Diagnostic accuracy of [-2]proPSA versus Gleason score and Prostate Health Index versus Gleason score for the determination of aggressive prostate cancer: a systematic review

Ruth Anyango[1,2] • Joel Ojwando[1,2] • Clifford Mwita[1,2] • Edward Mugalo[2]

[1]Afya Research Africa (ARA): A JBI Centre of Excellence, Nairobi, Kenya, [2]Moi University School of Medicine, Eldoret, Kenya

ABSTRACT

Objective: The objective of this review was to determine the diagnostic accuracy of [-2]proPSA (p2PSA) and the Prostate Health Index compared to the Gleason score in determining the aggressiveness of prostate cancer.

Introduction: Prostate cancer is the most commonly diagnosed cancer in men. However, the utility of currently available biomarkers for determining the aggressive form of the disease remains unknown. This review sought to determine the diagnostic accuracy of two new biomarkers in determining the aggressive form of prostate cancer.

Inclusion criteria: Diagnostic accuracy studies that enrolled men of any age and any prostate specific antigen (PSA) level with histologically confirmed prostate cancer in which Prostate Health Index and p2PSA were assessed in comparison to Gleason score for the determination of aggressive prostate cancer were considered for inclusion. There was no time limitation on study inclusion.

Methods: A three-step search strategy was utilized to identify both published and unpublished studies in the English language in the following sources: PubMed, Cochrane Central Register of Controlled Trials, CINAHL, Web of Science, Google Scholar, MedNar, and SIGLE. Databases were searched from inception to January 2019. Study selection, critical appraisal, data extraction, and data synthesis were done according to the approach recommended by JBI.

Results: A total of 12 studies (n = 8462) that recruited men with aggressive prostate cancer were considered in this review. The majority of included subjects had a total PSA level of 2 to 10ng/mL. The sensitivity of the Prostate Health Index ranged from 67% to 97% while specificity ranged from 6% to 64%. At a Prostate Health Index threshold of 25 and below (three studies, n = 3222), pooled sensitivity was 97% (95% confidence interval [CI], 95% to 98%) and specificity was 10% (95% CI, 6% to 16%). At a Prostate Health Index threshold of between 26 and 35 (six studies, n = 6030), pooled sensitivity was 87% (95% CI, 8% to 91%) and specificity was 45% (95% CI, 39% to 50%). At a Prostate Health Index threshold of 36 and above (five studies, n = 1476), pooled sensitivity was 72% (95% CI, 64% to 79%) and specificity was 74% (95% CI, 68% to 80%). Only one study assessed p2PSA. Sensitivity ranged from 80% to 95%, and specificity ranged from 9.9% to 27.9% with increasing threshold values from 7.9 to 10.9ng/mL.

Conclusions: Overall, both Prostate Health Index and p2PSA have acceptable accuracy for the determination of the likelihood of aggressive prostate cancer. However, the inverse relationship between sensitivity and specificity makes it difficult to determine an optimum cut-off value for positivity. Further research is warranted to determine their utility in the management of prostate cancer.

Keywords: diagnostic accuracy; p2PSA; prostate cancer; Prostate Health Index; systematic review

JBI Evid Synth 2021; 19(6):1263–1291.

Correspondence: Ruth Anyango, anyangoruth@gmail.com

The authors declare no conflicts of interest.

DOI: 10.11124/JBISRIR-D-19-00194

Summary of Findings

Question: Should Prostate Health Index threshold value <25 be used to diagnose aggressive prostate cancer?

Bibliography: Anyango R, Ojwando J, Mwita C, Mugalo E. Diagnostic accuracy of [-2]proPSA versus Gleason score and Prostate Health Index versus Gleason score for the determination of aggressive prostate cancer: a systematic review. JBI Evid Synth. 2021;19(6):1263–1291.

Sensitivity	0.97 (95% confidence interval: 0.95 to 0.98)
Specificity	0.10 (95% confidence interval: 0.06 to 0.16)
Prevalences	2.8%, 15.6%, 30.8%

Outcome	No. of studies (No. of patients)	Study design	Factors that may decrease certainty of evidence					Effect per 1000 patients tested			Test accuracy CoE
			Risk of bias	Indirectness	Inconsistency	Imprecision	Other considerations	Pre-test probability of 2.8%	Pre-test probability of 15.6%	Pre-test probability of 30.8%	
True positives (patients with aggressive prostate cancer)	3 studies (3222 patients)	cross-sectional (cohort type accuracy study)	serious[a]	not serious	serious[b]	not serious	none	27 (27 to 27)	151 (148 to 153)	299 (293 to 302)	⨁⨁◯◯ LOW
False negatives (patients incorrectly classified as not having aggressive prostate cancer)								1 (1 to 1)	5 (3 to 8)	9 (6 to 15)	
True negatives (patients without aggressive prostate cancer)	3 studies (3222 patients)	cross-sectional (cohort type accuracy study)	serious[a]	not serious	serious[b]	not serious	none	97 (58 to 156)	84 (51 to 135)	69 (42 to 111)	⨁⨁◯◯ LOW
False positives (patients incorrectly classified as having aggressive prostate cancer)								875 (816 to 914)	760 (709 to 793)	623 (581 to 650)	

CoE, certainty of the evidence
GRADE Working Group grades of evidence
High certainty: We are very confident that the true effect lies close to that of the estimate of the effect
Moderate certainty: We are moderately confident in the effect estimate: The true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different
Low certainty: Our confidence in the effect estimate is limited: The true effect may be substantially different from the estimate of the effect
Very low certainty: We have very little confidence in the effect estimate: The true effect is likely to be substantially different from the estimate of effect

Explanations
a. Methodological issues, particularly pertaining to blinding, were noted across studies included in these analyses.
b. Effect estimates (sensitivity and specificity) varied widely, particularly because different threshold values were reported/used in the analyses.

Question: Should Prostate Health Index threshold value of 26 to 35 be used to diagnose aggressive prostate cancer?

Bibliography: Anyango R, Ojwando J, Mwita C, Mugalo E. Diagnostic accuracy of [-2]proPSA versus Gleason score and Prostate Health Index versus Gleason score for the determination of aggressive prostate cancer: a systematic review. JBI Evid Synth. 2021;19(6):1263–1291.

Sensitivity	0.87 (95% confidence interval: 0.81 to 0.91)
Specificity	0.45 (95% confidence interval: 0.34 to 0.50)
Prevalences	2.8%, 15.6%, 30.8%

Outcome	No. of studies (No. of patients)	Study design	Factors that may decrease certainty of evidence					Effect per 1000 patients tested			Test accuracy CoE
			Risk of bias	Indirectness	Inconsistency	Imprecision	Other considerations	Pre-test probability of 2.8%	Pre-test probability of 15.6%	Pre-test probability of 30.8%	
True positives (patients with aggressive prostate cancer)	6 studies (6030 patients)	cohort type studies	serious[a]	not serious	serious[b]	not serious	none	24 (23 to 25)	136 (126 to 142)	268 (249 to 280)	⨁⨁◯◯ LOW
False negatives (patients incorrectly classified as not having aggressive prostate cancer)								4 (3 to 5)	20 (14 to 30)	40 (28 to 59)	
True negatives (patients without aggressive prostate cancer)	6 studies (6030 patients)	cohort type studies	serious[a]	not serious	serious[b]	not serious	none	437 (330 to 486)	380 (287 to 422)	311 (235 to 346)	⨁⨁◯◯ LOW
False positives (patients incorrectly classified as having aggressive prostate cancer)								535 (486 to 642)	464 (422 to 557)	381 (346 to 457)	

CoE, certainty of the evidence

GRADE Working Group grades of evidence
High certainty: We are very confident that the true effect lies close to that of the estimate of the effect
Moderate certainty: We are moderately confident in the effect estimate: The true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different
Low certainty: Our confidence in the effect estimate is limited: The true effect may be substantially different from the estimate of the effect
Very low certainty: We have very little confidence in the effect estimate: The true effect is likely to be substantially different from the estimate of effect

Explanations
a. Methodological issues, particularly pertaining to blinding, were noted across studies included in these analyses.
b. Effect estimates (sensitivity and specificity) varied widely, particularly because different threshold values were reported/used in the analyses.

Question: Should Prostate Health Index threshold value >36 be used to diagnose aggressive prostate cancer?

Bibliography: Anyango R, Ojwando J, Mwita C, Mugalo E. Diagnostic accuracy of [-2]proPSA versus Gleason score and Prostate Health Index versus Gleason score for the determination of aggressive prostate cancer: a systematic review. JBI Evid Synth. 2021;19(6):1263–1291.

Sensitivity	0.72 (95% confidence interval: 0.64 to 0.79)
Specificity	0.74 (95% confidence interval: 0.68 to 0.80)
Prevalences	2.8%, 15.6%, 30.8%

Outcome	No. of studies (No. of patients)	Study design	Factors that may decrease certainty of evidence					Effect per 1000 patients tested			Test accuracy CoE
			Risk of bias	Indirectness	Inconsistency	Imprecision	Other considerations	Pre-test probability of 2.8%	Pre-test probability of 15.6%	Ppre-test probability of 30.8%	
True positives (patients with aggressive prostate cancer)	5 studies (1476 patients)	cross-sectional (cohort type accuracy study)	serious[a]	not serious	serious[b]	not serious	none	20 (18 to 22)	112 (100 to 123)	222 (197 to 243)	⨁⨁◯◯ LOW
False negatives (patients incorrectly classified as not having aggressive prostate cancer)								8 (6 to 10)	44 (33 to 56)	86 (65 to 111)	
True negatives (patients without aggressive prostate cancer)	5 studies (1476 patients)	cross-sectional (cohort type accuracy study)	serious[a]	not serious	serious[b]	not serious	none	719 (661 to 778)	625 (574 to 675)	512 (471 to 554)	⨁⨁◯◯ LOW
False positives (patients incorrectly classified as having aggressive prostate cancer)								253 (194 to 311)	219 (169 to 270)	180 (138 to 221)	

CoE, certainty of the evidence

GRADE Working Group grades of evidence
High certainty: We are very confident that the true effect lies close to that of the estimate of the effect
Moderate certainty: We are moderately confident in the effect estimate: The true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different
Low certainty: Our confidence in the effect estimate is limited: The true effect may be substantially different from the estimate of the effect
Very low certainty: We have very little confidence in the effect estimate: The true effect is likely to be substantially different from the estimate of effect

Explanations
a. Methodological issues, particularly pertaining to blinding, were noted across studies included in these analyses.
b. Effect estimates (sensitivity and specificity) varied widely, particularly because different threshold values were reported/used in the analyses.

Question: Should [-2]proPSA be used to diagnose aggressive prostate cancer?

Bibliography: Anyango R, Ojwando J, Mwita C, Mugalo E. Diagnostic accuracy of [-2]proPSA versus Gleason score and Prostate Health Index versus Gleason score for the determination of aggressive prostate cancer: a systematic review. JBI Evid Synth. 2021;19(6):1263–1291.

Sensitivity	0.80 to 0.95
Specificity	0.10 to 0.28
Prevalences	2.8%, 15.6%, 30.8%

Outcome	No. of studies (No. of patients)	Study design	Factors that may decrease certainty of evidence					Effect per 1000 patients tested			Test accuracy CoE
			Risk of bias	Indirectness	Inconsistency	Imprecision	Other considerations	Pre-test probability of 2.8%	Pre-test probability of 15.6%	Pre-test probability of 30.8%	
True positives (patients with aggressive prostate cancer)	1 study (657 patients)	cross-sectional (cohort type accuracy study)	not serious	not serious	serious[a]	serious[b]	none	22 to 27	125 to 148	246 to 293	⨁⨁◯◯ LOW
False negatives (patients incorrectly classified as not having aggressive prostate cancer)								1 to 6	8 to 31	15 to 62	
True negatives (patients without aggressive prostate cancer)	1 study (657 patients)	cross-sectional (cohort type accuracy study)	not serious	not serious	serious[a]	serious[b]	none	96 to 271	84 to 235	69 to 193	⨁⨁◯◯ LOW
False positives (patients incorrectly classified as having aggressive prostate cancer)								701 to 876	609 to 760	499 to 623	

CoE, certainty of the evidence
GRADE Working Group grades of evidence
High certainty: We are very confident that the true effect lies close to that of the estimate of the effect
Moderate certainty: We are moderately confident in the effect estimate: The true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different
Low certainty: Our confidence in the effect estimate is limited: The true effect may be substantially different from the estimate of the effect
Very low certainty: We have very little confidence in the effect estimate: The true effect is likely to be substantially different from the estimate of effect

Explanations
a. Different cut-off values were employed across included studies, and this resulted in inconsistent estimates. Even where similar cut-off values were used, results still were not consistent.
b. There was a mix of wide and narrow confidence intervals noted.

