HEALTH TECHNOLOGY ASSESSMENT

VOLUME 19 ISSUE 87 OCTOBER 2015
ISSN 1366-5278

The clinical effectiveness and cost-effectiveness of the PROGENSA® prostate cancer antigen 3 assay and the Prostate Health Index in the diagnosis of prostate cancer: a systematic review and economic evaluation

Amanda Nicholson, James Mahon, Angela Boland, Sophie Beale, Kerry Dwan, Nigel Fleeman, Juliet Hockenhull and Yenal Dundar

DOI 10.3310/hta19870

National Institute for Health Research

The clinical effectiveness and cost-effectiveness of the PROGENSA® prostate cancer antigen 3 assay and the Prostate Health Index in the diagnosis of prostate cancer: a systematic review and economic evaluation

Amanda Nicholson, James Mahon, Angela Boland, Sophie Beale,* Kerry Dwan, Nigel Fleeman, Juliet Hockenhull and Yenal Dundar

Liverpool Reviews and Implementation Group, University of Liverpool, Liverpool, UK

*Corresponding author

Declared competing interests of authors: none

Published October 2015
DOI: 10.3310/hta19870

This report should be referenced as follows:

Nicholson A, Mahon J, Boland A, Beale S, Dwan K, Fleeman N, *et al*. The clinical effectiveness and cost-effectiveness of the PROGENSA® prostate cancer antigen 3 assay and the Prostate Health Index in the diagnosis of prostate cancer: a systematic review and economic evaluation. *Health Technol Assess* 2015;**19**(87).

Health Technology Assessment is indexed and abstracted in *Index Medicus*/MEDLINE, *Excerpta Medica*/EMBASE, *Science Citation Index Expanded* (SciSearch®) and *Current Contents®*/ Clinical Medicine.

Health Technology Assessment

NICE TAR and DAR

ISSN 1366-5278 (Print)

ISSN 2046-4924 (Online)

Impact factor: 5.116

Health Technology Assessment is indexed in MEDLINE, CINAHL, EMBASE, The Cochrane Library and the ISI Science Citation Index.

This journal is a member of and subscribes to the principles of the Committee on Publication Ethics (COPE) (www.publicationethics.org/).

Editorial contact: nihredit@southampton.ac.uk

The full HTA archive is freely available to view online at www.journalslibrary.nihr.ac.uk/hta. Print-on-demand copies can be purchased from the report pages of the NIHR Journals Library website: www.journalslibrary.nihr.ac.uk

Criteria for inclusion in the *Health Technology Assessment* journal

Reports are published in *Health Technology Assessment* (HTA) if (1) they have resulted from work for the HTA programme, and (2) they are of a sufficiently high scientific quality as assessed by the reviewers and editors.

Reviews in *Health Technology Assessment* are termed 'systematic' when the account of the search appraisal and synthesis methods (to minimise biases and random errors) would, in theory, permit the replication of the review by others.

HTA programme

The HTA programme, part of the National Institute for Health Research (NIHR), was set up in 1993. It produces high-quality research information on the effectiveness, costs and broader impact of health technologies for those who use, manage and provide care in the NHS. 'Health technologies' are broadly defined as all interventions used to promote health, prevent and treat disease, and improve rehabilitation and long-term care.

The journal is indexed in NHS Evidence via its abstracts included in MEDLINE and its Technology Assessment Reports inform National Institute for Health and Care Excellence (NICE) guidance. HTA research is also an important source of evidence for National Screening Committee (NSC) policy decisions.

For more information about the HTA programme please visit the website: http://www.nets.nihr.ac.uk/programmes/hta

This report

The research reported in this issue of the journal was commissioned and funded by the HTA programme on behalf of NICE as project number 14/03/01. The protocol was agreed in April 2014. The assessment report began editorial review in November 2014 and was accepted for publication in March 2015. The authors have been wholly responsible for all data collection, analysis and interpretation, and for writing up their work. The HTA editors and publisher have tried to ensure the accuracy of the authors' report and would like to thank the reviewers for their constructive comments on the draft document. However, they do not accept liability for damages or losses arising from material published in this report.

Published by the NIHR Journals Library (www.journalslibrary.nihr.ac.uk), produced by Prepress Projects Ltd, Perth, Scotland (www.prepress-projects.co.uk).

Abstract

The clinical effectiveness and cost-effectiveness of the PROGENSA® prostate cancer antigen 3 assay and the Prostate Health Index in the diagnosis of prostate cancer: a systematic review and economic evaluation

Amanda Nicholson, James Mahon, Angela Boland, Sophie Beale,* Kerry Dwan, Nigel Fleeman, Juliet Hockenhull and Yenal Dundar

Liverpool Reviews and Implementation Group, University of Liverpool, Liverpool, UK

*Corresponding author sophie.beale@liverpool.ac.uk

Background: There is no single definitive test to identify prostate cancer in men. Biopsies are commonly used to obtain samples of prostate tissue for histopathological examination. However, this approach frequently misses cases of cancer, meaning that repeat biopsies may be necessary to obtain a diagnosis. The PROGENSA® prostate cancer antigen 3 (PCA3) assay (Hologic Gen-Probe, Marlborough, MA, USA) and the Prostate Health Index (phi; Beckman Coulter Inc., Brea, CA, USA) are two new tests (a urine test and a blood test, respectively) that are designed to be used to help clinicians decide whether or not to recommend a repeat biopsy.

Objective: To evaluate the clinical effectiveness and cost-effectiveness of the PCA3 assay and the phi in the diagnosis of prostate cancer.

Data sources: Multiple publication databases and trial registers were searched in May 2014 (from 2000 to May 2014), including MEDLINE, EMBASE, The Cochrane Library, ISI Web of Science, Medion, Aggressive Research Intelligence Facility database, ClinicalTrials.gov, International Standard Randomised Controlled Trial Number Register and World Health Organization International Clinical Trials Registry Platform.

Review methods: The assessment of clinical effectiveness involved three separate systematic reviews, namely reviews of the analytical validity, the clinical validity of these tests and the clinical utility of these tests. The assessment of cost-effectiveness comprised a systematic review of full economic evaluations and the development of a de novo economic model.

Setting: The perspective of the evaluation was the NHS in England and Wales.

Participants: Men suspected of having prostate cancer for whom the results of an initial prostate biopsy were negative or equivocal.

Interventions: The use of the PCA3 score or phi in combination with existing tests (including histopathology results, prostate-specific antigen level and digital rectal examination), multiparametric magnetic resonance imaging and clinical judgement.

Results: In addition to documents published by the manufacturers, six studies were identified for inclusion in the analytical validity review. The review identified issues concerning the precision of the PCA3 assay measurements. It also highlighted issues relating to the storage requirements and stability of samples intended for analysis using the phi assay. Fifteen studies met the inclusion criteria for the clinical validity review. These studies reported results for 10 different clinical comparisons. There was insufficient evidence to enable the identification of appropriate test threshold values for use in a clinical setting. In addition, the implications of adding either the PCA3 assay or the phi to clinical assessment were not clear. Furthermore, the addition of the PCA3 assay or the phi to clinical assessment plus magnetic resonance imaging was not found to improve discrimination. No published papers met the inclusion criteria for either the clinical utility review or the cost-effectiveness review. The results from the cost-effectiveness analyses indicated that using either the PCA3 assay or the phi in the NHS was not cost-effective.

Limitations: The main limitations of the systematic review of clinical validity are that the review conclusions are over-reliant on findings from one study, the descriptions of clinical assessment vary widely within reviewed studies and many of the reported results for the clinical validity outcomes do not include either standard errors or confidence intervals.

Conclusions: The clinical benefit of using the PCA3 assay or the phi in combination with existing tests, scans and clinical judgement has not yet been confirmed. The results from the cost-effectiveness analyses indicate that the use of these tests in the NHS would not be cost-effective.

Study registration: This study is registered as PROSPERO CRD42014009595.

Funding: The National Institute for Health Research Health Technology Assessment programme.

Contents

List of tables

List of figures

List of boxes

Glossary

Accuracy A measure of the closeness of the experimental value to the actual amount of the substance in the matrix.

Active surveillance A form of monitoring patients with slow-growing prostate cancer. It differs from watchful waiting in that if the patient needs treatment the aim of the treatment will be curative, it is suitable for some men with cancer that is contained in the prostate (i.e. localised) and it usually involves more regular hospital tests such as biopsies and magnetic resonance imaging.

Analytical sensitivity A measure that represents the smallest amount of substance in a sample that can accurately be measured by an assay.

Analytical specificity The ability of an assay to measure a particular substance, rather than others, in a sample.

Area under the curve A measure of the diagnostic accuracy of a technology. The measure is based on the geometric inspection of a receiver operating characteristics curve. A receiver operating characteristics curve is a plot of the true-positive rate against the false-positive rate at different threshold settings. A technology with perfect diagnostic accuracy will have an area under the curve of 1, a technology which is no better than chance will have an area under the curve of 0.5 and a technology which miscategorises on every occasion will have an area under the curve of zero.

Atypical small acinar proliferation A collection of small prostatic glands, identified on prostate biopsy, whose significance is uncertain and cannot be determined to be benign or malignant.

Benign prostatic hyperplasia A common urological condition caused by the non-cancerous enlargement of the prostate gland in ageing men. Urinating symptoms can occur as the prostate enlarges.

Clinically significant/insignificant prostate cancer Prostate cancer that is unlikely to result in death. A cancer is said to be clinically significant if it is likely to be the cause of death.

Clinical utility A measure (preferably in a quantitative form) of the extent to which diagnostic testing improves health outcomes relative to the current best alternative, which could be some other form of testing or no testing at all.

Clinical validity The predictive value of a test for a given clinical outcome, for example the likelihood that cancer will develop in someone with a positive test.

Core Sample of material taken from the prostate during a biopsy.

Cost-effectiveness acceptability curve A curve that shows, for a range of maximum amounts of money, how much a decision-maker might be willing to pay for a particular unit change in outcome and the probability that (given the available data) one intervention is cost-effective compared with the alternative(s).

Cut-off See *Threshold (clinical)* and *Threshold (economics)*.

Decision curve analysis A graphical analysis showing the net benefit of various diagnostic models which take account of the benefit of diagnosed cases and harms of unnecessary biopsies.

Derived sensitivity Sensitivity estimates derived from a receiver operating characteristics curve rather than from a 2 × 2 table.

Derived specificity Specificity estimates derived from a receiver operating characteristics curve rather than from a 2 × 2 table.

Diagnostic accuracy The effectiveness of a diagnostic test to correctly categorise patients as either 'positive' or 'negative' for the presence of a disease. There are several ways this can be expressed, for example the area under the curve or as sensitivity and specificity.

Diagnostic odds ratio The ratio of the odds of a positive intervention test in those with the disease to the odds of a positive intervention test in those without the disease.

Direct head-to-head study A study in which participants receive both intervention and comparator tests, and the tests are therefore evaluated in the same population (also called a within-study comparison).

Discounting A method used to adjust the value of costs and outcomes which occur in different time periods into a common time period, usually the present.

End-to-end study A study following participants from early clinical investigation and the decision to have a repeat biopsy through to diagnosis, treatment and long-term follow-up for prostate cancer (same as *Test-to-treatment study*).

External Assessment Group An independent group of researchers commissioned to review the evidence on a group of diagnostic technologies. The Diagnostics Assessment Committee bases its discussions on the diagnostic assessment report produced by the External Assessment Group.

False negative In the case of prostate cancer, a negative intervention test in a man in who is found on biopsy to have prostate cancer.

False positive In the case of prostate cancer, a positive intervention test in man who is found on biopsy not to have prostate cancer.

Forest plot A graphical display designed to illustrate the relative strength of treatment effects in multiple quantitative scientific studies addressing the same question.

Gleason score A scoring system used to help evaluate the prognosis of men with prostate cancer. A score is given based on the cancer's microscopic appearance. Gleason scores range from 2 to 10; the higher the Gleason score, the more aggressive the cancer.

Healthcare Resource Group A grouping that consists of patient events that have been judged to consume a similar level of resource.

Incremental cost-effectiveness ratio The difference in the mean costs of two interventions in the population of interest divided by the difference in the mean outcomes in the population of interest.

Indirect (between-study) comparison An analysis comparing the performance of intervention and comparator tests using data from studies in which tests are evaluated in different study populations (also called a between-study comparison).

Intervention test The diagnostic test which is being evaluated.

Likelihood ratio A description of how much more likely it is that a person with a disease than one without that disease will have a particular test result.

Logistic regression models A statistical method for analysing a data set in which there are one or more independent variables that determine an outcome. The outcome is measured with a dichotomous variable (i.e. one with only two possible outcomes).

Negative predictive value The proportion of patients with negative test results who do not have the disease. The probability that a patient who is test negative on an intervention does not have prostate cancer detected on biopsy.

Nomogram Risk algorithms that combine multiple clinical and laboratory risk factors to create a cumulative risk score. Most nomograms aim to predict the probable course of a disease; however, some nomograms aim to predict the result of a biopsy in men suspected of having prostate cancer.

Positive predictive value The proportion of patients with positive test results who actually have the disease. The probability that a patient who tests positive on an intervention test has prostate cancer detected at biopsy.

Precision The extent to which individual measurements of a sample are close to each other.

Probabilistic sensitivity analysis A way to quantify the level of confidence that a decision-maker has in the conclusions of an economic evaluation.

Prostate biopsy A procedure in which small, hollow needle core samples are removed from a man's prostate gland to be examined microscopically for the presence of cancer.

Prostate-specific antigen An enzyme secreted by the epithelial cells of the prostate gland. It is present in small quantities in the serum of men with healthy prostates, but the level of prostate-specific antigen is often elevated in the presence of prostate cancer or other prostate disorders.

Quality-adjusted life-years An index of survival that is adjusted to account for the patient's quality of life which incorporates changes in both quantity (longevity/mortality) and quality (morbidity, psychological, functional, social and other factors) of life. They are used to measure benefits in cost–utility analysis. The number of quality-adjusted life-years gained is the mean number of quality-adjusted life-years associated with one intervention minus the mean number of quality-adjusted life-years associated with an alternative intervention.

Quality of life A concept incorporating all the factors that might impact on an individual's life, including factors such as the absence of disease or infirmity as well as other factors which might affect physical, mental and social well-being.

Radical prostatectomy The surgical removal of all of the prostate gland.

Receiver operating characteristics curve A plot of the true-positive rate against the false-positive rate of a test at different threshold settings.

Reference standard A diagnostic test used to estimate the sensitivity and specificity of another diagnostic test, known as an index test. The reference standard is assumed to have perfect sensitivity and specificity; thus, when both tests categorise something differently, the reference standard test categorisation is assumed to be correct (either true negative or true positive).

Saturation biopsy A type of biopsy that may be carried out transrectally or transperineally. A minimum of 20 cores are taken. This procedure may be carried out under general anaesthetic, particularly if a man has found the experience of a previous biopsy to be uncomfortable and/or distressing.

Sensitivity The proportion of those who actually have the disease and who are correctly identified with positive test results, that is the proportion of men with prostate cancer at biopsy who are identified by the intervention test (also called the *True-positive* rate).

Sensitivity analysis In health economics, the study of how the uncertainty in the magnitude of the output from the cost-effectiveness model (the incremental cost-effectiveness ratio per quality-adjusted life-year gained) can be apportioned to different sources of uncertainty in model inputs.

Specificity The proportion of those who do not have the disease who are correctly identified as having a negative test result, that is the proportion of men without prostate cancer at biopsy who are test negative on the intervention test (also called the *True-negative* rate).

Template biopsy A type of biopsy that involves taking 25–40 cores transperineally. A template or grid is used.

Test-to-treatment study See *End-to-end study*.

Threshold (clinical) A value, within a range of values, used to categorise observations into one of two mutually exclusive groups. For example, guidelines suggest that the decision whether or not to investigate for possible prostate cancer is influenced by prostate-specific antigen level, with a threshold of above 3 ng/ml used for men in their fifties, 4 ng/ml for men in their sixties and 5 ng/ml for men in their seventies.

Threshold (economics) The amount of variation needed in the parameter values of a model to achieve a specified outcome. In the context of cost-effectiveness analysis in the UK NHS, this specified outcome is usually the cost-effectiveness threshold of £20,000–30,000 per additional quality-adjusted life-year gained.

True negative In the case of prostate cancer, a negative intervention test in a man who does not, in fact, have prostate cancer.

True positive In the case of prostate cancer, a positive intervention test in a man who does, in fact, have prostate cancer.

Utility A measure of the strength of an individual's preference for a specific health state in relation to alternative health states. The utility scale assigns numerical values on a scale from 0 (death) to 1 (optimal or 'perfect' health). Health states can be considered worse than death and thus have a negative value.

Watchful waiting A form of cancer monitoring. It differs from active surveillance in that, if treatment is needed, its aim will be to control rather than cure the cancer. It is generally suitable for men with concomitant health problems who may be less able to cope with treatment or whose cancer may never cause a problem during their lifetime; it usually involves fewer tests and these usually take place at the general practitioner's surgery rather than at the hospital.

List of abbreviations

ASAP	atypical small acinar proliferation
AUC	area under the curve
CE	Conformité Européenne
CEAC	cost-effectiveness acceptability curve
CI	confidence interval
CV	coefficient of variation
DCE-MRI	dynamic contrast-enhanced magnetic resonance imaging
DRE	digital rectal examination
DW	diffusion weighted
DW-MRI	diffusion-weighted magnetic resonance imaging
EAG	External Assessment Group
FDA	Food and Drug Administration
FN	false negative
FP	false positive
fPSA	free prostate-specific antigen
GP	general practitioner
HGPIN	high-grade prostatic intraepithelial neoplasia
HRG	Healthcare Resource Group
HTA	Health Technology Assessment
LoB	limit of blank
LoD	limit of detection
LoQ	limit of quantitation
mpMRI	multiparametric magnetic resonance imaging
MRI	magnetic resonance imaging
mRNA	messenger ribonucleic acid
MRS	magnetic resonance spectroscopy
MRSI	magnetic resonance spectroscopy imaging
NICE	National Institute for Health and Care Excellence
OR	odds ratio
p2PSA	[–2]pro-prostate-specific antigen
PCA3	PROSTATE cancer antigen 3
PCPT	Prostate Cancer Prevention Trial
phi	Prostate Health Index
PSA	prostate-specific antigen
QALY	quality-adjusted life-year
QoL	quality of life
RCT	randomised controlled trial
REDUCE	Reduction by Dutasteride of Prostate Cancer Events trial
RNA	ribonucleic acid
ROC	receiver operating characteristics
SD	standard deviation
SSED	*Summary of Safety and Effectiveness Data*
T2-MRI	T2-weighted magnetic resonance imaging
TN	true negative
TP	true positive
tPSA	total prostate-specific antigen
TRUS	transrectal ultrasonography
WHO	World Health Organization

Note

This monograph is based on the Technology Assessment Report produced for NICE. The full report contained a considerable number of data that were deemed commercial-in-confidence. The full report was used by the Appraisal Committee at NICE in their deliberations. The full report with each piece of commercial-in-confidence data removed and replaced by the statement 'commercial-in-confidence information (or data) removed' is available on the NICE website: www.nice.org.uk.

The present monograph presents as full a version of the report as is possible while retaining readability, but some sections, sentences, tables and figures have been removed. Readers should bear in mind that the discussion, conclusions and implications for practice and research are based on all the data considered in the original full NICE report.

Plain English summary

It can be difficult to diagnose prostate cancer. Currently, men suspected of having cancer are sent for a prostate biopsy. This procedure involves removing a small part of a man's prostate and examining it under a microscope to find out if cancer is present. However, biopsies can miss cases of cancer. This means that some men may have several biopsies before the suspected cancer is found. The PROGENSA® prostate cancer antigen 3 assay (Hologic Gen-Probe, Marlborough, MA, USA) and the Prostate Health Index (Beckman Coulter Inc., Brea, CA, USA) are two new tests which may help avoid unnecessary biopsies. We reviewed all currently available information to find out whether or not these two new tests should be used in the NHS. We also built an economic model to find out whether or not the tests offer value for money to the NHS. We found no clear evidence that either of the two tests worked better than current practice (i.e. using only biopsies). Furthermore, results from the economic model showed that use of either of these tests would not represent value for money.

Scientific summary

Background

Prostate cancer is a leading cause of mortality and morbidity. Approximately 40,000 new cases are diagnosed each year in the UK, and in 2011 10,793 deaths in the UK were attributed to the disease. The major risk factors for prostate cancer are increasing age, family history in a first-degree relative (brother or father) and race (higher risk in black men). The disease shows a strong inverse social gradient, being more common in more affluent social groups.

There is no single definitive test to identify men with prostate cancer. In cases in which prostate cancer could be the cause of presenting symptoms, the general practitioner carries out a number of tests. If, after carrying out this exploratory work, the general practitioner considers that there is a risk of prostate cancer, then the patient is referred to a hospital consultant to discuss the options for further tests.

The most commonly used test to detect prostate cancer is a transrectal ultrasonography-guided biopsy. However, this biopsy can miss cancers altogether, it may identify small, low-risk cancers that do not need to be treated but that may cause anxiety, it is uncomfortable (sometimes painful) and there can be complications for the patient (including blood in the urine, rectal bleeding and acute urinary retention). In some cases, when prostate cancer has not been confirmed by the initial biopsy, a second biopsy may be recommended. However, there is no guarantee that the second biopsy will find cancers missed by the first biopsy and further biopsies may still be required. Techniques such as multiparametric magnetic resonance imaging have been introduced into diagnostic practice. Such techniques improve the diagnostic performance of biopsies, as they help identify the location of prostate cancer abnormalities. However, multiparametric magnetic resonance imaging is not available in all hospitals.

The PROGENSA® prostate cancer antigen 3 (PCA3) assay (referred to as the PCA3 assay; Hologic Gen-Probe, Marlborough, MA, USA) and the Prostate Health Index (phi; Beckman Coulter Inc., Brea, CA, USA) are two new tests (a urine test and a blood test, respectively) that are designed to be used to help a clinician decide whether or not a second biopsy should be recommended. The purpose of this assessment was to evaluate the clinical effectiveness and cost-effectiveness of these tests, used in combination with existing tests, scans and clinical judgement, in the diagnosis of prostate cancer in men who are suspected of having malignant disease and in whom the results of an initial prostate biopsy were negative or equivocal. The perspective of the evaluation was the NHS in England and Wales.

Objectives

The key objectives of this assessment were to address the following questions:

1. How well do the PCA3 and [–2]pro-prostate-specific antigen (p2PSA) tests measure the substances they are intended to measure?
2. How might the addition of the PCA3 assay or phi contribute to the diagnosis of prostate cancer?
3. How might the addition of PCA3 assay or phi to current diagnostic strategies affect patient outcomes?
4. Would the addition of PCA3 assay or phi to current diagnostic strategies be cost-effective?

Methods

The research comprised two elements: an assessment of clinical effectiveness (addressing objectives 1, 2 and 3) and an assessment of cost-effectiveness (addressing objective 4). Literature searches to inform both elements were undertaken in May 2014.

Assessment of clinical effectiveness

Assessing the clinical effectiveness of the PCA3 assay and the phi in the diagnosis of prostate cancer involved three separate systematic reviews:

1. a review of the analytical validity (how well laboratory tests measure the substances they are intended to measure) of the intervention tests to assess how accurately the tests measure PCA3 score/p2PSA level present in a sample
2. a review of the clinical validity (accuracy of the diagnostic tests) of comparator and intervention pathways to assess how the addition of the PCA3 score or the phi might contribute to the diagnosis of prostate cancer
3. a review of the clinical utility of the intervention test pathways to evaluate how the addition of the intervention tests might affect patient outcomes, including long-term outcomes such as mortality and morbidity from prostate cancer and intermediate outcomes such as side effects from tests.

The methods used followed the systematic review principles outlined in the Centre for Reviews and Dissemination guidance for undertaking reviews in health care, the National Institute for Health and Care Excellence Diagnostic Assessment Programme manual and publications from the Cochrane diagnostic test accuracy methods working group. The review of analytical validity was informed by the principles outlined in the Agency for Healthcare Research and Quality methods guide and the Evaluation of Genomic Applications in Practice and Prevention initiative.

Assessment of cost-effectiveness

The cost-effectiveness assessment included two components: a systematic review of existing full economic evaluations and the development of a de novo health economic model.

The review of cost-effectiveness literature was conducted in line with the Centre for Reviews and Dissemination guidance for undertaking reviews in health care.

A de novo economic model was constructed using data from the clinical validity review. The External Assessment Group (EAG) model used values for derived specificities at defined sensitivity levels. By modelling defined sensitivities, the only difference between testing strategies was the number of biopsies required to identify a given number of cancers. The cost-effectiveness results were, therefore, driven by the differences in costs and quality-adjusted life-years (QALYs) losses that accrued in the different testing strategies (as a result of differences in numbers of biopsies performed). Incremental cost-effectiveness ratios were presented. Model input values for resource use, costs and utility values were extracted from published sources. The time horizon, in the base case, was 3 years (extended to 6 years in a scenario analysis), the model perspective was that of the UK NHS, and costs and benefits were discounted at a rate of 3.5%.

Results

Analytical validity review

To inform the assessment of the two assays, the EAG relied on data that have been published primarily by the manufacturers in the form of pack inserts and/or to support their submissions for regulatory approval. The review highlighted some important issues concerning the precision of PCA3 assay measurements and the requirements for storage and stability of samples for phi.

Clinical validity review

The key findings from the clinical validity review are as follows:

- Ten studies consider the comparison of clinical assessment versus clinical assessment + PCA3. The findings indicate that the implications of adding the PCA3 assay to clinical assessment are not clear and it is not possible to identify a single-threshold value for use in a clinical setting.
- Four studies consider the comparison of clinical assessment versus clinical assessment + phi. The findings indicate that the implications of adding phi to clinical assessment are not clear and it is not possible to identify threshold values for use in a clinical setting.
- Two studies consider the comparison of clinical assessment + magnetic resonance imaging (MRI) versus clinical assessment + MRI + PCA3. The findings indicate that the addition of the PCA3 assay to clinical assessment + MRI does not have a noticeable impact on discrimination.
- Only one study assesses the comparison of clinical assessment + MRI versus clinical assessment + MRI + phi. The findings indicate that the addition of phi to clinical assessment + MRI does not have a noticeable impact on discrimination.

Clinical utility review

The EAG did not identify any published papers that met the inclusion criteria for the clinical utility review.

Cost-effectiveness literature review

The EAG did not identify any published papers that met the inclusion criteria for the cost-effectiveness literature review.

Cost-effectiveness modelling

The key results from the base-case analyses are as follows:

- Clinical assessment versus clinical assessment + PCA3: clinical assessment dominates clinical assessment + PCA3 (i.e. clinical assessment costs less and generates more QALYs than clinical assessment + PCA3).
- Clinical assessment versus clinical assessment + phi: clinical assessment dominates clinical assessment + phi (i.e. clinical assessment costs less and generates more QALYs than clinical assessment + phi).
- Clinical assessment + MRI versus clinical assessment + MRI + PCA3: clinical assessment + MRI costs less but is less effective than clinical assessment + MRI + phi, and the incremental cost-effectiveness ratio per QALY gained for clinical assessment + MRI + phi is £5,418,366 compared with clinical assessment + MRI.
- Clinical assessment + MRI versus clinical assessment + MRI + phi: clinical assessment + MRI costs less but is less effective than clinical assessment + MRI + phi and the incremental cost-effectiveness ratio per QALY gained for clinical assessment + MRI + phi is £2,500,530 compared with clinical assessment + MRI.

Results from the sensitivity and scenario analyses show that, other than in one scenario which employed an unrealistic prostate-specific antigen (PSA) monitoring strategy, the incremental cost-effectiveness ratios that were generated to test model uncertainty are all above £20,000 per QALY gained. The probabilistic sensitivity analyses confirm that alternative testing strategies using any test in addition to clinical assessment are not cost-effective, although it should be noted that QALY loss associated with a biopsy was not varied in the probabilistic analyses.

Discussion

Strengths of the assessment

Although the assessment of analytical validity relied on data that have been published primarily by the manufacturers, the EAG considers that the analytical validity of the two tests has been comprehensively documented.

The clinical validity review includes results for a wide range of outcome measures for 10 different clinical comparisons. Its key strength is its focus on four clinically relevant comparisons, that is those studies reporting the addition of the PCA3 assay or phi to clinical assessment (with or without MRI).

The de novo economic model is based on the best available clinical validity evidence (identified through the systematic review) and captures the trade-off between the high upfront costs of diagnostic tests and the reduction in subsequent biopsies and their costs. It captures all of the main factors relevant to the decision problem and calculations are transparent.

Limitations of the assessment

The limitations of the clinical validity assessment are as follows:

- The review conclusions are over-reliant on findings from one study; of the 10 clinically relevant comparisons described in the 17 studies, data from one study are used in nine comparisons.
- The clinical relevance of many of the reported outcome measures is unclear.
- Many of the reported results for the clinical validity outcomes include neither standard errors nor confidence intervals.
- Descriptions of clinical assessment vary widely within reviewed studies.
- There was no consistent use, in the literature, of threshold values for either PCA3 score or phi.

The limitations of the economic assessment were as follows:

- There was a lack of generalisable clinical validity data to inform the economic model.
- The model was unable to capture and/or value all the key factors that might influence cost-effectiveness. The main area where information was lacking was in relation to utility decrements associated with prostate biopsies.

Uncertainties

Owing to the lack of published literature, the assessment was unable to address three clinical issues outlined in the final scope: detection of clinically insignificant cancer, optimal order of the tests and the effect of using different forms of reference standard (biopsy).

Further uncertainties, which relate to the economic model, include:

- the extent to which the model reflects NHS clinical practice
- the best way to model the most representative PSA monitoring strategy employed after a negative or equivocal biopsy in NHS clinical practice
- lack of clarity around the extent to which, in practice, clinicians prioritise sensitivity over specificity or vice versa.

Generalisability of the findings

The target population is not homogeneous but appears to comprise three subpopulations, namely those for whom a second biopsy is clearly indicated, those for whom a second biopsy is unnecessary and those for whom the need for a second biopsy is unclear. Most of the study populations described in the included studies comprise men who were referred for a second biopsy because of clinical suspicions and the criteria for referral varied between studies. The EAG considers, therefore, that it is not appropriate to apply the results of this review to all men with negative or equivocal biopsy results.

In addition, the representation of clinical assessment varied in the included studies. Although this may reflect clinical practice, in which clinical assessment is not standardised, it is difficult to meaningfully compare the results of studies which have markedly different representations of clinical assessment.

The reference standard (prostate biopsy) is an imperfect diagnostic tool as it does not detect all cancers. Without a gold standard that offers 100% specificity and 100% sensitivity, it is difficult to confidently assess the accuracy of competing diagnostic strategies.

Conclusions

Overall, the EAG considers that the analytical validity of the PCA3 assay and the phi has been comprehensively documented. The EAG identified some important issues relating to the precision of PCA3 assay measurements. Issues highlighted in relation to the use of the p2PSA assay were sample handling and the thermal stability of samples.

The clinical benefit of using the PCA3 assay and the phi in combination with existing tests, scans and clinical judgement in the diagnosis of prostate cancer in men who are suspected of having malignant disease and in whom the results of an initial prostate biopsy are negative or equivocal has not yet been confirmed. Furthermore, results from the cost-effectiveness analyses indicate that the use of these tests in the NHS for men who are suspected of having prostate cancer and have had a negative or equivocal initial biopsy would not be cost-effective.

Implications for service provision

A number of issues may affect the successful implementation of the assays in the NHS:

- PCA3 assay: the urine sample required for the PCA3 assay needs to be transferred to specialist transport tubes within 4 hours. Primary care staff may need some training for this requirement to be met. In addition, the published precision estimates for the PCA3 assay raise concerns about the interpretation and use of the PCA3 score for detecting prostate cancer.
- phi: blood samples for the p2PSA assay need to centrifuged and the serum separated within 3 hours. This time limit may pose challenges to implementing the test throughout the NHS. Furthermore, it is not clear whether or not blood samples taken in a primary care setting could be routinely transported to a laboratory and processed as required within 3 hours.

Suggested research priorities

Longitudinal end-to-end studies following men from initial investigation through to diagnosis and treatment of prostate cancer are required. Ideally, these studies would be randomised controlled trials with men allocated to different diagnostic test pathways after an initial negative or equivocal biopsy. However, descriptive data from observational cohorts following men over several years from initial referral onwards could address some unanswered issues.

Study registration

This study is registered as PROSPERO CRD42014009595.

Funding

Funding for this study was provided by the Health Technology Assessment programme of the National Institute for Health Research.

Chapter 1 Background and definition of the decision problem

Brief description of the decision problem

There is no single definitive test for prostate cancer. In cases where prostate cancer could be the cause of presenting symptoms, the general practitioner (GP) carries out a number of tests. If, after carrying out this exploratory work, the GP feels that there is a risk of prostate cancer, then the patient will be referred to a hospital consultant to discuss the options for further tests.

The most commonly used test to detect prostate cancer is a transrectal ultrasonography (TRUS)-guided biopsy. However, this biopsy has a number of limitations. It can miss cancers altogether, it may identify small, low-risk cancers that do not need to be treated but the presence of which will cause anxiety, it is uncomfortable (sometimes painful) and there can be complications for the patient (including blood in semen and urine, rectal bleeding, voiding difficulties, and major and minor infections).[1] In some cases where prostate cancer has not been confirmed by the initial biopsy, a second biopsy may be recommended; however, there is no guarantee that the second biopsy will find cancers missed by the first biopsy and further biopsies may still be performed. Techniques such as magnetic resonance spectroscopy (MRS) and enhanced magnetic resonance imaging (MRI) have been introduced into diagnostic practice. Such techniques aid the localisation of prostate cancer abnormalities, thus improving the diagnostic performance of biopsies. However, MRS and MRI are not available in all hospitals.

The PROGENSA® prostate cancer antigen 3 (PCA3) assay (referred to as the PCA3 assay; Hologic Gen-Probe, Marlborough, MA, USA) and the Beckman Coulter Prostate Health Index (phi; Brea, CA, USA) are two new tests (a urine test and a blood test, respectively) that are designed to be used to help a clinician decide whether or not a repeat biopsy is necessary. The purpose of this assessment is to evaluate the clinical effectiveness and cost-effectiveness of these tests, in combination with existing tests, scans and clinical judgement, in the diagnosis of prostate cancer in men who are suspected of having malignant disease and in whom the results of an initial prostate biopsy were negative or equivocal. The perspective of this evaluation is the NHS in England and Wales.

This report contains reference to confidential information provided as part of the National Institute for Health and Care Excellence (NICE) appraisal process. This information has been removed from the report and the results, discussions and conclusions of the report do not include the confidential information. These sections are clearly marked in the report.

Epidemiology of prostate cancer

The prostate is a gland that is part of the urinary and reproductive system of males. Women do not have a prostate gland. It is located in the pelvic region, beneath the bladder, and surrounds the upper part of the urethra, the tube that carries urine from the bladder through the penis. It has two functions: first, muscle fibres squeeze the urethra slightly and help control the flow of urine, and, second, the prostate is the site of production of fluids that are added to the seminal fluid (semen).

The prostate starts to develop before birth and grows rapidly during puberty, staying the same size or growing slowly in healthy adults. In a normal young adult male the gland is approximately 3 cm long and weighs approximately 20 g.

The prostate has three glandular regions, namely the peripheral zone, the central zone and the transition zone.[2] The vast majority of prostate cancers are adenocarcinomas (meaning that they originate from glandular epithelial cells). Up to 70% of cancers arise in the peripheral zone, 15–20% arise in the central zone and 10–15% arise in the transition zone.[3]

The prognosis and natural history of prostate cancer vary depending on the extent of spread and the grade of cancer at diagnosis. The prognosis for men with disease localised to the prostate varies, and more aggressive changes on histopathology and higher prostate-specific antigen (PSA) levels are associated with a worse prognosis.[4] In the early stages, prostate cancer is localised to the prostate and its progression is driven by androgens. At this stage the disease may be cured with surgery or radiotherapy; alternatively, conservative management, that is active surveillance/watchful waiting, may be adopted.[5] Active surveillance involves regular tests to monitor the cancer. The tests are likely to vary by treatment centre but may include:

- a PSA test every 3–6 months
- a digital rectal examination (DRE) every 6–12 months
- a biopsy about a year after diagnosis and every few years thereafter
- a MRI scan if the patient's PSA level and/or DRE result suggest the cancer is growing.

