	<b>No. of citations in Google Scholar</b>	<b>No. of citations in Medline</b>	<b>No. of external validation found in Google Scholar</b>	<b>No. of external validation found in Medline</b>
Carlson, 1998	117	20	Null	Null
Babaian, 2000	128	13	Null	Null
Jansen, 2010	231	44	Null	Null
Hill, 2010	5	0	Null	Null
Lazzeri, 2013	195	0	Null	Null

**Supplementary File 3: Type of assays used to measure PSA and free PSA.**

Study	PSA Assay	Free PSA Assay
Carlson, 1998 <sup>1</sup>	PA immunoassay (Tosoh)	Investigational doubleantibody radioimmunometric assay (PSA II, DIANON Systems)
Babaian, 2000 <sup>2</sup>	PSA immunometric assay (Tosoh)	Tandem R assay (Beckman Coulter, San Diego, Calif)
Jansen, 2010 <sup>3</sup>	Access 2 Immunoassay System (Beckman Coulter, Brea, CA, USA)	Access 2 Immunoassay System (Beckman Coulter, Brea, CA, USA)
Hill, 2013 <sup>4</sup>	N/A	N/A
Lazzeri, 2013 <sup>5</sup>	Access 2 Immunoassay System (Beckman Coulter, Brea, CA, USA)	Access 2 Immunoassay System (Beckman Coulter, Brea, CA, USA)

## References:

1. 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. doi: 10.1016/S0090-4295(98)00205-2
2. 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.
3. 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. *European Urology* 2010;57:921-27. doi: 10.1016/j.eururo.2010.02.003
4. 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 urology* 2013;13 doi: 10.1186/1471-2490-13-6
5. 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. *European Urology* 2013;63:986-94. doi: 10.1016/j.eururo.2013.01.011

# Reporting checklist for systematic review and meta-analysis.

Based on the PRISMA guidelines.

## Instructions to authors

Complete this checklist by entering the page numbers from your manuscript where readers will find each of the items listed below.

Your article may not currently address all the items on the checklist. Please modify your text to include the missing information. If you are certain that an item does not apply, please write "n/a" and provide a short explanation.

Upload your completed checklist as an extra file when you submit to a journal.

In your methods section, say that you used the PRISMA reporting guidelines, and cite them as:

Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement

	Reporting Item	Page Number
<b>Title</b>		
	<a href="#">#1</a> Identify the report as a systematic review, meta-analysis, or both.	3
<b>Abstract</b>		
Structured summary	<a href="#">#2</a> Provide a structured summary including, as applicable: background; objectives; data sources; study eligibility criteria, participants, and interventions; study appraisal and synthesis methods; results; limitations; conclusions and implications of key findings; systematic review registration number	3-4
<b>Introduction</b>		
Rationale	<a href="#">#3</a> Describe the rationale for the review in the context of what is already known.	5-6
Objectives	<a href="#">#4</a> Provide an explicit statement of questions being addressed with	6

reference to participants, interventions, comparisons, outcomes, and study design (PICOS).

## Methods

Protocol and registration	<a href="#">#5</a>	Indicate if a review protocol exists, if and where it can be accessed (e.g., Web address) and, if available, provide registration information including the registration number.	n/a
Eligibility criteria	<a href="#">#6</a>	Specify study characteristics (e.g., PICOS, length of follow-up) and report characteristics (e.g., years considered, language, publication status) used as criteria for eligibility, giving rational	6
Information sources	<a href="#">#7</a>	Describe all information sources in the search (e.g., databases with dates of coverage, contact with study authors to identify additional studies) and date last searched.	6
Search	<a href="#">#8</a>	Present full electronic search strategy for at least one database, including any limits used, such that it could be repeated.	6
Study selection	<a href="#">#9</a>	State the process for selecting studies (i.e., for screening, for determining eligibility, for inclusion in the systematic review, and, if applicable, for inclusion in the meta-analysis).	7
Data collection process	<a href="#">#10</a>	Describe the method of data extraction from reports (e.g., piloted forms, independently by two reviewers) and any processes for obtaining and confirming data from investigators.	7
Data items	<a href="#">#11</a>	List and define all variables for which data were sought (e.g., PICOS, funding sources), and any assumptions and simplifications made.	8
Risk of bias in individual studies	<a href="#">#12</a>	Describe methods used for assessing risk of bias in individual studies (including specification of whether this was done at the study or outcome level, or both), and how this information is to be used in any data synthesis.	8
Summary measures	<a href="#">#13</a>	State the principal summary measures (e.g., risk ratio, difference in means).	7
Planned methods of analysis	<a href="#">#14</a>	Describe the methods of handling data and combining results of studies, if done, including measures of consistency (e.g., I <sup>2</sup> ) for each meta-analysis.	6-8
Risk of bias	<a href="#">#15</a>	Specify any assessment of risk of bias that may affect the	15-16