Introduction

Among men in high-income countries, prostate cancer (PCa) is the most commonly diagnosed cancer and the second most common cause of cancer-related death.[1,2] In low- and middle-income countries, the reported burden of PCa is less than that of high-income countries, and this difference is likely due to lower rates of PCa screening.[3,4] Ethnically, PCa has a higher incidence and higher mortality rates in men of African descent compared to men of Caucasian or Asian descent.[5]

Histological examination of prostatic tissue is the gold standard for the diagnosis of PCa. It is also useful in determining tumor aggressiveness by grading PCa according to the degree of cellular differentiation. The disadvantage of this technique is that obtaining prostate tissue for examination is invasive and carries a risk of infection and bleeding. This has necessitated the development and use of diagnostic biomarkers for PCa. One such marker is a prostate-specific antigen (PSA), a glycoprotein secreted by cells lining the prostate gland and currently the most widely used biomarker for the early detection of PCa.[6] However, it is non-specific and other prostatic conditions have been shown to cause elevated PSA levels.[7,8] Although PSA as a screening tool has been shown to lead to a reduction in mortality associated \with PCa, its poor specificity and inability to distinguish between lower-grade indolent tumors and higher-grade aggressive tumors means that patients are often subjected to unnecessary testing and therapy.[9]

The limitations of PSA have led to the introduction of PSA derivatives such as free PSA (fPSA), percentage free PSA (%fPSA), PSA density, and PSA velocity. However, these are only marginally better than total PSA (tPSA) screening at detecting PCa or determining its aggressiveness.[10] More recently, there has been the introduction of a zymogen precursor of PSA that comes in four different isoforms.[8] This marker is known as proPSA, and it has been shown to be preferentially elevated in the peripheral zone of prostatic tissue where the majority of PCa arises. As such, it has been proposed as a more specific biomarker for PCa.[9] One of its isoforms, [-2]proPSA (p2PSA), is mostly expressed in the peripheral zone of the gland and has been shown to have higher levels in the sera of men with PCa.[9,11] Additional efforts to improve the ability to detect PCa non-invasively have resulted in two other derivatives of p2PSA: percent p2PSA (%p2PSA) and the Beckman Coulter Prostate Health Index (PHI). The Beckman Coulter PHI is a mathematical combination of p2PSA, fPSA, and tPSA using the formula: ([p2PSA/fPSA] $\times$ $\sqrt{}$tPSA). This combination is thought to be better than its individual components at detecting PCa, and it has been shown to have an acceptable diagnostic accuracy for PCa.[12]

At present, as recommended by the European Association of Urology, tumor grading is performed on tissue obtained from a prostate biopsy using the Gleason scoring system.[13] This system is based on assessing the architecture of cells on tissue biopsy. Aggressive tumors have poorly differentiated cells and therefore grow faster and are more invasive. Less-aggressive tumors have well-differentiated cells and are slow growing, and as such, patients with such indolent tumors may live with the disease for a long period of time without therapy. High Gleason scores (>7) indicate a poorer prognosis and the need for more aggressive treatment. Compared to the Gleason score, %p2PSA has a sensitivity of 96% and specificity of 9% for detecting aggressive disease, whereas PHI has a sensitivity of 90% and a specificity of 17%.[12]

In contemporary urological practice, patients assessed for PCa first undergo a digital rectal examination. The presence of prostatic nodules, mucosal fixation, or obliteration of the prostatic median sulcus on digital rectal examination points to the possibility of PCa. In addition, the seminal vesicles may also be assessed for presence/involvement of tumor. Under such circumstances, a prostate biopsy (preferably with ultrasound guidance) is warranted. However, patients may have impalpable tumors, a situation that necessitates measurement of PSA levels. High PSA levels (10ng/mL or higher) warrant a biopsy. Owing to the poor accuracy of PSA, patients without PCa may undergo an unnecessary biopsy while others who have PCa (with normal PSA values) will not undergo further evaluation. It is therefore advantageous to have a test that can accurately and non-invasively determine the presence and aggressiveness of PCa.

Two systematic reviews on the diagnosis of aggressive PCa have been published. In the review by Russo *et al.*,[14] the accuracy of PHI for diagnosing both PCa and aggressive PCa was assessed. The pooled sensitivity for PHI for detecting PCa was 0.89 while specificity was 0.34. The pooled sensitivity for PHI for detecting aggressive PCa was 0.93 while specificity was 0.34. Although this review was

of acceptable quality, the authors did not report an *a priori* protocol nor did they consider the different cut-off values used in determining the positivity of PHI. The issue of PHI positivity cut-off level in this review has been raised in a published letter to the editor[15] in which the authors of the letter question the statistical methods employed by Russo and colleagues.[14] Further, since the publication of this review, there have been additional studies assessing the accuracy of PHI for determining aggressive PCa. In a more recent systematic review, Wang and colleagues[12] examined the accuracy of %p2PSA and PHI in determining PCa aggressiveness. The accuracy of PHI for aggressive PCa was reported as a sensitivity of 0.9 and specificity of 0.17. The authors concluded that PHI may be useful in diagnosing aggressive PCa. However, they did not perform an extensive search for relevant studies as only PubMed was searched for their review. Further, there was no mention of an *a priori* protocol to guide the conduct of the review. These factors point to a potentially high risk of bias associated with these two reviews.

This systematic review attempted to evaluate the role of p2PSA and PHI in detecting the aggressiveness of PCa using a more explicit methodology. The main outcome measures were sensitivity and specificity of the two tests compared to Gleason scores as the reference standard. A search in MEDLINE, PROSPERO, and the Cochrane library yielded no ongoing reviews with similar questions.

Review questions

i. What is the diagnostic accuracy of [-2]proPSA (p2PSA; index test) compared to Gleason score in determining the aggressiveness of PCa?
ii. What is the diagnostic accuracy of the Prostate Health Index (index test) compared to Gleason score in determining the aggressiveness of PCa?

Inclusion criteria

Participants

This review considered studies that enrolled men of any age who had a diagnosis of aggressive PCa as determined from biopsy specimens, and with any range of PSA levels.

Index test

Studies that evaluated the performance of p2PSA or PHI for the determination of the aggressiveness of PCa were included. Studies in which p2PSA and PHI were primarily evaluated for the diagnosis of PCa but also reported sufficient data on tumor aggressiveness were also considered for inclusion.

Reference test

This review considered studies that used the biopsy Gleason score as the reference standard for determining the aggressiveness of PCa.

Diagnosis of interest

Studies that considered the aggressiveness of biopsy-obtained histologically confirmed PCa as the diagnosis of interest were the focus of this review. Studies in which the confirmation of PCa was the main aim but that reported on the aggressiveness of disease were also considered.

Types of studies

This review considered diagnostic test accuracy studies that utilized observational cross-sectional, cohort, or case-control designs in which the index and reference tests were interpreted in the same group of participants. There was no time limitation on study inclusion. Studies that did not explicitly report diagnostic test accuracy outcomes (sensitivity and specificity), but that had sufficient information to calculate these outcomes, were also considered for inclusion.

Exclusion criteria

Systematic reviews, letters to the editor, literature reviews, conference abstracts, and studies published in languages other than English were excluded from the review. Studies that assessed aggressiveness of PCa using means other than the Gleason score were also not considered for inclusion in this review.

Methods

This review was conducted according to an *a priori* protocol.[16] The JBI methodology for systematic reviews of diagnostic test accuracy was utilized for this review.[17]

Search strategy

An extensive computerized literature search was conducted to locate both published and unpublished studies written in English. A three-step search strategy was implemented. First, an initial search of PubMed followed by analysis of text words in the title and

abstract were done. A second search using identified key words and index terms was conducted across all databases. Finally, the reference lists of included reports and articles were searched for additional studies. Databases searched were MEDLINE (PubMed), Cochrane Central Register of Controlled Trials, CINAHL (EBSCO), and Web of Science. The search for unpublished studies included Google Scholar, MedNar, and SIGLE (System for Information on Grey Literature in Europe). All databases were searched from inception to January 2019. Only studies published in English were considered for inclusion. Initial keywords used to construct the search were: PCa, pro2PSA, p2PSA, PHI, prostatic health index, accuracy, sensitivity, specificity, likelihood ratio, predictive value, Gleason score. The search strategies for all databases searched are shown in Appendix I.

Study selection

Records identified in the searches were collated and managed using EndNote v.X7 (Clarivate Analytics, PA, USA). After removal of duplicate records, identified titles and abstracts were screened and those that did not fit the inclusion criteria were excluded. Full-text articles for the remaining records were retrieved and assessed for inclusion. Studies that did not meet the inclusion criteria were excluded and particular reasons for exclusion were recorded (Appendix II). The reference lists of studies found eligible for inclusion were also searched for additional studies. Two independent reviewers were involved in the study selection processes. Discrepancies were resolved by discussion or by conferring with a third reviewer.

Assessment of methodological quality

Papers selected for inclusion in the review were assessed for methodological quality by two independent reviewers using the JBI critical appraisal checklist for diagnostic test accuracy studies adapted from the QUADAS-2 tool.[17] A critical appraisal item was designated ''Yes'' or ''No'' where the authors provided ample information for the reviewers to judge it as done or not done. In the event that such information was inadequate in order to make a judgment, the item was determined as ''Unclear.'' Where there were disagreements regarding study quality, resolution was reached by consensus or by involving a third reviewer. The results of methodological quality assessment did not affect the decision to include studies in the review; therefore, all studies that underwent critical appraisal were subsequently included in the review.

Data extraction

A JBI data extraction form was used to collect relevant data from included studies.[17] To minimize errors, data extraction was performed by two independent reviewers and discrepancies resolved by either consensus or adjudication by a third reviewer. The following parameters were obtained: date of publication, authors, study location, study design, number of study participants and their demographic characteristics, index test, and reference standard tests utilized in the study. Sensitivity and specificity, true positive (TP), false positive (FP), true negative (TN), and false negative (FN) data for tests at each cut-off level were obtained directly from the respective papers. Two-by-two contingency tables were utilized for the purpose of data extraction. For the purpose of this review (ie, ability of PHI and p2PSA to determine aggressive PCa), a positive reference standard was considered a Gleason score of 7 or above. A negative reference standard was considered a Gleason score of below 7 or a negative prostate biopsy. A positive index test was considered a PHI or p2PSA level above threshold while a negative index test was considered a PHI or p2PSA level below the threshold value. Patients with positive index and reference tests were considered TP, while those with a positive index and a negative reference test were considered FP. Those with a negative index and a positive reference test were considered FN, while those with negative index and reference tests were considered TN.

Where it was not possible to extract the aforementioned parameters directly from included studies, they were calculated from the data provided. This calculation was based on the two-by-two contingency tables constructed for each test and cut-off level as reported in each study. All studies where any of these parameters were missing reported both sensitivity and specificity at a certain threshold value. As such, for each threshold value, TP was determined by multiplying sensitivity by the number of participants with the target condition (ie, reference standard was positive), and FN by subtracting TP from the total number with the target condition. Similarly, for each threshold value, TN was arrived at by multiplying specificity by the number of participants without the target condition (ie, reference

standard was negative), and FP was determined by subtracting TN from the total number without the target condition.

Some of the included studies reported index test accuracy at various pre-determined threshold values. In such circumstances, TP, FP, TN, and FN were determined for each threshold value reported. Where no specific threshold value was reported but there were sufficient data to fill in the contingency table, the cut-off value recommended by Beckman Coulter was utilized. Where there were incomplete or missing data useful for the review, an attempt was made to contact the primary authors of included studies.

Data synthesis

The aim of data synthesis was to determine the diagnostic performance of p2PSA and PHI in determining aggressive PCa based on the Gleason score obtained from the results of tissue biopsy. Data synthesis was performed in three steps. First, for each study and at each threshold value specified, sensitivity and specificity were determined and results displayed on forest plots. Because included studies utilized different cut-off values, significant heterogeneity was bound to be present during data synthesis. Consequently, we did not determine summary measures of diagnostic accuracy for all included studies. Instead, for PHI, the results were categorized into three groups: PHI $\leq$25, PHI between 26 and 35, and PHI $\geq$36. Thereafter, the hierarchical summary receiver operator characteristic (HSROC) model was used to determine summary measures (sensitivity, specificity) within these categories and SROC curves with 95% confidence and prediction regions were generated. Based on the aforementioned categorization of PHI, there was considerable overlap of studies between the categories. As such, it was not possible to perform subgroup analysis. However, sensitivity analysis was done by omitting from each group the studies in which blinding was unclear. Data analysis was performed using STATA v.13 (Stata Corp LLC, Texas, USA) and RevMan v5.3.5 (Copenhagen: The Nordic Cochrane Center, Norway).

Assessing certainty in the findings

The Grading of Recommendations Assessment, Development, and Evaluation (GRADE) approach for assessing confidence in the quality of evidence was utilized in this review and the results presented in a Summary of Findings created using GRADEpro (McMaster University, ON, Canada).

Results

Study inclusion

Figure 1 presents the study selection process.[18] A total of 2878 potentially relevant titles were identified through database searching. After correcting for duplicates and screening titles and abstracts, 2772 studies were excluded and 106 potentially relevant studies were identified for full-text review. Of these, 94 were excluded and the remaining 12 studies[19-30] fulfilled the inclusion criteria and underwent critical appraisal. Reasons for exclusion for the 94 studies[31-123] are provided in Appendix II.

Methodological quality

All studies that met the inclusion criteria underwent methodological quality assessment by two independent reviewers. Discrepancies were resolved by either consensus or adjudication by a third reviewer. Overall, included studies were of good quality. In included studies, all patients who were recruited were included in the analysis (Q10). Similarly, every study reported that all patients received the same reference standard (Q9). The likelihood of interval bias across included studies was low because an appropriate time interval between index and reference tests was noted in majority of included studies and was only unclear in two studies (Q8).[19,20] The reference standard and index tests were both interpreted without knowledge of each other (Q4 and Q7) in seven studies,[19,21-26] meaning that blinding was adequate. In all studies, the reference standard used was likely to correctly classify the condition of interest (Q6), meaning that the likelihood of misclassification bias was low. The threshold value for the index test was pre-specified (Q5) in 10 studies while in two studies,[23,27] it was not specified. All included studies reported no inappropriate exclusions (Q3), and a consecutive/random sample of patients was recruited (Q1). Two included studies employed a case-control design (Q2).[19,23] No study was excluded on the basis of study quality. The methodological quality of included studies is presented in Table 1.

Characteristics of included studies

The key characteristics of the 12 studies included in this review are described below. A summary is provided in Table 2.