If the results of a test show that the cancer has grown, the patient will be offered curative treatment, for example surgery or radiotherapy.[6] Watchful waiting differs slightly from active surveillance. It is an approach that is generally suitable for men with other health problems who may be physically less able to cope with treatments or whose cancer may never cause major health problems during their lifetime. Active surveillance usually involves fewer tests, and these usually take place at the GP surgery rather than at a hospital.[6]

Patients who have inoperable locally advanced or metastatic disease at diagnosis or who have inoperable recurrent disease are treated with androgen deprivation therapy. As the disease progresses, the tumour ceases to respond to androgen deprivation therapy, but may respond to antiandrogens and oestrogenic agents.[7] Most patients receive two or more hormonal therapies and are then offered chemotherapy.[8]

Incidence

The most up-to-date figures (2011) indicate that prostate cancer is the most common cancer in men in the UK, accounting for 25% of all new cases of cancer in males.[9] In the same year, there were 35,567 new cases in England and 2346 new cases in Wales, giving a total of 37,913.[9] Age-standardised relative survival rates for prostate cancer in England during 2005–9 show that 93.5% of men with prostate cancer are expected to survive for at least 1 year, falling to 81.4% surviving 5 years or more. Survival rates in Wales are reported to be broadly similar to those in England.[10]

Prostate cancer incidence is strongly related to age, with the highest incidence rates being in older men. In the UK between 2009 and 2011, an average of 36% of cases were diagnosed in men aged 75 years and over, and only 1% were diagnosed in the under-fifties.[9] There is also evidence of an inverse association between prostate cancer incidence and deprivation in England, with prostate cancer being one of the few cancers with incidence rates lower among more-deprived males.[9] England-wide data for 2006–10 show that European age-standardised incidence rates are 17% lower for men living in the most deprived areas than for those in the least deprived areas.[9] In addition, there are links between prostate cancer and ethnicity. Age-standardised rates for white men with prostate cancer range from 96.0 to 99.9 per 100,000. Rates for Asian men are significantly lower, ranging from 28.7 to 60.6 per 100,000, while the rates for black men are significantly higher, ranging from 120.8 to 247.9 per 100,000.[9]

Mortality

Prostate cancer is the second most common cause of death due to cancer in men in England and Wales, second only to lung cancer.[11] Age-standardised mortality rates from prostate cancer declined by 13% between 2001 and 2012.[12] In 2012, there were 9133 from prostate cancer in England and 5556 deaths in Wales.

Quality of life of patients with prostate cancer

Glaser *et al.*[13] used a questionnaire survey to collect information about the quality of life (QoL) of patients with different types of cancer. Of the 1248 prostate cancer patients targeted, 866 (69.4%) returned completed questionnaires. The analysis indicated that patients who had surgery only (compared with radiotherapy and hormone treatment) had significantly higher QoL scores. The survey also revealed that:

- 38.5% reported some degree of urinary leakage
- 12.9% reported difficulty controlling their bowels
- 58.4% reported being unable to have an erection
- 11.0% reported significant difficulty in having or maintaining an erection.

The presence of urinary leakage was significantly associated with lower QoL scores, while erectile dysfunction and difficulty controlling bowels were not significantly associated with a reduction in QoL score.

Financial cost of prostate cancer

Biopsy cost

A study[14] was carried out to assess the diagnostic accuracy and cost-effectiveness of MRS and enhanced MRI techniques to aid the localisation of prostate abnormalities in a population undergoing repeat biopsy. Following this approach, assuming that approximately 25% of cancers are detected by repeat TRUS-guided needle biopsy[15] and that the cancer detection rate is approximately 25%,[16,17] then, based on a figure of 37,913 cases of prostate cancer in England and Wales, it can be assumed that 38,000 repeat biopsies are undertaken. The 2012–13 NHS reference costs[18] for the Healthcare Resource Group (HRG) of a needle biopsy of the prostate maps (LB27Z, outpatient procedure, urology) is £224, leading to a total cost to the NHS of approximately £8.5M in 2012–13. This figure should be considered as a lower limit for the cost of repeat biopsies, as it assumes that almost all men only receive a second biopsy and it takes little account of the cost of any subsequent biopsies.

First-year treatment cost

It has been estimated that the average first-year treatment cost per patient identified with prostate cancer is £2943.10 (2009 prices).[14] Inflating this cost to current prices (2012/13) results in a figure of £3167.72.[14] The number of cases in England and Wales in 2011 was 37,913, leading to an approximate first-year treatment cost of £120M. It should be noted that this is likely to be a conservative estimate, as the cost includes only active surveillance, radical prostatectomy and external beam radiation therapy. It does not include any other treatment costs, nor does it include any costs incurred by patients or the wider society. In addition, it is likely that this cost will rise even without any improvements in detection (and therefore incidence) because the population in the UK is ageing and, as the incidence of prostate cancer increases with age, it is likely that the number of cases of prostate cancer will increase over time. The number of patients treated and the cost of treatment are set to increase and this will lead to increased demand for resources (for example treatment facilities and trained specialists).

Current diagnostic practice

The recently updated NICE guideline,[11] *Prostate Cancer: Diagnosis and Treatment*, CG175, summarises current best practice for the diagnosis and management of prostate cancer.

Decision to perform initial biopsy

According to the updated NICE guideline,[11] men may initially present with clinical symptoms, such as difficulty with urination, or come to medical attention as the result of a raised PSA level. PSA is a protein produced in prostatic cell, which can be elevated in men with prostate cancer. However, it is also raised in

other benign prostatic conditions, such as infections (prostatitis) and hypertrophy. A raised PSA is not, therefore, specific to the presence of cancer and not all men with prostate cancer have increased PSA levels. The decision whether or not to investigate for possible cancer is influenced by age as well as PSA level. Men in their fifties with PSA levels above 3 ng/ml are considered for further investigation, with threshold levels being 4 ng/ml for men in their sixties and 5 ng/ml for men in their seventies.[19] The updated NICE guideline[11] recommends that the following factors should be taken into consideration when deciding to perform a biopsy: PSA level, DRE findings, comorbidities and individual risk factors such as increasing age, family history and Afro-Caribbean ethnicity. PSA level should not be used in isolation to guide clinician and patient decisions to biopsy.

Decision to perform a repeat biopsy

The NICE guideline[11] reviewed evidence supporting the efficacy of various prognostic factors when used to determine the need for further investigation in men with a negative initial biopsy. The recommendations are as follows:

Recommendation 1: a core member of the urological cancer multidisciplinary team should review the risk factors of all men who have had a negative first prostate biopsy, and discuss with the man that the risk of prostate cancer is increased if any of the following risk factors is present:

- the biopsy shows high-grade prostatic intraepithelial neoplasia (HGPIN)
- the biopsy shows atypical small acinar proliferation (ASAP)
- an abnormal DRE.

Recommendation 2: to consider multiparametric magnetic resonance imaging (mpMRI), using T2- and diffusion-weighted (DW) imaging, for men with a negative TRUS-guided 10- to 12-core biopsy, to determine whether or not another biopsy is needed.

Recommendation 3: do not offer another biopsy if the mpMRI, using T2-weighted and DW imaging, is negative, unless any of the risk factors listed in recommendation 1 are present.

However, in clinical practice, there may be considerable variation in the adherence to these recommendations.

Types of biopsy

Diagnosis usually relies on obtaining a biopsy for histopathological examination of prostate tissue. The prostate gland is situated deep in the pelvis and it is not easy to visualise. Needle biopsies of the prostate are obtained from the rectum under ultrasound control. The NICE guideline[11] recommends that prostate biopsies should be carried out following the procedure advocated by the Prostate Cancer Risk Management Programme (2006), 'undertaking a transrectal ultrasound (TRUS)[-]guided biopsy of the prostate' (p. 123).[20] This Programme advises that 'the prostate should be sampled through the rectum unless there is a specific condition that prevents this' and also that 'the scheme used at first biopsy should be a 10–12 core pattern that samples the mid-lobe peripheral zone and the lateral peripheral zone of the prostate only' (section 11, Biopsy Scheme, p. 5).

In the UK NHS these initial TRUS biopsies are usually carried out under local anaesthetic as an outpatient or day-case procedure.

Transrectal ultrasonography biopsies are poor at accessing, and hence detecting, anterior, apical and central lesions.[21] Foci of cancerous cells may therefore be missed. If an initial biopsy fails to detect cancerous cells and the clinician still believes that cancer may be present, one or more repeat biopsies may be performed.

The second biopsy may be another standard TRUS biopsy with 10–12 cores. However, more often, an increased number of samples are taken. Men may prefer to have a general anaesthetic when undergoing a second biopsy, especially if they found the experience of their initial biopsy to be uncomfortable and/or distressing. The biopsy options include:

- Saturation biopsy. A biopsy, which may be taken transrectally or transperineally, with an increased number of cores (minimum of 20).
- Template biopsy. 25–40 biopsy cores are taken transperineally using a template or grid to access more areas of the prostate, including anterior and apical zones. In the UK, this procedure is usually performed under general anaesthetic.
- Targeted biopsy. Information from a MRI is used to guide the biopsy to areas with disease (see *Clinical assessment plus magnetic resonance imaging*).

Prostate biopsies are recognised as being imperfect, and men with prostate cancer may have a negative prostate biopsy result. Prostate cancer detection rates vary by type of biopsy, number of cores taken and patient characteristics; published estimates are 14–22% for the initial biopsy, 10–28% for a second biopsy and 5–10% for a third biopsy.[17,22–24]

Prostate biopsies are painful and associated with side effects. Relatively common minor complications include haematospermia, haematuria and rectal bleeding which subsides after intervention, while major complications, which are comparatively rare, include prostatitis, fever, sepsis, urinary retention, epididymitis and rectal bleeding for longer than 2 days.[1]

Gleason score

A histopathologist reviews biopsy specimens. If cancerous cells are detected, the histopathology report includes the Gleason score;[25] the Gleason score is a measure of the aggressiveness of the tumour. The Gleason score[25] (range 2–10) describes the degree of abnormality of the tumour found in the biopsy. The higher the Gleason score, the more aggressive (and worse prognosis) the cancer.

The Gleason score[25] is calculated by first assessing (using a microscope) the biopsy specimen for the degree of abnormality in the prostate tissue, which is categorised as one of five different Gleason patterns. Gleason pattern 1 is the most differentiated and therefore the most favourable, and pattern 5 is the most disrupted and aggressive. Pattern 3 is the most common. The Gleason score is obtained by adding together the number of the most widespread pattern (primary grade) and the number of the second most prevalent pattern (secondary grade). If a tumour has patterns 3 and 2, the score would be 5. If the tumour has only one pattern, or less than 5% of a secondary pattern, the single pattern is added to itself (e.g. $3 + 3 = 6$). It is advised that the diagnosis of low-grade Gleason score 2–5 prostate carcinomas in the setting of needle biopsy should be made with extreme caution,[25] as such a diagnosis on final radical prostatectomy is proved wrong most of the time.[26] Recent consensus is that diagnosed prostate cancer must have a minimum score of 6.[27,28] Cancers with a Gleason score higher than 7 are considered to be aggressive.

Other reported abnormalities

Apart from cancerous cells, other abnormalities which may be reported on histopathology reports include:

- HGPIN. This is a premalignant change in glands which has been shown to be associated with increased risk of invasive cancer elsewhere in the prostate.
- ASAP. Atypical changes are present in cells but the pathologist is uncertain of their significance.

Clinically insignificant prostate cancer

The prognosis and natural history of prostate cancer vary with the extent of spread and grade of cancer at diagnosis. Clinically insignificant prostate cancer can be defined as a cancer which will not affect the patient during the natural course of his lifetime, meaning that he is likely to die from other causes.[29] The detection of these potentially clinically insignificant cancers on either initial or second biopsy is an

important issue and can lead to potentially invasive and unnecessary treatment as well as increased anxiety for men who live with a diagnosis of prostate cancer that may not affect their life expectancy.

There are a number of different definitions of the term 'clinically insignificant prostate cancer'. The definitions are based on observed survival rates after radical prostatectomy. These pathology-based definitions require that the disease is restricted to the prostate, with a Gleason score of 6 or less. In addition, some definitions include limits on the total tumour volume and/or largest individual tumour volume.[30,31] However, in clinical practice the challenge is to correctly identify men with clinically insignificant disease before any treatment or surgery, that is at diagnosis. There are several systems for predicting the risk of localised prostate cancer progressing.[32–34] However, recent data have suggested that these tools may be inaccurate[33] and the NICE guideline[11] includes a research recommendation for further research in this area.

Comparators

There are two main comparator pathways for men suspected of having prostate cancer whose initial biopsy result was negative or equivocal, as shown in *Box 1*.

Clinical assessment

Clinicians and patients may consider a number of factors to help inform decisions whether or not a second (or subsequent) biopsy should be undertaken. These include:

- DRE. This procedure involves a clinician inserting a finger into the patient's rectum to feel the prostate. The purpose is to identify any hard or irregular areas and to estimate the size of the prostate. A prostate gland with hard bumpy areas may suggest prostate cancer.
- PSA level. There are a number of different measures including:
  - total PSA (tPSA)
  - PSA density, the degree of elevation in relation to estimated prostate volume[35]
  - rising PSA levels, which can be expressed as PSA velocity (ng/ml increase over a time period)[36] or PSA doubling time.
- Patient's age. Prostate cancer is rare in men under the age of 50 years, and 86% of cases occur in men aged 65 years and over.[37]
- Family history. The family history of prostate cancer in first-degree relatives, such as father or brother, increases risk.[37]
- Nomograms. These are risk algorithms that combine multiple clinical and laboratory risk factors to create a cumulative risk score. Most nomograms aim to predict the likely course of a disease. However, some nomograms [e.g. risk calculator number four from the Prostate Cancer Research Foundation,[38,39] the Prostate Cancer Prevention Trial (PCPT)[40] and Montreal nomograms[41]] can predict the result of a biopsy in men suspected of having prostate cancer. It is not clear how often these tools are used to predict biopsy results in clinical practice, but they are used as a proxy for clinical decision-making in the research setting.

BOX 1 Comparator pathways

1. The use of established risk factors (including histopathology results of initial biopsy, PSA level and a DRE) to inform the decision to perform a second biopsy.
2. The use of established risk factors (including histopathology results of initial biopsy, PSA level and a DRE) followed by mpMRI to inform the decision to perform a second biopsy.

Clinical assessment plus magnetic resonance imaging

Clinical assessment may be combined with MRI when a repeat biopsy is being considered. MRI uses strong magnetic fields and radiowaves to form images of the body.[42] Standard anatomical imaging involves injection of a contrast agent and uses T2-weighted images to delineate the structures. The term mpMRI refers to the additional use of functional images including:

- Magnetic resonance spectroscopy imaging (MRSI) – MRSI or metabolic imaging which measures the concentration of various substances or metabolites within the body.
- Diffusion-weighted magnetic resonance imaging (DW-MRI) – DW-MRI is sensitive to the motion of water molecules in tissue and detects water.
- Dynamic contrast-enhanced magnetic resonance imaging (DCE-MRI) – dynamic contrast-enhanced MRI injects a different contrast agent. The uptake and washout of this contrast agent is increased in prostate cancer.

Results from a MRI scan can be used to decide whether or not to perform a repeat biopsy and/or to guide and target the cores taken during the biopsy. The role of MRI varies depending on the MRI facilities and radiological expertise available throughout the NHS. The exact role of MRI in guiding biopsies varies. In cognitive targeting, knowledge of the MRI scan result guides the freehand targeting of suspicious areas and requires no additional equipment. In direct MRI-guided biopsy, the biopsy is performed within the MRI tube. However, in fusion targeting, software is used to combine a pre-acquired MRI-derived target with real-time TRUS imaging to guide the biopsy.[43,44]

In current NHS practice, MRI may be prohibited for 6–12 weeks, or more, after a biopsy because of bleeding, as this can lead to imaging artefacts. This has important time implications for the diagnostic testing strategies involving MRI after a negative or equivocal initial biopsy and any subsequent treatment, and may lead to delays in investigation and treatment.

Clear definition of the interventions

PROGENSA prostate cancer antigen 3 assay

The PROGENSA PCA3 assay produced by Hologic Gen-Probe is an in vitro nucleic acid amplification test that is intended for the quantitative determination of *PCA3* messenger ribonucleic acid (mRNA) in urine. The *PCA3* gene (previously known as *DD3*) is overexpressed in prostate cancer cells and is, therefore, a potential biomarker for tumour cells. Prostatic cells are released into urine by prostatic massage, this leads to a general release of ribonucleic acid (RNA) and so the level of mRNA of another housekeeping gene is needed to correct for the overall level of prostatic cells in the urine. The gene which encodes PSA (*KK3* gene) has been selected as the housekeeping gene, as its mRNA expression is relatively constant in normal prostate cells with only a weak downregulation of PSA gene expression in prostate cancer cells. The PCA3 score report is a ratio of the *PCA3* mRNA copies/ml to PSA mRNA copies/ml multiplied by 1000. The score can be used as a continuous measure but studies[45–48] have used threshold scores of 20, 25 or 35 to identify men who are at higher risk of an underlying cancer. The manufacturers of the PCA3 assay have recommended a threshold score of 25, with values 25 and higher suggesting the presence of cancer and values under 25 suggesting the absence of cancer.[49]

The PCA3 assay requires 20–30 ml of first-catch urine after a DRE, which included a minimum of three strokes to each lobe of prostate. The manufacturers' documents[50,51] refer to the presence of prostatic cells in the urine and there is no literature to address whether the mRNA analysed in the urine samples is derived from prostatic cells or from prostatic secretions.

Urine must transferred within 4 hours to a transport specimen tube containing a urine transport medium that triggers lysis of prostatic cells and stabilises the RNA. The samples are then transferred to a laboratory within 5 days and are kept either at ambient temperature or frozen. Once at the laboratory, the samples can be kept for 14 days if stored at 2–8 °C, for 11 months if kept at –15 to –35 °C or for 36 months if kept below –65 °C. Samples may be subject to up to five freeze–thaw cycles.[51]

The PCA3 assay should be used with the Hologic Gen-Probe Direct Tube Sampling 400, 800 and 1600 molecular laboratory systems (Hologic Gen-Probe, Marlborough, MA, USA). It is not compatible with other analysers. The PCA3 assay is indicated[50] for use in conjunction with other patient information to inform the decision for repeat biopsy in men 50 years of age or older who have had one or more previous negative prostate biopsies and for whom a repeat biopsy would be recommended by a urologist based on current standard of care, before consideration of PCA3 assay results.

The PROGENSA PCA3 assay package insert[51] states that:

> *PROGENSA PCA3 assay should not be used for patients who are taking medications known to affect serum PSA levels such as finasteride (Proscar, Propecia), dutasteride (Avodart), and anti-androgen therapy (e.g. Lupron). The effect of these medications on PCA3 gene expression has not yet been evaluated.*[51]

Certain therapeutic and diagnostic procedures, including prostatectomy, radiation and prostate biopsy may affect the viability of prostatic tissue and, subsequently, an individual's PCA3 score. The effect of these procedures on assay performance has not yet been evaluated.

The assay has been granted US Food and Drug Administration (FDA) approval[52] and a Conformité Européenne (CE) mark for use in the European Union.

Beckman Coulter Prostate Health Index

The phi has been developed by Beckman Coulter to combine several different components of PSA, with the aim of creating a sensitive index of risk of prostate cancer. Total PSA is measured in the bloodstream where it occurs, both unbound [free prostate-specific antigen (fPSA)] and bound to other proteins (such as proteases). There is some evidence that the proportion of PSA that occurs unbound (%fPSA) is lower in men with cancer.[53,54] fPSA has been shown to include several isoforms, including [–2]pro-prostate-specific antigen (p2PSA), which is associated with cancerous cells. phi is calculated using the equation $(p2PSA/fPSA) \times \sqrt{tPSA}$;[55,56] p2PSA is the unique component of phi.

According to the manufacturer, the phi test is designed for prostate cancer detection in men aged 50 years and older, with tPSA levels between 2 ng/ml and 10 ng/ml and DRE findings that are not suspicious for cancer.[57] The phi score is a continuous measure. The manufacturer, however, suggests using three categories: 0–20 (low risk); 21–39.9 (moderate risk); and 40 and above (high risk). The manufacturer states that estimates of the risk of cancer being detected at biopsy are 8.7% for men with a phi score in the low-risk category, 20.6% for men in the moderate-risk category and 43.8% for men in the high-risk category.[58]

The phi score is not intended to be calculated using PSA or fPSA results from any other manufacturer's assay and the phi assay is compatible only with Beckman Coulter Access instruments (Access2, DxI600, DxI800, DxC600i, DxC680i, DxC800i, DxC880i; Beckman Coulter Inc., Brea, CA, USA). All PSA assays may be standardised to either the Hybritech or the World Health Organization (WHO) calibration with an approximate 22% difference in reported PSA levels (lower for WHO calibration).[59] It is important to use either the Hybritech or the WHO calibration consistently for PSA, fPSA or p2PSA measurements used in the phi calculation and to not mix measurement calibration systems.

The p2PSA molecule is not stable in coagulated blood. The manufacturer's draft pack insert[57] states that 'When left on a clotted sample at room temperature, the p2PSA concentration increases significantly after 3 hours, probably due to the degradation of other proPSA molecules'. However, the analyte is stable in serum at room temperature. Therefore, it is important that the serum sample is prepared (separated from the clot by centrifugation) within 3 hours of taking a blood sample. Blood taken for p2PSA specimens should be allowed to clot fully and the serum separated by centrifugation within 3 hours of collection. The serum can then be stored for 24 hours at 2–8 °C before assay or for up to 5 months at –20 °C or colder. Specimens requiring storage for longer than 5 months should be frozen at –70 °C.

Information provided by the manufacturer states that the effect of medication prescribed for benign prostate hyperplasia on the level of p2PSA is not known.[57] Specifically, the phi results cannot be interpreted in, and should not be offered to, patients receiving 5-α-reductase inhibitors medication.

The assay has been granted FDA approval[60] and a CE mark for use in the European Union.

Implementing PROGENSA prostate cancer antigen 3 assay and the Prostate Health Index testing in the NHS

Various practical issues will need to be considered before/when introducing these tests into the NHS. These include acceptability of the tests to patients and the need for a DRE before a urine sample is voided for PCA3 analysis. The stability of samples and any processing required before transport to the laboratory may pose logistic challenges to health services. The requirement that blood samples for p2PSA assay must be centrifuged and separated within 3 hours may mean that the blood sample must be taken at a hospital with laboratory facilities on site.

Place of the intervention in the treatment pathways

The intervention pathways considered in this report are summarised in *Box 2*.

BOX 2 Intervention pathways

1. The use of the PCA3 score/the phi alongside established risk factors (including histopathology results, PSA level and a DRE) to inform the decision to perform a second biopsy.
2. The use of the PCA3 score/phi alongside established risk factors (including histopathology results, PSA level and a DRE) to inform the decision to perform a mpMRI before second biopsy. If the mpMRI is positive, a second biopsy would be performed.
3. The use of the PCA3 score/phi alongside established risk factors (including histopathology results, PSA level and a DRE) to inform the decision to perform a second biopsy in men who have had a negative mpMRI.

Outcome measures

The aim of this review is to assess the impact of the use of two new tests (PCA3 assay and phi) on the health and well-being of men undergoing investigation for suspected prostate cancer and who had a negative or equivocal initial prostate biopsy. Analytical validity outcomes, diagnostic process outcomes, clinical outcomes and patient-reported outcomes can be useful when considering the impact of using PCA3 scores and phi, and are listed in *Box 3*. Further details of commonly used outcome measures to assess diagnostic tests are described in *Appendix 1*.

BOX 3 Outcome measures

Analytical validity outcomes

Pre-analytic variability.

Analytical specificity.

Analytical sensitivity.

Accuracy.

Precision.

Diagnostic process outcomes

Clinical validity/diagnostic test accuracy outcomes.

Test failure rate.

Time to TP diagnosis.

Number of repeat biopsies required.

Grade and stage of cancers detected.

Clinical outcomes

Morbidity and mortality from biopsies.

Morbidity and mortality from treatment of diagnosed cancer.

Adverse events from false test results including from treatment of clinically insignificant prostate cancer.

Health-related QoL.

Patient-reported outcomes

Patient anxiety associated with undergoing a biopsy (initial and repeated biopsies), waiting for diagnosis and living with the diagnosis of a clinically insignificant prostate cancer.

Patient distress and sequelae associated with the detection of clinically insignificant prostate cancer.

TP, true positive.

Methodological challenges

The External Assessment Group (EAG)'s review of clinical effectiveness has been designed to assess the incremental gain associated with the use of the PCA3 score or the phi in addition to standard clinical assessment (with or without MRI). The following issues pose challenges to achieving this aim.

Lack of evidence

Lack of long-term evidence

Ideally, clinical utility would be assessed in 'end-to-end' or 'test-to-treatment' studies and it would be possible to follow men from early clinical investigation through to diagnosis, treatment and long-term follow-up for prostate cancer. Such end-to-end studies of clinical utility are often not available. Published studies of clinical validity frequently focus on the diagnostic process and assess the performance of the different tests. Thus, although available studies provide some information on the effectiveness of the intervention tests, data describing the long-term impact of using new tests are often scarce.

Lack of clinically relevant comparisons

Many clinical validity studies focus on the use of (1) a new test or (2) a new test that is a replacement for an existing test. However, usually the comparator and intervention pathways involve combining multiple tests.

Study measurements of clinical assessment

In clinical validity studies, the intervention and comparator test pathways are compared with the results from the reference standard (biopsy). To assess the accuracy of the comparator pathways, the biopsy results must be available for men who 'test negative' on the comparator (e.g. clinical assessment or clinical assessment + MRI) as well as those who 'test negative' for the intervention test (e.g. clinical assessment + PCA3 or clinical assessment + MRI + PCA3). This means that the study design must include some form of clinical assessment of the entire study population and report the biopsy results for all participants, including those who tested positive or negative on clinical assessment. Differences in the methods used for clinical assessment may make comparing results from different studies problematic.

Heterogeneity in study populations and between-study comparisons

The target population is all men with a negative or equivocal initial prostate biopsy. It is, therefore, important to assess whether or not the study populations in the included studies are representative of this target population. There is likely to be some selection bias in the published studies because referral, or patient acceptance of a biopsy, is expected to be related to PSA level and/or abnormal clinical results; this means that the study populations are likely to be made up of men who are considered to be at higher than average risk of cancer.

Differences in the patient populations of published studies are likely to lead to considerable heterogeneity in estimates of diagnostic test accuracy. Any between-study comparisons that assume that tests perform equally in different populations may, therefore, give misleading results. Combinations of tests used in sequence are rarely reported in the literature and the reconstruction of such test pathways by combining summary measures for the various components assumes not only that the summary measures are constant across populations but also that the tests are independent.

Potential sources of bias

Sampling bias

Study recruitment may be restricted to, for example, men in the PSA 'grey zone' (i.e. with a PSA of 4–10 ng/ml) or to men with abnormal DRE findings in addition to a negative or equivocal initial biopsy. This means that the range of clinical assessment variables is restricted in the study population and hence the observed diagnostic accuracy of these clinical variables will be reduced. This sampling bias affects the generalisability of study findings to the population of interest to this review (i.e. all men with a negative or equivocal first biopsy).

Verification bias

Studies that consist of opportunistic cohorts of patients presenting at referral centres will include biopsy results for men who have been referred for, and have accepted, a repeat biopsy. Acceptance of biopsy is likely to be related to PSA level or PCA3/phi score; it is more likely that men with higher PSA levels will accept a repeat biopsy. This leads to so-called differential verification bias, when the availability of the reference standard result is dependent on the result of the intervention or comparator test.

Imperfect reference standard

In clinical validity studies the diagnostic accuracy of a new or intervention test is assessed against a reference standard. The reference standard is the best test available, that is the current preferred method of diagnosing a disease. In the case of prostate cancer, the reference standard is a biopsy. The diagnostic capabilities of all new tests need to be compared against the diagnostic accuracy of a biopsy.

In prostate cancer the reference standard (biopsy) does not detect all cancers and is considered to be imperfect. Some men with a negative biopsy result do have an undetected cancer. These men are indicated by *x* and *y* in the *Table 1*.

Different types of biopsy have different cancer detection rates and the sensitivity and specificity of the intervention pathways may therefore differ depending on the type of second biopsy that is carried out (see *Types of biopsy*). A different biopsy sampling scheme (such as saturation or extended) might mean that *x* and *y* are moved to the biopsy-positive column, as shown in *Table 2*.

The estimate of diagnostic accuracy for PCA3/phi scores will alter for different biopsy types if the proportion of *x*/false positive (FP) is not the same as the *y*/true negative (TN) value, that is if men in whom cancers were missed on a standard biopsy but would have been detected on a saturation biopsy are more likely or less likely than men with cancer detected on a standard biopsy to have raised PCA3/phi scores. This is plausible. For instance, a standard biopsy is more likely to detect widespread, rather than localised, cancers. If widespread cancers are also associated with higher PCA3/phi scores than localised disease, then a 'better' biopsy scheme which picks up more localised disease might reduce the diagnostic accuracy of the use of the PCA3 assay or phi.

TABLE 1 Biopsy results from standard biopsy

	Biopsy results (standard biopsy)	
Test result	**Prostate cancer**	**No prostate cancer**
Test positive	TP	FP (including *x*)
Test negative	FN	TN (including *y*)

FN, false negative; FP, false positive; TN, true negative; TP, true positive.

TABLE 2 Biopsy results from saturation biopsy

	Biopsy results (saturation biopsy)	
Test result	**Prostate cancer (positive)**	**No prostate cancer (negative)**
Test positive	TP + *x*	FP – *x*
Test negative	FN + *y*	TN – *y*

FN, false negative; FP, false positive; TN, true negative; TP, true positive.

Imaging used with biopsy: incorporation bias

A separate issue relating to biopsy is the type of imaging used. Using ultrasound or MRI to guide the biopsy in effect incorporates another test into the reference standard. Men with lesions detectable on ultrasound or MRI often have additional biopsy cores taken which have come from the area surrounding the identified lesions. This may well increase the chance of a positive biopsy result and so increase the observed diagnostic accuracy of MRI. However, this means that the type of reference standard used differs according to the MRI test result; if MRI is positive, more cores would be taken than if MRI was negative.

Chapter 2 Assessment of clinical effectiveness

Aims of the assessment of clinical effectiveness

Assessing the clinical effectiveness of the PCA3 assay and the phi in the diagnosis of prostate cancer involved three separate systematic reviews:

1. A review of the analytical validity of the intervention tests to assess how accurately the tests measure PCA3/phi level present in a sample. Analytical validity is the study of how well laboratory tests measure the substances they are intended to measure. As the p2PSA assay is the unique component of the phi, the analytical validity of the p2PSA assay was considered in this review. As the pre-analytical stability of samples may affect logistical issues concerning transport and storage before samples reach the laboratory for testing, this issue was also considered in the review.
2. A review of the clinical validity (diagnostic test accuracy) of comparator and intervention pathways to assess how the addition of the PCA3 assay or the phi might contribute to the diagnosis of prostate cancer.
3. A review of the clinical utility of the intervention test pathways to evaluate how the addition of the intervention tests might affect patient outcomes, including long-term outcomes such as mortality and morbidity from prostate cancer, and intermediate outcomes such as side effects from tests.

The methods used in each review followed the systematic review principles outlined in the Centre for Reviews and Dissemination guidance for undertaking reviews in health care,[61] the NICE *Diagnostics Assessment Programme Manual*[62] and publications from the Cochrane diagnostic test accuracy methods working group.[63] The review of analytical validity was informed by the principles outlined in the Agency for Healthcare Research and Quality methods guide[64] and the Evaluation of Genomic Applications in Practice and Prevention initiative.[65]

Analytical validity review

Search strategy: analytical validity review

Electronic databases

The following databases were searched on 28 April or 19 May 2014 for eligible studies:

- MEDLINE
- EMBASE
- Cochrane Central Register of Controlled Trials
- Health Technology Assessment (HTA) database
- Cochrane Database of Systematic Reviews
- Database of Abstracts of Reviews of Effectiveness
- ISI Web of Science
- Medion database for related diagnostic test accuracy reviews (www.mediondatabase.nl/)
- Aggressive Research Intelligence Facility database (www.birmingham.ac.uk/research/activity/mds/projects/HaPS/PHEB/ARIF/databases/index.aspx)
- PROSPERO systematic review register (www.crd.york.ac.uk/PROSPERO/).

No study design filters were applied and non-English-language reports were excluded. All databases were searched from 2000. The following types of report were excluded:

- editorials, opinion pieces and correspondence on journal articles
- conference abstracts.

Trial and research registers were searched on 24 July 2014 for ongoing trials and reviews including:

- ClinicalTrials.gov (http://clinicaltrials.gov/)
- *meta*Register of Current Controlled Trials and International Standard Randomised Controlled Trial Number (ISRCTN) Register (www.controlled-trials.com/mrct/)
- WHO International Clinical Trials Registry Platform (http://apps.who.int/trialsearch/).

Details of the search strategies used can be found in *Appendix 2*.

Searching other resources

Backward citation searching was undertaken on key review articles.[1,15,66–70] As in all searches, the US FDA website was searched for the following terms: PCA3, phi and p2PSA.

Study selection strategy: analytical validity review

Three reviewers (AN/AB/JH) independently screened all titles and abstracts identified via searching and obtained full-paper manuscripts that were considered relevant by any of the reviewers (stage 1). The relevance of each study was assessed (AN/AB/JH) in accordance with pre-specified inclusion criteria (stage 2). Studies that did not meet the criteria were excluded. Any discrepancies were resolved by consensus.

The analytical validity review focused on studies that addressed the ability of the intervention test to accurately and reliably measure the target analyte. Inclusion criteria are presented in *Table 3*.

Studies with precision or accuracy control data presented only as part of the methods section of a publication, in order to describe the test that was used, were not included in the review.

Data extraction and quality assessment strategy: analytical validity review

Data extraction and quality assessment were undertaken by two reviewers (AN/NF), with disagreements resolved by discussion. Data extraction included details of source population, number of samples, specific methods/platforms evaluated, number of positive samples and negative controls tested, as well as reported results. Quality assessment was informed by the checklist proposed by Teutsch *et al.*[65] and included the following:

- quality of description of test undertaken
- range of sample/study population tested representative of routine use
- definition of correct answer
- reporting of test failures.

A copy of the data extraction form used in the analytical validity review is included in *Appendix 3*.

TABLE 3 Inclusion criteria (analytical validity)

Item	Inclusion criteria
Patient population	All adult men
Intervention test	PCA3 assay or p2PSA or phi score
Outcomes	• Measures of consistency and accuracy between, and within, laboratories such as coefficient of variation • Sensitivity and specificity against external standard • Assay robustness • Test failure rate
Study design	All study designs including collaborative studies, external proficiency testing, peer-reviewed repeatability studies, internal reports and manufacturer data

Methods of data analysis/synthesis: analytical validity review

The design of the included studies and the types of outcomes reported were summarised in tabular form.

Results: analytical validity review

Search results

The results of the searches undertaken are summarised in *Figure 1*. A total of 2249 unique records were identified by database searching and via the use of additional resources (e.g. trial registers and backward citation searching). Of these, 2021 records were excluded at the title and abstract screening stage. Overall, 228 studies were reviewed in full text and six papers were considered to be relevant for inclusion in the analytical validity review.