1	across studies		cumulative evidence (e.g., publication bias, selective reporting within studies).	
2				
3				
4	Additional	<a href="#">#16</a>	Describe methods of additional analyses (e.g., sensitivity or	n/a
5	analyses		subgroup analyses, meta-regression), if done, indicating which	
6			were pre-specified.	
7				
8				
9	<b>Results</b>			
10				
11	Study selection	<a href="#">#17</a>	Give numbers of studies screened, assessed for eligibility, and	8-9
12			included in the review, with reasons for exclusions at each stage,	
13			ideally with a <a href="#">flow diagram</a> .	
14				
15				
16				
17	Study	<a href="#">#18</a>	For each study, present characteristics for which data were	8-9
18	characteristics		extracted (e.g., study size, PICOS, follow-up period) and provide	
19			the citation.	
20				
21				
22	Risk of bias	<a href="#">#19</a>	Present data on risk of bias of each study and, if available, any	15-17
23	within studies		outcome-level assessment (see Item 12).	
24				
25				
26	Results of	<a href="#">#20</a>	For all outcomes considered (benefits and harms), present, for	13-15
27	individual studies		each study: (a) simple summary data for each intervention group	
28			and (b) effect estimates and confidence intervals, ideally with a	
29			forest plot.	
30				
31				
32				
33	Synthesis of	<a href="#">#21</a>	Present the main results of the review. If meta-analyses are	12-15
34	results		done, include for each, confidence intervals and measures of	
35			consistency.	
36				
37				
38	Risk of bias	<a href="#">#22</a>	Present results of any assessment of risk of bias across studies	15-17
39	across studies		(see Item 15).	
40				
41				
42	Additional	<a href="#">#23</a>	Give results of additional analyses, if done (e.g., sensitivity or	n/a
43	analysis		subgroup analyses, meta-regression [see Item 16]).	
44				
45				
46	<b>Discussion</b>			
47				
48	Summary of	<a href="#">#24</a>	Summarize the main findings, including the strength of evidence	18-21
49	Evidence		for each main outcome; consider their relevance to key groups	
50			(e.g., health care providers, users, and policy makers	
51				
52				
53	Limitations	<a href="#">#25</a>	Discuss limitations at study and outcome level (e.g., risk of bias),	21-22
54			and at review level (e.g., incomplete retrieval of identified	
55			research, reporting bias).	
56				
57				
58				
59				
60				

1	Conclusions	<a href="#">#26</a>	Provide a general interpretation of the results in the context of	23
2			other evidence, and implications for future research.	
3				
4				
5	<b>Funding</b>			
6				
7	Funding	<a href="#">#27</a>	Describe sources of funding or other support (e.g., supply of	24
8			data) for the systematic review; role of funders for the systematic	
9			review.	
10				
11				

12 The PRISMA checklist is distributed under the terms of the Creative Commons Attribution License  
13 CC-BY. This checklist was completed on 30. September 2019 using <https://www.goodreports.org/>, a  
14 tool made by the [EQUATOR Network](#) in collaboration with [Penelope.ai](#)  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60