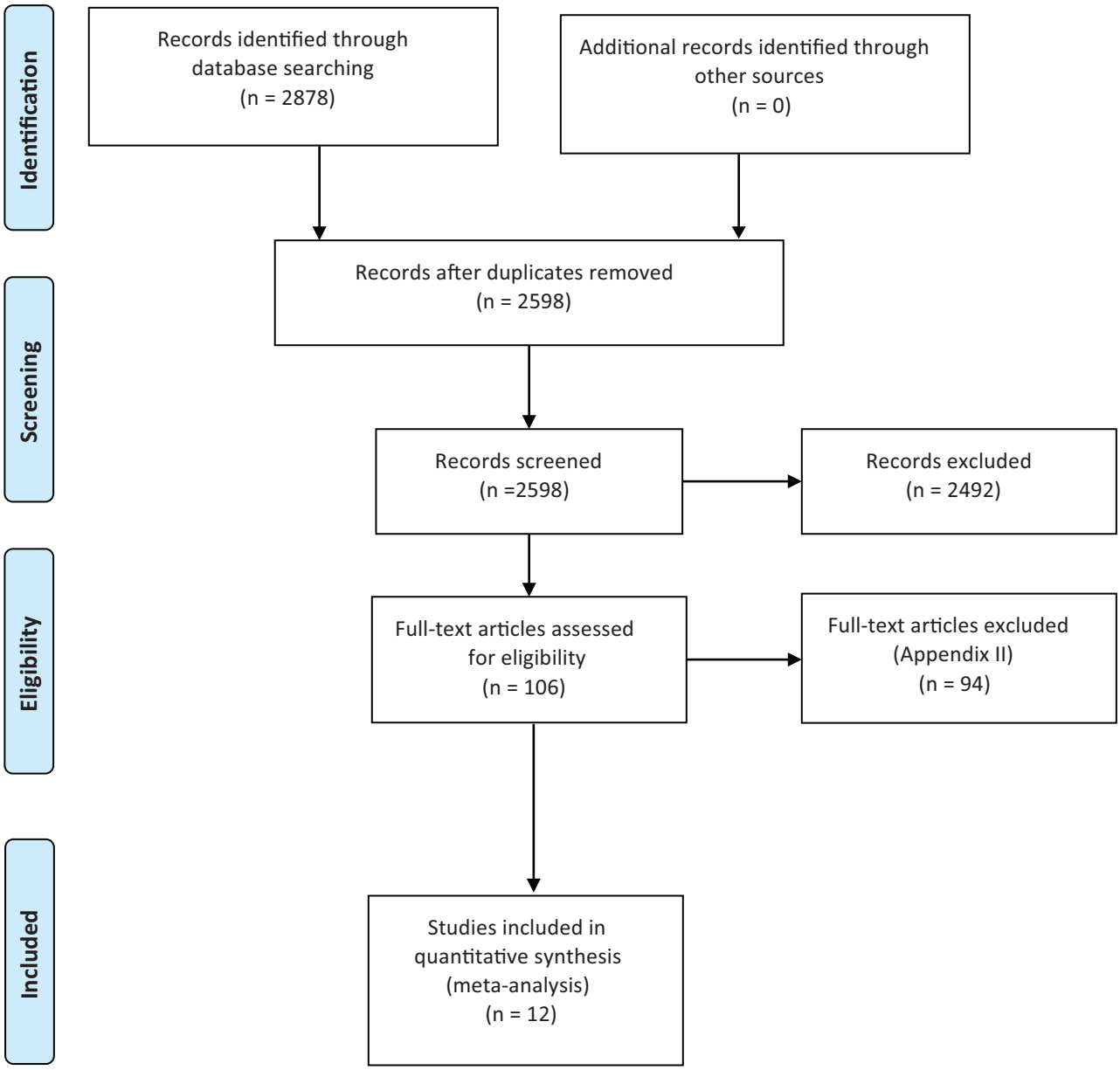


Figure 1: Search results and study selection and inclusion process[18]

Participants and setting

The included studies were performed in the period from 1994 to 2018. Four studies were conducted in the United States,[19,22-24] five in Asia,[20,21,26-28] one in Europe,[25] and one study was conducted in the Middle East.[29] One study was conducted in separate centers in Europe and Asia.[30] The number of participants enrolled in the studies ranged from 50 to 2488. Overall, a total of 8462 patients were considered in this review, with ages ranging from 35 to 90 years. However, three studies exclusively enrolled participants 45 years or older.[23,28,29] One study enrolled more than one set of participants.[22] The first set was a primary cohort used to determine the PHI cut-off value with the greatest specificity, while the second/validation set of participants was used to validate the PHI assay used in the primary cohort.

Table 1: Critical appraisal of included studies

Study	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Total (%) Yes
Al Saidi *et al.* 2017[29]	Y	Y	Y	U	Y	Y	U	Y	Y	Y	80%
Chiu *et al.* 2016[21]	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	100%
Chiu *et al.* 2018[30]	Y	Y	Y	U	Y	Y	U	Y	Y	Y	80%
De la Calle *et al.* 2015[22]	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	100%
Furuya *et al.* 2017[27]	Y	Y	Y	U	N	Y	U	Y	Y	Y	70%
Hsieh *et al.* 2018[28]	Y	Y	Y	Y	Y	Y	U	Y	Y	Y	90%
Loeb *et al.* 2015[23]	Y	N	Y	Y	N	Y	Y	Y	Y	Y	80%
Loeb *et al.* 2013[19]	Y	N	Y	Y	Y	Y	Y	U	Y	Y	80%
Loeb *et al.* 2017[24]	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	100%
Na *et al.* 2017[20]	Y	Y	Y	U	Y	Y	U	U	Y	Y	70%
Seisen *et al.* 2015[25]	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	100%
Tan *et al.* 2017[26]	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	100%

Y, yes; U, unclear; N, no; JBI critical appraisal checklist for diagnostic test accuracy studies
Q1. Was a consecutive or random sample of patients enrolled?
Q2. Was a case-control design avoided?
Q3. Did the study avoid inappropriate exclusions?
Q4. Were the index test results interpreted without the knowledge of the results of the reference standard?
Q5. If a threshold was used, was it pre-specified?
Q6. Is the reference standard likely to correctly classify the target condition?
Q7. Were the reference standard results interpreted without knowledge of the results in the index test?
Q8. Was there an appropriate interval between the index and reference standard?
Q9. Did all patients receive the same reference standard?
Q10. Were all patients included in the analysis?

Table 2: Characteristics of included studies

Author, setting	Study design	Patient characteristics	Results
Al Saidi, Oman[29]	Prospective cohort study	136 men, median age 64.5 years, TPSA >4ng/mL and in whom 12 core biopsies were performed. No mention of DRE findings. PHI was the index test and Gleason score was the reference standard.	Positive biopsy rate was 20.6% (28/136); 12.5% (17/136) had aggressive disease. At a PHI threshold value of 41.9, sensitivity was 82.1%, specificity was 80.6%, and area under the curve was 0.81 for detection of PCa.
Chiu, Hong Kong[21]	Prospective cohort study	569 men, median age of 66 years, TPSA 4-10ng/mL and non-suspicious findings on DRE. 10-core biopsy technique performed. PHI was the index test and Gleason score the reference standard.	Positive biopsy rate was 10.9% (62/569); 2.8% (16/569) had aggressive disease. At PHI threshold value of 35, the risk of PCa was 25% while the risk of aggressive disease was 8.6%.
Chiu, various sites in Europe and Asia[30]	Prospective multicenter study	1149 Asian and 503 European men with normal DRE and TPSA ranges of 2-20ng/mL. PHI was the index test and Gleason score was the reference standard.	Positive biopsy rate in the European population was 52% (262/503); 23% (115/503) had aggressive disease. Positive biopsy rate was 13% (151/1149) in the Asian population; 5.7% (66/1149) had aggressive disease.

Table 2: (*Continued*)

Author, setting	Study design	Patient characteristics	Results
De la Calle, USA[22]	Prospective cohort study	561 men (primary cohort), mean age 62.1 years, TPSA range and number of biopsies not mentioned. 395 men (validation cohort), mean age 62.8 years, TPSA range and number of biopsies not mentioned. PHI was the index test and Gleason score was the reference standard.	Positive biopsy rate 58.5% (328/561); 20.3% (114/561) had aggressive cancer. PHI threshold values ranged from 24-34.3; sensitivity ranged from 80-95.6%; specificity ranged from 34.9-64.8%. Aggressive cancer detected in 30.9% (122/395). At PHI threshold value of 24, sensitivity was 95.6% and specificity of 34.9%. In combined cohort, at a PHI threshold value of 22.9, the sensitivity was 95% and specificity was 30%.
Furuya, Japan[27]	Prospective cohort study	50 men, median age 68.5 years with TPSA 2-10ng/mL. 39/50 (78%) had negative DRE findings. 16- to 21-core biopsy was done after prostate magnetic resonance imaging. PHI was the index test and Gleason score was the reference standard.	Positive biopsy rate was 66% (33/50); 42% (21/50) had aggressive disease. At PHI threshold value $>$38.7 for PCa, sensitivity was 63.6%; specificity was 76.5%; PPV was 84%; and NPV was 52%.
Hsieh, Taiwan[28]	Prospective observational study	154 men with TPSA 4-10ng/mL with/without normal DRE findings and in whom 12-core biopsy was done. Mean age of 65.5 years. PHI was the index test and Gleason score was the reference standard.	Positive biopsy rate was 23.4% (36/154); 16.9% (26/154) had aggressive disease. The probability of prostate cancer at PHI threshold values of 0-26.9, 27-35.9, 36-54.9, and $\geq$55 was 10.26%, 20%, 43.75%, and 77.78%, respectively
Loeb, USA[23]	Observational study	658 men, median age 63 years, TPSA 2-10ng/mL, 97.8% had $>$10 core biopsies. Both PHI and p2PSA were assessed as index tests, and Gleason score was the reference standard.	Positive biopsy rate was 49.2% (324/658); 16.6% (109/658) had aggressive disease. PHI threshold values ranged from 28.1-31.9; sensitivity ranged from 80-95%; specificity ranged from 27.4-46.4%. p2PSA threshold values ranged from 7.9-10.9; sensitivity ranged from 80-95%; specificity ranged from 9.9-27.9%.

Table 2: (*Continued*)

Author, setting	Study design	Patient characteristics	Results
Loeb, USA[19]	Prospective multicenter study	892 men, 50 years or older with TPSA range 2-10ng/mL and benign findings on DRE. At least a 6-core biopsy was undertaken. PHI was the index test and Gleason score was the reference standard.	Positive biopsy rate was 48.2% (430/892); 15.6% (139/892) had aggressive disease. At PHI threshold value of 0-28.9, 29-39.9, 40-61.9, and $\geq$62, the risk of PCa was 11.3%, 18%, 34%, and 49.6%, respectively, while the risk of aggressive disease was 11%, 30%, 48%, and 50%, respectively.
Loeb, USA[24]	Prospective observational study	728 men, median age 62.8 years, TPSA 2-10 ng/mL, all had 6-core biopsies. PHI was the index test and Gleason score was the reference standard.	16.2% (118/728) had aggressive cancer. PHI threshold values ranged from 15-34; sensitivity ranged from 73.7-99.7%; specificity ranged from 15-55.9%; PPV ranged from 2-7.3%; and NPV ranged from 95.7-98.1%.
Na, China[20]	Prospective multicenter cohort study	1538 men, mean age 66.95 years with no TPSA restriction. PHI was the index test and Gleason score was the reference standard.	Positive biopsy rate was 40.2% (618/1538); 31.7% (488/1538) had aggressive disease. PHI threshold value ranged from 18-35. Sensitivity for detecting PCa ranged from 92-99.5%, while specificity ranged from 15.1-59.8%. Sensitivity for detecting aggressive disease ranged from 96.3-99.8%, while specificity ranged from 13.4-55.3%.
Seisen, France[25]	Prospective observational cohort study	138 men, median age 63.4 years, TPSA 4-20ng/mL, no mention of core biopsies. PHI was the index test and Gleason score was the reference standard.	Positive biopsy of 44.9% (62/138); aggressive cancer 28.3% (39/138). PHI threshold value $>$40, with sensitivity of 66.7%, specificity of 73.7%, NPV of 84.9%, and PPV of 50%.
Tan, Singapore[26]	Observational prospective study	157 men, median age 65 years, TPSA 4-10ng/mL, no report on biopsies. PHI was the index test and Gleason score was the reference standard.	Positive biopsy rate 19.1% (30/157); 12.1% (19/157) had aggressive cancer. At PHI threshold value of 26.75, sensitivity was 90% and specificity 55.1%.

DRE, digital rectal examination; NPV, negative predictive value; p2PSA, [-2]proPSA; PCa, prostate cancer; PHI, Prostate Health Index; PPV, positive predictive value; PSA, prostate specific antigen; TPSA, total prostate specific antigen.

Eleven studies considered tPSA values as part of the inclusion criteria.[19-21,23-30] In four of these studies, tPSA in the range 2 to 10ng/mL was considered,[19,23,24,27] while in three studies tPSA in the range 4 to 10ng/mL was considered.[21,26,28] Two studies had an expanded tPSA range of up to 20ng/mL: 2 to 20ng/mL for the study by Chiu *et al.*[30] and 4 to 20 ng/mL for the study by Seisen *et al.*[25] The remaining two studies specified no limits of tPSA.[20,29] Exclusion criteria were mentioned in 10 studies.[19-22,24-29] They included a history of acute bacterial prostatitis, initial prostate biopsy, previous

prostate surgery, and therapy with 5-alpha reductase inhibitors (5-ARI). Two studies did not explicitly state the exclusion criteria used.[23,30]

Index and reference tests

All 12 studies that met the inclusion criteria examined PHI as an index test. They all utilized the Beckman Coulter immunoassay analyzer. One study, however, recalculated the index test using World Health Organization calibration values.[19] One study explicitly mentioned that the PHI immunoassay was run by technologists.[22] Two studies assessed additional index tests (ie, PCA[25] and magnetic resonance imaging[27]). Among the studies that assessed PHI,[19-25,27-30] mean values were higher in men diagnosed with aggressive PCa compared to those who had indolent or no disease. Only one study assessed p2PSA. The Gleason score was the reference standard in all included studies as determined from prostate biopsy specimens taken before initiation of therapy. In nine studies, prostate biopsies were performed by transrectal ultrasound guidance,[20-22,25-30] while in three studies, there was no mention of how prostate biopsies were performed.[19,23,24] Eleven studies reported the number of core biopsies taken,[19-21,23-30] and they ranged from six to 21. This increases the accuracy of detecting PCa in biopsy samples. One study did not specify the number of core biopsies taken.[22] A pathologist was involved in determining the Gleason score in seven studies.[20,22,25-28,30] However, only four studies explicitly stated that a uropathologist was involved in determining Gleason scores.[21,25,26,28]

Diagnosis of interest

All included studies reported data on aggressive PCa as determined by the Gleason score. However, in addition to the Gleason score, one study also utilized the Epstein criteria to define significant/aggressive PCa.[23] Detection rates for aggressive PCa (using Gleason score) ranged from 2.8%[21] to 42.8%[20] of all patients enrolled in the included studies. Only one study reported exclusively on aggressive PCa.[24] The remaining 11 studies included data on both overall and aggressive PCa detection rates.