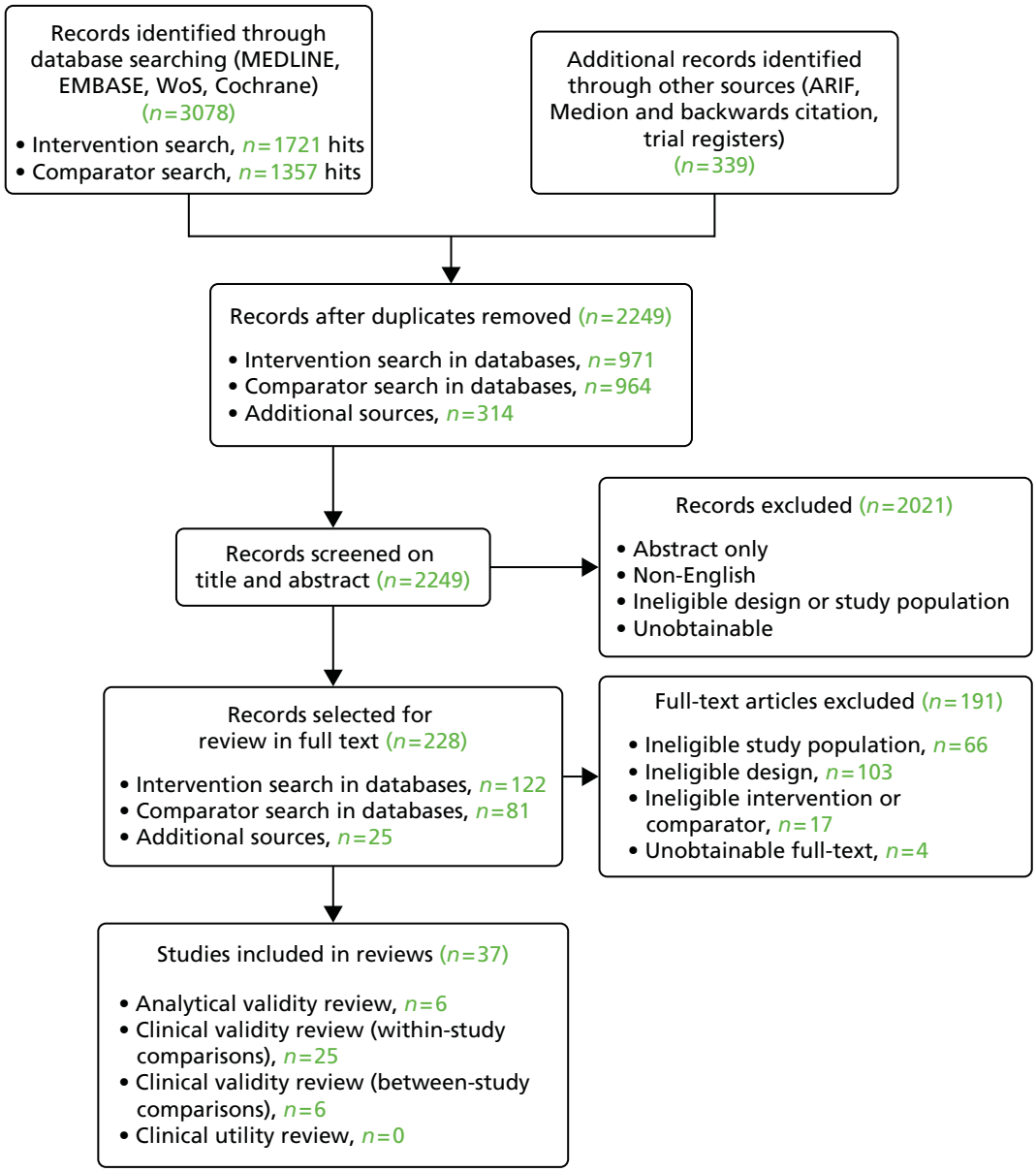

FIGURE 1 Flow chart of search results. ARIF, Aggressive Research Intelligence Facility database; WoS, Web of Science.

Six papers[48,71–75] reporting on analytical validity or pre-analytical effects were identified from the electronic databases: three related to the PCA3 assay[48,71,72] and three related to the p2PSA assay.[73–75] In addition, the *Summary of Safety and Effectiveness Data* (SSED) report to the US FDA for each test[50,58] was obtained. The manufacturers included a pack insert for the PCA3 assay[48] in their submission and the draft pack insert for the p2PSA assay was obtained from the FDA website.[57]

Data from two of the identified studies[73,74] for the p2PSA assay appear to be described in the SSED report[58] but, as citations were not stated, it was not possible to confirm this potential double use of data; the results from different data sources have been reported separately in the review and potential overlaps have been highlighted. The draft pack insert for p2PSA[57] does not appear to include any data that have not already been described in the SSED report[58] and so full data extraction using the draft pack insert[57] has not been undertaken; however, relevant data are reported in the study characteristics table for completeness.

A search of electronic trial databases found two trials[76,77] that were potentially relevant to the review of analytical validity of the PCA3 assay: one[76] was ongoing and the status of the other was unclear (*Table 4*). No relevant ongoing trials on phi or p2PSA were identified.[77]

Characteristics and quality assessment of included studies: analytical validity review

The study characteristics and outcomes reported in the included studies are summarised in *Appendix 4*. Quality assessment of all studies using the Teutsch *et al.* checklist[65] had been planned but, because of a lack of information in the included studies, it was not possible to use this checklist.[65] Instead, the EAG used a modified version of this checklist to assess the quality of the studies.

PROGENSA prostate cancer antigen 3 assay

All of the PCA3 studies[48,50,51,71,72] were conducted in the USA and reported data for clinical validity and analytical validity. All studies[48,50,51,71,72] measured precision. Four studies[48,50,51,71] measured accuracy, but only the SSED report[50] and pack insert[51] reported on five or more different outcomes that were all relevant to analytical validity. Pre-analytical effects were considered by all but one study.[72] None of the studies compared the PCA3 assay with a 'gold standard', as such a reference test does not exist for analytical validity. However, the PCA3 assay analyte quantitation was compared with in vitro transcripts which had been value-assigned using ultra violet spectroscopy in three studies.[48,50,71] Four of the studies[50,51,71,72] provided adequate descriptions of the test under study, that is reported specific methods/platforms evaluated and information on quality assurance measures. Although all studies used clinical samples, it was unclear whether or not the same population was used for both the analytical validity and clinical validity studies. Only in one study[72] did it appear that specimens represented routinely analysed clinical specimens in all aspects (e.g. collection, transport, processing). None of the studies provided sample size/power calculations, and the number of samples analysed varied by outcome both within and across studies (see *Appendix 4*, *Table 71*).

TABLE 4 Ongoing clinical trial relevant for analytical validity

Name	Details	Status	Registration number and URL
Comparing the Reliability of Expressed Prostatic Secretion (EPS) and Post Massage Urine (PMU) for the Prediction of Prostate Cancer Biopsy Outcome[76]	Randomised trial of expressed prostatic secretions vs. post-DRE urine in target population of 180 men undergoing first prostatic biopsy. Various biomarkers including PCA3 assessed in specimens	Ongoing	NCT01441687, http://clinicaltrials.gov/show/NCT01441687 (accessed 15 September 2014)
Pilot study: performance of the Progensa PCA3 test in post-oxytocin urine specimens[77]	To determine the yield of prostate cells and PCA3 score in the urine specimens from healthy male volunteers after oxytocin nasal spray, using a urine specimen with no manipulation as a reference method	Unclear	EUCTR2010-024649-61-NL, http://apps.who.int/trialsearch/Trial2.aspx?TrialID=EUCTR2010-024649-61-NL (accessed 15 September 2014)

[–2]pro-prostate-specific antigen assay

Studies of p2PSA were conducted in Germany[73,75] and in the USA;[74] one study[58] described results from studies that had been conducted in both of these countries. The manufacturers have confirmed that the assay that was research use only used in the study by Stephan *et al.*[75] is the same assay that is now commercially available. All studies reported analytical sensitivity, specificity, accuracy, precision, linearity and range. Pre-analytical effects were considered only in the SSED report to the FDA[58] and by Semjonow *et al.*[73] None of the studies compared the p2PSA test to a 'gold standard' as such a reference test does not exist for analytical validity. However, in the SSED report[58] recovery (a measure of accuracy) used internal reference preparation of p2PSA. Stephan *et al.*[75] was the only study that did not adequately describe the test under study, that is the study did not report the specific methods/platforms evaluated or present sufficient information on quality assurance measures. As with PCA3 studies, precise details about the population from which the samples were derived was not provided. The number of samples varied by outcome within and across studies (see *Appendix 4*, *Table 71*). None of the studies provided sample size/power calculations. In all instances, it was unclear if the specimens represented routinely analysed clinical specimens in all aspects (e.g. collection, transport, processing).

Prostate cancer antigen 3 results: analytical validity review

Impact of digital rectal examination

Sokoll *et al.*,[48] in a sample of 179 patients, found that 74.4% of urine samples taken before a DRE were informative, compared with 95.5% of urine samples that were taken after a DRE. First-morning void urine samples ($n = 56$) had an informative rate of 80.4%. The number of strokes per lobe in the DRE did not affect the informative rate of tests (98.7% for three strokes and 94.4% for eight strokes per lobe). There were no significant differences in the reported PCA3 score for those tests which were informative, regardless of whether or not the men had a prior DRE.

Storage of unprocessed urine samples

The SSED report[50] and pack insert[51] both described the effect of time spent at 30 °C and 2–8 °C on urine samples before processing into the transport tubes. The SSED report[50] included 12 specimens and the pack insert[51] reports on 10 specimens, and it is not clear whether or not these are the same samples. At 30 °C, the PCA3 score showed a 5% drift over 4 hours and at 2–8 °C a 2% drift over 4 hours. Estimates of drift are not presented for more than 4 hours' storage.[50]

Storage of processed urine samples

In Groskopf *et al.*,[71] three previously frozen processed urine specimens were thawed and held at 4 °C or 30 °C for 14 days. Degradation of mRNA was noted from day 1 at 30 °C; the PCA3 score remained within 20% of initial value for 14 days. The SSED[50] reported drift for 12 processed samples held for 6 days at varying ambient conditions between 30 °C and –70 °C and between 30 °C and 55 °C; both temperature ranges had a drift of 8%. The pack insert[50] reported that 12 specimens were stable for 21 days at 4 °C, for 5 days at 30 °C and for 90 days between –20 °C and –70 °C; no raw data were presented.

Groskopf *et al.*[71] and the pack insert[51] reported stable results in processed urine after five and six freeze–thaw cycles, respectively; neither of the studies presented raw data.

Analytical sensitivity

Limit of blank (LoB), limit of detection (LoD) and limit of quantitation (LoQ) were reported as shown in *Table 5*. The LoQ of both analytes (PCA3 and PSA) were the same as the corresponding LoD in Sokoll *et al.*[48] and in the SSED report.[50]

Analytical specificity

The assay did not detect unspliced *PCA3* RNA.[50,51] No assay interference was recorded in the SSED report,[50] with either 10 listed endogenous compounds or six micro-organisms; out of 27 exogenous compounds, only selenium and raw palmetto were reported to cause interference .[50] However, in the pack insert[51] it

TABLE 5 Analytical sensitivity: PCA3 assay

Study	Methods	Source	LoB (copies/ml)	LoD (copies/ml)	LoQ (copies/ml)
Sokoll 2008[48]	LoD – lowest measurable concentration of controls; LoB – 95th centile of zero calibrator; and LoQ – < 130% recovery and CV < 35%	PCA3	176	259	259
		PSA	831	2338	2338
SSED 2012[50]	Four blank female urine and four female urine spiked to calibrator 2 concentrations; LoD = LoB + 1.65 SD	PCA3	90	239	239
		PSA	254	3338	3338
Pack insert[51]	Diluted in vitro transcripts. LoQ assessed according to Clinical & Laboratory Standards Institute EP17-A	PCA3	NR	80	Calibrator 2 ≈ 750
		PSA	NR	1438	Calibrator 2 ≈ 7500

CV, coefficient of variation; NR, not reported; SD, standard deviation.

was reported that none of the 35 therapeutic substances tested (which included selenium and raw palmetto) interfered with the assay. In addition, the SSED report[50] states the effects of medications such as finasteride and dutasteride which are known to affect serum PSA levels were not evaluated. However, in the FDA pack insert[51] (p. 34) but not in the SSED report, these two drugs are clearly listed among the therapeutic substances tested. Nevertheless, the pack insert[51] states that 'The PROGENSA PCA3 assay should not be used for patients who are taking medications known to affect serum PSA levels such as finasteride (Proscar, Propecia), dutasteride (Avodart), and anti-androgen therapy (Lupron). The effect of these medications on *PCA3* gene expression has not yet been evaluated' (p. 31). Urine samples from men after a prostatectomy and from female participants were below the assay limit for PCA3.[51,71] RNA from 10 tissue types throughout the male urogenital tract was tested, and only prostate tissue RNA was detected in the assay.[51] The SSED report[50] included carryover studies with a 0% FP rate for negative samples interspersed with high-titre samples.[58]

Accuracy

No gold standard is available, and without a gold standard that offers 100% specificity and 100% sensitivity it is difficult to confidently assess the accuracy of competing diagnostic strategies. Four studies[48,50,51,71] assessed accuracy by calculating the percentage recovery of measured *PCA3* or *PSA* RNA copies/ml compared with ultra violet-determined copies/ml of female urine samples spiked with varying concentration of in vitro transcripts or with control samples. Across four studies,[48,50,51,71] accuracy varied from 90% to 118% for PCA3 and from 85% to 121% for PSA (*Table 6*).

TABLE 6 Per cent recovery: PCA3 assay

Study	Methods	Measurement	Minimum (%)	Maximum (%)
Sokoll 2008[48]	Three controls. Tested in two sites	PCA3	104.1	110.8
		PSA	93.2	108.8
SSED 2012[50]	Eight-member panel of female urine spiked with in vitro transcript	PCA3	90	118
		PSA	85	121
Pack insert[51]	Eight-member panel of female urine spiked with in vitro transcript	PCA3	94	108
		PSA	111	120
Groskopf 2006[71]	Three controls	PCA3	102	109
		PSA	94	111

Precision

Precision was assessed in four papers[48,50,51,71] by including only within-laboratory variation (including intra-run and inter-run variance, reagent, observer) and in two papers[50,71] by including both within- and between-laboratory variation. Multiple results were reported in some papers.[50,71] In six studies[48,50,51,71] the within-laboratory total coefficient of variation (CV)% ranged from 4% to 27% for PCA3 and from 7% to 18.7% for PSA; in two studies[50,51] the PCA3 score varied from 12% to 28% (*Table 7*).

Within- and between-laboratory total CV% was reported by two studies[48,50] and ranged from 5.9% to 17.2% for PCA3 and from 10.1% to 19.3% for PSA (*Table 8*). Only one study[50] reported within- and between-laboratory total CV% for the PCA3 score, which ranged from 12.3% to 25.0%. Most variation occurs within laboratory, with assays on different sites adding little extra variability.

Shappell *et al.*[72] reported between-laboratory precision from 50 clinical samples sent to two different laboratories in terms of concordance of PCA3 scores. When the PCA3 score was divided into three categories (indeterminate, < 35, ≥ 35), results were concordant in 47 out of 50 samples (94%). Correlation

TABLE 7 Total CV% for PCA3 assay: within-laboratory precision

Study	Methods	Total CV%: maximum and minimum		
		PCA3 (range)	PSA (range)	PCA3 score = PCA3/PSA × 1000 (range)
SSED 2012[50]	Four control samples Maximum of 80 results each sample Variation: within-run, between-run, day	5.2–18.3	9.5–18.7	14.0–20.7
Pack insert[51]	Three pooled samples and four control samples 36 results each sample Variation: within-run, between-run, operator, lot and equipment	7–27	9–14	12–28
Pack insert[51]	Three control samples 80 results each sample Variation: within-run, between-run, day	4–12	7–8	Not reported
Sokoll 2008[48]	Three control samples 100 results each sample Two different sites Variation: within-run, between-run	Within-run: 5.7–15.2 Between-run: 6.1–18.6	Within-run: 10.8–11.6 Between-run: 7.6–9.5	Not reported Not reported
Groskopf 2006[71]	Three control samples 54 results each sample Variation: within-run, between-run	6–19	8[a]	Not reported
Groskopf 2006[71]	Three patient samples 54 results each sample Variation: between-run	Total CV% not reported Between-run: 9–20	Total CV% not reported Between-run: 10–11	Total CV% not reported Between-run: 15–24

a Median presented as range not stated in text.

TABLE 8 Total CV% for PCA3 assay: within- and between-laboratory precision

Study	Methods	Total CV%: maximum and minimum		
		PCA3	PSA	PCA3 score = PCA3/PSA × 1000
SSED 2012[50]	Three control samples tested in three sites 360 results for each sample Variation: within-run, between-run, operator, lot and site	6.8–17.2	10.5–19.3	12.3–25.0
Sokoll 2008[48]	Three control samples 200 results each sample Variation: within-run, between-run, site	5.9–17.1	10.1–11.5	Not reported

was reported to be good for 48 informative samples ($r = 0.85$). When three outliers were omitted, this improved further ($r = 0.96$). The mean percentage difference in test values was 13.6% [standard deviation (SD) 42.5%].

The SSED report[50] also compared spiked female urine versus clinical samples (*Table 9*). Maximum variation in total CV% for PCA3, PSA and the PCA3 score appeared to be slightly less for clinical samples than for control samples. The sample precision in clinical samples was therefore reported to be comparable with that in control samples.

Linearity

The SSED report[50] assessed linearity using 11 samples of PCA3 and PSA in in vitro transcripts in processed female urine. Here the deviation from linearity for PCA3 was < 9% and for PSA deviation was < 7%. Linearity studies using 10 clinical specimens in specimen diluent or processed female urine were also reported. Deviation from linearity for PCA3 in specimen diluent and processed female urine was less than 6%. However, for PSA in specimen diluent, linearity was < 23% and in processed female urine deviation was < 30%. The higher than expected variance in PSA, although remaining within study acceptance criteria, may have been caused by variation in linearity panel preparation. The pack insert[51] reported a direct proportional relationship between dilutions tested and analyte copies/ml.

[–2]pro-prostate-specific antigen/Prostate Health Index: analytical validity review

Stability of [–2]pro-prostate-specific antigen in blood and serum

Semjonow *et al.*[73] examined the stability of 22 clinical samples stored as clotted blood at 21 °C or as serum at 4 °C and 21 °C and then frozen at –20 °C or –70 °C. Percentage recovery of the samples over time from each baseline measurement was reported. The stability criterion used was that mean change in recovery did not exceed 10%.

TABLE 9 Comparison of spiked female urine vs. clinical samples

Study	Methods	Total CV%: maximum and minimum		
		PCA3	PSA	PCA3 score = PCA3/PSA × 1000
SSED 2012[50]	Clinical specimens: 16 results for each specimen	6.8–12.5	8.0–15.9	8.3–16.0
	Control specimens: 16 results for each specimen	5.1–18.2	9.2–17.9	13.3–20.6

In clotted blood, the mean recovery at 1 hour 9 minutes was 105.6% [95% confidence interval (CI) 103.2% to 107.9%], compared with 100% at the 37-minute baseline. At 3 hours 1 minute, the mean recovery was 112.7% (95% CI 109.7% to 115.6%). These data show that, by 3 hours after drawing the blood sample, the stability criterion is not met. No data are available for storage times between 1 hour 9 minutes and 3 hours 1 minute and so it is not clear precisely when the stability criterion is breached. These data are the basis for the recommendation stipulated in SSED,[58] that is specimens should be spun and refrigerated within 3 hours. A regression equation extrapolated results to a baseline at 97% of time of specimen collection. The increase in value is considered to be because of proteolytic activity in the clot.

Samples in serum were within the stability criterion after 48 hours at either 4 °C or 21 °C and at least 12 months at –20 °C or –70 °C. Two freeze–thaw cycles did not result in < 10% variation compared with 21 °C.

Stability of reagents and calibration materials

The SSED report[58] included data confirming stability of reagents and calibration products, both as sealed packs and once opened.

Thermal sensitivity of assay

The effect of change in ambient temperature (18 °C, 23 °C and 31 °C) on assay performance was investigated for three different analysers (Access 2, DxI 800 and DxI 600; Beckman Coulter Inc., Brea, CA, USA) and reported in the SSED.[58] Results were compared with results at the centre-point ambient temperature. A thermal effect was noted, with 1.84–2.82% change in p2pSA per 1 °C ambient temperature. This suggests a maximum of 16.9% variation in p2PSA result for ± 6 °C change in ambient temperature compared with temperature at which the calibration curve performed.

Analytical sensitivity

Limit of blank, LoD and LoQ were reported as shown in *Table 10*. The results reported in the SSED[58] appear to be the same as the results reported in Sokoll *et al.*[74]

Analytical specificity

Potential interference with seven endogenous compounds was investigated by comparing test mean (with added compound) and control mean (without added compound) for three different concentrations of p2PSA.[74] Most recoveries were within 10% of the target 100%, with a mean of 93%, although the addition of 8.4 g/dl total protein reduced one recovery to 88.4%. The same seven compounds at the same concentrations were also reported to be analysed for interference.[58] The raw data (mean recoveries) were not reported, although a warning was given that protein levels greater than 8 g/dl may interfere with p2PSA measurement. It is unclear if the analyses reported in the SEED report[58] are the same as those reported by Sokoll *et al.*[74] Forty-nine commonly encountered medications and therapeutic drugs were also tested and the SEED report[58] concluded that they did not interfere with assay performance, although no raw data were presented.

TABLE 10 Analytical sensitivity: p2PSA assay

		p2PSA		
Study	Methods	LoB (pg/ml)	LoD (pg/ml)	LoQ (pg/ml)
SSED 2012[58]	LoB: 95th centile of zero analyte; LoD: LoB + 1.65 SD (SD from patient serum LoQ; dilutions of calibrators from LoD to 7 × LoD); LoQ: concentration with CV20% from quadratic model	0.50	0.69	3.23
Sokoll 2012[74]	Methods as for SSED. Appears to be same study results as in SSED	0.50	0.70	3.23
Stephan 2009[75]	LoD: repeat measurement of zero calibrator + 2 SD	Not reported	2.27	Not reported

Crossreactivity with other PSA isoforms, including α_1-antichymotrypsin-PSA, benign PSA, fPSA, (–4) PSA and (–5/–7) PSA, was tested in three studies.[58,74,75] Minimal cross-reactivity was detected (recovered test dose < 5%[58,74] or < 2.5%[75] of expected dose).

Carryover was reported by only one study[58] with no evidence of carryover from high concentration samples.

Accuracy

No gold standard is available and the reference material used is based on purified p2PSA. Accuracy was reported in three studies[58,74,75] by calculating the per cent recovery of measured p2PSA pg/ml in male serum samples spiked with varying concentration of purified p2PSA (*Table 11*). The data reported in SSED[58] and Sokoll *et al.*[74] appear to be from the same study.

Precision

Precision was assessed by including only within-laboratory variation (including intra-run and inter-run variance, reagent, observer) in three studies[58,74,75] and by including both within- and between-laboratory variation in one study.[58]

All studies reported CV% for p2PSA, but only the SSED report[58] included CV% for phi. Within-laboratory precision as measured by total CV% varied from 2.91% to 13.05% for p2PSA and from 8.5% to 12.0% for phi (*Table 12*). Within- and between-laboratory precision for p2PSA was reported as being between 5.39% and 9.39% for p2PSA and between 4.9% and 7.3% for phi (*Table 13*). These maximum estimates are lower than those for within-laboratory only precision, and it is likely that this reflects the different populations used. There appears to be an overlap in data for p2PSA variability between Sokoll *et al.*[74] and the SSED report.[58] Sokoll *et al.*[74] reported within-laboratory precision data from four sites, but one of these had higher than expected variability and no total variance across all was reported in this paper (see *Table 12*). The variance for p2PSA across three sites reported in the SSED report[58] may be from the three lower variance sites (see *Table 13*).

Linearity

The SSED[58] reported linearity studies using diluted known concentrations of p2PSA in 12 serum samples. Eleven out of the 12 samples had a slope of 1.0 ± 0.15. A linear range was confirmed to 4922 pg/ml. Sokoll *et al.*[74] assessed dilutions of three samples and Stephan *et al.*[75] assessed six samples; both confirmed a linear range to 5000 pg/ml and 4500 pg/ml, respectively. No hook effect to 15,000 pg/ml was found in the two studies.[58,74]

Discussion: analytical validity review

To inform the assessment of the analytical validity of the two assays, the EAG has relied on data that have been published, primarily by the manufacturers, in the form of pack inserts and/or reports included in submissions for regulatory approval. The EAG could not reject the premise that, for some results, the same analytical validity data had been reported in multiple publications. The EAG considered that the analytical validity of both the PCA3 and the p2PSA assays had been comprehensively documented. The EAG identified several areas where further consideration of the data might be merited for both the PCA3 assay (e.g. precision, single threshold) and the phi (e.g. sample handling and thermal sensitivity).

TABLE 11 Per cent recovery: p2PSA assay

		p2PSA		
Study	Methods	Minimum (%)	Maximum (%)	Other
Stephan 2009[75]	Six spiked patient serum samples	92.75	102.55	Regression range = 0.959–1.055
SSED 2012[58]	Six spiked patient serum samples	90	96	Mean 93%
Sokoll 2012[74]	Six spiked patient serum samples	90	96	–

TABLE 12 Total CV% for p2PSA and phi: within-laboratory precision

Study	Methods	Total CV%: maximum and minimum	
		p2PSA	phi
SSED 2012[58]	*For p2PSA*: three controls and six clinical samples Variation: within-run, between-run *For phi*: one control and four clinical samples Variation: within-run, between-run, day, lot, analyser	2.94–10.83	8.5–12.0
Sokoll 2012[74]	Three control and three clinical samples 80 runs each sample Variation: within-run, between-run, operator Four different sites reported separately – not combined. One site higher than expected variance	2.91–13.05[a] (2.91–7.10, with high variance site excluded)[a]	Not reported
Stephan 2009[75]	Four control and/or three control and one pooled clinical sample Variation: within-run, between-run	Total CV% not reported: within-run, 2.03–5.63; and between-run, 3.1–7.99	Not reported

a Data possibly same as reported in SSED.[58]

TABLE 13 Total CV% for p2PSA and phi: within- and between-laboratory precision

Study	Methods	Total CV%: maximum and minimum	
		p2PSA	phi
SSED 2012[58]	*For p2PSA*: three control and three clinical samples 80 runs each sample Three sites combined[a] Variation: within-run, between-run, reagent lot, site *For phi*: 10 clinical samples. Variation: within-run, between-run, day, site	5.53–9.39[a]	4.9–7.3

a Data possibly same as reported in Sokoll *et al.*[74]

PROGENSA prostate cancer antigen 3 assay

The analytical validity review has identified an important issue regarding the precision of the measurement of the PCA3 assay. Across the included studies, the CV% was estimated as being up to 25% for combined between- and within-laboratory variation and 28% for within-laboratory variation. Using a CV% of 25% means that, in a urine sample with a true PCA3 score of 25, the SD of the results obtained will be 6; this means that 67% of samples tested will have PCA3 scores between 19 and 31 and the remaining 33% will have PCA3 scores outside of this range. This uncertainty in the true PCA3 score is reflected in the SSED[50] report, which includes the following guidance: 'Due to normal assay variability, specimens with PCA3 Scores near the cut-off of 25 (i.e. 18 to 31) could yield a different overall interpretation of POSITIVE or NEGATIVE upon repeat testing. PCA3 Scores in the range from 18 to 31 should therefore be interpreted with caution' (p. 6). The consequences of this imprecision for the use of the PCA3 assay in routine NHS clinical practice are unknown.

There are no concerns regarding the stability of samples during storage once the samples have been processed. However, urine samples need to be transferred into specialist transport tubes within 4 hours of the urine being voided.

None of the papers included in the analytical validity review explored whether or not genotype affected PCA3 scores. However, the authors of a recent publication[78] have proposed that a single threshold for the PCA3 score may not be appropriate for all men and that multiple thresholds may be required, as the appropriate threshold may vary by genotype. This publication[78] did not meet the inclusion criteria for the analytical validity review. However, in this genome-wide association study[78] of the Reduction by Dutasteride of Prostate Cancer Events (REDUCE) trial population, two genotypes which were associated with PCA3 scores were identified. The study population included 278 subjects with prostate cancer detected on biopsy and 1371 without prostate cancer. The means of the PCA3 scores in the 1371 men with negative prostate biopsy varied from 13.35 to 20.76 depending on genotype. One of the genotypes (rs10992994 in the β-microseminoprotein gene) is a strong genetic marker for prostate cancer.[79,80] The authors calculated a personalised threshold score by adjusting the threshold of 35 by the relative genetic effect; the estimated personalised threshold scores varied between 24.9 and 60.6. Whether or not a single threshold for the PCA3 score is appropriate for all men with suspected prostate cancer is currently unknown.

[–2]pro-prostate-specific antigen/Prostate Health Index assay

Practical issues relating to the use of the p2PSA assay that may be important to consider are sample handling and thermal sensitivity. The draft pack insert[57] states that blood should be centrifuged and serum separated within 3 hours of the blood sample being taken; this guidance is based on the work of Semjonow *et al.*[73] However, the data in this paper suggest that by 3 hours 95% of samples will have breached the stability criterion of a 10% increase in p2PSA level. As neither the manufacturer nor Semjonow *et al.*[73] present a rationale for the use of the 10% stability criterion or a time period of 3 hours, the consequences of breaching the 3-hour time period are not clear. In addition, whether or not sample handing can be carried out in routine clinical practice as per the instructions set out in the draft pack insert[57] is not yet known; in particular, given the 3-hour time limit, only hospitals with on-site laboratory facilities may be able to offer this test.

Studies of the thermal sensitivity of the p2PSA assay indicated that there is a 16.9% variation in p2PSA result for a ± 6 °C change in ambient laboratory temperature.[58] However, the SSED[58] report suggests that any differences in results because of temperature change would not affect clinical validity results.

Clinical validity review

Search strategy and study selection strategy: clinical validity review

The same search strategy and study selection strategy were used for the analytical validity and clinical validity reviews. Full details are presented in *Search strategy: analytical validity review* and *Study selection strategy: analytical validity review*.

Inclusion criteria: clinical validity review

Comparisons between the performance of the intervention tests (PCA3 assay and phi) and the comparison tests (clinical assessment and MRI) can be made using either data from studies carried out in the same study population (within-study or direct comparisons) or data from studies in which intervention and comparator tests are carried out in different populations (between-study or indirect comparisons). The preferred data for this review are derived from within-study comparisons of intervention and comparator test pathways.

Within-study (direct) comparisons

Owing to uncertainty about the diagnostic pathways used in NHS clinical practice and the limited availability of MRI facilities, the EAG initially included all studies with a direct comparison of the PCA3 assay or the phi with any one or more of following component comparator tests:

- individual clinical risk factors such as age, a DRE
- standard clinical judgement/nomograms
- PSA levels
- MRI results: T2-weighted magnetic resonance imaging (T2-MRI)/DW-MRI.

As the intervention tests (PCA3 assay or phi) can be used as replacement, add-on or triage tests to the comparator tests, studies that have directly compared the clinical validity of the PCA3 assay with the clinical validity of phi, with or without other comparators, were also included. The inclusion criteria used to select eligible within-study comparisons are presented in *Table 14*.

Systematic reviews for use in between-study (indirect) comparisons

In the absence of any available within-study comparisons, the EAG would have considered carrying out between-study (indirect) comparisons of the intervention tests versus comparator tests. Given the probable large number of studies evaluating each of the intervention and comparator tests, estimates of the clinical validity of the intervention and comparator tests from good-quality systematic reviews with meta-analyses were sought to provide data for any between-study (indirect) comparisons undertaken. The inclusion criteria used to select eligible systematic reviews are presented in *Table 14*.

Data extraction strategy: clinical validity review

A paper-based data extraction form was created for the clinical validity review (see *Appendix 3*). These forms were revised after data had been extracted from three studies. Three reviewers (AN/JH/KD), who worked independently, extracted relevant data and the data extraction forms were cross-checked (AN/JH/KD). When more than one publication reported findings from a single study, a composite data form was created. In cases where reported data appeared to be missing or unclear, clarification was sought from study authors.

Within-study (direct) comparisons

Limited data [e.g. details relating to the comparator interventions, study population (including inclusion and exclusion criteria) and author conclusions] were extracted from studies that were eligible for inclusion but did not report data from a clinically relevant comparison, that is limited data were extracted from studies that reported the results of univariate PCA3 or univariate phi versus univariate PSA.

Complete data were extracted from all other eligible studies. Particular attention was paid to:

- how the intervention and comparator tests were used (replacement, add-on, triage or not stated)
- definition of positive biopsy, including grade and stage of tumour detected
- threshold values used for intervention tests.

The available data on all reported clinical validity outcomes were recorded including:

- 2 × 2 tables of true positive (TP), FP, false negative (FN) and TN values
- sensitivity, specificity, positive predictive value, negative predictive value, positive and negative likelihood ratios
- area under the curve (AUC) and sensitivity and specificity values derived from receiver operator characteristics (ROC) curves
- multivariate odds ratios (ORs) for logistic regression.

TABLE 14 Inclusion criteria (clinical validity – direct and indirect studies)

Item	Inclusion criteria
Patient population	Men suspected of having prostate cancer who had had at least one negative or equivocal biopsy. The review was restricted to studies where at least six cores were taken in initial biopsy. Studies of men taking medications known to affect serum PSA levels such as finasteride (Proscar®, Merck Sharp & Dohme Ltd; Propecia®, Merck Sharp & Dohme Ltd), dutasteride (Avodart®, GSK), and anti-androgen therapy or leuprorelin (Lupron, Takeda-Abbott Pharmaceuticals) were excluded
Intervention	Diagnostic test or test pathway including PCA3 and/or phi
Comparator	Diagnostic test or test pathway without PCA3 or phi and including one or more of following comparator tests: • individual clinical risk factors such as age or a DRE • standard clinical care/nomogram • PSA levels • MRI results: T2-MRI/DW-MRI Studies that directly compared the performance of PCA3 with that of phi, with or without other comparators, were also included
Reference standard	Eligible studies compared the performance of comparator or intervention pathways to a histological analysis of prostatic tissue. This could have been obtained from a second prostatic biopsy or from a prostatectomy specimen. Biopsy must have taken place within 1 year of the intervention test Studies with all types of second biopsy were included: • repeat standard TRUS-guided biopsy • saturation • template • MRI-targeted biopsies • use of prostatectomy specimens
Outcomes	Studies reporting any of the following were included: • estimates of the intervention or comparator test (means and SD, proportion positive) in men with positive and negative results on second biopsy • specificity and sensitivity for different threshold points of PCA3, phi or PSA • comparison of AUC for different tests or test combinations • gain in sensitivity and specificity estimates by adding intervention test as derived from ROC curves • results of logistic regression analyses • test failure rate • adverse effects of test or subsequent biopsies • risk group and stage of cancers detected
Study design	Studies reporting within-study comparison of interventions/comparators: • Paired design. Cross-sectional or longitudinal studies in which intervention test(s), comparator test(s) and reference standard test were performed in the same group of people • Unpaired design. Trials in which people were randomised to either the intervention or comparator test(s) and then all received the reference standard test Studies for inclusion in between-study comparisons of interventions/comparators: • systematic reviews with meta-analyses of the clinical validity of the intervention or any of the comparator tests

AUC, area under the curve; ROC, receiver operating characteristic.

Outcomes were recorded for every reported:

- threshold value
- combinations or sequence of tests
- grade of cancer.

Systematic reviews for use in between-study comparisons

Study extraction was limited to:

- details relating to the comparator interventions
- study population (including inclusion and exclusion criteria)
- number of studies and participants included in meta-analyses
- author conclusions.

Quality assessment strategy: clinical validity review

Quality assessment was not undertaken for studies which were eligible for inclusion but did not report data from a clinically relevant comparison. Quality assessment was not undertaken for systematic reviews for use in between-study comparisons as only the conclusions from these papers were included in the review.