Four studies attempted to determine clinical and biochemical factors associated with aggressive PCa.[22,23,25,26] Seisen *et al.* found that except for PHI, PCA-3, and PSA density, there were no significant clinical or biochemical differences between patients with or without significant PCa.[25] Both Tan and De la Calle reported that PHI was associated with a significant chance of detecting aggressive PCa.[22,26] In another study, Loeb and colleagues showed that a history of prostate biopsy as well as a larger prostate volume were significantly associated with a lower risk of overall and aggressive PCa.[23] Their study also pointed to PHI being a significant predictor of aggressive PCa regardless of whether aggressiveness was determined by Gleason score or Epstein criteria.

Review findings

The findings from the 12 included studies are presented below. They are stratified according to index test.

Accuracy of PHI for determining aggressive prostate cancer

The overall sensitivity of PHI for determining aggressive PCa ranged from 65% to 100% while specificity ranged from 1% to 81%. Generally, sensitivity decreased with increasing PHI values while specificity increased with increasing PHI values. Figure 2 presents forest plots depicting sensitivity and specificity from each included study stratified according to PHI threshold level. At a PHI value of 25 and below, overall sensitivity was 97% (95% confidence interval [CI], 95% to 98%); overall, specificity was 10% (95% CI, 6% to 16%). For PHI from 26 to 35, overall sensitivity was 87% (95% CI, 81% to 91%) and overall specificity was 45% (95% CI, 39% to 50%). For PHI of 36 and above, overall sensitivity was 72% (95% CI, 64% to 79%) and specificity was 74% (95% CI, 68% to 80%). For each of the three categories of PHI, sensitivity analysis by omitting studies in which blinding was unclear revealed similar results. The SROC curves with 95% confidence and prediction regions are presented in Figures 3–5.

Accuracy of p2PSA for determining aggressive prostate cancer

Only one study[23] examined the accuracy of p2PSA for determining the aggressiveness of PCa (Table 3). In this study, sensitivity ranged from 80% to 95% while specificity ranged from 9.9% to 27.9%. At a cut-off value of 10.9 for p2PSA, sensitivity was 80% while specificity was 27.9%. At a cut-off value of 10.2, sensitivity was 85% while specificity was 23%. At a cut-off value of 8.6, sensitivity was 90% while

Prostatic Health Index with threshold <25

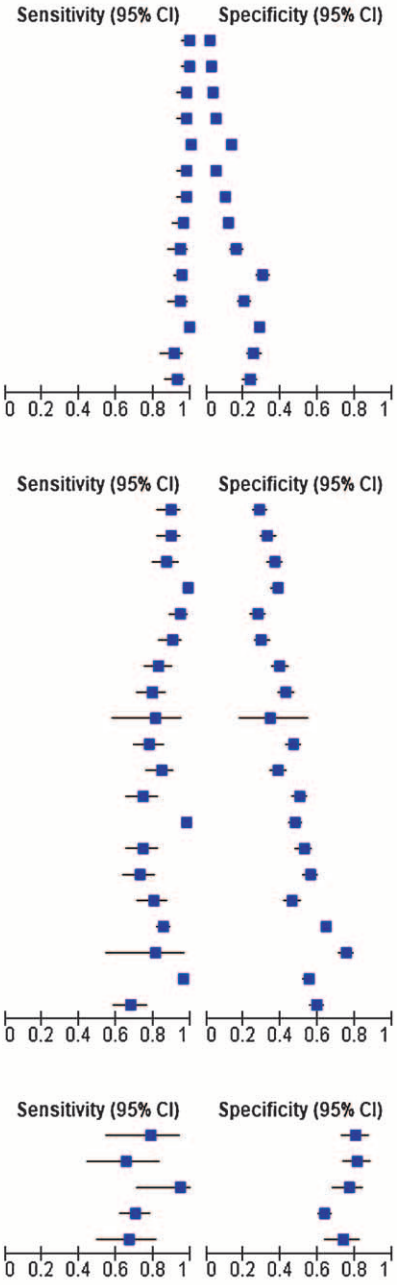
Study	TP	FP	FN	TN	Threshold	Sensitivity (95% CI)	Specificity (95% CI)
Loeb 2017(a)	117	601	1	9	15.0	0.99 [0.95, 1.00]	0.01 [0.01, 0.03]
Loeb 2017(b)	117	595	1	15	16.0	0.99 [0.95, 1.00]	0.02 [0.01, 0.04]
Loeb 2017(c)	115	589	3	21	17.0	0.97 [0.93, 0.99]	0.03 [0.02, 0.05]
Loeb 2017(d)	115	576	3	34	18.0	0.97 [0.93, 0.99]	0.06 [0.04, 0.08]
Na 2017(a)	487	909	1	141	18.0	1.00 [0.99, 1.00]	0.13 [0.11, 0.16]
Loeb 2017(e)	115	576	3	34	19.0	0.97 [0.93, 0.99]	0.06 [0.04, 0.08]
Loeb 2017(f)	115	548	3	62	20.0	0.97 [0.93, 0.99]	0.10 [0.08, 0.13]
Loeb 2017(g)	113	539	5	71	21.0	0.96 [0.90, 0.99]	0.12 [0.09, 0.14]
Loeb 2017(h)	111	512	7	98	22.0	0.94 [0.88, 0.98]	0.16 [0.13, 0.19]
De la Calle 2015	224	504	12	216	22.9	0.95 [0.91, 0.97]	0.30 [0.27, 0.33]
Loeb 2017(i)	111	488	7	122	23.0	0.94 [0.88, 0.98]	0.20 [0.17, 0.23]
Na 2017(b)	483	749	5	301	24.0	0.99 [0.98, 1.00]	0.29 [0.26, 0.32]
Loeb 2017(k)	107	454	11	156	24.0	0.91 [0.84, 0.95]	0.26 [0.22, 0.29]
Loeb 2017(j)	109	468	9	142	24.0	0.92 [0.86, 0.96]	0.23 [0.20, 0.27]

Prostatic Health Index with threshold 26-35

Study	TP	FP	FN	TN	Threshold	Sensitivity (95% CI)	Specificity (95% CI)
Loeb 2017(l)	105	435	13	175	26.0	0.89 [0.82, 0.94]	0.29 [0.25, 0.32]
Loeb 2017(m)	105	409	13	201	27.0	0.89 [0.82, 0.94]	0.33 [0.29, 0.37]
Loeb 2017(n)	103	387	15	223	28.0	0.87 [0.80, 0.93]	0.37 [0.33, 0.41]
Na 2017(c)	481	648	7	402	28.0	0.99 [0.97, 0.99]	0.38 [0.35, 0.41]
Loeb 2015(a)	103	398	6	150	28.1	0.94 [0.88, 0.98]	0.27 [0.24, 0.31]
Loeb 2015(b)	98	385	11	163	28.9	0.90 [0.83, 0.95]	0.30 [0.26, 0.34]
Loeb 2017(o)	98	368	20	242	29.0	0.83 [0.75, 0.89]	0.40 [0.36, 0.44]
Loeb 2017(p)	94	348	24	262	30.0	0.80 [0.71, 0.87]	0.43 [0.39, 0.47]
Furuya 2017	17	19	4	10	30.7	0.81 [0.58, 0.95]	0.34 [0.18, 0.54]
Loeb 2017(q)	92	326	26	284	31.0	0.78 [0.69, 0.85]	0.47 [0.43, 0.51]
Loeb 2015(c)	92	335	17	213	31.9	0.84 [0.76, 0.91]	0.39 [0.35, 0.43]
Loeb 2017(r)	88	305	30	305	32.0	0.75 [0.66, 0.82]	0.50 [0.46, 0.54]
Na 2017(d)	475	548	13	502	32.0	0.97 [0.95, 0.99]	0.48 [0.45, 0.51]
Loeb 2017(s)	88	291	30	319	33.0	0.75 [0.66, 0.82]	0.52 [0.48, 0.56]
Loeb 2017(t)	86	270	32	340	34.0	0.73 [0.64, 0.81]	0.56 [0.52, 0.60]
Loeb 2015(d)	87	294	22	254	34.4	0.80 [0.71, 0.87]	0.46 [0.42, 0.51]
Chiu 2018	381	731	64	1312	35.0	0.86 [0.82, 0.89]	0.64 [0.62, 0.66]
Chiu 2016	13	139	3	414	35.0	0.81 [0.54, 0.96]	0.75 [0.71, 0.78]
Na 2017(e)	469	470	19	580	35.0	0.96 [0.94, 0.98]	0.55 [0.52, 0.58]
Loeb 2017(u)	80	249	38	361	35.0	0.68 [0.59, 0.76]	0.59 [0.55, 0.63]

Prostatic Health Index with threshold 36+

Study	TP	FP	FN	TN	Threshold	Sensitivity (95% CI)	Specificity (95% CI)
Tan 2017	15	27	4	111	36.0	0.79 [0.54, 0.94]	0.80 [0.73, 0.87]
Hsieh 2018	17	24	9	104	36.0	0.65 [0.44, 0.83]	0.81 [0.73, 0.88]
Al Saidi 2017	16	28	1	91	40.0	0.94 [0.71, 1.00]	0.76 [0.68, 0.84]
Loeb 2013	98	274	41	478	40.0	0.71 [0.62, 0.78]	0.64 [0.60, 0.67]
Seisen 2015	26	26	13	73	40.0	0.67 [0.50, 0.81]	0.74 [0.64, 0.82]



CI, confidence interval; TP, true positive; FP, false positive; FN, false negative; TN, true negative

Figure 2: Forest plot of sensitivity and specificity of Prostate Health Index at various threshold values. Letters after the study author's name are used when a single study had multiple threshold values assessed.

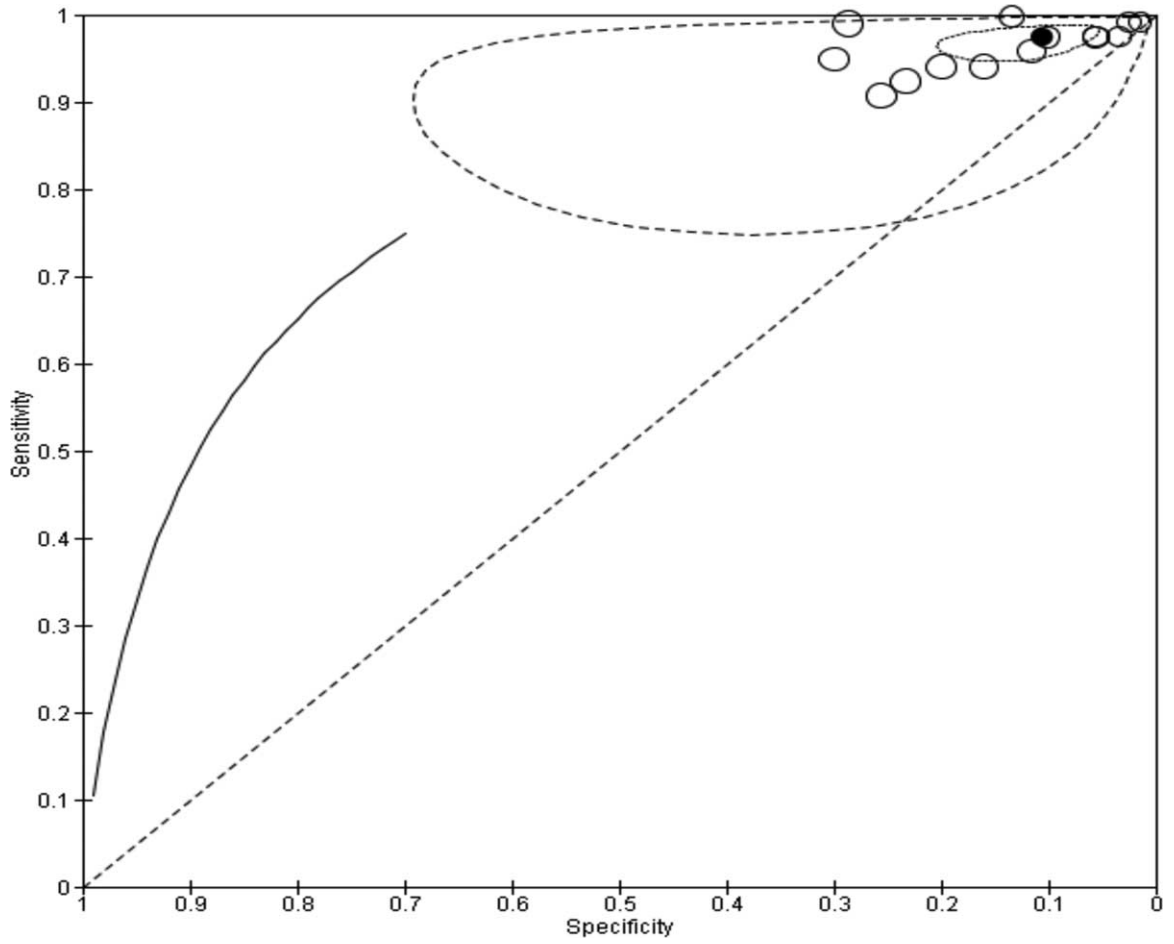


Figure 3: Summary receiver operator characteristics curve (solid curved line), summary point (shaded spot) with 95% confidence and prediction regions (inner and outer dotted lines) for Prostate Health Index cut-off of 25 and below. The unshaded spots represent individual studies included in the analysis.

specificity was 13.1%. At a cut-off value of 7.9, sensitivity was 95% while specificity was 9.9%.

Discussion

This review aimed to synthesize the best available evidence on the accuracy of p2PSA and PHI in detecting the aggressive form of PCa. While there have been previous attempts at determining the accuracy of these two tests for determining aggressive PCa, we endeavored to include the most recent relevant studies while utilizing a more rigorous systematic review methodology than previously reported. Overall, we found that test sensitivity was higher with lower index test values while specificity was higher with higher index test values (threshold effect). In the context of PCa, better specificity translates to a reduction in unnecessary biopsies and possibly unwarranted therapy. However, higher specificity is also accompanied by lower sensitivity meaning that the likelihood of missing an aggressive cancer is equally increased. As such, while both p2PSA and PHI are known to be better than PSA alone, they are not in themselves perfect tests and the dilemma surrounding detection of aggressive PCa remains.

These results are comparable to those of similar but older reviews. In the review by Russo *et al.*,[14] seven studies evaluated the accuracy of PHI for aggressive PCa. Although there was no explicit definition of aggressive PCa, the other inclusion criteria were similar to those of the present review. Further,

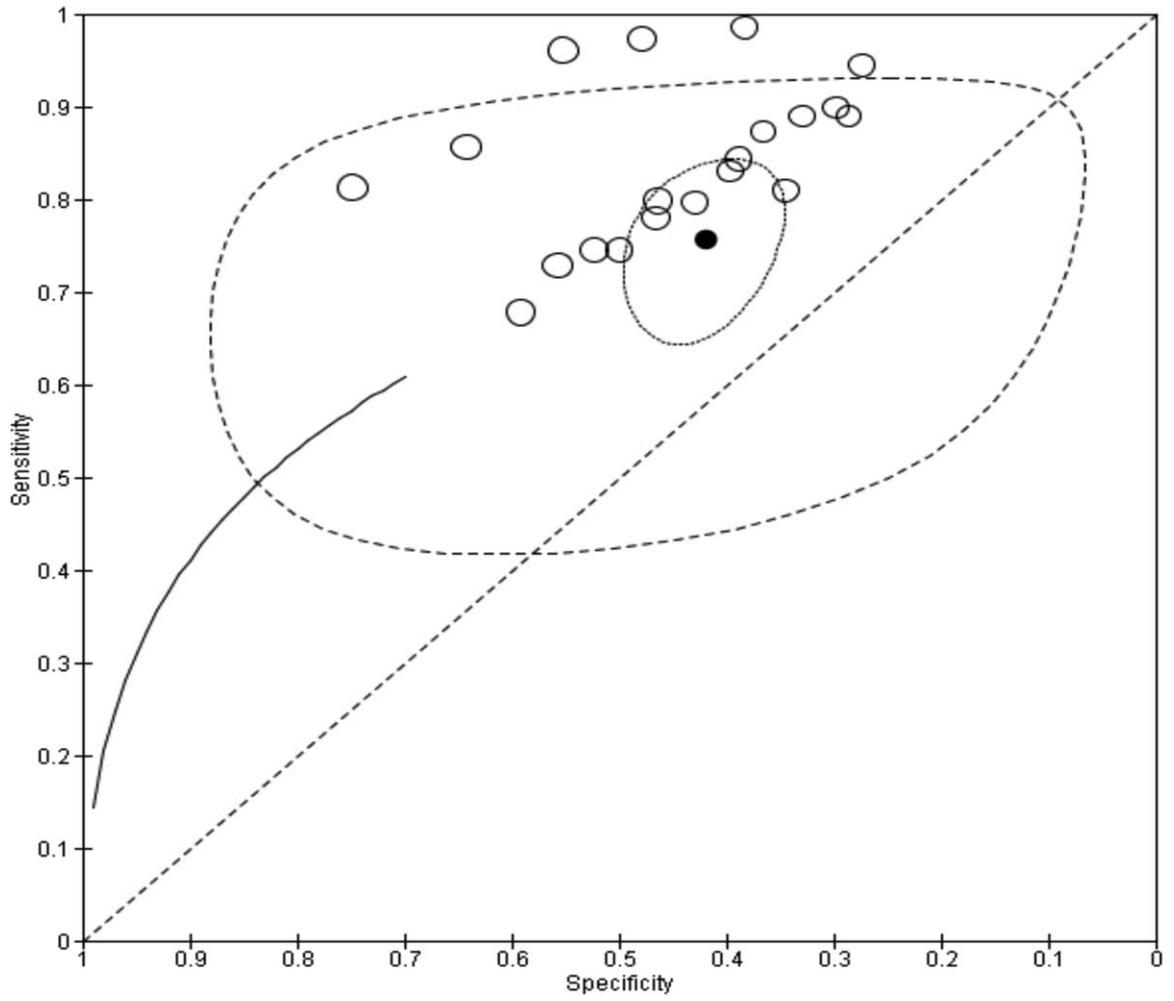


Figure 4: Summary receiver operator characteristics curve (solid curved line), summary point (shaded spot) with 95% confidence and prediction regions (inner and outer dotted lines) for Prostate Health Index cut-off of 26 to 35. The unshaded spots represent individual studies included in the analysis.

they also identified a large variability in threshold values utilized for positivity. They reported an overall sensitivity of 93% and specificity of 34%. Wang *et al.* reported on the outcome of meta-analysis of five studies assessing the utility of PHI for detecting aggressive PCa. Overall sensitivity was 90% while specificity was 17%. In our review, we managed to include more recent primary studies and opted to manage the varying thresholds present in included studies by stratifying the reported cut-off values into three subcategories (Figure 2). The main reasoning behind this decision was that in a number of studies, different cut-off values were investigated, resulting in an array of positivity thresholds. Subcategorizing afforded the ability to better demonstrate the inverse relationship between sensitivity and specificity, attempt to reduce heterogeneity, and perform a less biased meta-analysis.

The majority of included studies in this review recruited men with tPSA levels between 2 and 20ng/mL (ie, the diagnostic gray area). This is similar to the other reviews on this topic. This population is particularly troublesome because this is the range in which tPSA has poorest specificity, especially when combined with an unremarkable rectal examination. It is already known that both PHI and p2PSA are better than PSA alone at detecting PCa. However, high-risk PCa accounts for less than 15% of newly

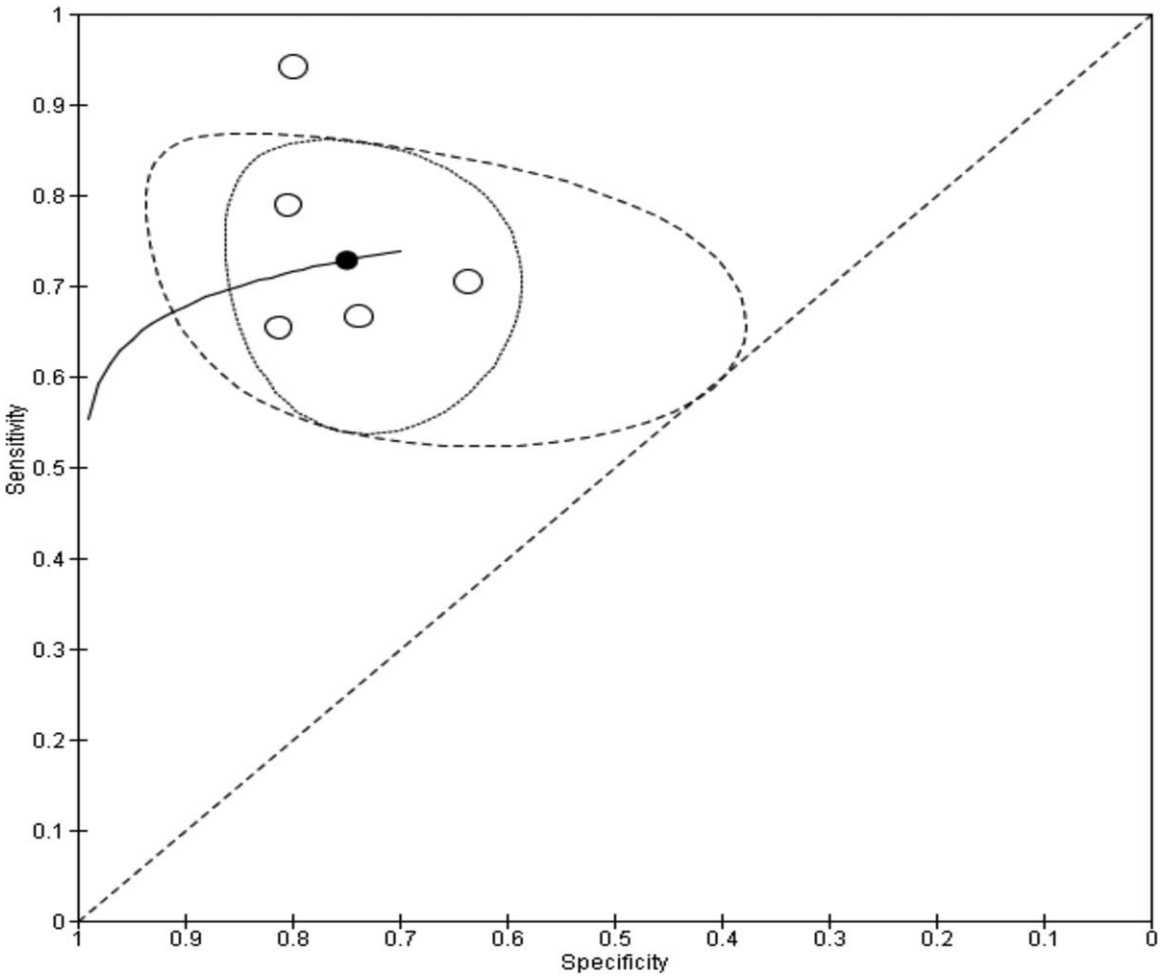


Figure 5: Summary receiver operator characteristics curve (solid curved line), summary point (shaded spot) with 95% confidence and prediction regions (inner and outer dotted lines) for Prostate Health Index cut-off of 36 and above. The unshaded spots represent individual studies included in the analysis.

diagnosed cases, meaning the vast majority of cancers of the prostate are indolent and best served with minimal intervention. As such, and as stated by Russo and colleagues,[14] a useful biomarker for PCa is one that is sufficiently capable of identifying the aggressive form of disease. Indeed, this has been the main push behind the ongoing efforts to identify new biomarkers for prostatic cancer.

Table 3: Accuracy of [-2]proPSA for detecting aggressive prostate cancer at different threshold values

Study	Threshold	Sensitivity	Specificity
Loeb/USA[23]	7.9	0.95	0.099
	8.6	0.9	0.131
	10.2	0.85	0.23
	10.9	0.8	0.279

Based on the findings of this review and the current recommendations for clinical practice, we contend that the sole use of these newer biomarkers may be inappropriate because they are only marginally better than PSA. Rather, they may be more useful when employed using an algorithmic approach to patient management. At present, digital rectal examination and PSA levels are employed in the detection of PCa. An abnormal rectal examination irrespective of PSA levels warrants biopsy.

Similarly, highly elevated PSA levels warrant biopsy and are more likely to signal advanced disease. However, for patients with PSA levels within the aforementioned diagnostic gray area and normal rectal examination, there may be a role for the newer biomarkers. Stattin and colleagues[124] in Sweden showed that 4-Kallikrein was able to discriminate men with PSA levels in the gray area who need further testing due to high risk of PCa progression. Various studies have shown that both 4-Kallikrein and PHI are equally accurate in detecting aggressive PCa. Thus, for men with slight to modest elevations in PSA levels and who are unlikely to have aggressive PCa based on the newer biomarkers, a less aggressive approach to therapy may be employed because the risk of disease progression in this group of men is low. However, such an approach warrants more research.

We found only one study that attempted to determine the utility of p2PSA for the determination of aggressive PCa. The most probable reason for this is that p2PSA is one parameter needed for the determination of PHI and, therefore, researchers in this field may find it superfluous to study a marker they are already researching, albeit indirectly. Nonetheless, from the available data, p2PSA seems to have acceptable accuracy in detecting aggressive PCa, although it also displays an inverse relationship between sensitivity and specificity. At present, it is not prudent to compare it directly to PHI because the available data are from one study that simultaneously studied PHI.

The overall risk of bias in this review was low. All included studies scored 70% and above on the JBI critical appraisal checklist for diagnostic test accuracy. Blinding was unclear among a selection of included studies, and this formed the basis for sensitivity analysis. However, there was not much difference in summary measures between studies that performed blinding and those in which it was unclear (results not shown). Therefore, we contend that the impact of unclear blinding among some of the included studies is likely to be small and, thus, the impact on the results is equally likely to be small. In addition to blinding, the studies included in this review had differences in terms of the population characteristics, threshold values, and number of core biopsies taken. These factors may affect the accuracy of both the index and reference tests, and ultimately affect the internal validity of our findings. This review also only included studies published in English thereby introducing the risk of publication bias and affecting external validity. Nonetheless, our extensive search did not reveal any relevant studies in other languages that would have forced exclusion based on language. As a strength, we included more recent studies in this review and attempted to consider the effect of variable thresholds of PHI in the overall accuracy of this biomarker.

Conclusion

This review shows that PHI and p2PSA are potential biomarkers for the detection of aggressive PCa. However, the inverse relationship between sensitivity and specificity makes it difficult to determine an optimum cut-off value for positivity. Nonetheless, further research is warranted to determine their utility in the management of PCa.

Recommendations for practice

While the available evidence shows that both p2PSA and PHI may have acceptable accuracy for the detection of aggressive PCa, this observation is based on low-quality evidence (GRADE ranking). However, a weak recommendation for their use may be made, and they should be employed in conjunction with other clinical parameters such as rectal examination and imaging findings to help guide clinical decision-making.

Recommendations for research

There is wide variability in the cut-off values utilized in determining the positivity of p2PSA and PHI for aggressive PCa. As such, further research is warranted to determine the optimum cut-off value that can be applied in clinical practice as well as determine the true role of these tests in the management of aggressive PCa. In particular, future research should focus on the utility of these biomarkers in aiding determination of therapy (surgery versus radiotherapy versus hormonal therapy) and post-therapy follow-up.

References

1. Al-Khalil S, Ibilibor C, Cammack JT, de Riese W. Association of prostate volume with incidence and aggressiveness of prostate cancer. Res Rep Urol 2016;8:201–5.
2. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2015. CA Cancer J Clin 2015;65(1):5–29.