The Quality Assessment of Diagnostic Accuracy Studies – version 2,[81] a modified version of the Quality Assessment of Diagnostic Accuracy Studies tool,[82,83] was used to assess the quality of included studies. This tool considers four domains: patient selection; index and comparator tests; reference standard; and flow and timing. These domains were assessed both for risk of bias (whether the conduct or design of the study led to a distortion of results) and for applicability issues (whether or not the study reflected the population and tests used in practice). The tool content was tailored to meet the requirements for this review and a copy of the tool is displayed in *Appendix 3*. The following issues were of particular importance to this review:

- Patient selection: the extent to which the study population was pre-selected on variables such as PSA level, a DRE, ethnicity or family history. This is important as it affects both risk of bias for these variables and applicability.
- Intervention test (PCA3 assay or phi): whether or not the tests were conducted and interpreted without knowledge of other comparators and of the reference standard; whether or not any lack of blinding posed an important risk of bias given the automated and objective nature of the test; whether or not thresholds used were determined in advance or selected to maximise the diagnostic power of the test; and whether or not the conduct of the test in the study was comparable to that used in standard clinical practice.
- Comparator test [clinical and PSA (variables included in the multivariate analyses or nomogram were considered)]: whether or not the assessment was independent of, and blinded to, the results of the intervention tests, MRI and the reference standard; whether or not attempts to standardise assessment were carried out; whether or not methods used were a fair reflection of clinical practice.
- Comparator test (MRI): whether or not a definition of abnormality was given and whether or not the radiologist interpreting the scan was blind to results of intervention tests, MRI and the reference standard.
- Reference standard (biopsy): whether biopsy cores taken were standardised; or whether number and pattern of cores were affected by results such as clinical findings, TRUS result or MRI.

Methods of data analysis/synthesis: clinical validity review

Extracted data, grouped by type of outcome, were tabulated for each comparison. Measures of difference between the comparator test pathways were calculated for the following measures:

- comparison of AUCs
- sensitivity at set values of specificity
- specificity at set values of sensitivity.

Odds ratios from multivariate logistic regression analyses were recorded as a measure of the independence of the effect of the intervention tests.

The following sensitivity analyses were considered:

- type of second biopsy (saturation, template or guided)
- threshold value used for intervention test
- different risk groups (grades or stages) of tumour detected by the second biopsy.

Within-study comparisons: search results

The results of the searches undertaken are summarised in *Figure 1*. A total of 2249 unique records were identified by database searching and via the use of additional resources (e.g. trial registers and backward citation searching). Of these, 2021 records were excluded at the title and abstract screening stage because of ineligible study population (e.g. initial biopsy population only) or ineligible design. If the study population was unclear or a mixed biopsy population was reported in the abstract, the studies were retained and the full text obtained. Similarly, studies in which the design was unclear were retained. Studies were not excluded at this stage if comparators were not mentioned; comparisons with PSA and other clinical variables are not always highlighted in the abstract. Overall, 228 studies were reviewed in full text and 25 papers were considered to be eligible for inclusion in the review of clinical validity.

Clinical trials search results

A search of electronic trial databases found one ongoing randomised trial[84] that was possibly relevant for the clinical validity review of the PCA3 assay; summary details of this trial are shown in *Table 15*. No relevant ongoing trials that included phi were identified.

Excluded studies

The studies excluded from the review at stage 2 (including any listed in the manufacturer submissions which were not eligible for inclusion in the review) are listed in *Appendix 5* with the reasons for their exclusion. This list contains both excluded primary studies and excluded systematic reviews and meta-analysis papers. The most frequent reason for exclusion was ineligible or unclear study population.

Within-study comparison results: clinical validity review

A total of 25 papers[45,46,85–106] met the inclusion criteria for the within-study comparisons. A total of 21 papers[45,46,85,86,88–92,94–99,101–106] were identified which reported within-study comparisons between clinical assessment + PCA3 and/or clinical assessment + phi versus a comparator.

The results from 17 papers[45,46,85,86,89–92,94,96,97,99,102–106] reporting 15 different study populations were included in the review; results from two study populations were published in two publications each (European cohort[46,85] and the REDUCE trial[86,105]). Full data extraction and quality assessment were undertaken on these papers. Three other publications[88,98,101] from the European cohort study and one additional publication by Pepe and Aragona[95] were eligible for inclusion in the review but did not present additional study results and are included in the number of eligible studies ($n = 21$) for information only.

TABLE 15 Ongoing clinical trial relevant for clinical validity review

Name of trial	Details	Status	Registration number, URL
Medical Economics of Urinary PCA3 Test for Prostate Cancer Diagnosis[84]	Randomised trial of men undergoing prostate biopsy. Intervention group will have PCA3 results available and control group will not. Outcomes include number of inappropriate biopsy, costs of management	Ongoing. Estimated completion 2021	NCT01632930, http://clinicaltrials.gov/show/NCT01632930 (accessed 15 September 2014)

Four papers[87,93,100,107] reported only univariate assessments of the PCA3 assay or phi versus univariate PSA and the limited data extracted from these studies are presented in *Appendix 6*.

The 17 included studies[45,46,85,86,89–92,94,96,97,99,102–106] reported various comparisons of intervention and comparator tests and these are listed in *Table 16*.

Intervention pathways

Three intervention pathways are of interest to this review; these pathways are repeated here from *Box 2*.

TABLE 16 Summary of comparisons reported

Number	Comparison	Studies reporting data on comparison
1	Clinical assessment vs. clinical assessment + PCA3	European cohort[46,85]
		REDUCE placebo[86,105]
		Perdonà 2011[97]
		Gittelman 2013[45]
		Busetto 2013[90]
		Scattoni 2013[102]
		Porgpiglia 2014[99]
		Pepe 2013[96]
		Goode 2013[91]
		Wu 2012[106]
		Bollito 2012[89]
2	Clinical assessment vs. clinical assessment + phi	Stephan 2013[104]
		Lazzeri 2012[92]
		Scattoni 2013[102]
		Porpiglia 2014[99]
3	Clinical assessment + MRI vs. clinical assessment + MRI + PCA3	Busetto 2013[90]
		Porpiglia 2014[99]
4	Clinical assessment + MRI vs. clinical assessment + MRI + phi	Porpiglia 2014[99]
5	Clinical assessment + PCA3 vs. clinical assessment + phi	Scattoni 2013[102]
		Porpiglia 2014[99]
6	Clinical assessment + MRI + PCA3 vs. clinical assessment + MRI + phi	Porpiglia 2014[99]
7	Clinical assessment vs. clinical assessment + PCA3 + phi	Scattoni 2013[102]
		Porpiglia 2014[99]
8	Clinical assessment + PCA3 vs. clinical assessment + MRI	Busetto 2013[90]
		Porpiglia 2014[99]
		Panebianco 2011[94]
9	Clinical assessment + PCA3 vs. clinical assessment + MRI + PCA3	Sciarra 2012[103]
10	Clinical assessment + phi vs. clinical assessment + MRI	Porpiglia 2014[99]

BOX 2 Intervention pathways

1. The use of the PCA3 score/phi alongside established risk factors (including histopathology results, PSA level and a DRE) to inform the decision to perform a second biopsy.
2. The use of the PCA3 score/phi alongside established risk factors (including histopathology results, PSA level and a DRE) to inform the decision to perform mpMRI before second biopsy. If the mpMRI image is positive a second biopsy would be performed.
3. The use of the PCA3 score/phi alongside established risk factors (including histopathology results, PSA level and a DRE) to inform the decision to perform a second biopsy in men who have had a negative mpMRI.

Of the three intervention pathways described, data are available from the included studies to address only the first pathway. As the results of tests are most often presented as outputs from logistic regression models, it is not possible to determine from the data available whether or not carrying out the diagnostic tests in one order is better than carrying out the tests in a different order. Nor is it possible to determine whether diagnostic accuracy is improved if the PCA3 assay (or phi) test is carried out before or after a MRI. Therefore, there are no included studies which explicitly address the second or third pathways.

Within-study comparisons: baseline characteristics

The study characteristics of the 15 included study populations (17 papers) are summarised in *Table 17*. The EAG did not group the studies by PCA3 or phi as some studies included both the PCA3 assay and the phi, and often the tests were assessed using different combinations of tests and different criteria for assessment within a single publication.

Study designs and populations

Fourteen studies[45,46,85,86,89–92,94–97,99,102–106] were observational cohort studies and one was a randomised controlled trial (RCT).[103] Eleven studies were of a prospective cohort design,[45,46,85,86,89,90,92,94,97,99,102,103] three studies[91,95,96,106] were of a retrospective cohort design and one study[104] was of mixed design. None of the studies was based in the UK; the relevance of the included study results to UK clinical practice is, therefore, uncertain.

In all but one trial, the study population was made up of men who had been referred for repeat prostatic biopsy for clinical indications; the exception was the REDUCE study,[86,105] in which men were participants in the placebo arm of a clinical trial.

The criteria for referral for repeat biopsy were often unclear and differed across studies. Some studies were restricted to men with normal[89,94,103] or abnormal DREs.[96] The terms 'positive DRE' and 'negative DRE' were often used, and we assumed that 'positive' meant abnormal and 'negative' meant normal. The percentage of men with abnormal DRE scores in the studies therefore varied from 0% to 100%.[47,96] The mean (or median) age of study populations was between 60[94] and 67 years.[45,89,102] Mean or median PSA, when stated, ranged from 4.8[91] to 11.0 ng/ml.[106] Seven studies recruited only men with PSA scores within the grey zone of PSA (PSA levels between 4 and 10 ng/ml).[47,86,90,94,96,97,105,107] The prevalence of cancer detected on repeat biopsy varied from 11.4%[76] to 68.3%.[94]

Recruitment to the placebo arm of the REDUCE trial[86,105] did not rely on referral for repeat biopsy. Men aged 50–75 years were recruited to the main REDUCE trial on the basis of increased PSA levels (2.5–10 ng/ml) and a negative initial biopsy.[86,105] Participants were then scheduled to receive a repeat biopsy at 2 and 4 years regardless of clinical indications. A subsample of these men, based on whether or not the trial centre was able to process urine samples for PCA3 testing, were included in this study.[86] However, this study excluded all biopsy results that were indicated by abnormal clinical assessment, such as rising PSA or an abnormal DRE, and used only the results from the biopsies that were mandated by the trial protocol. This study population is, in effect, the reverse selection of the clinically selected population seen in most studies and represents a low-risk population.

TABLE 17 Within-study comparisons: study characteristics

Study	Study design	Manufacturer funding/financial interest	*n*	Selection criteria	Age (years)	PSA (ng/ml)	Abnormal DRE (%)	Type of repeat biopsy (% positive)
European cohort (Ankerst 2008[85])	Prospective cohort; six centres in five European countries	Yes	443	Men scheduled for repeat biopsy with one or two previous negative biopsies (≥ 6 cores performed at ≥ 3 months prior to enrolment)	Median 66.0 (range 11–83)	Median 7.0 (range 0.3–85)	18.7	Minimum 10 cores from peripheral zone (27.8%)
European cohort (Haese 2008[46])		Yes	463		Mean 64.4 (SD 6.6)	Mean 8.9 (SD 7.6)	19.0	Minimum 10 cores from peripheral zone (27.6%)
REDUCE trial – placebo arm (Aubin 2010[86])	Prospective cohort of patients from placebo arm of REDUCE trial. International multicentre trial	Yes	1072	Selection to trial based on PSA level of 2.5–10 ng/ml < 60 years, 3–10 ng/ml > 60 years and a negative initial biopsy. Selection for this study depended on trial site being able to process urine sample for PCA3. Only routine scheduled biopsies used	NR	Range 0.30–33.9	NR	10-core TRUS (17.7%)
REDUCE trial – placebo arm (Tombal 2013[105])		Yes	1024		Mean 65.5 (SD 6.0)	Mean 6.4 (SD 3.0)	3.5	10-core TRUS (17.9%)
Perdonà 2011[97]	Prospective cohort; two centres in Italy	No	84	Men referred for prostate biopsy because of abnormal PSA and/or suspicious DRE. No PSA > 10 ng/ml	Median 66.0 (IQR 60–72)[a]	Median 6.7 (IQR 5.0–9.0)[a]	22.5[a]	> 12-core TRUS. Median 12 core (IQR 12–16) (34.5%)
Panebianco 2011[94]	Prospective cohort; one centre in Italy	No	41	Men with first random TRUS-guided prostate biopsy negative for prostate adenocarcinoma or high-grade PIN; persistent elevated PSA levels (tPSA ≥ 4 ng/ml and < 10 ng/ml) and a negative DRE	Mean 60.3 (range 48–69)	Mean 6.37 (range 4–10)	0	10-core TRUS + three additional cores if MRSI suspicious (68.3%)
Bollito 2012[89]	Prospective cohort; three centres in Italy	No	509	Men receiving PCA3 test and referred for repeat biopsy based on persistent PSA elevation. Men with a positive DRE or ASAP on initial biopsy excluded	Median 67 (range 42–89)[a]	Median 6.7 (range 2.5–48)[a]	0	14–18 peripheral and transition zone core (24.2%)

continued

TABLE 17 Within-study comparisons: study characteristics (*continued*)

Study	Study design	Manufacturer funding/financial interest	*n*	Selection criteria	Age (years)	PSA (ng/ml)	Abnormal DRE (%)	Type of repeat biopsy (% positive)
Sciarra 2012[103]	RCT in Italy. Men randomly assigned (1 : 1) to PCA3 only (arm A) or PCA3 plus MRI (arm B) before repeat biopsy	No	168	Men with first negative prostate biopsy to cancer and HGPIN, persistent tPSA > 4 ng/ml and a negative DRE	Given by study arm A/B. A: mean 63.2 (SD 7.1); B: mean 64.1 (SD 7.4)	A: mean 6.9 (SD 2.1); B: mean 7.1 (SD 3.5)	0	10-core TRUS + MRI targeted in MRI arm if positive (32.7%)
Wu 2012[106]	Retrospective cohort; one centre in USA	No	103	Indications for repeat prostate biopsy were based on a suspicious DRE, persistently elevated PSA, previous suspicious histology (such as HGPIN or ASAP) and/or patient preference	Mean 63.5 (SD 7.4)	Mean 11.0 (SD 5.5)	13	TRUS, at least 12 cores. Additional cores taken if TRUS abnormal (35.9%)
Busetto 2013[90]	Prospective cohort; one centre in Italy	No	163	Men with first random TRUS prostate biopsy negative for prostate carcinoma or high-grade PIN; persistent elevated PSA levels (tPSA ≥ 4 ng/ml and < 10 ng/ml)	Mean 66.4 (SD 5.3)	Mean 6.8 (SD 1.6)	29.4	10-core TRUS + two extra cores if MRI abnormal (41.7%)
Gittelman 2013[45]	Prospective cohort in 14 centres in USA	Yes	466	Participants were men without prostate cancer and with one or more previous negative prostate biopsies who were recommended by their physician for repeat biopsy	Mean 67.0 (SD 8.1)	Mean 7.0 (SD 5.6)	27.3	TRUS, 12 cores or more (21.9%)
Goode 2013[91]	Retrospective cohort; one centre in USA	Not clear	167	Men with no known personal history of prostate cancer who underwent a prostate biopsy because of an elevated PSA level, abnormal DRE, or abnormal previous prostate biopsy PIN or ASAP	Median 66 (range 41–90)[a]	Mean 4.8 (range 0.1–54.2)[a]	NR	12 cores (11.4%)

Study	Study design	Manufacturer funding/financial interest	*n*	Selection criteria	Age (years)	PSA (ng/ml)	Abnormal DRE (%)	Type of repeat biopsy (% positive)
Pepe 2011,[95] 2013[96]	Retrospective cohort; one centre in Italy	Not clear	100	All men had negative family history, a negative DRE, PSA 4.1–10 ng/ml or 2.6–4 ng/ml. All Caucasian	Median 66	Median 7.9 (range 3.7–10)	100	Transperineal saturation biopsy, median 30 (range 24–38) cores; US (28%)
Scattoni 2013[102]	Prospective cohort; two centres in Italy	Yes	95	Indication for repeat biopsy, ASAP, plurifocal HGPIN, PSA 2–15 ng/ml and/or a positive DRE	Mean 67.7 (SD 7.3)	Mean 9.8 (SD 3.9)	8.5[a]	Saturation TRUS, 14–24 cores, mean 18.7 (SD 3.2) cores (31.5%)
Porpiglia 2014[99]	Prospective cohort; one centre in Italy	No	170	Negative initial biopsy, 12 cores. Persistently elevated PSA levels, and/or a positive DRE	Median 65 (IQR 60–70)	Median 6.9 (IQR 5.2–9.8)	7.6	TRUS, 18/24 cores depending on prostate volume, blind to MRI results (30.6%)
Stephan 2013[104]	Described as both case–control and cohort. Patients enrolled prospectively and retrospectively; four centres in France and Germany	Yes	280	Men scheduled for a repeated biopsy, tPSA level 1.6–8.0 ng/ml (WHO calibration)	Negative biopsy: median 61 (95% CI 59 to 63); positive biopsy: median 64 (95% CI 63 to 66)	Negative biopsy: median 5.2 (95% CI 4.9 to 5.6); positive biopsy: median 5.5 (95% CI 5.1 to 6.0)	Negative biopsy: 7; positive biopsy: 20	TRUS, 10–22 cores (41%)
Lazzeri 2012[92]	Prospective cohort; one centre in Italy	Yes, analysers and reagents only	222	Men in whom a first biopsy was negative but in whom suspicion of prostate cancer persisted and who were scheduled for repeat biopsy in accordance with the European Association of Urology guidelines of increasing and/or persistently elevated PSA, a suspicious DRE, ASAP and HGPIN	Mean 63.9 (SD 7.1)	Median 7.6 (range 0.3–46.4)	14	24-core saturation, TRUS. Median 20 (range 12–26) cores (31.9%)

IQR, interquartile range; NR, not reported; PIN, prostatic intraepithelial neoplasia; RCT, randomised controlled trial; US, ultrasound guided.
a Characteristics of entire study population not just repeat biopsy population. Means reported if both means and medians given in article.

The details of the reference standards used in studies were poorly reported. An added complication was the fact that the number of biopsy cores taken often differed across patients within a study. Among cases for which details were provided, 10- or 12-core biopsies were the most common. Three studies[96,99,102] exploring the efficacy of the PCA3 score and all four studies[92,99,102,104] exploring the efficacy of phi used saturation biopsies or reported that the mean or median number of cores taken was 20 or more. The repeat biopsy was usually performed transrectally under ultrasonography guidance, taking 10–20 cores. Two studies described the repeat biopsy as saturation biopsies,[96,102] and in one of these the biopsy route was transperineal[96] and in the other transrectal.[102]

Within-study comparisons: quality assessment

The results of the Quality Assessment of Diagnostic Accuracy Studies – version 2 assessment[81] are presented in *Figures 2* and *3*, with the full assessments documented in *Appendix 7*.

Patient selection

Risk of bias from patient selection was assessed as being unclear for 10 studies,[45,46,85,86,90–92,97,99,102,105,106] as the type of clinical pre-selection operating was not explicit, and, therefore, it was impossible to assess how this might have biased the assessment of the clinical variables within the diagnostic models. Four studies[89,94,96,103] were assessed as having a high risk of bias because of pre-selection on a DRE or other clinical variables, and these studies[89,94,96,103] also had high concerns regarding applicability of patient selection. The study by Stephan *et al.*[104] was assessed as having a high risk of bias from patient selection as, in this mixed prospective and retrospective study, 29% of patients were excluded from the analyses because it was unclear whether the biopsy was initial or repeat.

Intervention tests

The majority of studies provided no details of whether the intervention tests (PCA3 assay or phi) were conducted with or without knowledge of other important considerations, for example results of comparator tests or the reference standard. Studies that did record this information included those by Gittelman *et al.*[45] and Goode *et al.*,[91] the REDUCE trial placebo arm[86,105] and the study by Stephan *et al.*[104] However, given the objective nature of the intervention tests, all studies were assessed as having a low risk of bias. As all study authors reported using the PROGENSA PCA3 assay (or, in the case of four studies,[85,86,105,106] this was confirmed by the manufacturer) and/or the Beckman Coulter assay systems, we had little concern regarding the applicability of the intervention tests.

Comparison tests

The risk of bias and concerns regarding applicability arising from the comparison tests were considered separately for clinical assessment, including PSA levels, and for MRI. Clinical assessment, including PSA levels, was often poorly described with no criteria given for an abnormal DRE. It was often not clear when the data used for clinical assessment had been recorded in relation to the timing of the biopsy or the intervention tests and if these had been collected or recorded by the analyst without knowledge of intervention tests or reference standard results. In multicentre studies there was no description of how clinical variables were standardised across centres. All studies were assessed as having an unclear risk of bias from the clinical comparator tests.

Concern regarding applicability of clinical assessment was assessed on the variables used in the multivariate models, algorithm or nomogram. Studies which pre-selected patients who had undergone DREs[89,94,96,103] (which meant that a DRE was not included in the model) were marked as being of high concern, as this does not reflect clinical assessment in routine practice. Most studies, including studies of phi, included standard PSA measures such as tPSA or fPSA in the clinical comparator model and added p2PSA to the intervention test model. However, Porpiglia *et al.*[99] and Goode *et al.*[91] did not include PSA in the clinical comparator model, and so we had a high degree of concern about the applicability of these study results to UK clinical practice.

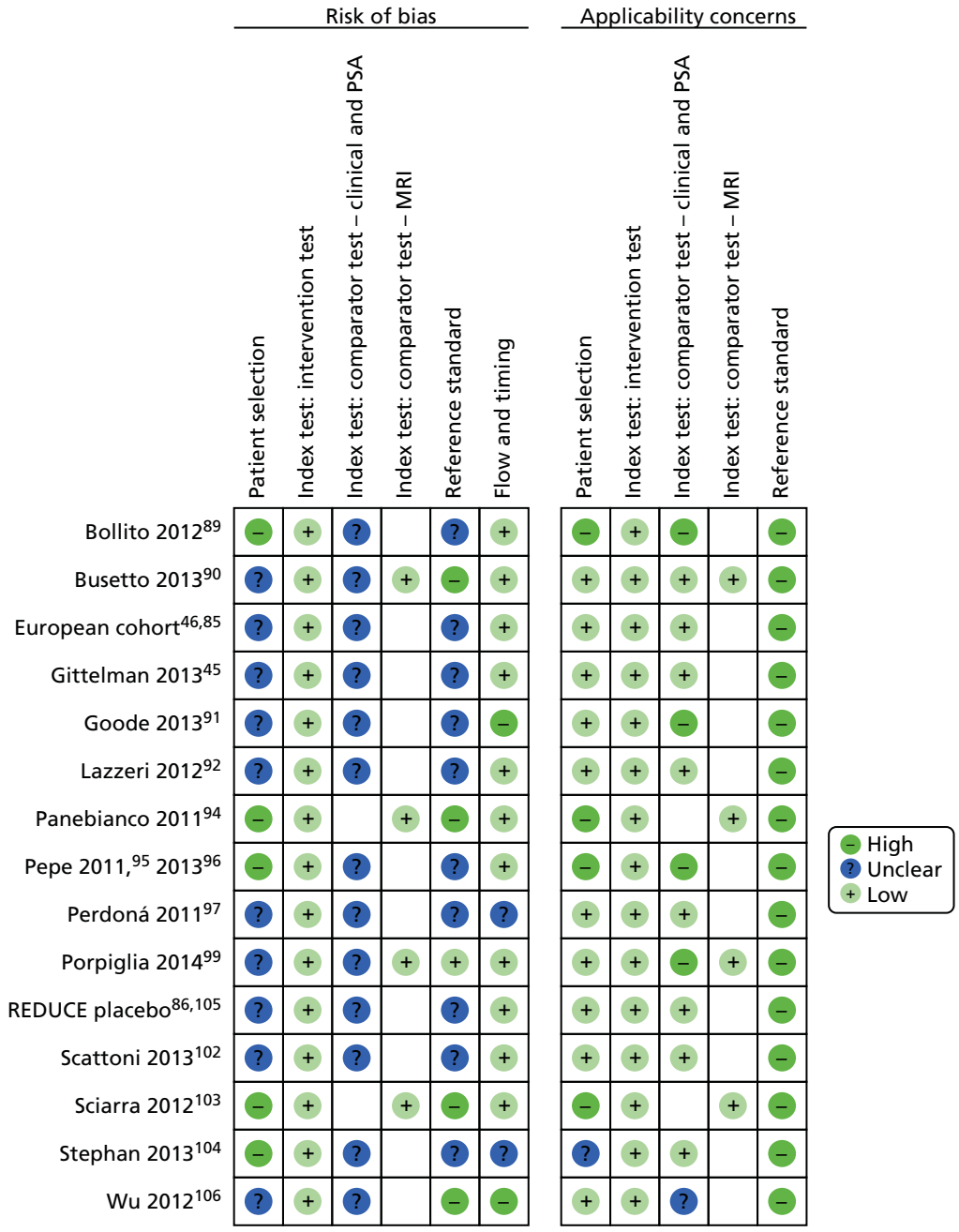

FIGURE 2 Risk of bias and applicability concerns summary: review authors' judgements about each domain for each included study population. Empty cells shown where study did not report this comparison.

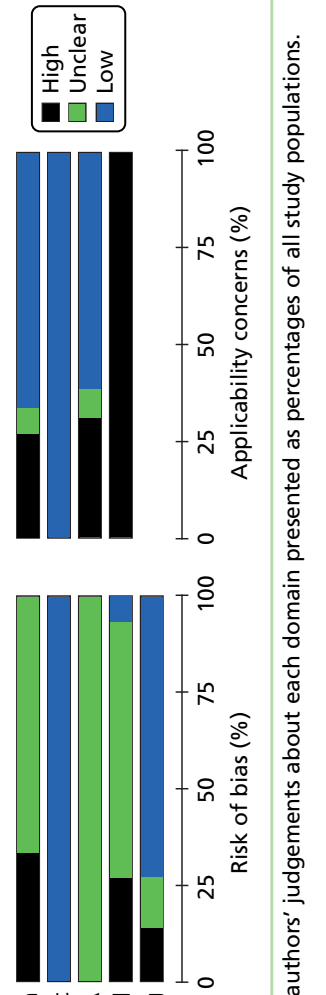

FIGURE 3 Risk of bias and applicability concerns graph: review authors' judgements about each domain presented as percentages of all study populations.

Four studies[90,94,99,103] which compared the PCA3 assay or phi with MRI were all assessed as having a low risk of bias and concerns regarding applicability of the MRI were low. MRI was either performed before repeat biopsy[90,94,103] or the radiologist was blinded to the biopsy result.[99] Diagnostic criteria were described in Porpiglia *et al.*[99] and Panebianco *et al.*[94] mpMRI with T2-imaging, DW imaging and DCE-MRI were performed in Porpiglia *et al.*[99] and mpMRI with T2-imaging, MRS, DW imaging and DCE-MRI were performed in the other studies.[90,94,103]

Reference standard

The reference standard involves two procedures, both of which are prone to bias: the targeting and selection of the biopsy cores and the pathology reporting of the cores taken. In many studies,[45,46,85,86,91,97,105,106] the reporting of the details of type and pattern of repeat biopsy performed was poor. In seven studies,[46,85,89,96,97,102,104,106] the number of cores taken varied, indicating that the number and locations of cores taken were affected by clinical findings such as a DRE or TRUS. In two studies,[92,97] although the methods specified a set number of cores to be taken, the number of cores actually taken was reported as a range. In addition to variation in the number and site of biopsy cores taken, pathology reporting is a potential source of bias. In three multicentre studies,[45,46,85,104] pathology reporting was not centralised, with potential for differences in biopsy processing protocols and pathology reporting. By contrast, in the REDUCE study,[86,105] all cores were processed at a single central laboratory. Four studies indicated that the pathologists were blinded to the clinical status of the patient,[45,91,92,99] and one study reported blinding to MRI.[103] Owing to these uncertainties, eight studies[45,46,85,86,89,96,97,102,104,105] were assessed as having an unclear risk of bias. If it was clear that additional cores were taken because of abnormalities identified on MRI[90,94,103] or TRUS,[106] the study was assessed as having a high risk of bias. The study by Porpiglia *et al.*[99] had a low risk of bias, as it was stated that the biopsies taken were not affected by MRI results or biomarker results and that a constant number of cores (depending on prostate volume) and pattern were taken.

In all studies, the applicability of the reference standard used was an area of high concern. Although the TRUS prostate biopsy is the usual method of diagnosing prostate cancer, this type of biopsy is inaccurate and often misses cancers.

Five studies had funding from, or financial links to, companies which produced the assays.[45,46,85,86,102,104,105] In another study, the manufacturer had supplied reagents.[92]

Overall, the results of the quality assessment exercise revealed that none of the studies was free from the risk of bias. The main areas for concern were related to the applicability of the study populations, variation in clinical assessment and whether or not choice and use of the reference standard were linked to previous clinical results. None of the studies was carried out in the UK and so the results of the studies were not directly relevant to NHS decision-makers.

Within-study comparisons: outcome measures reported

Results were most frequently reported using multivariate logistic regression models using AUC statistics, ROC curves, multivariate ORs, and derived sensitivity and specificity values. In these logistic regression models a range of clinical variables were entered into the models separately and were sometimes formulated as a nomogram,[85,91,97] sometimes using Bayesian methods.[85] These analyses relied on probabilities generated by the statistical model to classify patients as at risk or not, and generated ROC curves by varying the threshold probability. Intervention tests were added to the baseline clinical models either as a continuous variable or as a binary variable dichotomised at the reported threshold. Studies reported several models using intervention variables continuously and then dichotomously, or used different threshold values to create a dichotomous variable. Where appropriate, these models have been entered separately into the results tables. The EAG notes that only one study[105] presented independent sensitivity and specificity estimates.

Two studies[96,105] used the clinical variables that were incorporated into a decision tool that classified patients as test positive or negative rather than based on a continuous risk score. In the study by Tombal *et al.*,[105] 'best clinical judgement' was based on expert recommendations which had been formulated using a RAND/University of California, Los Angeles (UCLA) appropriateness model in a previous study.[108] Using variables of life expectancy, a DRE, prior biopsy, prostate volume and PSA score, recommendations for biopsy for each study participant (in the placebo arm of the REDUCE trial) were classified as appropriate or uncertain versus inappropriate. The PCA3 score was then incorporated into the decision tool, grouped into the following score levels: < 20, 20–34, 35–50 and > 50. As the decision tool combined all the variables and produced an overall assessment of test positive or not, conventional sensitivity and specificity were reported.

In Pepe and Aragona,[96] a case-finding protocol was used to identify the study population and was tested within the population; results for the case-finding protocol were not included in the results as the case definition altered between analyses. Results for the PCPT risk calculator in this study population are included in results.

Six studies[90,92,97,99,102,106] reported results using decision curve analysis. Decision curve analysis calculates the net benefit of a diagnostic model by subtracting the harm of unnecessary biopsies from the benefit of diagnosed cases of prostate cancer. Unlike the conventional trade-off between sensitivity and specificity, in decision curve analysis there is an attempt to weight the relative harms and benefits using the threshold probability of cancer at which the patient or clinician will opt for a biopsy. Further details describing this analysis method are summarised in *Appendix 1*. The net benefit of various diagnostic models was presented for threshold probability of cancer between 0% and 70%, with all studies reporting results between 10% and 40% threshold probability. The graphs of the decision curve analysis reported in the included studies are presented in *Appendix 8*.

Seven studies[45,46,85,86,89,90,96,103,105] reported diagnostic accuracy results for the PCA3 assay for the detection of more aggressive cancers – usually based on a Gleason score of 7 or higher. In six studies,[45,46,86,89,90,103,105] the authors employed univariate analyses and showed the ability of the PCA3 score to predict a Gleason score of 7 or higher. Only one study[105] reported how the use of the PCA3 score in combination with clinical assessment contributed to the prediction of more aggressive cancers. Only one study[92] considered the relationship between phi and the Gleason score. The results of these analyses are presented in *Within-study comparisons: Gleason score*.

Within-study comparisons: definition of clinical assessment

There was considerable variation in the definition of clinical assessment used in the included studies. Three studies used the PCPT nomogram,[85,96,97] which includes age, ethnicity, a DRE, prostate volume and PSA family history. One study[97] used the Chun nomogram,[109] which includes age, ethnicity, PSA, a DRE previous biopsies and prostate volume. In other studies, the base model of a series of logistic regression models was taken to represent clinical assessment. For one study[99] this included only age and a DRE, and for another[89] age and PSA alone, but for others a wider range of clinical risk factors were included, such as PSA, prostate volume, family history and ethinicity.[45,46,85,86,91,92,97,102,104,105] Studies of phi differed according to whether or not tPSA and/or fPSA were included in the clinical assessment definition. Porpiglia *et al.*[99] did not include PSA in the definition of clinical assessment, whereas other studies[92,102,104] included PSA.

Tombal *et al.*[105] used a clinical decision algorithm which had been developed using RAND/UCLA appropriateness methods and by consulting 12 European urologists. This included measures of life expectancy, a DRE and previous biopsy prostate volume. Pepe and Aragona[96] used a case-finding protocol to identify their study population and also included this measure in some analyses, but these results are not included in this review because of differences in definition.

Within-study comparisons: study results

The order in which the results of the comparisons (*Tables 18–27*) are presented reflects the relevance of the results to health-care professionals who are likely to use the tests in routine clinical practice in the NHS. The EAG considers the four most clinically relevant comparisons to be:

- clinical assessment versus clinical assessment + PCA3
- clinical assessment versus clinical assessment + phi
- clinical assessment + MRI versus clinical assessment + MRI + PCA3
- clinical assessment + MRI versus clinical assessment + MRI + phi.

Comparison 1: clinical assessment versus clinical assessment + PROGENSA prostate cancer antigen 3 assay

Area under the curve

Eight AUC results were reported from six study populations[45,46,85,86,90,99,102] for the comparison of clinical assessment versus clinical assessment + PCA3; one study[86] reported the results from two models, one using the PCA3 score as a continuous variable and one employing a threshold value of 35. Results from the same study population were reported in two separate papers.[46,85] The studies showed an increase in discrimination of between 1% and 19% when the PCA3 score was added to the clinical assessment model, either as a continuous or binary variable.

In addition, two studies[91,97] reported AUC results only for models of clinical assessment + PCA3, and these results were similar to the AUC results reported in other studies; Goode *et al.*[91] reported an AUC of 0.61 for a multivariate logistic regression model and Perdonà *et al.*[97] reported an AUC of 0.74 for the Chun nomogram and an AUC of 0.74 for the PCPT nomogram.

Multivariate odds ratios for PROGENSA prostate cancer antigen 3 assay

Five studies[45,86,89,99,106] reported seven multivariate ORs for clinical assessment + PCA3. Four studies[45,86,89,99] presented statistically significant results (ORs were above 1 and CIs did not include 1). One study had an OR above 1 with a CI that included 1.[106] Haese *et al.*[46] reported that the multivariate OR for the PCA3 score was significant ($p = 0.006$) in the model but did not report the effect size. These results are consistent with the AUC results and indicate that the addition of the PCA3 score to the clinical assessment model increases discrimination. Two studies[86,106] reported ORs for PCA3 using the PCA3 score as a continuous variable; in the remaining studies[45,86,89,99] various different thresholds were used to divide the PCA3 scores into a dichotomous variable.

Sensitivity and specificity

Only one study[105] presented independent sensitivity and specificity estimates. In this study, the addition of PCA3 scores to best clinical judgement reduced sensitivity from 75% to 66% and increased specificity from 26% to 71%. In this population (prevalence of all cancers = 17.9%) adding the PCA3 score to clinical assessment meant that 18 cancers would have been missed and 371 biopsies would have been avoided compared with clinical assessment alone. However, when the analyses were repeated for cancers with a Gleason score of 7 or higher (prevalence = 5.4%), the addition of PCA3 the score increased sensitivity from 75% to 85% and specificity from 26% to 67%, meaning that six more cancers would have been detected and 395 biopsies would have been avoided compared with clinical assessment alone.