3. Ferlay J, Soerjomataram I, Dikshit R, Eser S, Mathers C, Rebelo M, *et al*. Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. Int J Cancer 2015;136(5):E359–86.
4. Hassanipour-Azgomi S, Mohammadian-Hafshejani A, Ghoncheh M, Towhidi F, Jamehshorani S, Salehiniya H. Incidence and mortality of prostate cancer and their relationship with the Human Development Index worldwide. Prostate Int 2016;4(3):118–24.
5. Bhardwaj A, Srivastava SK, Khan MA, Prajapati VK, Singh S, Carter JE, *et al*. Racial disparities in prostate cancer: a molecular perspective. Front Biosci 2017;22:772–82.
6. Placer J, Planas J, Celma A, Morote J. [Current role of protatic specific antigen (PSA) and its by-products in the diagnosis of prostate cancer]. Arch Esp Urol 2015;68(3): 210–28; Spanish.
7. Dasgupta A, Wahed A. Clinical chemistry, immunology and laboratory quality control. A comprehensive review for board preparation, certification and clinical practice. San Diego: Elsevier Inc; 2014.
8. Hori S, Blanchet JS, McLoughlin J. From prostate-specific antigen (PSA) to precursor PSA (proPSA) isoforms: a review of the emerging role of proPSAs in the detection and management of early prostate cancer. BJU Int 2013;112(6): 717–28.
9. Boegemann M, Stephan C, Cammann H, Vincendeau S, Houlgatte A, Jung K, *et al*. The percentage of prostate-specific antigen (PSA) isoform [-2]proPSA and the Prostate Health Index improve the diagnostic accuracy for clinically relevant prostate cancer at initial and repeat biopsy compared with total PSA and percentage free PSA in men aged </=65 years. BJU Int 2016;117(1):72–9.
10. Mearini L, Ferri C, Lazzeri M, Bini V, Nunzi E, Fiorini D, *et al*. Evaluation of prostate-specific antigen isoform p2PSA and its derivates, %p2PSA, prostate health index and prostate dimension-adjusted related index in the detection of prostate cancer at first biopsy: an exploratory, prospective study. Urol Int 2014;93(2):135–45.
11. Ruzhanskaya AV, Evgina SA, Skibo II. [The practical application of marker -2proPSA and health index of prostate phi in diagnostics of prostate cancer]. Klin Lab Diagn (1): 2014:4–8; Russian.
12. Wang W, Wang M, Wang L, Adams TS, Tian Y, Xu J. Diagnostic ability of %p2PSA and prostate health index for aggressive prostate cancer: a meta-analysis. Sci Rep 2014;4:5012.
13. Ham WS, Chalfin HJ, Feng Z, Trock BJ, Epstein JI, Cheung C, *et al*. The impact of downgrading from biopsy Gleason 7 to prostatectomy Gleason 6 on biochemical recurrence and prostate cancer-specific mortality. J Urol 2017;197(4): 1060–7.
14. Russo GI, Regis F, Castelli T, Favilla V, Privitera S, Giardina R, *et al*. A Systematic review and meta-analysis of the diagnostic accuracy of Prostate Health Index and 4-Kallikrein Panel Score in predicting overall and high-grade prostate cancer. Clin Genitourin Cancer 2017;15(4):429–39; e1.
15. Zappala SM, Dong Y, Re: Russo GI, Regis F, Castelli T, *et al*. A systematic review and meta-analysis of the diagnostic accuracy of Prostate Health Index and 4-kallikrein Panel Score in predicting overall and high-grade prostate cancer. Clin Genitourin Cancer 2017;15:429–39.
16. Anyango R, Ojwando J, Mwita C, Mugalo E. Diagnostic accuracy of the [-2]Pro-PSA and Prostate Health Index versus the Gleason score for determining the aggressiveness of prostate cancer: a systematic review protocol. JBI Database System Rev Implement Rep 2018;16(11): 2066–71.
17. Campbell JM, Kulgar M, Ding S, Carmody DP, Hakonsen SJ, Jadotte YT, *et al*. Chapter 9: Diagnostic test accuracy systematic reviews. In: Aromataris E, Munn Z, editors. JBI Reviewer's Manual [internet]. Adelaide: JBI, 2017. [cited 2015 Sep 13]. Available from: https://synthesismanual.jbi.global.
18. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med 2009;6(7):e1000097.
19. Loeb S, Sokoll LJ, Broyles DL, Bangma CH, van Schaik RHN, Klee GG, *et al*. Prospective multicenter evaluation of the Beckman Coulter Prostate Health Index using WHO calibration. J Urol 2013;189(5):1702–6.
20. Na R, Ye DW, Qi J, Liu F, Helfand BT, Brendler CB, *et al*. Prostate Health Index significantly reduced unnecessary prostate biopsies in patients with PSA 2-10 ng/mL and PSA > 10 ng/mL: results from a Multicenter Study in China. Prostate 2017;77(11):1221–9.
21. Chiu PK, Roobol MJ, Teoh JY, Lee WM, Yip SY, Hou SM, *et al*. Prostate health index (PHI) and prostate-specific antigen (PSA) predictive models for prostate cancer in the Chinese population and the role of digital rectal examination-estimated prostate volume. Int Urol Nephrol 2016;48(10): 1631–7.
22. de la Calle C, Patil D, Wei JT, Scherr DS, Sokoll L, Chan DW, *et al*. Multicenter evaluation of the Prostate Health Index to detect aggressive prostate cancer in biopsy naive men. J Urol 2015;194(1):65–72.
23. Loeb S, Sanda MG, Broyles DL, Shin SS, Bangma CH, Wei JT, *et al*. The Prostate Health Index selectively identifies clinically significant prostate cancer. J Urol 2015;193(4): 1163–9.
24. Loeb S, Shin SS, Broyles DL, Wei JT, Sanda M, Klee G, *et al*. Prostate Health Index improves multivariable risk prediction of aggressive prostate cancer. BJU Int 2017;120(1): 61–8.
25. Seisen T, Rouprêt M, Brault D, Léon P, Cancel-Tassin G, Compérat E, *et al*. Accuracy of the prostate health index versus the urinary prostate cancer antigen 3 score to predict overall and significant prostate cancer at initial biopsy. Prostate 2015;75(1):103–11.

26. Tan LG, Tan YK, Tai BC, Tan KM, Gauhar V, Tiong HY, *et al*. Prospective validation of %p2PSA and the Prostate Health Index, in prostate cancer detection in initial prostate biopsies of Asian men, with total PSA 4-10 ng ml-1. Asian J Androl 2017;19(3):286–90.
27. Furuya K, Kawahara T, Narahara M, Tokita T, Fukui S, Imano M, *et al*. Measurement of serum isoform [–2]proPSA derivatives shows superior accuracy to magnetic resonance imaging in the diagnosis of prostate cancer in patients with a total prostate-specific antigen level of 2-10 ng/ml. Scand J Urol 2017;51(4):251–7.
28. Hsieh PF, Chang CH, Yang CR, Huang CP, Chen WC, Yeh CC, *et al*. Prostate Health Index (PHI) improves prostate cancer detection at initial biopsy in Taiwanese men with PSA 4-10 ng/mL. Kaohsiung J Med Sci 2018;34(8):461–6.
29. Al Saidi SS, Al Riyami NB, Al Marhoon MS, Al Saraf MS, Al Busaidi SS, Bayoumi R, *et al*. Validity of Prostate Health Index and percentage of [-2] Pro-prostate-specific antigen as novel biomarkers in the diagnosis of prostate cancer: Omani Tertiary Hospitals Experience. Oman Med J 2017;32(4):275–83.
30. Chiu PK, Ng CF, Semjonow A, Zhu Y, Vincendeau S, Houlgatte A, *et al*. A multicentre evaluation of the role of the Prostate Health Index (PHI) in regions with differing prevalence of prostate cancer: adjustment of PHI reference ranges is needed for European and Asian settings. Eur Urol 2018;75(4). 558-.
31. Cantiello F, Russo GI, Ferro M, Cicione A, Cimino S, Favilla V, *et al*. Prognostic accuracy of Prostate Health Index and urinary prostate cancer antigen 3 in predicting pathologic features after radical prostatectomy. Urol Oncol 2015;33(4). 163.e15-.e23.
32. Chiu PK, Lai FM, Teoh JY, Lee WM, Yee CH, Chan ES, *et al*. Prostate Health Index and %p2PSA predict aggressive prostate cancer pathology in Chinese patients undergoing radical prostatectomy. Ann Surg Oncol 2016;23(8): 2707–14.
33. Dolejsova O, Fuchsova R, Topolcan O, Kucera R, Hora M, Svobodova H, *et al*. Can PHI better separate Gleason score 6 tumors and facilitate decision for right management for patients with prostate cancer? Tumor Biol 2017;39(12): 9–10.
34. Druskin SC, Tosoian JJ, Young A, Collica S, Srivastava A, Ghabili K, *et al*. Combining Prostate Health Index density, magnetic resonance imaging and prior negative biopsy status to improve the detection of clinically significant prostate cancer. BJU Int 2018;121(4):619–26.
35. Eminaga O, Bogemann M, Breil B, Titze U, Wotzel F, Eltze E, *et al*. Preoperative prostate-specific antigen isoform p2PSA </= 22.5 pg/ml predicts advanced prostate cancer in patients undergoing radical prostatectomy. Urol Oncol 2014;32(8):1317–26.
36. Ferro M, Bruzzese D, Perdona S, Mazzarella C, Marino A, Sorrentino A, *et al*. Predicting prostate biopsy outcome: prostate health index (phi) and prostate cancer antigen 3 (PCA3) are useful biomarkers. Clin Chim Acta 2012;413(15–16):1274–8.
37. Ferro M, Lucarelli G, Bruzzese D, Perdona S, Mazzarella C, Perruolo G, *et al*. Improving the prediction of pathologic outcomes in patients undergoing radical prostatectomy: the value of prostate cancer antigen 3 (PCA3), Prostate Health Index (PHI) and sarcosine. Anticancer Res 2015;35(2):1017–23.
38. Ferro M, Terracciano D, Mazzarella C, Marino A, Mariano A, Di Carlo A, *et al*. Prostate Health Index (PHI) is able to discriminate benign and precancerous conditions from prostate cancer. Anticancer Res 2011;31(5):1916–7.
39. Filella X, Foj L, Augé JM, Molina R. Evaluation of the P2PSA and prostate health index to improve prostate cancer detection. Tumor Biol 2014;35:S13.
40. Friedl A, Stangl K, Bauer W, Kivaranovic D, Schneeweiss J, Susani M, *et al*. Prostate-specific antigen parameters and Prostate Health Index enhance prostate cancer prediction with the In-bore 3-T magnetic resonance imaging-guided transrectal targeted prostate biopsy after negative 12-core biopsy. Urology 2017;110:148–53.
41. Le BV, Griffin CR, Loeb S, Carvalhal GF, Kan DH, Baumann NA, *et al*. [-2] proenzyme prostate specific antigen is more accurate than total and free prostate specific antigen in differentiating prostate cancer from benign disease in a prospective prostate cancer screening study. J Urol 2010;183(4):1355–9.
42. Na R, Ye DW, Liu F, Chen HT, Qi J, Wu YS, *et al*. Performance of serum Prostate-Specific Antigen Isoform -2 proPSA (p2PSA) and the Prostate Health Index (PHI) in a Chinese hospital-based biopsy population. Prostate 2014;74(15): 1569–75.
43. Osredkar J, Kumer K, Fabjan T, Hlebič G, Podnar B, Lenart G, *et al*. The performance of [-2]proPSA and prostate health index tumor markers in prostate cancer diagnosis. LaboratoriumsMedizin 2016;40(6):419–24.
44. Perdonà S, Bruzzese D, Ferro M, Autorino R, Marino A, Mazzarella C, *et al*. Prostate Health Index (PHI) and prostate cancer antigen 3 (PCA3) significantly improve diagnostic accuracy in patients undergoing prostate biopsy. Prostate 2013;73(3):227–35.
45. Sanchis-Bonet A, Barrionuevo-González M, Bajo-Chueca A, Morales-Palacios N, Sanchez-Chapado M. Does [-2]Pro-Prostate Specific Antigen Meet the Criteria to Justify Its Inclusion in the Clinical Decision-Making Process? Urol Int 2018;100(2):146–54.
46. Schwen ZR, Tosoian JJ, Sokoll LJ, Mangold L, Humphreys E, Schaeffer EM, *et al*. Prostate Health Index (PHI) Predicts High-stage Pathology in African American Men. Urology 2016;90:136–40.
47. Stephan C, Kahrs AM, Cammann H, Lein M, Schrader M, Deger S, *et al*. A [-2] pro PSA-based artificial neural network significantly improves differentiation between prostate

cancer and benign prostatic diseases. Prostate 2009;69(2): 198–207.
48. Tosoian JJ, Loeb S, Feng ZY, Isharwal S, Landis P, Elliot DJ, *et al*. Association of -2 proPSA with Biopsy Reclassification During Active Surveillance for Prostate Cancer. J Urol 2012;188(4):1131–6.
49. Lazzeri M, Haese A, de la Taille A, Palou Redorta J, McNicholas T, Lughezzani G, *et al*. Serum isoform [-2]proPSA derivatives significantly improve prediction of prostate cancer at initial biopsy in a total PSA range of 2-10 ng/ml: a multicentric European study. Eur Urol 2013;63(6):986–94.
50. Chiu PK, Teoh JY, Lee WM, Yee CH, Chan ES, Hou SM, *et al*. Extended use of Prostate Health Index and percentage of [-2]pro-prostate-specific antigen in Chinese men with prostate specific antigen 10-20 ng/mL and normal digital rectal examination. Investig Clin Urol 2016;57(5):336–42.
51. Boegemann M, Stephan C, Cammann H, Vincendeau S, Houlgatte A, Jung K, *et al*. The percentage of prostate-specific antigen (PSA) isoform -2 proPSA and the Prostate Health Index improve the diagnostic accuracy for clinically relevant prostate cancer at initial and repeat biopsy compared with total PSA and percentage free PSA in men aged $<=$ 65 years. Bju International 2016;117(1):72–9.
52. Guazzoni G, Nava L, Lazzeri M, Scattoni V, Lughezzani G, Maccagnano C, *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.
53. Fossati N, Lazzeri M, Haese A, McNicholas T. Clinical performance of serum isoform [-2]proPSA (p2PSA), and its derivatives %p2PSA and the Prostate Health Index, in men aged $<$60 years: results from a multicentric European study. BJU Int 2015;115(6):913–20.
54. Fujizuka Y, Ito K, Oki R, Suzuki R, Sekine Y, Koike H, *et al*. Predictive value of different prostate-specific antigen-based markers in men with baseline total prostate-specific antigen $<$2.0 ng/mL. Int J Urol 2017;24(8):602–9.
55. Stephan C, Jung K, Semjonow A, Schulze-Forster K, Cammann H, Hu X, *et al*. Comparative assessment of urinary prostate cancer antigen 3 and TMPRSS2:ERG gene fusion with the serum [-2]proprostate-specific antigen-based prostate health index for detection of prostate cancer. Clin Chem 2013;59(1):280–8.
56. Jansen FH, van Schaik RH, Kurstjens J, Horninger W, Klocker H, Bektic J, *et al*. Prostate-specific antigen (PSA) isoform p2PSA in combination with total PSA and free PSA improves diagnostic accuracy in prostate cancer detection. Eur Urol 2010;57(6):921–7.
57. Ng CF, Chiu PK, Lam NY, Lam HC, Lee KW, Hou SS. The Prostate Health Index in predicting initial prostate biopsy outcomes in Asian men with prostate-specific antigen levels of 4-10 ng/mL. Int Urol Nephrol 2014;46(4):711–7.
58. Nordström T, Vickers A, Assel M, Lilja H, Grönberg H, Eklund M. Comparison between the four-kallikrein panel and prostate health index for predicting prostate cancer. Eur Urol 2015;68(1):139–46.
59. Sriplakich S, Lojanapiwat B, Chongruksut W, Phuriyaphan S, Kitirattakarn P, Jun-Ou J, *et al*. Prospective performance of the Prostate Health Index in prostate cancer detection in the first prostate biopsy of men with a total prostatic specific antigen of 4-10 ng/mL and negative digital rectal examination. Prostate Int 2018;6(4):136–9.
60. Abrate A, Lazzeri M, Buffi N, Haese A, De La Taille A, McNicholas T, *et al*. Accuracy of p2PSA and derivatives (%p2PSA and PHI) in predicting prostate cancer in obese men from a multicenter European study. Eur Urol Suppl 2014;13(1):e341.
61. Barisiene M, Stanciute D, Bakavicius A, Jurkeviciene J, Zelvys A, Ulys A, *et al*. Diagnostic accuracy of [-2]proPSA, %p2PSA and Prostate Health Index for prostate cancer detection. Eur Urol Suppl 2018;17(5):e2184.
62. Basso DCf, Padoan A, Prayer-Galetti T, Secco S, Zattoni F, *et al*. PCA3 and Prostate Health Index (PHI) do not challenge F/TPSA in prostate cancer (PCA) diagnosis. Biochim Clin 2013;37:S157.
63. Bektic J, Darte C, Skradski V, Steiner E, Schaefer G, Bartsch G, *et al*. Access [-2] propsa and beckman coulterprostate health index (PHI) and early detection of aggressive prostate cancers. Eur Urol Suppl 2010;9(2):309.
64. Bektic J, Darte C, Steiner E, Stenzel B, Skradski V, Schaefer G, *et al*. [-2] proPSA is an early marker for prostate cancer aggressiveness. Eur Urol Suppl 2011;10(2):205–6.
65. Blanchet J, Stephan C, Vincendeau S, Houlgatte A, Semjonow A. [-2]proPSA and Prostate Health Index (PHI) improve detection of prostate cancer at initial and repeated biopsies in young men ($\leq$ 60 year old) preferentially detecting clinically significant cancer. Biochim Clin 2013;37:S158.
66. Blanchet JS, Durand X, Houlgatte A, Ramirez JN, Bensalah K, Guille B, *et al*. The Beckman Coulter Prostate Health Index (PHI) increases the specificity of detection of prostate cancer and reduces the number of negative biopsies. Tumor Biol 2010;31:S120.
67. Blanchet JS, Vincendeau S, Durand X, Ramirez JN, Bensalah K, Guille B, *et al*. Detection of aggressive prostate cancer using [-2] proPSA and the prostate health index. Tumor Biol 2010;31:S64.
68. Bordás N, Szalay I, Farkas G, Salgó L. The diagnostic value of the [-2]proPSA and PHI markers in the diagnosis of prostate cancer. Biochim Clin 2013;37:S159.
69. Dolejsova O, Kucera R, Fuchsova R, Topolcan O, Svobodova H, Hes O, *et al*. The ability of Prostate Health Index (PHI) to predict Gleason score in patients with prostate cancer and discriminate patients between Gleason score 6 and Gleason score higher than 6: a study on 320 patients after radical prostatectomy. Technol Cancer Res Treat 2018; 17:1–6.

70. Fiala V, Capoun O, Soukup V, Novak V, Stejskal J, Zalesky M, *et al*. The impact of prostate specific antigen density and Prostate Health Index assessment for prediction of prostate cancer in prostate biopsy. Eur Urol Suppl 2017;16(11): e2872.
71. Filella X, Foj L, Augé JM, Molina R, Alcover J. Clinical utility of %p2PSA and Prostate Health Index in the detection of prostate cancer. Clin Chem Lab Med 2014;52(9):1347–55.
72. Fillee C, Philippe M, Tombal B. Clinical evaluation of Prostate Health Index in a general population in order to detect prostate cancer. Clin Chem Lab Med 2011; 49:S237.
73. Foley R, Gorman L, Lundon D, Sharifi N, Murphy K, Tuzova A, *et al*. Analysis of pro [-2] PSA: a novel biomarker that significantly improves the detection of prostate cancer. Int J Surg 2015;23:S15.
74. Gnanapragasam VJ, Burling K, George A, Kastner C, Doble A, Barret T, *et al*. The Prostate Health Index (PHI) predicts positive cancer biopsies in men with a negative mpMRI in a repeat biopsy population. Eur Urol Suppl 2016;15(3):e384.
75. Guery T, Forzy G, Bonnal JL. Has the Prostate Health Index (PHI) a role in biological screening of prostate cancer? Preliminary study. Immuno-anal Biol spec 2011;26(3): 143–6.
76. Ito K, Miyakubo M, Sekine Y, Koike H, Matsui H, Shibata Y, *et al*. Diagnostic significance of [-2]pro-PSA and prostate dimension-adjusted PSA-related indices in men with total PSA in the 2.0-10.0 ng/mL range. World J Urol 2013;31(2): 305–11.
77. Ito K, Miyakubo M, Yamamoto T, Suzuki K. Diagnostic significance of [-2]proPSA and volume adjusted PSA-related indices in japanese men with total PSA in the 2.0 to 10. 0 ng/mL range J Urol 2010;183(4):e669–70.
78. Ito K, Yokomizo A, Tokunaga S, Arai G, Sugimoto M, Akakura K, *et al*. Importance of [-2] proPSA measurement in a diagnosis of prostate cancer: prostate health index trial (prophet)-diagnostic impacts of clinical laboratory-based indices on Gleason grade group †2 or †3 cancer. J Urol 2018;199(4):e152–3.
79. Klecka J, Behounek P, Topolcan O, Hora M, Fuchsova R, Karlíková M, *et al*. Is pro PSA more cancer specific form of prostate specific antigen for the early detection of prostate cancer? Urology 2011;78(3):S149–50.
80. Kucera R, Topolcan O, Dolejsova O, Hora H, Fuchsova R, Kinkorova J. PHI and prostate cancer: optimal management. Tumor Biol 2017;39(12):13.
81. Lalic N, Vukovic I, Glisic B, Djordjevic D, Durutovic O, Milenkovic-Petronic D, *et al*. Evaluation (-2)proprostete-specific antigen and Prostate Health Index for detection aggressive prostate cancer. Urology 2014;84(4):S303–4.
82. Larcher A, Lazzeri M, Lughezzani G, Gadda G, Abrate A, Scattoni V, *et al*. Serum isoform -2 proPSA (p2PSA) and its derivates, %p2PSA and PHI (Prostate Health Index), are more accurate than the reference standard test (PSA) in men scheduled for repeat biopsy. Anticancer Res 2012;32(5):1850–1.
83. Lazzeri M. Prostate cancer: the Prostate Health Index (PHI) may guide diagnostic and treatment. Tumor Biol 2011; 32:S40.
84. Lazzeri M. Impact of Prostate Health Index on the management of patients with prostate cancer. Biochim Clin 2013;37:S53–4.
85. Lazzeri M, Lughezzani G, Larcher A, Gadda G, Scattoni V, Sangalli M, *et al*. Serum isoform -2 proPSA (p2PSA) and its derivates, %p2PSA and PHI (Prostate Health Index), are more accurate than the reference standard test (PSA) in men scheduled for repeat biopsy. Eur Urol Suppl 2012;11(1):E259-E.
86. Lazzeri M, Lughezzani G, Nava L, Cestari A, Larcher A, Losa A, *et al*. 2proPSA, %-2proPSA and PHI (Prostate Health Index) correlate with cancer aggressiveness in patients who underwent radical prostatectomy. Eur Urol Suppl 2011;10(2):66.
87. Loeb S. PROSTATE CANCER Prostate Health Index-improving screening in men with family history. Nat Rev Urol 2013;10(9):497–8.
88. Lughezzani G, Lazzeri M, Hurle R, Buffi NM, Casale P, Fiorini G, *et al*. Clinical utility of PHI (Prostate Health Index) in men with tPSA >10 ng/mL. Results from a multicentric European study. J Urol 2016;195(4):e9.
89. Lukić I, Mandić S, Horvat V, Rolić T, Pavlović O, Šerić V. Use of serum isoform [-2]proPSA and Prostate Health Index for early prostate cancer detection. Biochem Med 2018;28: S150–1.
90. McNicholas T, Lazzeri M, Haese A, De La Taille A, Palou J, Lughezzani G, *et al*. Serum isoform [-2]proPSA (p2PSA) and its derivatives, namely %p2PSA and PHI (Prostate Health Index) in men younger than 60 years of age for prediction of prostate cancer. Results from a multicentric european study (prometheus project). J Urol 2013;189(4):e790.
91. McNicholas T, Lazzeri M, Haese A, De La Taille A, Palou J, Lughezzani G, *et al*. PHI (Prostate Health Index) and %p2PSA for prediction of prostate cancer in men younger than 60 years of age. A nested-case control study from PROpsa Multicentric European Study (PROMEtheuS project). Eur Urol Suppl 2013;12(1):e137–8.
92. Miyakubo M, Ito K, Yamamoto T, Suzuki K. Diagnostic significance of [-2]proPSA, total and transition zone prostate volume adjusted PSA-related indices in Japanese men with total PSA in the 2.0 to 10.0 ng/ml range. Eur Urol Suppl 2011;10(2):65.
93. Nordström T, Vickers AJ, Lilja HG, Grönberg H, Eklund M. A head to head comparison between a 4-kallikrein panel and a [-2]proPSA derivative model. Eur Urol Suppl 2014;13(1): e339.
94. Park H, Lee SW. Diagnostic performance of %[-2]proPSA and Prostate Health Index for prostate cancer: prospective, multi-institutional study. J Korean Med Sci 2018;33(11):e94.

95. Pepdjonovic L, Huang S, Mann S, Frydenberg M, Moon D, Snow R, *et al*. What is the role of the Prostate Health Index (PHI) in the diagnostic work-up of prostate cancer? Our preliminary experience. BJU Int 2017;119:113.
96. Perdona S, Quarto G, De Domenico R, Sorrentino D, Marino A, Sorrentino A, *et al*. Preoperative p2PSA and Prostate Health Index predict pathological outcomes in radical prostatectomy for prostate cancer. Anticancer Res 2012;32(5):1914.
97. Perdonà S, Quarto G, Sorrentino D, De Domenico R, Cariati F, Sorrentino A, *et al*. PHI and PCA3 significantly improve diagnostic accuracy in patients undergoing prostate biopsy. Anticancer Res 2012;32(5):1920.
98. Porpiglia F, Russo F, Manfredi M, Mele F, Poggio M, Grande S, *et al*. The roles of multiparametric MRI, PCA3, and PHI in the prediction of prostate cancer after an initial negative biopsy: results of a prospective study. Eur Urol Suppl 2014;13(1):e952–b.
99. Porpiglia F, Russo F, Manfredi M, Mele F, Poggio M, Grande S, *et al*. MRI, PCA3, or PHI to predict prostate cancer after an initial negative biopsy? Results of a prospective study. J Urol 2014;191(4):e819.
100. Regis L, Gasanz C, Miret E, Celma A, Planas J, Placer J, *et al*. Diagnostic accuracy of prostate health index for aggressive prostate cancer. An institutional validation study. J Urol 2016;195(4):e646.
101. Roobol MJ, Semjonow A, Bangma CH. Performance of the european randomized study of screening for prostate cancer (ERSPC) risk calculator for high grade PC in a clinical setting and the additional value of the Prostate Health Index (PHI). Clin Chem Lab Med 2011;49:S268.
102. Sanchis-Bonet A, Morales-Palacios N, Barrionuevo-González M, Ortega-Polledo LE, Tamayo-Ruiz JC, Sanchez-Chapado M. Clinical performance of [-2] pro-prostate-specific antigen and Prostate Health Index for prediction of prostate cancer in a cohort of Spanish men. Eur Urol Suppl 2016;15(13):e1562.
103. Sanda M, Wei J, Broyles D, Shin S, Partin A, Klee G, *et al*. Prostate Health Index (PHI) for reducing overdetection of indolent prostate cancer and unnecessary biopsy while improving detection of aggressive cancers. J Urol 2013;189(4):e843.
104. Scattoni V, Lazzeri M, De Luca S, Bollito E, Randone D, Lughezzani G, *et al*. Comparison of PHI (Prostate Health Index) and PCA3 assay in the prediction of prostate biopsy (PBx) outcome in patients who have undergone initial and repeated prostatic biopsies. Eur Urol Suppl 2012;11(1): E912–7.
105. Scattoni V, Lazzeri M, Lughezzani G, De Luca S, Passera R, Bollito E, *et al*. Comparison between PCA3 and phi specificty and sensitivity in predicting the presence of cancer at initial or repeat biopsy. Anticancer Res 2013;33(5):2293.
106. Scattoni V, Villa L, De Luca S, Capitanio U, Porpiglia F, Lazzeri M, *et al*. Prostate Health Index (PHI) is more accurate than PCA3 assay in the prediction of aggressive characteristics at initial prostate biopsy (PBX). Eur Urol Suppl 2014;13(1):e825.
107. Semjonow A, Bögemann M, Eminaga O, Hinkelammert R, Abbas M, Eltze E, *et al*. [-2] proPSA improves prediction of repeat prostate biopsy results. Urologe A 2011;50:73.
108. Sokoll LJ, Marks L, Sanda M, Wei J, Klee G, Bangma C, *et al*. Value of the prostate health index (PHI)1 for prostate cancer detection in men undergoing first or repeat biopsy. A multi-center prospective clinical study. Clin Chem 2011;57(10):A6.
109. Stephan C. Improving PSA-based prostate cancer detection with the Prostate Health Index (PHI). Clin Chem Lab Med 2011;49:S156.
110. Stephan C, Jung K, Semjonow A. Prospective head-to-head evaluation of [-2]proPSA, PCA3 and TMPRSS2:ERG in patients referred for prostate biopsy. Tumor Biol 2012; 33:S49.
111. Stephan C, Semjonow A, Schulze-Forster K, Cammann H, Hu X, Miller K, *et al*. Parallel measurement of urinary PCA3 and TMPRSS2:ERG with serum [-2]proPSA based PHI for prostate cancer detection. Eur Urol Suppl 2013;12(1):e859.
112. Stephan C, Vincendeau S, Houlgatte A, Miller K, Semjonow A. [-2]PROPSA and Prostate Health Index (PHI) improve detection of prostate cancer at initial and repeated biopsies in young men (<60 year old) preferentially detecting clinically significant cancer. J Urol 2013;189(4):e873–4.
113. Stephan C, Vincendeau S, Houlgatte A, Semjonow A. Prostate Health Index (PHI) using [-2]proPSA improves detection of prostate cancer preferentially identifying aggressive cancers. Eur Urol Suppl 2011;10(2):65–6.
114. Sugimoto M, Hirama H, Ito K, Shiraishi T, Kakehi Y. Prediction of reclassification at repeat biopsy with [-2]proPSA during active surveillance for low-risk prostate cancer: prospective longitudinal analysis of a Japanese multicenter study cohort. J Clin Oncol 2013;31(15).
115. Tan S, Xu Z, Wu J, Liang K, Jai R. The value of PHI/PCA3 in the early diagnosis of prostate cancer. Natl Med J China 2016;96(2):100–3.
116. Thompson J, Shnier R, Moses D, Brenner P, Delprado W, Tran M, *et al*. Prospective study of pre-biopsy magnetic resonance imaging (MRI) versus PCA3 and PHI for detection of clinically significant prostate cancer in men with previous negative biopsy: which is most accurate to guide selection of men for repeat biopsy? BJU Int 2015;115:65–6.
117. Tsang CF, Yiu MK, Cheung FK, Chiu Y, Ho BSH, Ng ATL, *et al*. Prospective study on the performance of the prostate-specific antigen isoform p2PSA and its derivative prostate health index in prediction of prostate cancer in local Chinese men with total prostate-specific antigen levels of 4-10 ng/mL. Int J Urol 2016;23:62.
118. Veltri RW. Serum marker %[-2]proPSA and the Prostate Health Index improve diagnostic accuracy for clinically relevant prostate cancer. BJU Int 2016;117(1):12–3.