Derived sensitivity and specificity

Pepe and Aragona[96] reported sensitivity and specificity for various risk thresholds in the logistic regression model. At a 25% risk threshold, models of the PCPT nomogram alone and of the PCPT + PCA3 had 100% sensitivity and low specificity (1% and 8%, respectively). Using a 40% risk threshold, the model with PCPT alone had 75% sensitivity and 26% specificity, whereas PCPT + PCA3 had 85.8% sensitivity and 25% specificity. This study population comprised Caucasian men with an abnormal DRE and negative family history; the diagnostic power of the PCPT was therefore likely to be reduced.

TABLE 18 Clinical assessment vs. clinical assessment + PCA3

AUC					
	Clinical assessment		Clinical assessment + PCA3		Difference and *p*-value if given
Study	Variables included	Result	Threshold	Result	
[a]European cohort (Haese 2008)[46]	Age, prostate volume, DRE, tPSA and %fPSA	0.67	Continuous	0.71	+0.04; $p < 0.001$
[a]European cohort (Ankerst 2008)[85]	PCPT nomogram: family history, number of previous biopsies, DRE and PSA	0.65 (95% CI 0.59 to 0.71)	Continuous	0.70 (95% CI 0.64 to 0.75)	+0.04; $p < 0.05$
REDUCE placebo (Aubin 2010)[86]	Age, family history, prostate volume, PSA and %fPSA	0.72 (95% CI 0.68 to 0.76)	Continuous	0.75 (95% CI 0.71 to 0.79)	+0.04; $p = 0.0009$
Scattoni 2013[102]	Age, DRE, prostate volume, tPSA and fPSA	0.75 (95% CI 0.64 to 0.87)	Continuous	0.76 (95% CI 0.64 to 0.88)	+0.01; $p = 0.719$
Busetto 2013[90]	Age, DRE and PSA	0.55 (95% CI 0.46 to 0.64)	Continuous	0.74 (95% CI 0.66 to 0.82)	+0.19; $p = 0.0002$
Porpiglia 2014[99]	Age and DRE	0.62 (95% CI 0.53 to 0.72)	Continuous	0.69 (95% CI 0.60 to 0.78)	+0.06
Gittelman 2013[45]	Age, family history, race, number of previous biopsies and DRE	0.65	25	0.74	+0.09 (95% CI 0.04 to 0.14); $p = 0.0007$
REDUCE placebo (Aubin 2010)[86]	Age, family history, prostate volume, PSA and %fPSA	0.72 (95% CI 0.68 to 0.76)	35	0.74 (95% CI 0.70 to 0.78)	+0.02; $p = 0.0558$
Goode 2013[91]	Age, DRE, prostate volume, race and family history	–	Unclear	0.61	–
Perdonà 2011[97]	PCPT nomogram: age, race, PSA, family history, DRE and previous biopsies	–	Continuous	0.74 (95% CI 0.63 to 0.83)	–
Perdonà 2011[97]	Chun nomogram: age, PSA, DRE, previous biopsies and prostate volume	–	Continuous	0.74 (95% CI 0.64 to 0.83)	–

Multivariate OR for PCA3					
	Clinical assessment			Clinical assessment + PCA3	
Study	Variables included			Threshold	Result
REDUCE placebo (Aubin 2010)[86]	Age, family history, prostate volume, tPSA and %fPSA			Continuous	1.02 (95% CI 1.01 to 1.02)
Wu 2012[106]	DRE, TRUS, PSA and PSAD			Continuous	1.02 (95% CI 1.00 to 1.03)
Gittelman 2013[45]	Age, family history, race, number of previous biopsies, DRE and tPSA			25	4.56 (95% CI 2.65 to 7.84)
Porpiglia 2014[99]	Age and DRE			Unclear	3.88 (95% CI 1.27 to 12.95)
REDUCE placebo (Aubin 2010)[86]	Age, family history, prostate volume, tPSA and %fPSA			35	2.65 (95% CI 1.86 to 3.79)
Bollito 2012[89]	Age, PSA and %fPSA			39	9.44 (95% CI 5.15 to 17.31)
Bollito 2012[89]	Age, PSA and %fPSA			50	9.29 (95% CI 5.11 to 16.89)
Sensitivity and specificity					
	Clinical assessment			Clinical assessment + PCA3	
Study	Variables included	Sensitivity (%)	Specificity (%)	Threshold	Sensitivity (%)	Specificity (%)
REDUCE placebo (Tombal 2013)[105]	Best clinical judgement (life expectancy, DRE, prior biopsy, prostate volume and PSA): all cancers	75 (95% CI 68 to 81)	26 (95% CI 23 to 30)	Grouped: < 20, 20–34, 35–50, > 50	66 (95% CI 58 to 72)	71 (95% CI 67 to 74)
REDUCE placebo (Tombal 2013)[105]	Best clinical judgement: Gleason score of ≥ seven	75 (95% CI 61 to 85)	26 (95% CI 23 to 29)	Grouped: < 20, 20–34, 35–50, > 50	85 (95% CI 73 to 93)	67 (95% CI 64 to 70)

continued

TABLE 18 Clinical assessment vs. clinical assessment + PCA3 (*continued*)

Derived sensitivity and specificity at various risk thresholds						
	Clinical assessment			Clinical assessment + PCA3		
Study	Variables included	Sensitivity (%)	Specificity (%)	Threshold	Sensitivity (%)	Specificity (%)
Pepe 2013[96]	PCPT nomogram: age, race, family history, PSA, DRE and prior biopsy. 25% risk threshold	100	1	PCPT + continuous PCA3. 25% risk threshold	100	8
Pepe 2013[96]	PCPT nomogram: age, race, family history, PSA, DRE and prior biopsy. 40% risk threshold	75	26	PCPT + continuous PCA3. 40% risk threshold	85.8	25
Derived sensitivity: for various set specificity levels						
	Clinical assessment			Clinical assessment + PCA3		
Study	Variables included		Result (%)	Threshold	Result (%)	Difference (%) and *p*-value if given
80% specificity						
European Cohort (Ankerst 2008)[85]	PCPT nomogram: age, family history, number of previous biopsies, DRE and PSA		43.9	Continuous	46.3	+2.4
Porpiglia 2014[99]	Age and DRE		48.0	Continuous	38.5	–9.5
90% specificity						
European Cohort (Ankerst 2008)[85]	PCPT nomogram: age, family history, number of previous biopsies, DRE and PSA		24.4	Continuous	28.5	+4.1
Porpiglia 2014[99]	Age and DRE		23.0	Continuous	26.9	+3.9
95% specificity						
European Cohort (Ankerst 2008)[85]	Age, family history, number of previous biopsies, DRE and PSA		11.4	Continuous	17.1	+5.7
Porpiglia 2014[99]	Age and DRE		17.3	32.5	19.2	+1.9

Derived specificity: for various set sensitivity levels					
	Clinical assessment		Clinical assessment + PCA3		Difference (%) and *p*-value if given
Study	Variables included	Result (%)	Threshold	Result (%)	
80% sensitivity					
Scattoni 2013[102]	Age, DRE, tPSA, fPSA and prostate volume	49	Continuous	47	–2
Porpiglia 2014[99]	Age and DRE	27.1	Continuous	37.3	+10.2
90% sensitivity					
Scattoni 2013[102]	Age, DRE, tPSA, fPSA and prostate volume	35	Continuous	25	–10
Gittelman 2013[45]	Age, family history, race, number of previous biopsies and DRE	18.9 (95% CI 10.3 to 36.9)	25	41.5 (95% CI 32.5 to 49.9)	+22.6 (90% CI 9.0 to 33.1)
Porpiglia 2014[99]	Age and DRE	12.7	Continuous	11.0	–1.7
95% sensitivity					
Porpiglia 2014[99]	Age and DRE	0.8	Continuous	8.5	+7.7

PSAD, prostate-specific antigen density.
a Based on same study population.
Differences in difference (%) are due to rounding.

TABLE 19 Clinical assessment vs. clinical assessment + phi

Study	Clinical assessment: Variables included	Clinical assessment: Result	Clinical assessment + phi: Threshold	Clinical assessment + phi: Result	Difference and *p*-value if given
AUC					
Stephan 2013[104]	Age, DRE, prostate volume, tPSA and %fPSA	0.74 (95% CI 0.67 to 0.80)	Continuous	0.80 (95% CI 0.74 to 0.85)	+0.06
Lazzeri 2012[92]	DRE, prostate volume, tPSA, %fPSA and PSAD	0.68 (95% CI 0.60 to 0.74)	Continuous	0.78 (95% CI 0.71 to 0.84)	+0.10
Scattoni 2013[102]	Age, DRE, tPSA, fPSA and prostate volume	0.75 (95% CI 0.64 to 0.87)	Continuous	0.81 (95% CI 0.70 to 0.92)	+0.06; $p = 0.137$
Porpiglia 2014[99]	Age and DRE	0.62 (95% CI 0.53 to 0.72)	Continuous	0.65 (95% CI 0.55 to 0.74)	+0.02
Multivariate OR for phi					
Lazzeri 2012[92]	DRE, prostate volume, tPSA, %fPSA and PSAD		Continuous	1.05 (95% CI 1.02 to 1.07)	
Porpiglia 2014[99]	Age and DRE		Unclear	3.52 (95% CI 1.04 to 14.14)	
Derived sensitivity: for various set specificity levels					
80% specificity					
Porpiglia 2014[99]	Age and DRE	48.0	Continuous	42.3	–5.7
90% specificity					
Porpiglia 2014[99]	Age and DRE	23.0	Continuous	25	+2.0
95% specificity					
Porpiglia 2014[99]	Age and DRE	17.3	Continuous	19.2	+1.9
Derived specificity: for various set sensitivity levels					
80% sensitivity					
Scattoni 2013[102]	Age, DRE, tPSA, fPSA and prostate volume	49	Continuous	66	+17
Porpiglia 2014[99]	Age and DRE	27.1	Continuous	24.6	–2.5
90% sensitivity					
Scattoni 2013[102]	Age, DRE, tPSA, fPSA and prostate volume	35	Continuous	37	+2
Porpiglia 2014[99]	Age and DRE	12.7	Continuous	2.5	–10.2
95% sensitivity					
Porpiglia 2014[99]	Age and DRE	0.8	Continuous	1.7	+0.9

PSAD, prostate-specific antigen density.

TABLE 20 Clinical assessment + MRI vs. clinical assessment + MRI + PCA3

AUC					
	Clinical assessment + MRI		Clinical assessment + MRI + PCA3		Difference and *p*-value if given
Study	MRI type and biopsy	Result	Threshold	Result	
Porpiglia 2014[99]	T2, DWI and DCE-MRI. No targeted biopsy	0.94 (95% CI 0.90 to 0.98)	Continuous	0.93 (95% CI 0.89 to 0.98)	–0.04
Busetto 2013[90]	T2, MRS, DCE-MRI and DWI. MRI-targeted biopsy	0.78 (0.71 to 0.85)	Continuous	0.81 (0.74 to 0.87)	+0.03
Multivariate ORs for PCA3					
	Clinical assessment + MRI		Clinical assessment + MRI + PCA3		
Study	MRI type and biopsy	Result	Threshold	Result	
Porpiglia 2014[99]	T2, DWI and DCE-MRI. No targeted biopsy	99.52 (95% CI 34.00 to 365.17)	PCA3 – unclear	1.85 (95% CI 0.26 to 9.90)	
			MRI	94.55 (95% CI 32.14 to 346.54)	
Derived sensitivity: for various set specificity levels					
	Clinical assessment + MRI		Clinical assessment + MRI + PCA3		Difference (%) and *p*-value if given
Study	MRI type and biopsy	Result (%)	Threshold	Result (%)	
80% specificity					
Porpiglia 2014[99]	T2, DWI and DCE-MRI. No targeted biopsy	94.2	Continuous	94.2	0
90% specificity					
Porpiglia 2014[99]	T2, DWI and DCE-MRI. No targeted biopsy	90.4	Continuous	90.7	+0.3
95% specificity					
Porpiglia 2014[99]	T2, DWI and DCE-MRI. No targeted biopsy	55.8	Continuous	55.8	0
Derived specificity: for various set sensitivity levels					
	Clinical assessment + MRI		Clinical assessment + MRI + PCA3		Difference (%) and *p*-value if given
Study	MRI type and biopsy	Result (%)	Threshold	Result (%)	
80% sensitivity					
Porpiglia 2014[99]	T2, DWI and DCE-MRI. No targeted biopsy	93.2	Continuous	93.2	0
90% sensitivity					
Porpiglia 2014[99]	T2, DWI and DCE-MRI. No targeted biopsy	89.0	Continuous	89.8	+0.8
95% sensitivity					
Porpiglia 2014[99]	T2, DWI and DCE-MRI. No targeted biopsy	64.4	Continuous	58.5	–5.9

TABLE 21 Clinical assessment + MRI vs. clinical assessment + MRI + phi

AUC					
	Clinical assessment + MRI		Clinical assessment + MRI + phi		Difference and *p*-value if given
Study	MRI type and biopsy	Result	Threshold	Result	
Porpiglia 2014[99]	Age, DRE, T2, DWI and DCE-MRI. No targeted biopsy	0.94 (95% CI 0.90 to 0.98)	Continuous	0.94 (95% CI 0.90 to 0.98)	0
Multivariate ORs for phi					
	Clinical assessment + MRI		Clinical assessment + MRI + phi		
Study	MRI type and biopsy	Result	Threshold	Result	
Porpiglia 2014[99]	T2, DWI and DCE-MRI. No targeted biopsy	99.52 (95% CI 34.00 to 365.17)	phi – unclear	0.76 (95% CI 0.17 to 4.40)	
			MRI	103.47 (95% CI 34.49 to 387.45)	
Derived sensitivity: for various set specificity levels					
	Clinical assessment + MRI		Clinical assessment + MRI + phi		Difference (%) and *p*-value if given
Study	MRI type and biopsy	Result (%)	Threshold	Result (%)	
80% specificity					
Porpiglia 2014[99]	T2, DWI and DCE-MRI. No targeted biopsy	94.2	Continuous	94.2	0
90% specificity					
Porpiglia 2014[99]	T2, DWI and DCE-MRI. No targeted biopsy	90.4	Continuous	90.7	+0.3
95% specificity					
Porpiglia 2014[99]	T2, DWI and DCE-MRI. No targeted biopsy	55.8	Continuous	55.8	0
Derived specificity: for various set sensitivity levels					
	Clinical assessment + MRI		Clinical assessment + MRI + phi		Difference (%) and *p*-value if given
Study	MRI type and biopsy	Result (%)	Threshold	Result (%)	
80% sensitivity					
Porpiglia 2014[99]	T2, DWI and DCE-MRI. No targeted biopsy	93.2	Continuous	93.2	0
90% sensitivity					
Porpiglia 2014[99]	T2, DWI and DCE-MRI. No targeted biopsy	89.0	Continuous	89.8	+0.8
95% sensitivity					
Porpiglia 2014[99]	T2, DWI and DCE-MRI. No targeted biopsy	64.4	Continuous	65.3	+0.9

TABLE 22 Clinical assessment + PCA3 vs. clinical assessment + phi

AUC					
	Clinical assessment + PCA3		Clinical assessment + phi		Difference and *p*-value if given
Study	Threshold	Result	Threshold	Result	
Porpiglia 2014[99]	Continuous	0.69 (95% CI 0.60 to 0.78)	Continuous	0.65 (95% CI 0.55 to 0.74)	–0.04
Scattoni 2013[102]	Continuous	0.76 (95% CI 0.64 to 0.88)	Continuous	0.81 (95% CI 0.70 to 0.92)	+0.05
Multivariate ORs for PCA3/phi					
	Clinical assessment + PCA3		Clinical assessment + phi		
Study	Threshold	Result	Threshold	Result	
Porpiglia 2014[99]	Unclear	3.88 (95% CI 1.28 to 12.95)	Unclear	3.52 (95% CI 1.04 to 14.14)	
Derived sensitivity: for various set specificity levels					
	Clinical assessment + PCA3		Clinical assessment + phi		Difference (%) and *p*-value if given
Study	Threshold	Result (%)	Threshold	Result (%)	
80% specificity					
Porpiglia 2014[99]	Continuous	38.5	Continuous	42.3	+3.8
90% specificity					
Porpiglia 2014[99]	Continuous	26.9	Continuous	25	–1.9
95% specificity					
Porpiglia 2014[99]	Continuous	19.2	Continuous	19.2	0
Derived specificity: for various set sensitivity levels					
	Clinical assessment + PCA3		Clinical assessment + phi		Difference and *p*-value if given
Study	Threshold	Result (%)	Threshold	Result (%)	
80% sensitivity					
Scattoni 2013[102]	Continuous	47	Continuous	66	+19
Porpiglia 2014[99]	Continuous	37.3	Continuous	24.6	–12.7
90% sensitivity					
Scattoni 2013[102]	Continuous	25	Continuous	37	+12
Porpiglia 2014[99]	Continuous	11	Continuous	2.5	–8.5
95% sensitivity					
Porpiglia 2014[99]	Continuous	8.5	Continuous	1.7	–6.8

TABLE 23 Clinical assessment + MRI + PCA3 vs. clinical assessment + MRI + phi

AUC					
	Clinical assessment + MRI + PCA3		Clinical assessment + MRI + phi		
Study	Threshold	Result	Threshold	Result	Difference
Porpiglia 2014[99]	Continuous	0.93 (95% CI 0.89 to 0.98)	Continuous	0.94 (95% CI 0.90 to 0.98)	+0.01
Multivariate ORs for PCA3/phi					
	Clinical assessment + MRI + PCA3		Clinical assessment + MRI + phi		
Study	Threshold	Result	Threshold	Result	
Porpiglia 2014[99]	PCA3 – unclear	1.85 (95% CI 0.26 to 9.90)	phi – unclear	0.76 (95% CI 0.17 to 4.40)	
	MRI	94.55 (95% CI 32.14 to 346.54)	MRI	103.47 (95% CI 34.49 to 387.45)	
Derived sensitivity: for various set specificity levels					
	Clinical assessment + MRI + PCA3		Clinical assessment + MRI + phi		Difference and *p*-value if given
Study	Threshold	Result (%)	Threshold	Result (%)	
80% specificity					
Porpiglia 2014[99]	Continuous	94.2	Continuous	94.2	0
90% specificity					
Porpiglia 2014[99]	Continuous	90.7	Continuous	90.7	0
95% specificity					
Porpiglia 2014[99]	Continuous	55.8	Continuous	55.8	0
Derived specificity: for various set sensitivity levels					
	Clinical assessment + MRI + PCA3		Clinical assessment + MRI + phi		Difference and *p*-value if given
Study	Threshold	Result (%)	Threshold	Result (%)	
80% sensitivity					
Porpiglia 2014[99]	Continuous	93.2	Continuous	93.2	0
90% sensitivity					
Porpiglia 2014[99]	Continuous	89.8	Continuous	89.8	0
95% sensitivity					
Porpiglia 2014[99]	Continuous	58.5	Continuous	65.3	+6.8

TABLE 24 Clinical assessment vs. clinical assessment + PCA3 + phi

AUC					
	Clinical		Clinical + PCA3 + phi		
Study	Variables included	Result	Threshold	Result	Difference and *p*-value if given
Scattoni 2013[102]	Age, DRE, tPSA, fPSA and prostate volume	0.75 (95% CI 0.64 to 0.87)	Continuous	0.81 (95% CI 0.70 to 0.92)	+0.06; $p = 0.17$
Porpiglia 2014[99]	Age and DRE	0.62 (95% CI 0.53 to 0.72)	Continuous	0.69 (95% CI 0.60 to 0.78)	+0.07
Multivariate ORs for PCA3 and phi					
	Clinical		Clinical + PCA3 + phi		
Study	Variables included		Threshold	Result	
Porpiglia 2014[99]	Age and DRE		PCA3 – unclear	3.87 (95% CI 1.25 to 13.23)	
			phi – unclear	3.44 (95% CI 1.01 to 13.87)	
Derived sensitivity: for various set specificity levels					
	Clinical		Clinical + PCA3 + phi		Difference (%) and *p*-value if given
Study	Variables included	Result (%)	Threshold	Result (%)	
80% specificity					
Porpiglia 2014[99]	Age and DRE	48.0	Continuous	51.9	+3.9
90% specificity					
Porpiglia 2014[99]	Age and DRE	23.0	Continuous	26.9	+3.9
95% specificity					
Porpiglia 2014[99]	Age and DRE	17.3	Continuous	19.2	+1.9
Derived specificity: for various set sensitivity levels					
	Clinical		Clinical + PCA3 + phi		Difference (%) and *p*-value if given
Study	Variables included	Result (%)	Threshold	Result (%)	
80% sensitivity					
Scattoni 2013[102]	Age, DRE, tPSA, fPSA and prostate volume	49	Continuous	49	0
Porpiglia 2014[99]	Age and DRE	27.1	Continuous	39.8	+12.7
90% sensitivity					
Scattoni 2013[102]	Age, DRE, tPSA, fPSA and prostate volume	35	Continuous	33	–2
Porpiglia 2014[99]	Age and DRE	12.7	Continuous	22.9	+10.2
95% sensitivity					
Porpiglia 2014[99]	Age and DRE	0.8	Continuous	7.6	6.8

TABLE 25 Clinical assessment + PCA3 vs. clinical assessment + MRI

Study	Clinical + PCA3 Threshold	Clinical + PCA3 Result	Clinical + MRI MRI type and biopsy	Clinical + MRI Result	Difference and *p*-value if given
AUC					
Porpiglia 2014[99]	Continuous	0.69 (95% CI 0.60 to 0.78)	T2, DWI and DCE-MRI. No targeted biopsy	0.94 (95% CI 0.90 to 0.98)	+0.25
Busetto 2013[90]	Continuous	0.74 (0.66 to 0.82)	T2, MRS, DCE-MRI and DWI. MRI-targeted biopsy	0.78 (0.71 to 0.85)	+0.04
[a]Panebianco 2011[94]	35	0.76 (0.60 to 0.88)	T2, MRS, DCE-MRI and DWI. MRI-targeted biopsy	0.86 (0.73 to 0.95)	+0.10
Multivariate OR					
Porpiglia 2014[99]	Unclear	3.88 (95% CI 1.27 to 12.95)	T2, DWI and DCE-MRI. No targeted biopsy	99.52 (95% CI 34.00 to 363.17)	
Derived sensitivity					
80% specificity					
Porpiglia 2014[99]	Continuous	38.5	T2, DWI and DCE-MRI. No targeted biopsy	94.2	+55.7
90% specificity					
Porpiglia 2014[99]	Continuous	26.9	T2, DWI and DCE-MRI. No targeted biopsy	90.4	+63.5
95% specificity					
Porpiglia 2014[99]	Continuous	19.2	T2, DWI and DCE-MRI. No targeted biopsy	55.8	+36.6
Derived specificity					
80% sensitivity					
Porpiglia 2014[99]	Continuous	37.3	T2, DWI and DCE-MRI. No targeted biopsy	93.2	+55.9
90% sensitivity					
Porpiglia 2014[99]	Continuous	11	T2, DWI and DCE-MRI. No targeted biopsy	89.0	+78.0
95% sensitivity					
Porpiglia 2014[99]	Continuous	8.5	T2, DWI and DCE-MRI. No targeted biopsy	64.4	+55.9

a It is unclear from the text whether or not the clinical assessment was included in the model.

TABLE 26 Clinical assessment + PCA3 vs. clinical assessment + MRI + PCA3

AUC					
	Clinical + PCA3		Clinical + MRI + PCA3		
Study	MRI type and biopsy	Result	Threshold	Result	Difference and *p*-value if given
[a]Sciarra 2012[103]	T2, MRS, DCE-MRI and DWI. MRI targeted biopsy	0.83 (95% CI 0.73 to 0.90)	35	0.86 (95% CI 0.76 to 0.92)	+0.03; $p < 0.001$

a It is unclear from the text whether or not the clinical assessment was included in the models.

TABLE 27 Clinical assessment + phi vs. clinical + MRI

AUC					
	Clinical + phi		Clinical + MRI		
Study	Threshold	Result	MRI type and biopsy	Result	Difference and *p*-values if given
Porpiglia 2014[99]	Continuous	0.65 (95% CI 0.55 to 0.74)	T2, DWI and DCE-MRI. No targeted biopsy	0.94 (95% CI 0.90 to 0.98)	+0.29
Multivariate ORs					
	Clinical + phi		Clinical + MRI		
Study	Threshold	Result	MRI type and biopsy	Result	
Porpiglia 2014[99]	Unclear	3.52 (95% CI 1.04 to 14.14)	T2, DWI and DCE-MRI. No targeted biopsy	99.52 (95% CI 34.00 to 363.17)	
Derived sensitivity – for various set specificity levels					
	Clinical + phi		Clinical + MRI		Difference (%) and *p*-value if given
Study	Threshold	Result (%)	MRI type and biopsy	Result (%)	
80% specificity					
Porpiglia 2014[99]	Continuous	42.3	T2, DWI and DCE-MRI. No targeted biopsy	94.2	+51.9
90% specificity					
Porpiglia 2014[99]	Continuous	25.0	T2, DWI and DCE-MRI. No targeted biopsy	90.4	+65.4
95% specificity					
Porpiglia 2014[99]	Continuous	19.2	T2, DWI and DCE-MRI. No targeted biopsy	55.8	+36.6
Derived specificity – for various set sensitivity levels					
	Clinical + phi		Clinical + MRI		Difference (%) and *p*-value if given
Study	Threshold	Result (%)	MRI type and biopsy	Result (%)	
80% sensitivity					
Porpiglia 2014[99]	Continuous	24.6	T2, DWI and DCE-MRI. No targeted biopsy	93.2	+68.6
90% sensitivity					
Porpiglia 2014[99]	Continuous	2.5	T2, DWI and DCE-MRI. No targeted biopsy	89.0	+86.5
95% sensitivity					
Porpiglia 2014[99]	Continuous	1.7	T2, DWI and DCE-MRI. No targeted biopsy	64.4	+62.7

Two studies[85,99] reported derived sensitivity values for specificity levels set at 80%, 90% and 95%. At 90% and 95% specificity, both studies show an improvement in sensitivity when the PCA3 score is added to clinical assessment. However, the derived sensitivity results for 80% specificity are conflicting: Porpiglia *et al.*[99] shows a 9.5% decrease, whereas Ankerst *et al.*[85] shows a 2.4% increase in discrimination.

Three studies[45,99,102] reported derived specificity for sensitivity set at 80%, 90% or 95%. The results are conflicting. When sensitivity is set at 80% or 90%, Scattoni *et al.*[102] shows that derived specificity decreases when PCA3 score is added to clinical assessment. Gittelman *et al.*[45] reports increased derived specificity when sensitivity is set at 90%, when the PCA3 score is added to clinical assessment specificity increases from 18.9% to 41.5%. Porpiglia *et al.*[99] reports that adding the PCA3 score to clinical assessment increases derived specificity when sensitivity is set at 80% and 95% and reduces derived specificity when sensitivity is set at 90%.

Decision curve analysis

Three studies[90,99,102] presented decision curve analyses comparing net benefit for clinical assessment and for clinical assessment + PCA3. The results are presented graphically with no statistical significance testing. The graphs are included in *Appendix 7*. Visual review of the published graphs in Busetto *et al.*[90] and Porpiglia *et al.*[99] suggest that no benefit is gained from adding the PCA3 score to clinical assessment at a threshold probability between 10% and 20%. Net benefit was greater for the model including the PCA3 score in Busetto *et al.*[90] from 25% to 50% threshold; in Porpiglia *et al.*[99] the increase in net benefit for the model including the PCA3 score appears only between 20% and 35%, and then the curves are similar. In Scattoni *et al.*[102] net benefit was reduced when the PCA3 score was added to the clinical assessment at a threshold probability between 10% and 40%. At 40% the curves then reversed with increased net benefit associated with the clinical assessment + PCA3 model from 50% to 90% threshold probability.

Comparison 2: clinical assessment versus clinical assessment + Prostate Health Index

Area under the curve

Four studies[92,99,102,104] reported AUC for the comparisons of clinical assessment versus clinical assessment plus phi. All studies showed an increase in discrimination of between 2% and 10% when phi was added to the clinical assessment model as a continuous variable.

Multivariate odds ratios for PROGENSA prostate cancer antigen 3 assay

Two studies[92,99] reported multivariate ORs for phi. Both studies presented statistically significant results indicating that an increase in phi score was associated with an increased probability of cancer on biopsy (ORs were above 1 and CIs did not include 1). These results are consistent with the AUC results and indicate that the addition of phi to the clinical assessment model increases discrimination.

Derived sensitivity and specificity

One study[99] reported derived sensitivity values for 80%, 90% and 95% specificity. The results were mixed. Adding phi to clinical assessment is associated with either a small increase (2% at 90% and 95% specificity) or a decrease (–5.7% at 80% specificity) in derived sensitivity. Two studies[99,102] reported derived specificity for 80% and 90% sensitivity. Scattoni *et al.*[102] showed that adding phi to clinical assessment increased derived specificity at 80% and 90% sensitivity by 17% and 2%, respectively. Porpiglia *et al.*[99] showed that adding phi to clinical assessment reduced derived specificity by –2.5% and –10.2% at 80% and 90% sensitivity, respectively; at 95% sensitivity, adding phi to clinical assessment increased derived specificity by 0.9%.

Decision curve analysis

Three studies[92,99,102] presented decision curve analyses comparing net benefit for clinical assessment and clinical assessment + phi. Lazzeri[92] showed that net benefit was greater for the clinical assessment model at threshold probabilities from 20% to 25% and showed that the clinical assessment + phi model had a

greater net benefit at threshold probabilities between 25% and 40%. Scattoni *et al.*[102] showed increased net benefit for the clinical assessment + phi model at threshold probabilities from 10% to 50%. Porpiglia *et al.*[99] demonstrated that estimates of net benefit for both models were similar at threshold probabilities between 10% and 70%.

Comparison 3: clinical assessment + magnetic resonance imaging versus clinical assessment + magnetic resonance imaging + PROGENSA prostate cancer antigen 3 assay

Area under the curve

Two studies[90,99] investigated the addition of the PCA3 score to a diagnostic model which comprised clinical assessment + MRI. Adding PCA3 score to clinical assessment + MRI had very little effect on the size of the AUC reported. Porpiglia *et al.*[99] found a slight decrease (–1%) in AUC and Busetto *et al.*[90] reported a slight increase (3%) in AUC. Only small changes in AUC were expected as models of clinical assessment + MRI give very high estimates of AUC and so adding to these models is not likely to generate substantial gains or losses.

Multivariate odds ratios

Multivariate ORs for clinical assessment + MRI versus clinical assessment + MRI + PCA3 were reported in one study.[99] In the model containing both MRI and PCA3 score, the OR for MRI was much larger (OR 94.55, 95% CI 32.14 to 346.54) than that for PCA3 score (OR1.85, 95% CI 0.26 to 9.90); in this model, the OR for PCA3 score was not statistically significant.

Derived sensitivity and specificity

At 80%, 90% and 95% specificity, Porpiglia *et al.*[99] reported minimal changes in derived sensitivity for clinical assessment + MRI compared with clinical assessment + MRI + PCA3; derived sensitivity increased by 0%, 0.3% and 0%, respectively.

At 80% and 90% sensitivity, Porpiglia *et al.*[99] reported minimal changes in derived specificity for clinical assessment + MRI compared with clinical assessment + MRI + PCA3; derived specificity increased by 0%, and 0.8%, respectively. At 95% sensitivity, Porpiglia *et al.*[99] reported a change in derived specificity of –5.9% when PCA3 score was added to clinical assessment + MRI.

Decision curve analysis

Decision curve analysis results for two studies[90,99] demonstrate that the addition of PCA3 score does not improve diagnostic accuracy when added to the clinical assessment + MRI at threshold probabilities between 10% and 50%.

Comparison 4: clinical assessment + magnetic resonance imaging versus clinical assessment + magnetic resonance imaging + Prostate Health Index

Area under the curve

One study[99] reported the results of a head-to-head comparison of clinical assessment + MRI versus clinical assessment + MRI + phi. Porpiglia *et al.*[99] demonstrated that the addition of phi to a model comprising clinical assessment + MRI had no effect on the size of the AUC.

Multivariate odds ratios

Multivariate ORs for clinical assessment + MRI + phi compared with clinical assessment + MRI were reported in one study.[99] In the model containing both MRI and phi, the OR for MRI was larger (OR 103.45, 95% CI 34.49 to 387.45) than the OR for phi (OR 0.76, 95% CI 0.17 to 4.40). In this model, the OR for phi was not statistically significant.

Derived sensitivity and specificity

At 80%, 90% and 95% specificity, Porpiglia *et al.*[99] reported minimal change in derived sensitivity for clinical assessment + MRI + phi compared with clinical assessment + MRI; derived sensitivity increased by 0%, 0.3% and 0%, respectively.

At 80%, 90% and 95% sensitivity, Porpiglia *et al.*[99] reported minimal change in derived specificity for clinical assessment + MRI compared with clinical assessment + PCA3; derived specificity increased by 0%, 0.8% and 0.9%, respectively. The addition of phi to diagnostic models incorporating clinical assessment + MRI had a negligible effect on outcome measures.

Decision curve analyses

The decision curve analysis graphs in the study by Porpiglia *et al.*[99] demonstrate that the addition of phi does not improve diagnostic accuracy when added to clinical assessment + MRI at threshold probabilities between 10% and 60%.

Comparison 5: clinical assessment + PROGENSA prostate cancer antigen 3 assay versus clinical assessment + Prostate Health Index

Area under the curve

Two studies[99,102] reported the results of a head-to-head comparison of clinical assessment + PCA3 and clinical assessment + phi. The AUC results of the two studies[99,102] were conflicting. Porpiglia *et al.*[99] reported a 4% decrease in AUC with the use of clinical assessment + phi compared with clinical assessment + PCA3. In contrast, Scattoni *et al.*[102] demonstrated a 5% increase in AUC with the use of clinical assessment + phi compared with clinical assessment + PCA3.

Multivariate odds ratios

Multivariate ORs for phi and PCA3 scores in separate models were reported in one study.[99] Both statistically significant ORs indicated an increased risk of cancer on biopsy for increases in phi or PCA3 score, with CIs that did not cross 1.

Derived sensitivity and specificity

One study[99] showed derived sensitivity for specificity set at 80%, 90% and 95%. At 80% specificity, derived sensitivity was 3.8% higher when using clinical assessment + phi compared with clinical assessment + PCA3. At 90% specificity, derived sensitivity was 1.9% lower when using clinical assessment + phi compared with clinical assessment + PCA3. At 95% specificity, derived sensitivity was the same for both models.

Two studies[99,102] reported derived specificity for sensitivity set at 80% and 90%. Scattoni *et al.*[102] found higher derived specificity for clinical assessment + phi compared with clinical assessment plus PCA3 score for sensitivity set at 80% and 90%. In contrast, Porpiglia *et al.*[99] reported higher derived specificity for clinical assessment + PCA3 compared with clinical assessment + phi for sensitivity set at 80% and 90%. At 95% sensitivity, Porpiglia *et al.*[99] showed higher derived specificity for clinical assessment + PCA3 compared with clinical assessment + phi.

Decision curve analysis

The decision curve analyses results reflect the derived sensitivity and specificity results for clinical assessment + PCA3 and clinical assessment + phi. Porpiglia *et al.*[99] shows a larger net benefit for clinical assessment + PCA3 than for clinical assessment + phi between 15% and 35% threshold probability of cancer. In contrast, Scattoni *et al.*[102] found that clinical assessment + phi had greater net benefit than clinical assessment + PCA3 at threshold probabilities between 10% and 45%.

Comparison 6: clinical assessment + magnetic resonance imaging + PROGENSA prostate cancer antigen 3 assay versus clinical assessment + magnetic resonance imaging + Prostate Health Index

Area under the curve

Porpiglia *et al.*[99] reported the results of a head-to-head comparison of clinical assessment + MRI + PCA3 with clinical assessment + MRI + phi. Porpiglia *et al.*[99] demonstrated that using phi instead of PCA3 score alongside clinical assessment + MRI led to a 1% increase in the AUC.

Multivariate odds ratio

The multivariate OR results from the study by Porpiglia *et al.*[99] confirm that MRI remains a significant predictor of biopsy outcome when used in addition clinical assessment + PCA3 or clinical assessment + phi, but neither PCA3 score nor phi is a significant predictor in these models.

Derived sensitivity and specificity

Data from the study by Porpiglia *et al.*[99] suggest that, at 80%, 90% and 95% specificity, derived sensitivity values for clinical assessment + MRI + PCA3 are identical to the derived sensitivity values for clinical assessment + MRI + phi.

Data from the study by Porpiglia *et al.*[99] suggest that, at 80% and 90% specificity, derived sensitivity values for clinical assessment + MRI + PCA3 are identical to the derived sensitivity values for clinical assessment + MRI + phi. At 95% sensitivity, use of clinical assessment + MRI + phi leads to a 6.8% gain in derived specificity over clinical assessment + MRI + PCA3.

Decision curve analysis

The decision curve analysis graphs for all models containing MRI overlapped at threshold probabilities 10% and 60%, which means that there is no additional increase in net benefit from adding either PCA3 score or phi to clinical assessment + MRI.

Comparison 7: clinical assessment versus clinical assessment + PROGENSA prostate cancer antigen 3 assay + Prostate Health Index

Area under the curve

The effect of adding both PCA3 score and phi to clinical assessment was assessed in two studies;[99,102] both studies reported a 6–7% increase in AUC.

Multivariate odds ratios

Multivariate ORs for phi and PCA3 score used together were reported in one study.[99] Both ORs were statistically significant and indicated an increased risk of cancer on biopsy for increases in phi or PCA3 score, with CIs that did not cross 1.

Derived sensitivity and specificity

When adding PCA3 score and phi to clinical assessment, Porpiglia *et al.*[99] demonstrated small improvements (1.9–3.9%) in derived sensitivity when specificity was set at 80%, 90% and 95%. When sensitivity was set at 80% and 90%, the addition of the PCA3 score and phi to clinical assessment increased derived specificity by 12.7% and 10.2%, respectively; at 95% sensitivity there was a 6.8% increase in derived specificity.

In the study by Scattoni *et al.*,[102] the addition of PCA3 score and phi to clinical assessment led to no change in derived specificity at 80% sensitivity and a very small decrease (–2%) in derived specificity at 90% sensitivity.

Decision curve analysis

Two studies[92,99] reported decision curve analysis results, and both studies indicate that net benefit increases when phi and PCA3 score are added to clinical assessment for threshold probabilities between 15% and 40%.

Comparison 8: clinical assessment + PROGENSA prostate cancer antigen 3 assay versus clinical assessment + magnetic resonance imaging

Area under the curve

Three studies[90,94,99] reported AUC results for the head-to-head comparison of clinical assessment + PCA3 and clinical assessment + MRI. All of the studies[90,94,99] demonstrated an increase in AUC for clinical assessment + MRI compared with clinical assessment + PCA3; Porpiglia *et al.*[99] showed an increase of 25% and Busetto *et al.*[90] showed an increase of 4%. In Panebianco *et al.*,[94] the AUC for clinical assessment + PCA3, for a threshold of 35, was estimated to be 0.76, and the AUC for clinical assessment + MRI was estimated to be 0.86. The data showed an increase in the AUC of 10%; however, it is not clear from the data presented in the study to what extent clinical assessment had been undertaken.

Multivariate odds ratios

Multivariate ORs for PCA3 score and for MRI were reported in one study.[99] The OR for clinical assessment + MRI was high (OR 99.52, 95% CI 34.00 to 363.17) compared with the OR for clinical assessment + PCA3 (OR 3.88, 95% CI 1.27 to 12.95); both ORs were statistically significant. The data indicate that a positive MRI is a better predictor of cancer detected on biopsy than a raised PCA3 score.

Derived sensitivity and specificity

At 80%, 90% and 95% specificity, Porpiglia *et al.*[99] reported substantial increases in derived sensitivity for clinical assessment + MRI compared with clinical assessment + PCA3; derived sensitivity increased by 55.7%, 63.8% and 36.6%, respectively.

At 80%, 90% and 95% sensitivity Porpiglia *et al.*[99] reported substantial increases in derived specificity for clinical assessment + MRI compared with clinical assessment + PCA3; derived specificity increased by 55.9%, 78% and 55.9%, respectively.

Decision curve analysis

Decision curve analysis results for two studies[90,99] also showed a sustained increase in net benefit for clinical assessment + MRI compared with clinical assessment + PCA3 at threshold probabilities between 10% and 60%.

Comparison 9: clinical assessment + PROGENSA prostate cancer antigen 3 assay versus clinical assessment + magnetic resonance imaging + PROGENSA prostate cancer antigen 3 assay

Area under the curve

The RCT reported in Sciarra *et al.*[103] randomised participants to PCA3 score alone or PCA3 + MRI. It is not clear from the published paper to what extent clinical assessment was included in any of the analyses. This study demonstrated that the addition of MRI to clinical assessment + PCA3 improved discrimination, as the reported AUC increased by 3 percentage points from 0.83 in the clinical assessment + MRI group to 0.86 in the clinical assessment + MRI + PCA3 group ($p < 0.001$).

Comparison 10: clinical assessment + Prostate Health Index versus clinical assessment + magnetic resonance imaging

Area under the curve

One study[99] reported the results of a head-to-head comparison of clinical assessment + phi versus clinical assessment + MRI. Porpiglia *et al.*[99] demonstrated a 29% gain in AUC when clinical assessment + MRI was used instead of clinical assessment + phi.

Multivariate odds ratios

Multivariate ORs for clinical assessment + phi and for clinical assessment + MRI were reported for separate models in one study.[99] The OR for MRI in the clinical assessment + MRI model was much larger (OR 99.52, 95% CI 34.00 to 363.17) than that for phi in the clinical assessment + phi model (OR 3.52, 95% CI 1.04 to 14.14).

Derived sensitivity and specificity

Porpiglia *et al.*[99] showed that large differences in derived sensitivity and derived specificity were achieved when using clinical assessment + MRI compared with clinical assessment + phi.

At 80%, 90% and 95% specificity, Porpiglia *et al.*[99] reported substantial increases in derived sensitivity for clinical assessment + MRI compared with clinical assessment + phi; derived sensitivity increased by 51.9%, 65.4% and 36.6%, respectively.

At 80%, 90% and 95% sensitivity, Porpiglia *et al.*[99] reported substantial increases in derived specificity for clinical assessment + MRI compared with clinical assessment + PCA3; derived specificity increased by 68.6%, 86.5% and 62.7%, respectively.

Decision curve analysis

The decision curve analysis graphs in the study by Porpiglia *et al.*[99] showed a sustained increase in net benefit for clinical assessment + MRI compared with clinical assessment + phi at threshold probabilities between 10% and 60%.

Within-study comparisons: additional data analyses

As the included studies were heterogeneous in many ways (e.g. study population, outcomes reported, threshold used and type of analysis), it was not appropriate, from a clinical or statistical perspective, to carry out a meta-analysis of sensitivity or specificity.

There were insufficient data reported in the included studies in any one subgroup for any of the sensitivity analyses that were considered, as listed in *Methods of data analysis/synthesis: clinical validity review*, to be undertaken.

Within-study comparisons: Gleason score

Seven studies[45,46,86,89,90,92,103,105] reported diagnostic accuracy results for PCA3 score for the detection of more aggressive cancers, usually based on a Gleason score of ≥ 7. In six studies[45,46,86,89,90,103] the authors employed univariate analyses and showed the ability of PCA3 score to predict a Gleason score of ≥ 7.

Two studies[46,86] reported median PCA3 scores for detected cancers with Gleason score of > or < 7. Both found that the PCA3 scores were higher in cancers with higher Gleason scores. In Haese *et al.*,[46] the median PCA3 scores were 28.1 for cancers with a Gleason score of < 7 and 45.3 for cancers with a Gleason score of ≥ 7 ($p = 0.04$). In Aubin *et al.*,[86] the corresponding median PCA3 scores were 31.8 and 49.5, respectively ($p = 0.002$). In addition, Busetto *et al.*[90] reported a statistically significant association ($p < 0.001$, $\chi^2 = 71.27$) between the Gleason score and PCA3 score. Hease *et al.*[46] also reported significant differences in the median PCA3 scores for clinical stage T1c cancers compared with T2 cancers (26.8 vs. 61.7; $p = 0.005$) and for indolent cancers (defined as clinical stage T1c, PSA density < 0.15, Gleason score of ≤ 6 and percentage of positive cores $\leq 33\%$) versus significant cancers (21.4 vs. 42.1; $p = 0.006$).

Gittelman *et al.*[45] reported the sensitivity and specificity and the AUC for PCA3 using a score of 25 as the threshold for the detection of all cancers, of cancers with a Gleason score of ≥ 7 and of significant cancers (defined as clinical stage T2 or above, PSA density > 0.15, Gleason score of ≥ 7 and three or more cores positive for cancer). The AUC values reported were 0.707 for all cancers, 0.638 for cancers with a Gleason score of ≥ 7 and 0.689 for significant cancers. The sensitivity values were 77.5 (95% CI 68.4 to 84.5), 76.5 (95% CI 60.0 to 87.6) and 78.9 (95% CI 68.5 to 86.6), respectively, and specificity values were 57.1

(95% CI 52.0 to 62.1), 51.6 (95% CI 46.9 to 56.3) and 55.1 (95% CI 50.2 to 60.0), respectively. There was no evidence that the sensitivity or specificity of the PCA3 assay varied between the groups. Sciarra *et al.*[103] also reported that there was no statistically significant difference in the predictive accuracy of PCA3 score for cancers with a Gleason score of ≤ 7 or less ($3 + 4$) and cancers with a Gleason score of ≥ 7 ($4 + 3$) ($p = 0.089$).

Bollito *et al.*[89] and Haese *et al.*[46] report the numbers of 'missed' cancers that would have been missed using PCA3 screening alone and would have had a Gleason score of ≥ 7. In Haese *et al.*,[46] using a PCA3 score of 20 as the threshold for detection of cancer, 35 out of 128 cancers would have been missed, and 12 of these 35 missed cancers would have had a Gleason score of ≥ 7. Using a PCA3 score of 35 as the threshold for the detection of cancer, 68 out of 128 cancers would have been missed, and 27 of these 68 cancers would have had a Gleason score of ≥ 7. In Bollito *et al.*,[89] using a PCA3 score of 39 for the threshold for detection of cancer, 22 out of 281 cancers would have been missed and none of these would have had a Gleason score of > 7 ($4 + 3$). Using a threshold of 50, 29 out of 281 cancers would have been missed and 5 of these 29 would have had a Gleason score of > 7 ($4 + 3$).

Only one study[105] reported how the use of the PCA3 score in combination with clinical assessment contributed to the prediction of more aggressive cancers. It found that PCA3 score had higher sensitivity and specificity for the detection of cancers with a Gleason score of ≥ 7 than for the detection of all cancers. This study[105] presented independent sensitivity and specificity estimates for all cancers and demonstrated that the addition of the PCA3 score to best clinical judgement reduced sensitivity from 75% to 66% and increased specificity from 26% to 71%. In this population (prevalence of all cancer = 17.9%), adding the PCA3 score to clinical assessment meant that 18 cancers would have been missed and 371 biopsies would have been avoided compared with clinical assessment alone. However, when the analyses were repeated for cancers with a Gleason score of ≥ 7 (prevalence = 5.4%), the addition of the PCA3 score increased sensitivity from 75% to 85% and specificity from 26% to 67%, meaning that six more cancers would have been detected and 395 biopsies would have been avoided compared with clinical assessment alone.

Only one study[92] considered the relationship between phi and the Gleason score. The authors found a significant correlation, with increased phi score being associated with a higher Gleason score (Spearman's rho 0.299; $p = 0.013$). It is not clear from the published paper[92] whether these findings are for all biopsies or for repeat biopsy only.

Between-study comparisons: search results

Six papers[1,15,66,67,70,110] reporting five systematic reviews and meta-analyses were identified which met the inclusion criteria. As data from within-study comparisons were available, these reviews are summarised for completeness in *Table 28*. The EAG notes that none of these reviews considers clinically relevant comparisons.

Between-study comparisons: summary of systematic reviews

Two reviews[66,67,110] assessed the clinical validity of using the PCA3 score to predict prostate cancer. Luo *et al.*[110] considered a repeat biopsy population and included studies without a comparator. Luo *et al.*[110] concluded that use of the PCA3 score improved the accuracy of prostate cancer detection and the authors claimed that unnecessary biopsies could be avoided using a PCA3 threshold of 20%. The Bradley *et al.*[66,67] review restricted inclusion to studies which compared the PCA3 score with a comparator of either clinical assessment or PSA and concluded that, although the PCA3 score appeared to be more discriminatory for detecting prostate cancer than tPSA, the strength of evidence was low. Bradley *et al.*[66,67] included initial and repeat biopsy study populations in the full review; the repeat biopsy results were reported separately.

TABLE 28 Between-study comparisons

Study	Tests reviewed	Comparator	Number of studies (patients)	Inclusion criteria for studies included in review	Author conclusions
Luo 2014[110]	PCA3	None	11 (3373)	Population consisted of adult men who had undergone a repeat biopsy for PCa. The intervention must have consisted of a quantitative determination of *PCA3* gene expression in urine samples by molecular biology methods. The prostate biopsy was the gold standard with which to assess the technique. The results had to include the specific values of the diagnostic tests, such as sensitivity, specificity, positive predictive value, negative predictive value and receiver operating characteristic curves, which must have been calculated using TPs, FPs, FNs and TNs	PCA3 can be used for repeat biopsy of the prostate to improve accuracy of PCa detection. Unnecessary biopsies can be avoided by using a PCa cut-off score of 20
Bradley 2014[66,67]	PCA3	Clinical nomograms; PSA	7 (2586)	Appropriate study design; study subjects at increased prostate cancer risk based on increased tPSA and/or an abnormal DRE; and reported results for PCA3, comparator test(s) and prostate biopsy. Studies of initial and repeat population included and repeat studies reported separately	Seven matched studies addressed diagnostic accuracy for PCA3 in populations where all men were having a repeat biopsy. Five studies reported on PCA3 and tPSA, four on %fPSA, one on PSA velocity and two on externally validated nomograms. However, the numbers of comparisons possible for each of these matched analyses remained small. For example, one of three tPSA studies providing AUC data restricted recruitment to men with tPSA levels in the 'grey zone'. No studies addressed other comparators or outcomes. Strength of evidence was insufficient for all comparisons

continued

TABLE 28 Between-study comparisons (*continued*)

Study	Tests reviewed	Comparator	Number of studies (patients)	Inclusion criteria for studies included in review	Author conclusions
Mowatt 2013[14]	MRS; dynamic contrast-enhanced MRI; DW-MRI	T2-MRI; TRUS prostate biopsy	51 (10,264)	The population considered will be men with suspected prostate cancer and elevated PSA levels up to 20 ng/ml but previously negative biopsy	MRS had higher sensitivity and specificity than T2-MRI. Relative cost-effectiveness of alternative strategies was sensitive to key parameters/assumptions. Under certain circumstances, T2-MRI may be cost-effective compared with systematic TRUS. If MRS and DW-MRI can be shown to have high sensitivity for detecting moderate-/high-risk cancer, while negating patients with no cancer/low-risk disease to undergo biopsy, their use could represent a cost-effective approach to diagnosis. However, owing to the relative paucity of reliable data, further studies are required. In particular, prospective studies are required in men with suspected PCa and elevated PSA levels but previously negative biopsy, comparing the utility of the individual and combined components of a multiparametric magnetic resonance approach (MRS, DCE-MRI and DW-MRI) with both a magnetic resonance-guided/-directed biopsy session and an extended 14-core TRUS-guided biopsy scheme against a reference standard of histopathological assessment of biopsied tissue obtained via saturation biopsy, template biopsy or prostatectomy specimens
Zhang 2014[70]	MRI	Subgroup analysis of MRI vs. MRS vs. MRI + MRSI	14 (698)	First, only prospective studies with patients having MRI and prior negative prostate biopsies and persistently elevated PSA levels were selected. Second, a histopathological analysis was used as the reference standard. Third, sufficient data were reported to construct 2 × 2 contingency tables consisting of the TP, FP, FN and TN values. Fourth, 10 or more patients had to be included	A limited number of studies suggest that the value of MRI to target prostate cancer in patients with previous negative biopsies and elevated PSA levels appears significant. MRI combined with MRSI is particularly accurate. Further studies are necessary to confirm the eventual role of DW-MRI in this field

Study	Tests reviewed	Comparator	Number of studies (patients)	Inclusion criteria for studies included in review	Author conclusions
Eichler 2005[1]	Systematic prostate biopsy	None	11 (1071)	Prospective studies were included that compared the cancer yield of a systematic prostate biopsy scheme (index test) with a systematic reference test. Sufficient information had to be available to construct a 2 × 2 table. Excluded were studies that did not compare the index test with the reference test in the same population, non-systematic biopsy schemes (e.g. lesion-directed biopsies), and computer simulation studies. Included participants were men of all age groups with suspected prostate cancer scheduled for a prostate biopsy. Men with already proven prostate cancer were excluded. Studies of initial and repeat population included and repeat studies reported separately when possible	Schemes which apply additional laterally directed cores showed a higher cancer yield. It still has to be demonstrated that extended biopsy schemes with a higher cancer yield do lead to a survival benefit as a result of early cancer detection. The influence first vs. repeated biopsies on the pooled results could not be assessed systematically

PCa, prostate cancer.

Two reviews[14,70] assessed the clinical validity of MRI in the detection of prostate cancer. The review by Mowatt *et al.*[14] was the most comprehensive, including DW-MRI, dynamic contrast-enhanced MRI and MRS, compared with Zhang *et al.*,[70] which included MRI and MRS. The number of studies and patients included differed significantly across the reviews; Mowatt *et al.*[14] included 51 studies with 10,264 patients, and Zhang[70] included 14 studies with 698 patients. Zhang *et al.*[70] concluded that there was some evidence for the effectiveness of MRI in detecting prostate cancer and Mowatt *et al.*[14] concluded that MRS had higher sensitivity and specificity than T2-MRI. The authors of both reviews highlighted the lack of reliable data and the need for further research.

The review by Eichler *et al.*[1] assessed the effectiveness of various systematic prostate biopsy schemes to diagnose prostate cancer; entry was not restricted to studies with a comparator. This review[1] of 11 studies and 1071 patients included initial and repeat biopsy study populations. Results were reported separately for the repeat biopsy population for some biopsy schemes only. The authors concluded that schemes which apply additional laterally directed cores showed a higher cancer yield and conclude that the impact this has on patient survival has yet to be determined.

Two systematic reviews assessed the clinical validity of PCA3 scores in the diagnosis of prostate cancer,[66,67,110] two described the effectiveness of MRI-guided biopsies[70,15] and one investigated the effect of systematic protocols for prostate biopsies.[1] No reviews were identified that assessed the effectiveness of phi in a repeat biopsy population.

Discussion of results: clinical validity review

Quality assessment

Generalisability of study population

The clinical validity review addressed the potential use of the PCA3 assay and the phi, in combination with clinical assessment, to assess the need for a second biopsy in men suspected of having prostate cancer whose initial biopsy result was negative or equivocal. This population can be considered to be made up of three different groups of men: first, men who have signs which are strongly suggestive of prostate cancer (such as abnormal DRE findings and/or abnormal histopathology) and who would be referred for second biopsy by most, if not all, clinicians; second, men who no longer display signs of prostate cancer, for example men whose PSA levels have returned to normal and who have no other evident risk factors – many clinicians would not refer this group of men for a second biopsy; and, third, men who fall between these two positions, that is men who have some signs of prostate cancer. In the case of this last group of men, referral to second biopsy may vary between clinicians. The use of PCA3 scores and phi may contribute to the diagnostic process in all three of these populations by providing clinicians with an additional source of clinical information which they can consider before the decision regarding referral to a second biopsy is made.

All but two[78,79] of the populations described in the included studies comprise men who were referred for a second biopsy because, following a negative initial biopsy result, clinicians still suspected that malignant prostate cancer was present. In these cases, the effect of adding the PCA3 assay or phi to clinical assessment was tested in populations of men whose clinician had already made the decision that a second biopsy was necessary. The study entry criteria differ across the included studies, suggesting that the disease characteristics of the study populations may be heterogeneous. Therefore, some of these study populations include patients for whom the reason for referral for a second biopsy is clear, while in other studies the reason is unclear.

The two papers[78,79] in which the populations had not been referred for a second biopsy report diagnostic accuracy outcomes for participants in the placebo arm of the REDUCE trial. Participants in this trial constitute a low-risk population of men who do not exhibit any clinical signs to suggest that a second biopsy might be appropriate.

These differences in patient selection criteria mean that it may not make clinical sense to apply the results of this review, without clearly stated caveats, to all men with negative or equivocal biopsy results. Furthermore, none of the included studies was conducted in the UK.

Variation in clinical assessment

An additional issue that makes it difficult to draw firm conclusions from the data is that the representation of clinical assessment varies in the included studies. Although clinical assessment is not standardised in practice, it is difficult to meaningfully compare the results of studies which have markedly different representations of clinical assessment. For example, it may be inappropriate to compare the results of clinical assessment (a DRE, age) + phi versus clinical assessment (a DRE, age, family history, prostate volume, ethnicity and PSA level) + PCA3.

Reference standard

How the reference standard was used was unclearly reported in some of the studies, and this is an indication of a poor-quality study. As the choice and conduct of the reference standard may be affected by the results of the comparator or intervention test, studies that stated explicitly that the MRI results influenced the choice of reference standard were assessed as having a high risk of bias. Finally, the reference standard (prostate biopsy) is an imperfect diagnostic tool as it does not detect all cancers. Without a gold standard that offers 100% specificity and 100% sensitivity, it is difficult to confidently assess the accuracy of competing diagnostic strategies. The applicability of all of the included studies was, therefore, assessed as being of high concern and so the quality of the studies was inevitably low.

Summary of key findings

Seventeen[45,46,85,86,89–92,94,96,97,99,102–106] relevant studies of within-study comparisons were identified for inclusion in this review of the clinical effectiveness of the PCA3 assay and the phi in combination with existing tests, scans and clinical judgement in the diagnosis of prostate cancer in men who are suspected of having malignant disease and in whom the results of an initial prostate biopsy were negative or equivocal. Data were available from the included studies to compare 10 distinct sets of comparisons. The following four comparisons are most relevant to NHS clinicians:

- clinical assessment versus clinical assessment + PCA3
- clinical assessment versus clinical assessment + phi
- clinical assessment + MRI versus clinical assessment + MRI + PCA3
- clinical assessment + MRI versus clinical assessment + MRI + phi.

Addition of PROGENSA prostate cancer antigen 3 assay score to clinical assessment

Study findings varied depending on the outcome metric used in the analysis.

Eight[45,46,85,86,90,99,102] efficacy comparisons of clinical assessment versus clinical assessment + PCA3 using AUC data demonstrated that the addition of PCA3 score to clinical assessment led to improvements (1–19%) in diagnostic accuracy. Two studies[45,86] used the PCA3 score as a dichotomous variable using thresholds of 25 and 35; the remainder used PCA3 score as a continuous variable. The representation of clinical assessment varied across the studies.

Seven[45,86,89,99,106] efficacy comparisons using multivariate ORs also showed that the addition of PCA3 score to clinical assessment increased diagnostic accuracy compared with clinical assessment alone; six of the seven ORs for PCA3 score were statistically significant and one OR was borderline. Two studies[86,106] reported ORs for unit increase in PCA3 score, four studies[45,86,89] used the PCA3 score as a dichotomous variable (25, 35, 39 and 50), and in one study[99] the threshold used was unclear. The representation of clinical assessment varied across the studies.

Diagnostic performance was assessed in terms of sensitivity and specificity in only one study.[105] The study by Tombal *et al.*[105] showed that the addition of PCA3 score to clinical assessment led to a small decrease in sensitivity (from 75% to 66%) but led to a marked increase in specificity (from 26% to 71%).

Studies[45,99,102] which fixed sensitivity at 80% or 90% and derived specificity from logistic regression models also reported mixed results. The results from studies[90,99,102] that assessed efficacy using decision curve analyses were also mixed, with no clear benefit associated with adding PCA3 score to clinical assessment; increased net benefit was shown in two studies[90,99] when the risk threshold was set at 25%. The implications of adding the PCA3 score to clinical assessment are not clear and it is not possible to identify a single-threshold value for use in a clinical setting.

Addition of Prostate Health Index to clinical assessment

Four studies[92,99,102,104] compared clinical assessment versus clinical assessment plus phi, and demonstrated higher AUC estimates (2–6%) when phi was included. All of the studies used the phi result as a continuous variable. The representation of clinical assessment varied across studies. Two studies[92,99] reported multivariate ORs for clinical assessment + phi and the results indicated that the addition of phi to clinical assessment led to a small improvement in diagnostic accuracy (OR > 1). No studies reported sensitivity and/or specificity and the studies[92,99,102] reporting results for derived sensitivity and specificity or decision curve analysis have conflicting results. The implications of adding phi to clinical assessment are not clear and it is not possible to identify threshold values for use in a clinical setting.

The addition of PCA3 score to clinical assessment + MRI to the diagnostic process did not have a noticeable impact on discrimination.

Addition of PROGENSA prostate cancer antigen 3 assay score to clinical assessment + magnetic resonance imaging

Two studies[90,99] assessed the incremental gain in diagnostic accuracy resulting from adding PCA3 score to clinical assessment + MRI using AUC estimates; the addition of PCA3 score in both studies had negligible impact. Both studies used PCA3 score as a continuous variable. The study by Busetto *et al.*[90] employed a MRI-targeted biopsy, whereas the study by Porpiglia *et al.*[99] did not. The OR for PCA3 score when added to clinical assessment + MRI was not statistically significant.

No studies reported sensitivity or specificity. Porpiglia *et al.*[99] reported results at set levels of specificity and showed that adding PCA3 score had minimal effect on derived sensitivity; Porpiglia *et al.*[99] also demonstrated that adding PCA3 score to clinical assessment + MRI at set levels of sensitivity had minimal effect on derived specificity (–5.9% to 0.8%). In these two studies,[90,99] the results of decision curve analyses showed that the addition of the PCA3 score to clinical assessment + MRI did not improve diagnostic accuracy when added to clinical assessment + MRI at threshold probabilities between 10% and 50%.

Addition of Prostate Health Index to clinical assessment + magnetic resonance imaging

Only the study by Porpiglia *et al.*[99] assessed the gain associated with adding phi to clinical assessment + MRI. Adding phi to clinical assessment + MRI had no effect on AUC. The OR for phi when added to clinical assessment + MRI was not statistically significant. At set levels of sensitivity and specificity, the addition of phi had a minor effect on derived specificity and sensitivity. In this study,[90,99] the results of decision curve analyses showed that the addition of phi to clinical assessment + MRI did not improve diagnostic accuracy at threshold probabilities between 10% and 60%.

The addition of phi to clinical assessment + MRI to the diagnostic process did not have a noticeable impact on discrimination.

Systematic reviews

None of the systematic reviews identified for inclusion in the clinical validity review included comparisons that assessed the addition of the PCA3 assay or phi to clinical assessment with or without MRI.

Clinical utility review

The planned methods for the clinical utility review were as described in the protocol.[111] No studies were identified for inclusion in the clinical utility review and therefore no results can be reported.

Chapter 3 Assessment of cost-effectiveness

There are two distinct elements to this section on cost-effectiveness. First, the methods and results of a literature search for economic evidence are presented. Second, the EAG's independent de novo economic model is described alongside comprehensive interpretation of the results generated by the model.

This report contains reference to confidential information provided as part of the NICE appraisal process. This information has been removed from the report and the results, discussions and conclusions of the report do not include the confidential information. These sections are clearly marked in the report.

Systematic review of existing cost-effectiveness evidence

Search strategy

Full details of the main search strategy conducted by the EAG are presented in *Chapter 2*, *Search strategy: analytical validity review*. The EAG did not use specific economics-related search terms in the main strategy, as all of the potential references were scanned for studies containing economic evidence.

Inclusion and exclusion criteria

Three reviewers (AN, AB and JH) independently screened all titles and abstracts identified via searching and set aside the subset of records with the term 'cost' or 'economic' included in the title or abstract (stage 1). At stage 2, two reviewers (AB and SB) independently screened the titles and abstracts of the records that were potentially relevant to the cost-effectiveness review. Full-paper manuscripts of any titles and abstracts that were considered relevant by either reviewer were obtained. The relevance of each study was then assessed (AB and SB) in accordance with the criteria set out in *Table 29*. Studies that did not meet the criteria were excluded. Any discrepancies were resolved by consensus and, where necessary, a third reviewer was consulted.

Data extraction and quality assessment strategy

The EAG planned to extract data relating to both study design and quality by two reviewers (AB and SB) into an Excel spreadsheet (Excel Software, Henderson, NV, USA). The EAG planned to quality assess all economic evaluations identified for inclusion in the review according to the Drummond and Jefferson[112] 10-point checklist.

Results: quantity and quality of research available

After deduplication, the 2249 remaining titles and abstracts (when available) were screened for inclusion at stage 1. Of these, 2146 references were immediately excluded because they did not include the term 'cost' or 'economic' in either the title or the abstract. The remaining 103 records were assessed for eligibility and 99 were excluded because they did not include the relevant comparators or did not consider an eligible

TABLE 29 Inclusion and exclusion criteria

Item	Inclusion criteria	Exclusion criteria
Intervention or comparator	PCA3, phi	–
Study design	Full economic evaluation	Methodological paper, letter[a] or abstract[b]
Perspective	UK or European perspective	Non-European perspective
Population	Men suspected of having prostate cancer who had had at least one negative or equivocal biopsy	Screening population

a Letters were included if they were related to a study already included in the review.
b Abstracts were judged for inclusion at the very end of the inclusion process in order to ascertain whether or not sufficient information was available for the abstract to be included in the review.

study population. Full texts were obtained for four references.[113–116] However, none of the four references met the study inclusion criteria and they were, therefore, excluded from the systematic review. The PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram for the cost-effectiveness review is shown in *Figure 4*.

The search carried out by the EAG identified the two studies[114,116] (one considering the PCA3 assay and the other phi) that had been summarised in the NICE scope.[49] One of these studies[116] focused on a screening population and was carried out from a US health-care perspective. The second study was carried out in France, but only 21.1% of the population had had a prior negative biopsy.[114] Both of these studies were, therefore, excluded from the EAG's review. The two further studies (both abstracts) identified by the EAG's search were a study that focused on patients with a prior negative biopsy that was carried out from a US perspective[115] and a study that considered a screening population.[113] Both of these studies were, therefore, also excluded from the EAG's review.

Details of the four studies identified by the EAG search and the reasons for their exclusion from the review are provided in *Table 30*.

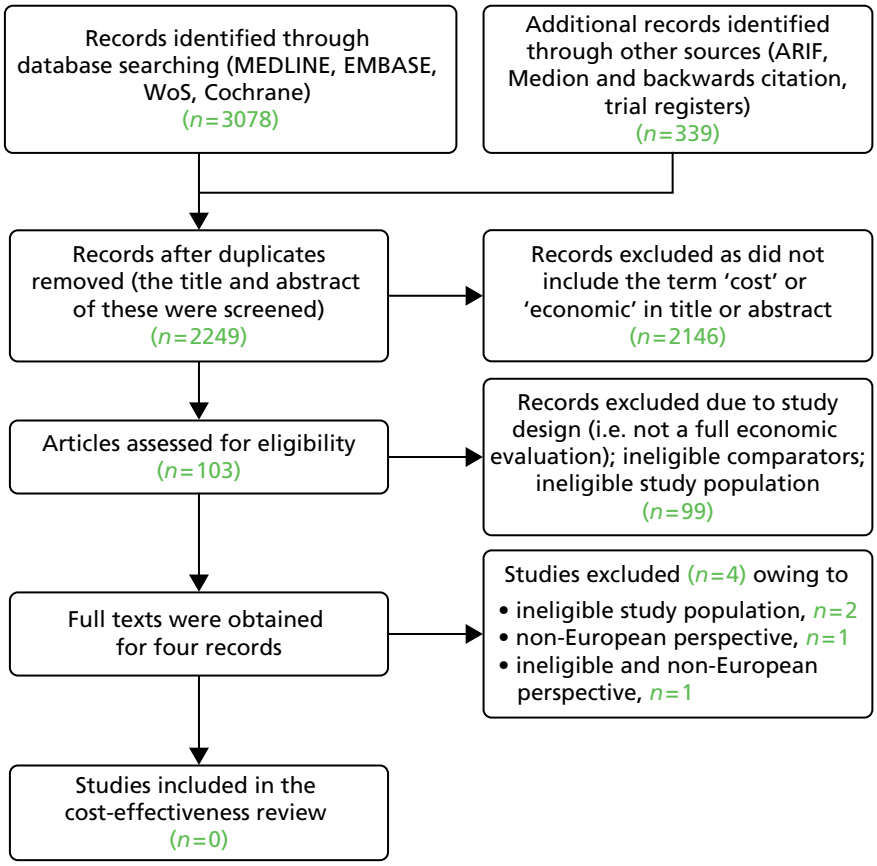

FIGURE 4 The PRISMA flow diagram for the systematic review of cost-effectiveness evidence. ARIF, Aggressive Research Intelligence Facility database; WoS, Web of Science.

TABLE 30 List of the four excluded studies

Study	Title	Reason for exclusion
Excluded studies		
Heijnsdijk 2012[113]	The cost-effectiveness of prostate cancer detection using Beckman Coulter Prostate Health Index	A screening population
Malavaud 2013[114]	Impact of adoption of a decision algorithm including PCA3 for repeat biopsy on the cost for prostate cancer in France	Only 13.2% of the population had had one negative biopsy (an additional 7.9% had had ≥ 2 negative biopsies)
Nepple 2012[115]	Cost-analysis of PCA3 vs. PSA in the detection of prostate cancer in men with a prior negative biopsy	Non-European perspective. Additionally, not a full economic evaluation
Nichol 2011[116]	Cost-effectiveness of prostate health index for prostate cancer detection	A screening population

Conclusions of the External Assessment Group cost-effectiveness literature review

The EAG did not identify any published papers that met the inclusion criteria for the review.

Independent economic assessment

Approach to modelling

The search for economic literature did not identify any studies that evaluated the cost-effectiveness (from a UK NHS perspective) of PCA3 assay or phi, in combination with existing tests, scans and clinical judgement, in the diagnosis of prostate cancer in men suspected of having malignant disease in whom the results of an initial prostate biopsy were negative or equivocal. A de novo economic analysis was therefore undertaken by the EAG.

Modelling effectiveness

A number of different measures are used by researchers to show the relative efficacy of different diagnostic strategies being considered in this assessment (including sensitivity, specificity, AUC, multivariate ORs and decision curve analyses results). Of these measures, those that are the most readily useable in an economic model are sensitivity and specificity. This is because, in combination, these metrics allow a simple comparison to be made of the number of cancers that are correctly identified and the number of unnecessary biopsies that are undertaken when using competing diagnostic strategies.

The differences in benefits and costs arising from the diagnostic strategies can, therefore, be separated into the benefits and costs arising from differences in:

- undetected, untreated cancers (the higher the sensitivity, the higher the rate of detected cancer)
- unnecessary repeat biopsies for patients without cancer (the higher the specificity, the lower the rate of unnecessary biopsies for those without cancer).

Only one of the studies included in the review of clinical validity reported sensitivity and specificity estimates; it was more common for the studies to report derived sensitivity and derived specificity values for a range of different intervention and comparator diagnostic strategies. It was, therefore, not possible for the EAG to undertake a meta-analysis of sensitivity and specificity across trials as the data were unavailable, as explained in *Chapter 2*, *Within-study comparisons: additional data analyses*.

In the review of clinical validity, the included studies present sensitivity and specificity results either for specific test thresholds or, more often, as estimates that are derived from logistic regression models.

The EAG's de novo economic model uses the derived specificities for stated sensitivity levels. The questions that this approach address are 'Given a desired cancer detection rate for the target population, what proportion of the population would need a second biopsy?' and 'What proportion of these second biopsies would be unnecessary?'.