119. Vincendeau S, Ramirez J, Durand X, Deligne E, Houlgatte A. The beckman coulter Prostate Health Index (PHI) improves diagnostic accuracy in prostate cancer detection. Eur Urol Suppl 2010;9(2):309.
120. Vincendeau S, Stephan C, Houlgatte A, Semjonow A. The Beckman Coulter Prostate Health Index (PHI) Increases the specificity of detection of prostate cancer and reduces the number of negative biopsies. Clin Chem Lab Med 2011;49:S270.
121. Xie Y, Iskakova EE, Yang Q, Sakenova N, Chen Y, Tse ZTH, *et al*. Over diagnosis of prostate cancer in Eurasian men by PSA and PHI: example of heterosis. Ann Oncol 2016; 27:ix92.
122. Yu ATO, Yeung VHW, Chu SK, Man CW. The application of Prostate Health Index improves positive prostate biopsy rate. BJU Int 2017;119:12.
123. Yuwono A, Tan TW, Yeow Y, Lee CH, Tan CH, Chong KT, *et al*. Evaluation of prostate health index (PHI) in detecting clinically significant prostate cancer in men who underwent MRI-ultrasound fusion prostate biopsy. BJU Int 2017;119:27.
124. Stattin P, Vickers AJ, Sjoberg DD, Johansson R, Granfors T, Johansson M, *et al*. Improving the specificity of screening for lethal prostate cancer using prostate-specific antigen and a panel of kallikrein markers: a nested case-control study. Eur Urol 2015;68(2):207–13.

Appendix I: Search strategy

Database	Terms	Records retrieved
MEDLINE (PubMed; Last search January 2019)	((((((prostate cancer[Title/Abstract]) OR prostate cancer[MeSH Terms]) OR significant prostate cancer[Title/Abstract]) OR aggressive prostate cancer[Title/Abstract])) AND (((((((predictive value* [Title/Abstract]) OR likelihood ratio[Title/Abstract]) OR specificity[MeSH Terms]) OR specificity[Title/Abstract]) OR sensitivity[MeSH Terms]) OR sensitivity[Title/Abstract]) OR accuracy[MeSH Terms]) OR accuracy[Title/Abstract])) AND (((((prostatic health index[MeSH Terms]) OR prostatic health index[Title/Abstract]) OR PHI[Title/Abstract]) OR p2PSA[Title/Abstract]) OR p2PSA)	155
Embase (Elsevier; Last search January 2019)	(('prostate cancer'/exp/mj OR 'prostate cancer':ab,ti OR 'significant prostate cancer':ab,ti OR 'aggressive prostate cancer':ab,ti) AND ('predictive value'/exp/mj OR 'predictive value':ab,ti OR 'likelihood ratio'/exp/mj OR 'likelihood ratio':ab,ti OR 'sensitivity and specificity'/exp/mj OR 'specificity':ab,ti OR 'sensitivity':ab,ti OR 'accuracy':ab,ti) AND ('prostatic health index':ab,ti OR 'PHI':ab,ti OR 'p2PSA':ab,ti))	269
Cochrane Central Register of Controlled Trials (CENTRAL; Last search January 2019)	MeSH descriptor: [Prostatic Neoplasms] AND (prostate health index:ti,ab,kw OR p2PSA:ti,ab,kw OR pro-PSA:ti,ab,kw) AND (MeSH descriptor: [Sensitivity and Specificity] OR MeSH descriptor: [Predictive Value of Tests] OR accuracy:ti,ab,kw)	9
CINAHL (EBSCO; Last search January 2019)	(((TI "prostate cancer") OR (AB "prostate cancer") OR (MH "Prostatic Neoplasms") OR (TI "significant prostate cancer") OR (AB "significant prostate cancer") OR (TI "aggressive prostate cancer") OR (AB "aggressive prostate cancer")) AND ((TI "predictive value") OR (AB "predictive value") OR (MH "predictive validity") OR (TI "likelihood ratio") OR (AB "likelihood ratio") OR (MH "Sensitivity and Specificity") OR (TI "specificity") OR (AB "specificity") OR (TI "sensitivity") OR (AB "sensitivity") OR (MH "Validity") OR (TI "accuracy") OR (AB "accuracy")) AND ((TI "prostatic health index") OR (AB "prostatic health index") OR (TI "PHI") OR (AB "PHI") OR (TI "p2PSA") OR (AB "p2PSA")))	14
Web of Science (Last search January 2019)	Prostatic health index OR p2psa	893
MedNar (Last search January 2019)	Prostatic health index P2PSA	1344 192
System for Information on Grey Literature (SIGLE) (Last search January 2018)	Prostate health index Prostatic health index P2PSA	3 0 0
Total		2879

Appendix II: Studies ineligible following full-text review

	Author	Reason for exclusion
1.	Cantiello *et al.*[31]	Reference standard not biopsy Gleason score
2.	Chiu *et al.*[32]	Reference standard not biopsy Gleason score
3.	Dolejsova *et al.*[33]	Reference standard not biopsy Gleason score
4.	Druskin *et al.*[34]	No Prostate Health Index threshold defined
5.	Eminaga *et al.*[35]	Reference standard not biopsy Gleason score
6.	Ferro *et al.*[36]	Insufficient data on aggressive prostate cancer
7.	Ferro *et al.*[37]	Reference standard not biopsy Gleason score
8.	Ferro *et al.*[38]	Conference abstract/proceedings
9.	Fillela *et al.*[39]	Insufficient data on aggressive prostate cancer
10.	Friedl *et al.*[40]	Insufficient data on aggressive prostate cancer
11.	Le *et al.*[41]	Insufficient data on aggressive prostate cancer
12.	Mearini *et al.*[10]	Insufficient data on aggressive prostate cancer
13.	Na *et al.*[42]	Insufficient data on aggressive prostate cancer
14.	Osredkar *et al.*[43]	Insufficient data on aggressive prostate cancer
15.	Perdona *et al.*[44]	Insufficient data on aggressive prostate cancer
16.	Sanchis-bonet *et al.*[45]	Insufficient data on aggressive prostate cancer
17.	Schwen *et al.*[46]	Reference standard not biopsy Gleason score
18.	Stephan *et al.*[47]	Reference standard not biopsy Gleason score
19.	Tosoian *et al.*[48]	Reference standard not biopsy Gleason score
20.	Lazzeri *et al.*[49]	Insufficient data on aggressive prostate cancer
21.	Chiu *et al.*[50]	Insufficient data on aggressive prostate cancer
22.	Boegemann *et al.*[51]	Insufficient data on aggressive prostate cancer
23.	Guazzoni *et al.*[52]	Insufficient data on aggressive prostate cancer
24.	Fossati *et al.*[53]	Insufficient data on aggressive prostate cancer
25.	Fujizuka *et al.*[54]	Insufficient data on aggressive prostate cancer
26.	Stephan *et al.*[55]	Insufficient data on aggressive prostate cancer
27.	Jansen *et al.*[56]	Insufficient data on aggressive prostate cancer
28.	Ng *et al.*[57]	Insufficient data on aggressive prostate cancer
29.	Nordstrom *et al.*[58]	Insufficient data on aggressive prostate cancer
30.	Sriplakich *et al.*[59]	Insufficient data on aggressive prostate cancer
31.	Abrate *et al.*[60]	Conference abstract/proceedings
32.	Barisiene *et al.*[61]	Conference abstract/proceedings
33.	Basso *et al.*[62]	Conference abstract/proceedings
34.	Bektic *et al.*[63]	Conference abstract/proceedings

(*Continued*)		
	Author	**Reason for exclusion**
35.	Bektic *et al.*[64]	Conference abstract/proceedings
36.	Blanchet *et al.*[65]	Conference abstract/proceedings
37.	Blanchet *et al.*[66]	Conference abstract/proceedings
38.	Blanchet *et al.*[67]	Conference abstract/proceedings
39.	Bordas *et al.*[68]	Conference abstract/proceedings
40.	Dolejsova *et al.*[69]	Conference abstract/proceedings
41.	Fiala *et al.*[70]	Conference abstract/proceedings
42.	Filella *et al.*[71]	Conference abstract/proceedings
43.	Fillee *et al.*[72]	Conference abstract/proceedings
44.	Foley *et al.*[73]	Conference abstract/proceedings
45.	Gnanapragasam *et al.*[74]	Conference abstract/proceedings
46.	Guery *et al.*[75]	Conference abstract/proceedings
47.	Ito *et al.*[76]	Insufficient data on aggressive prostate cancer
48.	Ito *et al.*[77]	Conference abstract/proceedings
49.	Ito *et al.*[78]	Conference abstract/proceedings
50.	Klecka *et al.*[79]	Conference abstract/proceedings
51.	Kucera *et al.*[80]	Conference abstract/proceedings
52.	Lalic *et al.*[81]	Conference abstract/proceedings
53.	Larcher *et al.*[82]	Conference abstract/proceedings
54.	Lazzeri *et al.*[83]	Conference abstract/proceedings
55.	Lazzeri *et al.*[84]	Conference abstract/proceedings
56.	Lazzeri *et al.*[85]	Conference abstract/proceedings
57.	Lazzeri *et al.*[86]	Conference abstract/proceedings
58.	Loeb *et al.*[87]	Conference abstract/proceedings
59.	Lughezzani *et al.*[88]	Conference abstract/proceedings
60.	Lukic *et al.*[89]	Conference abstract/proceedings
61.	McNicholas *et al.*[90]	Conference abstract/proceedings
62.	McNicholas *et al.*[91]	Conference abstract/proceedings
63.	Miyakubo *et al.*[92]	Conference abstract/proceedings
64.	Nordstrom *et al.*[93]	Conference abstract/proceedings
65.	Park *et al.*[94]	Insufficient data on aggressive prostate cancer
66.	Pepdjonovic *et al.*[95]	Conference abstract/proceedings
67.	Perdona *et al.*[96]	Conference abstract/proceedings
68.	Perdona *et al.*[97]	Conference abstract/proceedings
69.	Porpiglia *et al.*[98]	Conference abstract/proceedings

(*Continued*)		
	Author	**Reason for exclusion**
70.	Porpiglia *et al.*[99]	Conference abstract/proceedings
71.	Regis *et al.*[100]	Conference abstract/proceedings
72.	Roobol *et al.*[101]	Conference abstract/proceedings
73.	Sanchis-bonet *et al.*[102]	Conference abstract/proceedings
74.	Sanda *et al.*[103]	Conference abstract/proceedings
75.	Scattoni *et al.*[104]	Conference abstract/proceedings
76.	Scattoni *et al.*[105]	Conference abstract/proceedings
77.	Scattoni *et al.*[106]	Conference abstract/proceedings
78.	Semjonow *et al.*[107]	Conference abstract/proceedings
79.	Sokoll *et al.*[108]	Conference abstract/proceedings
80.	Stephan *et al.*[109]	Conference abstract/proceedings
81.	Stephan *et al.*[110]	Conference abstract/proceedings
82.	Stephan *et al.*[111]	Conference abstract/proceedings
83.	Stephan *et al.*[112]	Conference abstract/proceedings
84.	Stephan *et al.*[113]	Conference abstract/proceedings
85.	Sugimoto *et al.*[114]	Conference abstract/proceedings
86.	Tan *et al.*[115]	Conference abstract/proceedings
87.	Thompson *et al.*[116]	Conference abstract/proceedings
88.	Tsang *et al.*[117]	Conference abstract/proceedings
89.	Veltri[118]	Conference abstract/proceedings
90.	Vincendeau *et al.*[119]	Conference abstract/proceedings
91.	Vincendeau *et al.*[120]	Conference abstract/proceedings
92.	Xie *et al.*[121]	Conference abstract/proceedings
93.	Yu et *et al.*[122]	Conference abstract/proceedings
94.	Yuwono *et al.*[123]	Conference abstract/proceedings