The EAG acknowledges that this approach to modelling may not precisely reflect clinical practice in the NHS in England and Wales. As stated in *Appendix 1*, derived sensitivity or derived specificity estimates are calculated from ROC curves, and it is often not possible to associate a stated sensitivity/specificity combination with a particular threshold of the intervention test.

As sensitivity and specificity are required to populate the model, the EAG also considered using sensitivity values for the tests that were based on the estimated cancer rates associated with the different threshold levels recommended by the manufacturers. However, to translate these values into estimates of sensitivity and specificity, several other pieces of information would be required:

- the proportion of patients that would be at each threshold
- how clinical assessment of other patient information (PSA level, DRE findings, etc.) would influence whether or not a biopsy would be recommended at different thresholds
- the proportion of patients who, on being recommended to receive a second biopsy, choose to receive one.

As this information is not readily available, the values for the above would have had to be assumed to generate sensitivity–specificity combinations. The EAG considered that such an approach would generate considerable uncertainty and that a more robust approach would be to focus on the available evidence on derived sensitivity–specificity combinations that is underpinned by findings from clinical studies, if not from clinical practice.

The use of derived specificity at stated sensitivity levels allows a fair comparison to be made between different testing strategies. Using this approach, the percentage of cancers that are detected is always the same regardless of the strategy chosen, but the number of biopsies required to detect these cancers differs. This simplifies the decision problem, negating issues such as which test threshold values to use in the model and how test results interplay with patient and clinician risk preferences.

As the percentage of detected underlying cancers is the same for all diagnostic strategies in the EAG model, the proportion of patients with treated and untreated cancers is also the same for all diagnostic strategies. Consequently, patient benefits and costs from cancer detection and treatment are the same for all diagnostic strategies. Therefore, as specificity levels for a given level of sensitivity differ across the comparator diagnostic strategies, the differences in patients' benefits and costs between strategies are driven only by the difference in unnecessary biopsies carried out on patients without cancer. Although there is some evidence that biopsies may be linked to increased mortality in the short term, this is as yet unproven.[117] The EAG model, therefore, only considers the short-term impact of a biopsy on QoL and associated complications.

Population

As stated in *Chapter 2*, *Within-study comparisons: baseline characteristics*, the populations described in the studies included in the clinical validity review are mainly made up of men who have been referred for a second biopsy because, following a negative initial biopsy result, clinicians still suspect that malignant prostate cancer is present. Data from one study population were reported in two publications that included men in a clinical trial who were scheduled for a second biopsy without obvious clinical signs of prostate cancer and a negative or equivocal initial biopsy. However, data from these two studies[78,79] did not provide evidence on sensitivity and specificity in a form that could be incorporated into the economic model and so the model population comprises those for whom a suspicion of cancer remains despite negative or equivocal results following their initial biopsy.

In the EAG model, the assumed prevalence of undetected cancer after initial biopsy is 24%. This is based on a study of cancer detection rates using a saturation biopsy[17] for a cohort of patients with a previous negative or equivocal biopsy result but persistently elevated PSA levels (> 4 ng/ml) and/or an abnormal DRE.

Comparators

Clinical validity data were available for the following diagnostic strategies and these are the diagnostic strategies that have, therefore, been included in the economic model:

- clinical assessment
- clinical assessment + PCA3
- clinical assessment + phi
- clinical assessment + PCA3 + phi
- clinical assessment + mpMRI
- clinical assessment + mpMRI + PCA3
- clinical assessment + mpMRI + phi
- clinical assessment + mpMRI + PCA3 + phi.

Model structure

A schematic of the diagnostic strategy used in the model is shown in *Figure 5*. Following an initial negative biopsy, clinical assessment alone, or results from an alternative diagnostic strategy are used by the clinician to decide whether or not to recommend a second biopsy.

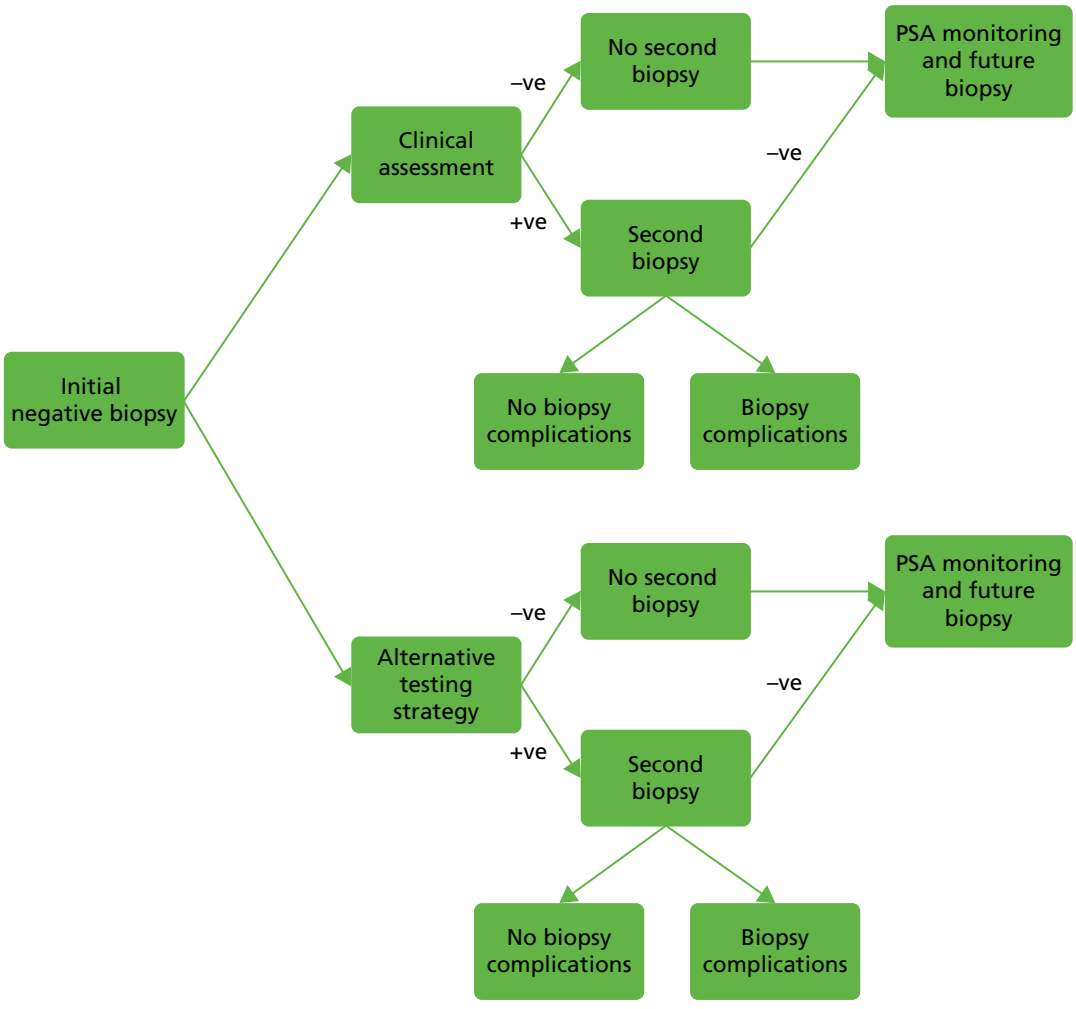

FIGURE 5 Model pathway. –ve, negative; +ve, positive.

As part of the development of the NICE clinical guideline CG175,[11] an economic model was produced which explored the use of mpMRI before TRUS-guided prostate biopsy in men with suspected prostate cancer. Following the approach taken in the CG175 MRI model,[11] the EAG has assumed that all patients who are recommended for a second biopsy choose to have a biopsy and all those for whom a second biopsy is not recommended do not demand one. Patients having a biopsy may experience a short-term deterioration in QoL; in addition, biopsies may result in complications.

In the CG175[11] MRI model, patients whose second biopsy results are negative or equivocal do not immediately have a third biopsy; instead they enter a PSA monitoring phase. When choosing the most appropriate monitoring and future biopsy strategy to employ, the EAG considered the PSA monitoring strategy used in the CG175[11] MRI model and also drew on the content of a recently published HTA report by Mowatt *et al.*,[14] which included a model that explored the cost-effectiveness of mpMRI to aid the localisation of prostate abnormalities for biopsy; the authors of CG175[11] also drew on data reported in the HTA report.

The CG175[11] MRI model, which assessed the use of TRUS biopsy with or without mpMRI, included the following assumptions:

- If a first TRUS biopsy is negative, 50% of patients are offered, and accept, a second TRUS biopsy, which is undertaken 3 months after the first biopsy.
- None of the patients in whom the first biopsy, a mpMRI-targeted-biopsy, is negative is offered a second biopsy.
- All patients in whom a first mpMRI-targeted biopsy is negative or in whom a first or second TRUS biopsy is negative (if a second biopsy is undertaken) are assumed to remain at risk of cancer and their PSA level is monitored. It is assumed that after 1 year, 2 years and 3 years, respectively, 25%, 50% and 100% of these patients are sent for a second investigation. Thus, after 3 years, all patients will have had two investigations. The second investigation is either a repeat TRUS biopsy or mpMRI followed by a biopsy (if the mpMRI indicates that a biopsy should be carried out).
- Patients with negative findings after a second investigation continue to have their PSA level monitored but, after 1 year, 2 years and 3 years, 25%, 50% and 100% of these patients, respectively, are sent for a saturation biopsy.

Under these assumptions, all patients with cancer have a correct diagnosis after 6 years, with the majority of those whose cancer was originally missed having a correct diagnosis after 3 years.

The population in the Mowatt *et al.*[14] model comprised patients who had already been selected to undergo a second biopsy. The Mowatt *et al.*[14] model includes the following assumptions:

- Following a second biopsy, patients who are classified as having no cancer have their PSA level monitored every 6 months for a year. Those with undetected cancer are assumed to have a rising PSA level at the end of the year, whereas the PSA level of those without cancer is assumed to be stable.
- Patients with an elevated PSA level are offered a saturation biopsy and 90% agree to undergo this procedure.

The PSA monitoring assumption used in the CG175[11] MRI model is that every man shown to be cancer free at the time of his initial biopsy, and who is not sent for a second biopsy, requires PSA monitoring and goes on to have one, possibly two, further biopsies over the next 3–6 years. Those in whom the results of a second biopsy are negative or equivocal do not enter PSA monitoring and so do not incur PSA monitoring costs or have the potential to undergo a third biopsy. This means that, at best, the comparator diagnostic strategies can achieve a reduction in the number of second biopsies at the expense of up to 6 years of PSA monitoring and at least one further biopsy. In the EAG model, such an assumption would result in the optimal strategy being to immediately carry out a further biopsy on everyone whose initial biopsy was negative or equivocal and undertake no additional PSA monitoring.

In the base case, the EAG has, therefore, adopted the assumption used in the Mowatt *et al.*[14] model, that is that patients with undiagnosed cancer, whether or not they have undergone a second biopsy, will continue to have elevated PSA levels. In addition, the EAG has assumed that 25% of men without cancer will also continue to have a rising PSA level and that, at 1, 2 and 3 years, 25%, 50% and 100% of patients, respectively, with a rising PSA level will have a saturation biopsy. The EAG has included sensitivity analyses to explore the impact of 0%, 25%, 50% and 75% of men with a negative second biopsy entering PSA monitoring.

In addition, the following two scenario analyses have been undertaken by the EAGA:

- the monitoring and second biopsy strategy used in the CG175[11] MRI model
- the monitoring strategy used in the Mowatt *et al.* model.[14]

Time horizon

The NICE reference case[118] states that the time horizon of economic models should be:

> *Long enough to reflect all important differences in costs or outcomes between the technologies being compared.*
>
> *p. 57*[118]

In the EAG economic assessment, the differences in costs and outcome are limited to:

- the differences in costs and complication-related outcomes from the additional biopsies indicated by the testing strategies
- the costs and outcomes of any monitoring of patients who are either indicated as negative for cancer by the testing strategy or who have a negative repeat biopsy.

As a PSA monitoring strategy can run for several years, the time horizon of the model is limited to the time that patients spend within any such strategy. The monitoring strategy is independent of the diagnostic strategies assessed in the model, so unless there is a lifetime PSA monitoring strategy the model does not require a lifetime horizon. In the base case, the PSA monitoring strategy runs for 3 years so the time horizon is also 3 years. The time horizons for the scenario analyses exploring the impact of the PSA monitoring strategies used in the CG175[11] MRI model and the Mowatt *et al.*[14] model are 6 years and 1 year, respectively.

There is currently no unequivocal evidence that the biopsy procedure increases mortality. In the EAG model, the proportion of cancers identified and treated is assumed to be identical regardless of testing strategy. Thus, overall mortality rates will also be identical across testing strategies and so were not included in the model. Mortality could influence costs during the monitoring phase, but Bill-Axelson *et al.*,[119] who collected data on 348 patients with localised cancer being monitored by watchful waiting, found very low mortality rates over 3 years (under 5%). As almost all of the population in the EAG model who enter PSA monitoring do not have prostate cancer, and those who do are picked up and treated in a relatively short period of time, mortality rates would be even lower for patients in the EAG model than the levels reported by Bill-Axelson *et al.*[119] Given this, the impact on cost from introducing mortality into the model would be negligible and so has been excluded.

In addition, the following two scenario analyses were undertaken by the EAG:

- the monitoring and second biopsy strategy used in the CG175[11] MRI model
- the monitoring strategy used in the Mowatt *et al.* model.[14]

These two scenarios can be considered to represent the 'least costly' (Mowatt *et al.*[14]) and 'most costly' (CG175 MRI model[11]) PSA monitoring scenarios.

A sensitivity analysis which involved varying the percentage of patients without cancer who had persistently elevated PSA levels, and so would require re-biopsy while under PSA monitoring, was considered. However, the Mowatt *et al.*[14] scenario considers the case that no men without cancer continued to have an elevated PSA level and the CG175[11] MRI model scenario considered the case where all men without cancer continued to have an elevated PSA level. These two scenarios represent the two extremes of the percentage of men with persistently elevated PSA levels. Thus, the EAG felt that the inclusion of a sensitivity analysis varying the percentage of patients without cancer who had a persistently elevated PSA level would be uninformative and so such an analysis was not undertaken.

Model parameters

Clinical effectiveness

The clinical effectiveness estimates for different diagnostic testing strategies have been taken from the available published clinical evidence (see *Chapter 2*). As it has not been possible to carry out between-trial analysis and pool effectiveness data, the data in each study have been considered independently. Three studies provide information on derived specificity at differing sensitivity levels: Porpiglia *et al.*,[99] Scattoni *et al.*[102] and Gittelman *et al.*[45] Of these, Porpiglia *et al.*[99] provide data for all of the diagnostic testing strategies considered by the other two studies[102,45] plus additional strategies that include mpMRI. Therefore, results from Porpiglia *et al.*[99] have been used in the base case, while data from the other two studies[102,45] have been used in scenario analyses to explore the effect that different levels of effectiveness (and elements of clinical assessment) might have on conclusions.

Clinical validity data reported by Tombal *et al.*[105] were considered for incorporation into the model. However, while sensitivity and specificity values are reported they are only reported for a specific PCA3 threshold value. Reported results, therefore, do not allow comparisons of specificity rates (at stated sensitivity levels) for the PCA3 assay against alternative strategies and so have not been used in the model.

Clinical advice to the EAG is that it is very difficult to pinpoint a precise sensitivity estimate that most clinicians use in clinical practice. Furthermore, clinical decisions regarding biopsy referral are made with sensitivity implicitly in mind but not explicitly stated. Choosing a sensitivity level to use in the EAG base case was necessarily arbitrary; 90% sensitivity was chosen as it is the middle estimate of the three levels of sensitivity data that were provided in the study reported by Porpiglia *et al*.[99] The impact of using sensitivity levels of 80% and 95% was explored in scenario analyses. Only the Gittelman *et al.*[45] paper included data on the variance, or range, of the derived specificity estimates and, therefore, it has not been possible to vary these values in the probabilistic sensitivity analysis.

Studies reporting clinical validity data use a biopsy as the reference standard despite biopsies not being 100% sensitive or specific. To check the impact that this assumption has on model findings, sensitivity analyses were undertaken in which the proportion of cancers detected at biopsy were set at 50% and 100%.

As stated in CG175,[11] mpMRI-targeted biopsy has greater sensitivity and specificity than TRUS biopsy alone. However, in the Porpiglia *et al.* study,[99] which includes data on the sensitivity and specificity of mpMRI in combination with PCA3 assay or phi, the urologists were blinded to the mpMRI results before a biopsy was taken. Although mpMRI can influence biopsy sensitivity and specificity, the EAG has assumed that a biopsy after mpMRI does not influence the final diagnostic accuracy of the second biopsy. This assumption will put downwards pressure on the efficacy of mpMRI, but the decision question is ultimately about the addition of PCA3 assay or phi with or without MRI. Thus, this assumption will influence only the comparison of mpMRI with strategies without mpMRI, biasing the findings against mpMRI.

The sensitivity/derived specificity values used for the different diagnostic strategies are presented in *Table 31*.

TABLE 31 Sensitivity and derived specificity values for different diagnostic strategies

Study	Sensitivity (%)	Derived specificity
Clinical assessment		
Porpiglia 2014[99]	80	27.1
	90	12.7
	95	0.8
Scattoni 2013[102]	80	49.0%
	90	35.0%
Gittelman 2013[45]	90	18.9%
Clinical assessment + PCA3		
Porpiglia 2014[99]	80	37.3
	90	11.0
	95	8.5
Scattoni 2013[102]	80	47.0%
	90	25.0%
Gittelman 2013[45]	90	41.5%
Clinical assessment + phi		
Porpiglia 2014[99]	80	24.6
	90	2.5
	95	1.7
Scattoni 2013[102]	80	66.0%
	90	37.0%
Clinical assessment + phi + PCA3		
Porpiglia 2014[99]	80	39.8
	90	22.9
	95	7.6
Scattoni 2013[102]	80	49.0%
	90	33.0%
Clinical assessment + mpMRI		
Porpiglia 2014[99]	80	93.2
	90	89.0
	95	64.4
Clinical assessment + mpMRI + PCA3		
Porpiglia 2014[99]	80	93.2
	90	89.8
	95	58.5

continued

TABLE 31 Sensitivity and derived specificity values for different diagnostic strategies (*continued*)

Study	Sensitivity (%)	Derived specificity
Clinical assessment + mpMRI + phi		
Porpiglia 2014[99]	80	93.2
	90	89.8
	95	65.3
Clinical assessment + mpMRI + phi + PCA3		
Porpiglia 2014[99]	80	93.2
	90	89.8
	95	56.8

Biopsy complications

The CG175[11] MRI model provides a detailed description of biopsy complication rates identified by a literature review. As CG175[11] was published in the same year as the EAG model was constructed, it was postulated that it might be appropriate to use the complication rates used in CG175[11] in the EAG model. Citation searches were carried on the relevant studies.[120,121] These searches failed to identify more up-to-date rates and so the complication rates used in the CG175[11] MRI model were also used in the EAG model.

Biopsy complication rates are shown in *Table 32*. The costs associated with biopsy complications that are included in the model should be considered as conservative, as literature searches failed to identify any published studies reporting the costs associated with sepsis or antibiotic resistance. To establish whether or not this omission could affect findings, a sensitivity analysis was undertaken in which all complication costs were increased by 100%.

Values for the upper and lower CIs have been used to model pessimistic and optimistic resource use scenarios.

TABLE 32 Biopsy complication rates

Event	Probability (95% CI)	Distribution for PSA	Source
Biopsy complication	0.117 (0.100 to 0.137)	Beta distribution: alpha 134; beta 1013	Rosario 2012[121]
Probability of hospital admission given biopsy complication	0.112 (0.069 to 0.176)	Beta distribution: alpha 15; beta 119	Rosario 2012[121]
Reasons for hospital admission			
Urinary infection	0.716 (0.675 to 0.738)	Dirichlet distribution: alpha 556	Nam 2010[120]
Urinary bleeding	0.194 (0.166 to 0.221)	Dirichlet distribution: alpha 151	Nam 2010[120]
Urinary obstruction	0.090 (0.081 to 0.124)	Dirichlet distribution: alpha 79	Nam 2010[120]
Biopsy-related consultation after complication	0.888 (0.824 to 0.931)	Beta distribution: alpha 119; beta 15	Rosario 2012[121]
Location of consultation			
GP	0.773 (0.690 to 0.839)	Dirichlet distribution: alpha 92	Rosario 2012[121]
Urology department nurse	0.118 (0.071 to 0.138)	Dirichlet distribution: alpha 14	Rosario 2012[121]
Other – NHS Direct	0.109 (0.065 to 0.178)	Dirichlet distribution: alpha 13	Rosario 2012[121]

Cost year
Unless otherwise stated, the costs are in 2014 GBP.

Cost of clinical assessment
The diagnostic strategies included in the EAG model comprise one or more of four separate components: clinical assessment, phi, PCA3 assay and mpMRI. While the nature of clinical assessment varies between studies, within studies it is the same for all participants. As the model does not pool data but looks at evidence from studies individually, and clinical assessment is required in all diagnostic strategies, there is no requirement to model the cost of clinical assessment, as it will make no difference to costs between strategies.

Cost of PROGENSA prostate cancer antigen 3 assay
The PCA3 assay costs provided by the manufacturer have been calculated by applying UK costs to resource use obtained from a US study. (Commercial-in-confidence information has been removed.) The estimated cost of the PCA3 testing kit was given as £164.67 including value-added tax (VAT) and (commercial-in-confidence information has been removed). This higher cost of £175.11 has been used in a scenario analysis.

The cost of the PCA3 assay has not been varied in the probabilistic sensitivity analysis.

Cost of Prostate Health Index
The manufacturer provided the cost of a single phi test. This was £89.83 including VAT. (Commercial-in-confidence information has been removed.) Deterministic sensitivity analysis has been used to explore the impact of the number of tests being 50% lower and 50% higher than this figure, effectively changing the cost of the test by $\pm$ 50%.

With no evidence available on the distribution of tests conducted in a year, the cost of phi testing has not been varied in the probabilistic sensitivity analysis.

Cost of multiparametric magnetic resonance imaging
The CG175[11] report provided detailed costings of mpMRI and these costings were used in the EAG model, updated where necessary with up-to-date unit or NHS Reference Costs (2012/13).[18]

The unit costs for staff time and equipment costs used in the CG175[11] MRI model were taken directly from Mowatt *et al.*[14] However, staffing costs, which were based on bottom-up calculations of staff time, were increased in the CG175[11] MRI model, as the NICE Guidelines Development Group considered that they were an underestimate. Resource use and costs of mpMRI are provided in *Table 33*.

As no measures of dispersion were available on the cost per hour, or time, per patient no sensitivity analyses were undertaken around this cost.

TABLE 33 Resource use and costs associated with using mpMRI

Resource	Time per patient (minutes)	Cost per hour (£)	Total cost (£)
Radiographer 1	43.33	48.33	34.91
Radiographer 2	43.33	50.00	36.11
Radiologist – consultant	45.00	162.00	121.50
Equipment cost per patient	–	–	88.42
Administration and consumable cost	–	–	34.62
Total mpMRI cost	–	–	315.56

Costs of biopsy

Biopsy costs are dependent on the type of biopsy undertaken. In the base case, the EAG has assumed that the second biopsy will be a TRUS biopsy carried out as an outpatient appointment.

The assumption used in the CG175[11] MRI model is that mpMRI results influence whether a TRUS or a transperineal biopsy is performed. Incorporating this assumption into the EAG model is difficult as it contradicts the model assumption that mpMRI does not influence the nature, or accuracy, of the second biopsy, an assumption made because the evidence[99] from which efficacy data were drawn explicitly blinded clinicians performing the biopsy to the mpMRI results. Therefore, while in the base case the model assumption is that all patients have a TRUS biopsy as an outpatient procedure, a scenario analysis has explored the impact on results of a situation in which 50% of second biopsies are transperineal biopsies carried out as day-case procedures.

The CG175[11] MRI model uses NHS reference costs as the basis for costing the different biopsy procedures. The clinical experts advising on the CG175[11] model considered that NHS reference costs did not take adequate account of pathology costs and, therefore, that they underestimate the true cost of biopsy. The developers of the CG175[11] MRI model therefore increased the cost of the HRG for histopathology by adding an estimate provided by a NHS Pathology Department in Bristol. A saturation biopsy was assumed to have a higher cost than a routine biopsy, as the former procedure generates a greater number of cores for analysis. The biopsy costs used in the CG175[11] MRI model were based on NHS reference costs from 2011/12. These have been updated to 2012/13 prices by the EAG (*Table 34*).

Costs of biopsy complications

In line with the CG175[11] MRI model, the model developed by the EAG uses the costs of biopsy complications reported by Mowatt *et al.*[14] (updated to 2012/13 prices). However, some HRG codes have changed slightly since the Mowatt *et al.*[14] model was developed; where appropriate, the HRG codes that appear most relevant to the codes reported by Mowatt *et al.*[14] have been used in the EAG model. Biopsy complication costs are presented in *Table 35*.

TABLE 34 Biopsy costs incorporated into the economic model

Cost element	Cost (£)	Probabilistic sensitivity analysis distribution	Source
TRUS (standard) biopsy			
Outpatient	224	Log-normal	Department of Health 2013,[18] NHS reference cost LB27Z[a] in outpatient procedures – urology
Histopathology	112.79		NCCC 2014[11]
Total	336.79		–
Transperineal (standard) biopsy			
Day case	595	Log-normal	Department of Health 2013,[18] NHS reference cost LB27Z[a] in day-case procedures – urology
Histopathology	112.79		NCCC 2014[11]
Total	707.79		–
Saturation biopsy			
Day case	595	Log-normal	Department of Health 2013,[18] NHS reference cost LB27Z[a] in day-case procedures – urology
Histopathology	281.97		NCCC 2014[11]
Total	876.97		–

NCCC, National Collaborating Centre for Cancer.
a LB27Z = minor endoscopic prostate or bladder neck procedures (male) – urology.

TABLE 35 Costs of biopsy complications

Event	Cost (£)	PSA distribution	Source
Hospital stay			
Urinary tract infection	445	Log-normal	Department of Health 2013.[18] HRG LA04Q (Inpatient short-stay general medicine) kidney or urinary tract infections, without interventions, with CC score of 4–7
Urinary bleeding	483	Log-normal	Department of Health 2013.[18] HRG LB18Z (Inpatient short-stay urology). Attention to suprapubic catheter
Urinary obstruction	1504	Log-normal	Department of Health 2013.[18] HRG LB09D (Inpatient short-stay urology) intermediate endoscopic ureter procedures, 19 years and over; HRG LB15E (Inpatient short-stay urology) minor bladder procedures, 19 years and over plus cost of catheter bags at £19.08
Consultation			
GP	45	Not varied	Curtis 2013.[122] 11.7-minute consultation with qualification costs
Urology department nurse	78	Log-normal	NICE CG175 2013[11]
Other – NHS Direct	£20	Not varied	Mowatt 2013[14]

CC, complications.

The costs associated with biopsy complications that are included in the model should be considered as conservative, as literature searches failed to identify any published studies reporting the costs associated with sepsis or antibiotic resistance. To establish if this omission could affect findings, a sensitivity analysis was undertaken in which all complication costs were increased by 100%.

Costs of prostate-specific antigen monitoring

In CG175[11] and in Mowatt *et al.*[14] PSA monitoring was assumed to occur twice a year and to be carried out by a GP practice nurse. A targeted literature search was undertaken by the EAG to identify any additional information that could indicate alternative PSA monitoring strategies, but no information was found that invalidated this assumption.

The cost, estimated to be £19.60, is based on a PSA test cost provided by Newcastle upon Tyne Hospitals NHS Foundation Trust (reported in CG175[11]) and the cost of a consultation with a practice nurse reported by Curtis.[122]

Utility values

The only utility values required in the model are the disutilities associated with a biopsy. A targeted search of the literature was undertaken but no primary studies collecting disutility values specifically associated with prostate biopsy were identified. Neither the CG175[11] MRI model nor the Mowatt *et al.*[14] model considers any disutility from the actual biopsy. Both studies include a discussion of the impact of including the disutility associated with urinary incontinence, should this occur as a biopsy complication or as a result of treatment; however, it is not clear how this is applied, as urinary tract infection, urinary bleeding and urinary obstruction do not necessarily lead to urinary incontinence.

The literature search identified one study (Heijnsdijk *et al.*[117]) that investigated the utility values associated with a PSA screening programme. This study reported a utility decrement of 0.1 that lasted 3 weeks following a biopsy. This utility value was taken from an earlier study[123] that focused on breast cancer biopsy and the duration of decrement was an assumption based on clinical opinion. Although this utility value is not ideal, in the absence of any other evidence, it has been incorporated into the EAG base-case model as a quality-adjusted life-year (QALY) loss of 0.0058 from a prostate biopsy.

It is not clear if the decrement used in the Heijnsdilk *et al.*[117] study is an average value that includes disutility as a result of complications of biopsy [including those listed in the table presenting the costs of biopsy complications (see *Table 35*) and additional complications such as sepsis]. In the absence of evidence to the contrary, the EAG has assumed that there was no additional QALY loss as a result of biopsy-associated complications. This assumption favours less specific strategies and therefore it is likely that it would bias results against the proposed tests.

Heijnsdilk *et al.*[117] also report lower and upper bounds of biopsy-related disutility of 0.06 and 0.13, respectively (QALY losses of 0.00346 and 0.0075, respectively), based on minimum and maximum values reported in the breast biopsy study.[123] As no measure of dispersion of disutility was provided in the published paper beyond these minimum and maximum values, the disutility has not been varied in the probabilistic sensitivity analysis. However, as suggested by the NICE Decision Support Unit report,[124] the uncertainty has been explored using sensitivity analysis.

Discount rates

Both costs and benefits have been discounted at 3.5% per annum, as suggested in the NICE guide to the methods of technology appraisal.[118] A scenario analysis exploring the impact of a discount rates of 0% and 5% per annum was considered. However, as the model was based on a decision tree with linear transition through a pathway, changing the discount rate would change the costs and QALYs in each arm by exactly the same proportion and so leave the incremental cost-effectiveness ratio per QALY gained unchanged.

Uncertainty

Uncertainty in parameter values and the impact this could have on results have been explored both through the scenario analyses and sensitivity analyses previously described and also through probabilistic sensitivity analysis, varying those parameters for which probability distributions could be derived from, or were provided in, the literature. Probabilistic sensitivity analysis results have been presented as cost-effectiveness acceptability curves (CEACs) where different willingness-to-pay thresholds for a QALY are used to show which strategy is likely to have the largest net benefit for that threshold.

Interpreting results

Incremental cost-effectiveness ratios

The results of cost-effectiveness analysis are presented as incremental cost-effectiveness ratios per QALY gained. These are calculated by dividing the difference in costs associated with two alternative strategies by the difference in QALYs:

$$ICER = \frac{Cost\ of\ B - Cost\ of\ A}{QALY\ of\ B - QALY\ of\ A}. \quad (1)$$

Where more than two strategies are being compared, the incremental cost-effectiveness ratio is calculated according to the following process:

- The strategies are ranked in terms of cost, from least to most expensive.
- If a strategy is more expensive and less effective than the preceding strategy it is said to be 'dominated' and is excluded from further analysis.
- Incremental cost-effectiveness ratios are then calculated for each strategy compared with the next most expensive non-dominated option. If the incremental cost-effectiveness ratio for a strategy is higher than that of the next most effective strategy, then it is ruled out by 'extended dominance'.
- Incremental cost-effectiveness ratios are recalculated excluding any strategy subject to dominance or extended dominance.
- The non-dominated strategies form an 'efficiency frontier' of strategies that are cost-effective and can then be judged against the value of an incremental cost-effectiveness ratio that is generally considered cost-effective by NICE, that is £20,000–30,000 per QALY gained.

Base-case results

The model was executed with a hypothetical cohort of 1000 patients. The results throughout this section are the values generated from the model for this cohort.

Total number of biopsies

The different number of biopsies under each diagnostic strategy drives the different patient outcomes in the model. In the base case, the total number of biopsies is split into second biopsies recommended by the testing strategy and biopsies undertaken during PSA monitoring (*Table 36*).

Under the base-case PSA monitoring scenario, all patients without a second biopsy, or with a negative second biopsy, enter PSA monitoring. The total number of these patients is the same regardless of the strategy; therefore, the number of patients undergoing a repeat biopsy during PSA testing is independent of the strategy chosen and is always the same.

Mean costs and benefits

Costs and QALYs generated using the base-case parameter values are shown in *Table 37*.

TABLE 36 Total number of biopsies undertaken (base case: 90% sensitivity)

Strategy	Second biopsies	Biopsies while under PSA monitoring	Total biopsies
Clinical assessment	879	220	1099
Clinical assessment + phi	957	220	1177
Clinical assessment + PCA3	892	220	1112
Clinical assessment + phi + PCA3	802	220	1022
Clinical assessment + mpMRI	300	220	520
Clinical assessment + phi + mpMRI	294	220	514
Clinical assessment + PCA3 + mpMRI	294	220	514
Clinical assessment + phi + PCA3 + mpMRI	294	220	514

TABLE 37 Costs and QALYs (base case: 90% sensitivity)

Strategy	Test costs (£)	Biopsy costs (£)	Biopsy complication costs (£)	PSA monitoring costs (£)	Total costs (£)	Total QALY loss
Clinical assessment	0	481,088	15,168	83,007	579,264	6.29
Clinical assessment + phi	89,830	507,196	16,079	83,007	696,113	6.74
Clinical assessment + PCA3	164,670	485,440	15,320	83,007	748,437	6.36
Clinical assessment + phi + PCA3	254,500	454,980	14,257	83,007	806,745	5.84
Clinical assessment + mpMRI	315,560	285,791	8355	83,007	692,713	2.94
Clinical assessment + phi + mpMRI	405,390	283,743	8284	83,007	780,424	2.91
Clinical assessment + PCA3 + mpMRI	480,230	283,743	8284	83,007	855,264	2.91
Clinical assessment + phi + PCA3 + mpMRI	570,060	283,743	8284	83,007	945,094	2.91

Total costs may be higher than the sum of individual costs owing to rounding.

Incremental analysis

The incremental results from the base-case analysis are presented in *Table 38*.

Summary of base-case results

The incremental analysis shows that the testing strategies that lie on the efficiency frontier are clinical assessment + mpMRI and clinical assessment + phi + mpMRI. However, the incremental cost-effectiveness ratio per QALY gained for both strategies exceeds the £20,000–30,000 threshold that NICE generally considers cost-effective.

Scenario analysis

Full results are presented in tables for each of the scenario analyses that alter (from the base case) the number of biopsies undertaken.

Results are not shown for the scenario in which 50% of patients who have mpMRI have a transperineal rather than a TRUS biopsy. This is because this scenario will result in an increase in the cost of mpMRI but will not alter effectiveness. The consequence of this is that the incremental cost-effectiveness ratio per QALY gained will be greater than in the base case. As the base-case incremental cost-effectiveness ratio per QALY gained is already above the £20,000–30,000 threshold that NICE generally considers cost-effective, results from this scenario would be uninformative.

Varying derived sensitivity

The total numbers of biopsies performed if sensitivity is set at 80% or 95%, using data from the Porpiglia *et al.* study,[99] are shown in *Table 39*.

TABLE 38 Incremental analysis (base case: 90% sensitivity)

Strategy	Discounted costs	Discounted QALYs	Incremental costs (£)	Incremental QALYs	Incremental cost-effectiveness ratio (£)
Clinical assessment	579,264	–6.29	–	–	–
Clinical assessment + mpMRI	692,713	–2.94	113,449	3.35	33,911
Clinical assessment + phi	696,113	–6.74	3399	–3.79	Dominated
Clinical assessment + PCA3	748,437	–6.36	55,724	–3.42	Dominated
Clinical assessment + phi + mpMRI	780,424	–2.91	87,711	0.04	2,500,530
Clinical assessment + phi + PCA3	806,745	–5.84	26,321	–2.93	Dominated
Clinical assessment + PCA3 + mpMRI	855,264	–2.91	74,840	0	Dominated
Clinical assessment + phi + PCA3 + mpMRI	945,094	–2.91	164,670	0	Dominated

TABLE 39 Total number of biopsies undertaken (varying sensitivity)

Strategy	80% sensitivity			95% sensitivity		
	Second biopsies	Biopsies while under PSA monitoring	Total biopsies	Second biopsies	Biopsies while under PSA monitoring	Total biopsies
Clinical assessment	746	250	996	982	205	1187
Clinical assessment + phi	765	250	1015	975	205	1180
Clinical assessment + PCA3	669	250	919	923	205	1128
Clinical assessment + phi + PCA3	650	250	900	930	205	1135
Clinical assessment + mpMRI	244	250	494	499	205	704
Clinical assessment + phi + mpMRI	244	250	494	492	205	697
Clinical assessment + PCA3 + mpMRI	244	250	494	543	205	748
Clinical assessment + phi + PCA3 + mpMRI	244	250	494	556	205	761

Mean costs and benefits

Costs and QALYs generated by varying sensitivity values are shown in *Tables 40* and *41*.

Incremental analysis

The incremental results from varying the sensitivity level are presented in *Tables 42* and *43*.

TABLE 40 Costs and QALYs (80% sensitivity)

Strategy	Test costs (£)	Biopsy costs (£)	Biopsy complication costs (£)	PSA monitoring costs (£)	Total costs (£)	Total QALY loss
Clinical assessment	0	461,359	14,260	84,902	560,521	5.69
Clinical assessment + phi	89,830	467,758	14,483	84,902	656,974	5.80
Clinical assessment + PCA3	164,670	435,251	13,349	84,902	698,173	5.24
Clinical assessment + phi + PCA3	254,500	428,852	13,126	84,902	781,380	5.13
Clinical assessment + mpMRI	315,560	292,169	8358	84,902	700,990	2.79
Clinical assessment + phi + mpMRI	405,390	292,169	8358	84,902	790,820	2.79
Clinical assessment + PCA3 + mpMRI	480,230	292,169	8358	84,902	865,660	2.79
Clinical assessment + phi + PCA3 + mpMRI	570,060	292,169	8358	84,902	955,490	2.79

Total costs may be higher than the sum of individual costs owing to rounding.

TABLE 41 Costs and QALYs (95% sensitivity)

Strategy	Test costs (£)	Biopsy costs (£)	Biopsy complication costs (£)	PSA monitoring costs (£)	Total costs (£)	Total QALY loss
Clinical assessment	0	502,983	16,042	82,060	601,085	6.80
Clinical assessment + phi	89,830	500,679	15,962	82,060	688,531	6.76
Clinical assessment + PCA3	164,670	483,274	15,354	82,060	745,358	6.46
Clinical assessment + phi + PCA3	254,500	485,578	15,435	82,060	837,572	6.50
Clinical assessment + mpMRI	315,560	340,192	10,363	82,060	748,175	4.01
Clinical assessment + phi + mpMRI	405,390	337,889	10,283	82,060	835,621	3.97
Clinical assessment + PCA3 + mpMRI	480,230	355,294	10,890	82,060	928,474	4.27
Clinical assessment + phi + PCA3 + mpMRI	570,060	359,645	11,042	82,060	1,022,807	4.34

Total costs may be higher than the sum of individual costs owing to rounding.

TABLE 42 Incremental analysis (80% sensitivity)

Strategy	Discounted costs (£)	Discounted QALYs	Incremental costs (£)	Incremental QALYs	Incremental cost-effectiveness ratio (£)
Clinical assessment	560,521	–5.69	–	–	–
Clinical assessment and phi	656,974	–5.80	96,452	–0.11	Dominated
Clinical assessment + PCA3	698,173	–5.24	137,651	0.45	Extendedly dominated
Clinical assessment + mpMRI	700,990	–2.79	140,468	2.90	48,467
Clinical assessment + phi + PCA3	781,380	–5.13	80,390	–2.341	Dominated
Clinical assessment + phi + mpMRI	790,820	–2.79	89,830	0	Dominated
Clinical assessment + PCA3 + mpMRI	865,660	–2.79	164,670	0	Dominated
Clinical assessment + phi + PCA3 + mpMRI	955,490	–2.79	254,500	0	Dominated

TABLE 43 Incremental analysis (95% sensitivity)

Strategy	Discounted costs (£)	Discounted QALYs	Incremental costs (£)	Incremental QALYs	Incremental cost-effectiveness ratio (£)
Clinical assessment	601,085	–6.80	–	–	–
Clinical assessment and phi	688,531	–6.76	87,446	0.04	Extendedly dominated
Clinical assessment + PCA3	745,358	–6.46	144,274	0.34	Extendedly dominated
Clinical assessment + mpMRI	748,175	–4.01	147,090	2.79	52,747
Clinical assessment + phi + mpMRI	835,621	–3.97	87,446	0.04	2,215,980
Clinical assessment + phi + PCA3	837,572	–6.50	1951	–2.53	Dominated
Clinical assessment + PCA3 + mpMRI	928,474	–4.27	92,852	–0.30	Dominated
Clinical assessment + phi + PCA3 + mpMRI	1,022,807	–4.34	187,186	–0.37	Dominated

Different prostate-specific antigen monitoring assumptions

The total numbers of biopsies performed when the PSA monitoring strategies assumed in CG175[11] and Mowatt *et al.*[14] are adopted are shown in *Table 44*.

TABLE 44 Total number of biopsies undertaken (different PSA monitoring assumptions)

	CG175[11]			Mowatt *et al.*[14]		
Strategy	Second biopsies	Biopsies while under PSA monitoring	Total biopsies	Second biopsies	Biopsies while under PSA monitoring	Total biopsies
Clinical assessment	879	241	1122	879	22	901
Clinical assessment + phi	957	86	1043	957	22	979
Clinical assessment + PCA3	892	215	1107	892	22	914
Clinical assessment + phi + PCA3	802	396	1198	802	22	824
Clinical assessment + mpMRI	300	1401	1701	300	22	322
Clinical assessment + phi + mpMRI	294	1413	1707	294	22	316
Clinical assessment + PCA3 + mpMRI	294	1413	1707	294	22	316
Clinical assessment + phi + PCA3 + mpMRI	294	1413	1707	294	22	317

Total biopsies may be higher than the sum of biopsies owing to rounding.

Mean costs and benefits

Costs and QALYs generated when the PSA monitoring strategies assumed in CG175[11] and Mowatt *et al.*[14] are adopted are shown in *Tables 45* and *46*, respectively.

TABLE 45 Costs and QALYs (CG175[11] PSA monitoring assumptions)

Strategy	Test costs (£)	Biopsy costs (£)	Biopsy complication costs (£)	PSA monitoring costs (£)	Total costs (£)	Total QALY loss
Clinical assessment	0	448,471	15,423	20,511	484,405	6.35
Clinical assessment + phi	89,830	381,638	13,093	5693	490,254	5.99
Clinical assessment + PCA3	164,670	437,332	15,034	18,042	635,078	6.29
Clinical assessment + phi + PCA3	254,500	515,304	17,752	35,329	822,886	6.72
Clinical assessment + mpMRI	315,560	948,410	32,851	131,354	1,428,175	9.11
Clinical assessment + phi + mpMRI	405,390	953,652	33,034	132,516	1,524,592	9.14
Clinical assessment + PCA3 + mpMRI	480,230	953,652	33,034	132,516	1,599,432	9.14
Clinical assessment + phi + PCA3 + mpMRI	570,060	953,652	33,034	132,516	1,689,262	9.14

Total costs may be higher than the sum of individual costs owing to rounding.

TABLE 46 Costs and QALYs (Mowatt *et al.*[14] PSA monitoring assumptions)

Strategy	Test costs (£)	Biopsy costs (£)	Biopsy complication costs (£)	PSA monitoring costs (£)	Total costs (£)	Total QALY loss
Clinical assessment	0	315,143	10,828	30,733	356,703	5.20
Clinical assessment + phi	89,830	341,251	11,739	30,733	473,552	5.65
Clinical assessment + PCA3	164,670	319,494	10,980	30,733	525,877	5.27
Clinical assessment + phi + PCA3	254,500	289,035	9917	30,733	584,185	4.75
Clinical assessment + mpMRI	315,560	119,845	4015	30,733	470,153	1.85
Clinical assessment + phi + mpMRI	405,390	117,797	3944	30,733	557,864	1.82
Clinical assessment + PCA3 + mpMRI	480,230	117,797	3944	30,733	632,704	1.82
Clinical assessment + phi + PCA3 + mpMRI	570,060	117,797	3944	30,733	722,534	1.82

Total costs may be higher than the sum of individual costs owing to rounding.

Incremental analysis

The incremental results when the PSA monitoring strategies assumed in CG175[11] and Mowatt *et al.*[14] are adopted are presented in *Tables 47* and *48*, respectively.

As the Mowatt *et al.*[14] PSA monitoring scenario results in an identical reduction in biopsy numbers (and therefore cost and QALY loss) across the testing strategies, this scenario does not affect the base-case incremental costs and QALYs between strategies. Therefore, the resultant incremental cost-effectiveness ratios per QALY gained for this scenario are the same as those for the base case.

TABLE 47 Incremental analysis (CG175[11] PSA monitoring assumptions)

Strategy	Discounted costs (£)	Discounted QALYs	Incremental costs (£)	Incremental QALYs	Incremental cost-effectiveness ratio (£)
Clinical assessment	484,405	–6.35	–	–	–
Clinical assessment + phi	490,254	–5.99	5849	0.37	15,898
Clinical assessment + PCA3	635,078	–6.29	144,824	–0.31	Dominated
Clinical assessment + phi + PCA3	822,886	–6.72	332,632	–0.74	Dominated
Clinical assessment + mpMRI	1,428,175	–9.11	937,921	–3.12	Dominated
Clinical assessment + phi + mpMRI	1,524,592	–9.14	1,034,338	–3.15	Dominated
Clinical assessment + PCA3 + mpMRI	1,599,432	–9.14	1,109,178	–3.15	Dominated
Clinical assessment + phi + PCA3 + mpMRI	1,689,262	–9.14	1,199,008	–3.15	Dominated

TABLE 48 Incremental analysis (Mowatt *et al.*[14] PSA monitoring assumptions)

Strategy	Discounted costs (£)	Discounted QALYs	Incremental costs (£)	Incremental QALYs	Incremental cost-effectiveness ratio (£)
Clinical assessment	356,703	–5.20	–	–	–
Clinical assessment + mpMRI	470,153	–1.85	113,449	3.35	33,911[a]
Clinical assessment + phi	473,552	–5.65	3399	–3.79	Dominated
Clinical assessment + PCA3	525,877	–5.27	55,724	–3.42	Dominated
Clinical assessment + phi + mpMRI	557,864	–1.82	87,711	0.03	2,500,530[a]
Clinical assessment + phi + PCA3	584,185	–4.75	26,321	–2.933	Dominated
Clinical assessment + PCA3 + mpMRI	632,704	–1.82	74,840	0	Dominated
Clinical assessment + phi + PCA3 + mpMRI	722,534	–1.82	164,670	0	Dominated

a Rounding used.

Alternative effectiveness data sources

The total number of biopsies performed if the derived sensitivity values presented in Scattoni *et al.*[102] (80% and 90%) or Gittelman *et al.*[45] (90%) are employed in the model are shown in *Tables 49* and *50*, respectively.

Mean costs and benefits

Costs and QALYs generated if the derived sensitivity values presented in Scattoni *et al.*[102] (80% and 90%) or Gittelman *et al.*[45] (90%) are employed in the model are shown in *Tables 51–53*, respectively.

Incremental analysis

The incremental results if the derived sensitivity values presented in Scattoni *et al.*[102] (80% and 90%) or Gittelman *et al.*[45] (90%) are employed in the model are shown in *Tables 54–56*, respectively.

TABLE 49 Total number of biopsies undertaken (Scattoni *et al.*:[102] 80% and 90% sensitivity)

	80% sensitivity			90% sensitivity		
Strategy	Second biopsies	Biopsies while under PSA monitoring	Total biopsies	Second biopsies	Biopsies while under PSA monitoring	Total biopsies
Clinical assessment	580	250	830	710	220	930
Clinical assessment + PCA3	450	250	700	695	220	915
Clinical assessment + phi	595	250	845	786	220	1006
Clinical assessment + phi + PCA3	580	250	830	725	220	945

TABLE 50 Total number of biopsies undertaken (Gittelman *et al.*:[45] 90% sensitivity)

Strategy	Second biopsies	Biopsies while under PSA monitoring	Total biopsies
Clinical assessment	832	220	1052
Clinical assessment + PCA3	661	220	881

TABLE 51 Costs and QALYs (Scattoni *et al.*:[102] 80% sensitivity)

Strategy	Test costs (£)	Biopsy costs (£)	Biopsy complication costs (£)	PSA monitoring costs (£)	Total costs (£)	Total QALY loss
Clinical assessment	0	405,304	12,304	84,902	502,511	4.73
Clinical assessment + PCA3	89,830	361,791	10,786	84,902	547,309	3.98
Clinical assessment + phi	164,670	410,423	12,483	84,902	672,478	4.81
Clinical assessment + phi + PCA3	254,500	405,304	12,304	84,902	757,011	4.73

Total costs may be higher than the sum of individual costs owing to rounding.

TABLE 52 Costs and QALYs (Scattoni *et al.*:[102] 90% sensitivity)

Strategy	Test costs (£)	Biopsy costs (£)	Biopsy complication costs (£)	PSA monitoring costs (£)	Total costs	Total QALY loss
Clinical assessment	0	424,009	13,177	83,007	520,194	5.31
Clinical assessment + PCA3	89,830	418,890	12,998	83,007	604,726	5.22
Clinical assessment + phi	164,670	449,605	14,070	83,007	711,352	5.75
Clinical assessment + phi + PCA3	254,500	429,128	13,356	83,007	779,991	5.40

Total costs may be higher than the sum of individual costs owing to rounding.

TABLE 53 Costs and QALYs (Gittelman *et al.*:[45] 90% sensitivity)

Strategy	Test costs (£)	Biopsy costs (£)	Biopsy complication costs (£)	PSA monitoring costs (£)	Total costs (£)	Total QALY loss
Clinical assessment	0	465,219	14,615	83,007	562,841	6.02
Clinical assessment + PCA3	164,670	407,372	12,597	83,007	667,646	5.03

TABLE 54 Incremental analysis (Scattoni *et al.*:[102] 80% sensitivity)

Strategy	Discounted costs (£)	Discounted QALYs	Incremental costs (£)	Incremental QALYs	Incremental cost-effectiveness ratio
Clinical assessment	502,511	–4.73	–	–	–
Clinical assessment + PCA3	547,309	–3.98	44,799	0.75	59,732
Clinical assessment + phi	672,478	–4.81	169,968	–0.44	Dominated
Clinical assessment + phi + PCA3	757,011	–4.73	254,500	–0.09	Dominated

TABLE 55 Incremental analysis (Scattoni *et al.*:[102] 90% sensitivity)

Strategy	Discounted costs (£)	Discounted QALYs	Incremental costs (£)	Incremental QALYs	Incremental cost-effectiveness ratio
Clinical assessment	520,194	–5.31	–	–	–
Clinical assessment + PCA3	604,726	–5.22	84,532	0.09	963,964
Clinical assessment + phi	711,352	–5.75	106,627	–0.53	Dominated
Clinical assessment + phi + PCA3	779,991	–5.40	175,266	–0.18	Dominated

TABLE 56 Incremental analysis (Gittelman *et al.*:[45] 90% sensitivity)

Strategy	Discounted costs (£)	Discounted QALYs	Incremental costs (£)	Incremental QALYs	Incremental cost-effectiveness ratio
Clinical assessment	562,841	–6.02	–	–	–
Clinical assessment + PCA3	667,646	–5.03	104,805	0.99	105,765

Deterministic sensitivity analysis

Some of the parameters varied in the deterministic sensitivity analyses could only increase the incremental cost-effectiveness ratios per QALY gained for any of the diagnostic strategies compared with clinical assessment alone. As the incremental cost-effectiveness ratios per QALY gained in the base case are already above the threshold (£20,000 to £30,000) generally considered cost-effective by NICE, the results of these analyses are not shown. For this reason, results from the following sensitivity analysis have been excluded:

- increasing the cost of PCA3 assay or phi
- lower bound of biopsy complication rates
- QALY loss from biopsy reduced by 50%.

Where a sensitivity analysis does not change the number of biopsies, only the incremental analysis is shown. Where biopsy numbers are also changed, full results are provided.

Upper bound of biopsy complication rates

Table 57 shows the incremental analysis if the upper bound of complication rates from *Table 32* are used.

Lower price of Prostate Health Index

Table 58 shows the incremental analysis if the cost of phi test decreased by 50% (i.e. £44.92 as opposed to £89.83).

TABLE 57 Incremental analysis (base case with upper bound of biopsy complication rates)

Strategy	Discounted costs (£)	Discounted QALYs	Incremental costs (£)	Incremental QALYs	Incremental cost-effectiveness ratio (£)
Clinical assessment	591,003	–6.29	–	–	–
Clinical assessment + mpMRI	698,401	–2.94	107,397	3.35	32,102
Clinical assessment + phi	708,661	–6.74	10,260	–3.79	Dominated
Clinical assessment + PCA3	760,311	–6.36	61,911	–3.42	Dominated
Clinical assessment + phi + mpMRI	786,048	–2.91	87,647	0.035	2,498,721
Clinical assessment + phi + PCA3	817,676	–5.84	31,627	–2.933	Dominated
Clinical assessment + PCA3 + mpMRI	860,888	–2.91	74,840	0	Dominated
Clinical assessment + phi + PCA3 + mpMRI	950,718	–2.91	164,670	0	Dominated

TABLE 58 Incremental analysis (base case with lower bound for phi cost)

Strategy	Discounted costs (£)	Discounted QALYs	Incremental costs (£)	Incremental QALYs	Incremental cost-effectiveness ratio (£)
Clinical assessment	579,264	–6.32	–	–	–
Clinical assessment + phi	651,203	–6.77	71,939	–0.45	Dominated
Clinical assessment + mpMRI	692,713	–2.96	113,449	3.36	33,732
Clinical assessment + phi + mpMRI	735,514	–2.93	42,801	0.04	1,213,727
Clinical assessment + PCA3	748,437	–6.40	12,923	–3.47	Dominated
Clinical assessment + phi + PCA3	761,835	–5.87	26,321	–2.91	Dominated
Clinical assessment + PCA3 + mpMRI	855,264	–2.93	119,750	0	Dominated
Clinical assessment + phi + PCA3 + mpMRI	900,184	–2.93	164,670	0	Dominated

Quality-adjusted life-year loss from biopsy

Table 59 shows the incremental analysis if the QALY loss from biopsy was at the upper bound suggested in the literature[117] (i.e. 0.0075 as opposed to 0.0058).

One hundred per cent increase in biopsy complication costs

Table 60 shows the incremental analysis if the costs of biopsy complications are increased by 100%.

TABLE 59 Incremental analysis (highest estimated QALY loss)

Strategy	Discounted costs (£)	Discounted QALYs	Incremental costs (£)	Incremental QALYs	Incremental cost-effectiveness ratio (£)
Clinical assessment	579,264	–8.18	–	–	–
Clinical assessment + mpMRI	692,713	–3.83	113,449	4.35	26,086
Clinical assessment + phi	696,113	–8.76	3399	–4.93	Dominated
Clinical assessment + PCA3	748,437	–8.27	55,724	–4.44	Dominated
Clinical assessment + phi + mpMRI	780,424	–3.78	87,711	0.05	1,923,484
Clinical assessment + phi + PCA3	806,745	–7.60	26,321	–3.82	Dominated
Clinical assessment + PCA3 + mpMRI	855,264	–3.78	74,840	0	Dominated
Clinical assessment + phi + PCA3 + mpMRI	945,094	–3.78	164,670	0	Dominated

TABLE 60 Incremental analysis (100% increase in biopsy complication costs)

Strategy	Discounted costs (£)	Discounted QALYs	Incremental costs (£)	Incremental QALYs	Incremental cost-effectiveness ratio
Clinical assessment	592,073	–6.29	–	–	–
Clinical assessment + mpMRI	698,710	–2.94	106,637	3.35	31,875
Clinical assessment + phi	709,833	–6.74	11,123	–3.79	Dominated
Clinical assessment + PCA3	761,398	–6.36	62,688	–3.42	Dominated
Clinical assessment + phi + mpMRI	786,350	–2.91	87,639	0.035	2,498,493
Clinical assessment + phi + PCA3	818,644	–5.84	32,294	–2.933	Dominated
Clinical assessment + PCA3 + mpMRI	861,190	–2.91	74,840	0	Dominated
Clinical assessment + phi + PCA3 + mpMRI	951,020	–2.91	164,670	0	Dominated

Fifty per cent of cancers are missed on second biopsy

Table 61 shows the incremental analysis if the sensitivity of second biopsy indicated by a testing strategy is 50% rather than 100% (as assumed in the base case).

As was the case with the Mowatt *et al.*[14] PSA monitoring scenario, reducing the sensitivity of biopsy by 50% changes biopsy numbers by the same amount for all strategies. Thus, it alters overall costs for strategies, but incremental costs and incremental cost-effectiveness ratios remain the same as in the base case.

Variation in the proportion of patients with negative second biopsies entering prostate-specific antigen monitoring

Reducing the percentage of patients with negative second biopsies entering PSA monitoring should favour those testing strategies with lower specificity. This is because such testing strategies result in more second biopsies and under this sensitivity analysis fewer patients receive PSA monitoring. Therefore, the results of only the extreme end of the sensitivity analysis are shown, that is where 0% of patients with negative second biopsies enter PSA monitoring. The different numbers of biopsies associated with each diagnostic strategy, assuming 0% of patients with negative second biopsy enter PSA monitoring, are shown in *Table 62*.

TABLE 61 Incremental analysis (50% of cancers missed on second biopsy)

Strategy	Discounted costs (£)	Discounted QALYs	Incremental costs (£)	Incremental QALYs	Incremental cost-effectiveness ratio
Clinical assessment	704,213	–7.04	–	–	–
Clinical assessment + mpMRI	817,663	–3.69	113,449	3.35	33,911
Clinical assessment + phi	821,062	–7.48	3399	–3.79	Dominated
Clinical assessment + PCA3	873,386	–7.11	55,724	–3.42	Dominated
Clinical assessment + phi + mpMRI	905,374	–3.66	87,711	0.04	2,500,530
Clinical assessment + phi + PCA3	931,694	–6.59	26,321	–2.93	Dominated
Clinical assessment + PCA3 + mpMRI	980,214	–3.66	74,840	0	Dominated
Clinical assessment + phi + PCA3 + mpMRI	1,070,044	–3.66	164,670	0	Dominated

TABLE 62 Total number of biopsies undertaken (0% of negative second biopsies entering PSA monitoring)

Strategy	Second biopsies	Biopsies while under PSA monitoring	Total biopsies
Clinical assessment	879	54	934
Clinical assessment + phi	957	35	992
Clinical assessment + PCA3	892	51	943
Clinical assessment + phi + PCA3	802	74	875
Clinical assessment + mpMRI	300	199	499
Clinical assessment + phi + mpMRI	294	201	494
Clinical assessment + PCA3 + mpMRI	294	201	494
Clinical assessment + phi + PCA3 + mpMRI	294	201	494

Total biopsies may be higher than the sum of biopsies owing to rounding.

This sensitivity analysis shows that testing strategies with higher specificity result in more biopsies being undertaken during PSA monitoring than testing strategies with lower specificity.

Mean costs and benefits

Costs and QALYs generated by each diagnostic strategy, assuming that 0% of patients with negative second biopsy enter PSA monitoring, are shown in *Table 63*.

Incremental analysis

The incremental results, assuming that 0% of men with a negative result from a second biopsy enter PSA monitoring, are shown in *Table 64*.

Probabilistic sensitivity analysis

Probabilistic sensitivity analysis was undertaken using (1) the base-case evidence and assumptions and (2), individually, the alternative evidence sources and sensitivity rates. The probabilistic analysis was undertaken by running 1000 iterations of the model, with each iteration choosing a value at random for each variable in the model, where applicable, from the distributions shown in *Model parameters*.

TABLE 63 Costs and QALYs (0% of negative second biopsies entering PSA monitoring)

Strategy	Test costs (£)	Biopsy costs (£)	Biopsy complication costs (£)	PSA monitoring costs (£)	Total costs (£)	Total QALY loss
Clinical assessment	0	341,691	11,522	12,196	365,410	5.37
Clinical assessment + phi	89,830	351,512	12,007	3923	457,272	5.71
Clinical assessment + PCA3	164,670	343,328	11,603	10,817	530,418	5.43
Clinical assessment + phi + PCA3	254,500	331,870	11,038	20,470	617,877	5.03
Clinical assessment + mpMRI	315,560	268,226	7896	74,085	665,767	2.83
Clinical assessment + phi + mpMRI	405,390	267,456	7858	74,734	755,438	2.80
Clinical assessment + PCA3 + mpMRI	480,230	267,456	7858	74,734	830,278	2.80
Clinical assessment + phi + PCA3 + mpMRI	570,060	267,456	7858	74,734	920,108	2.80

Total costs may be higher than the sum of individual costs owing to rounding.

TABLE 64 Incremental analysis (0% of negative second biopsies entering PSA monitoring)

Strategy	Discounted costs (£)	Discounted QALYs	Incremental costs (£)	Incremental QALYs	Incremental cost-effectiveness ratio (£)
Clinical assessment	365,410	5.37	–	–	–
Clinical assessment + phi	457,272	5.71	91,862	–0.34	Dominated
Clinical assessment + PCA3	530,418	5.43	165,009	–0.06	Dominated
Clinical assessment + phi + PCA3	617,877	5.03	252,468	0.34	Extendedly dominated
Clinical assessment + mpMRI	665,767	2.83	300,358	2.54	118,066
Clinical assessment + phi + mpMRI	755,438	2.80	89,671	0.03	3,361,804
Clinical assessment + PCA3 + mpMRI	830,278	2.80	74,840	0	Dominated
Clinical assessment + phi + PCA3 + mpMRI	920,108	2.80	164,670	0	Dominated

Base-case analysis

The CEAC for the base-case analysis is shown in *Figure 6*. It demonstrates that the most cost-effective strategy at £20,000 per QALY gained is clinical assessment alone in 100% of model iterations. At about £33,500 per QALY gained approximately half of the iterations suggest that clinical assessment alone is the most cost-effective strategy, whereas the remaining iterations suggest that it is clinical assessment + mpMRI that is the most cost-effective strategy. At a threshold of £37,000 per QALY gained, all iterations suggest that clinical assessment + mpMRI dominates all other strategies.

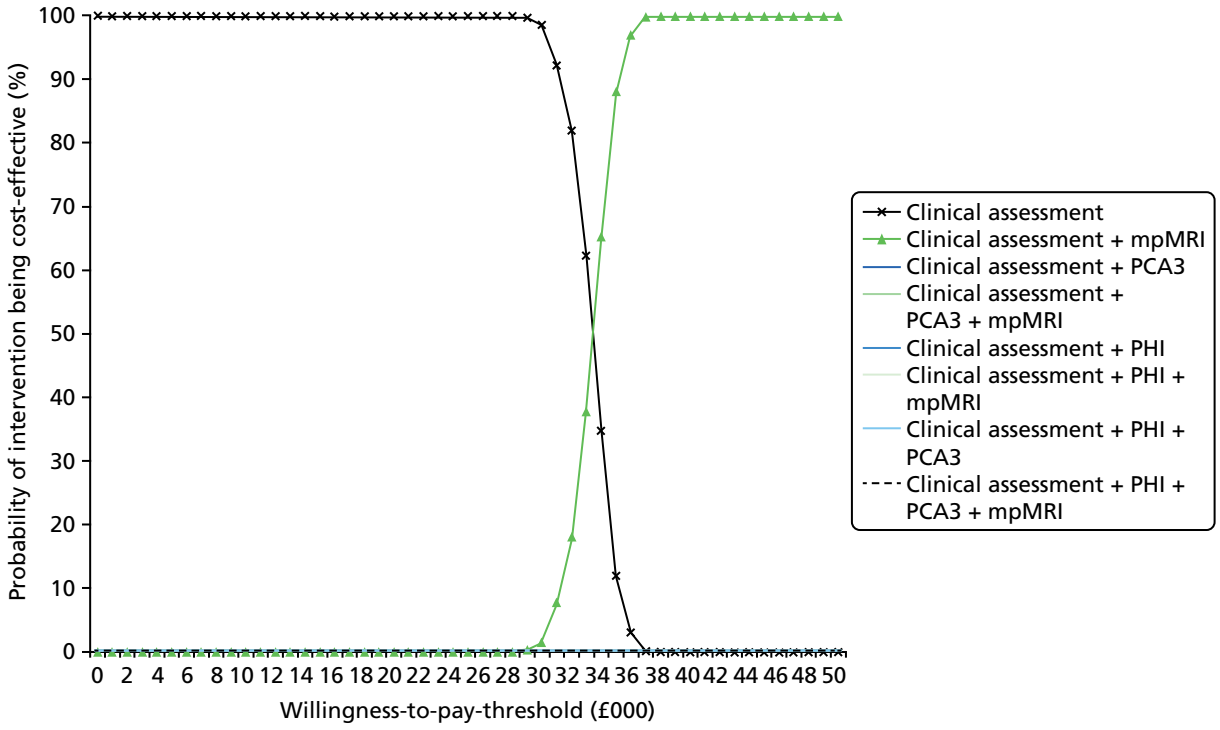

FIGURE 6 Cost-effectiveness acceptability curve (base case).

Figures 7 and *8* show the CEACs for the base case using 80% and 95% sensitivity estimates, respectively. As with sensitivity at 90%, both CEACs show that no testing strategy other than clinical assessment was cost-effective in any model iterations at threshold values below £30,000 per QALY gained.

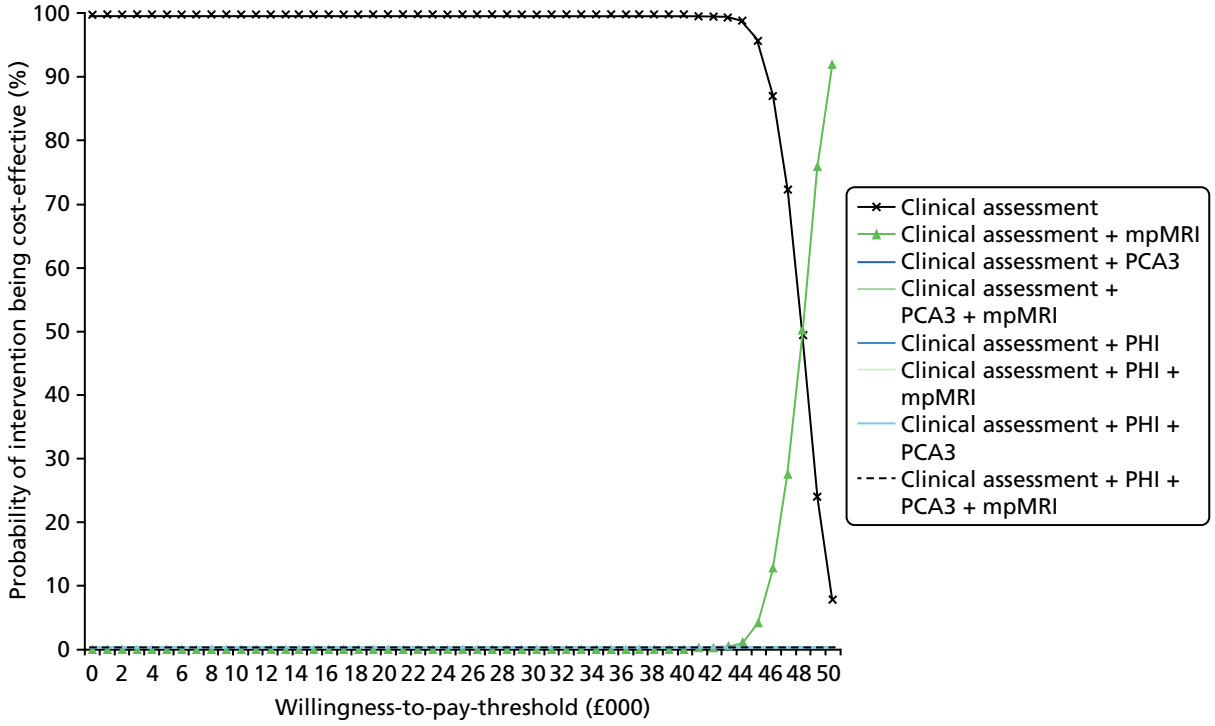

FIGURE 7 Cost-effectiveness acceptability curve (base case: 80% sensitivity).

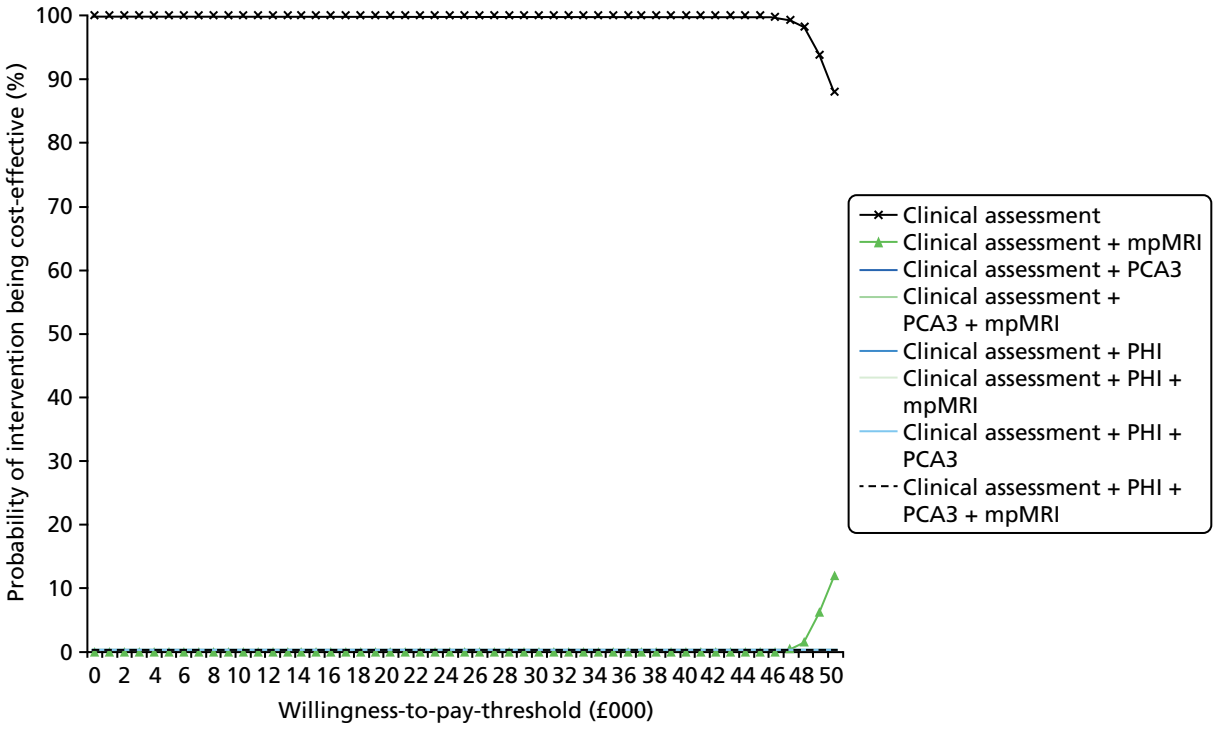

FIGURE 8 Cost-effectiveness acceptability curve (base case: 95% sensitivity).

Alternative effectiveness data sources

Cost-effectiveness acceptability curves using alternative effectiveness data reported by Scattoni *et al.*[102] and Gittelman *et al.*[45] are shown in *Figures 9–11*. All three CEACs show that there were no model iterations in which the PCA3 assay or phi was cost-effective compared with clinical assessment alone at threshold values at, or below, £30,000 per QALY gained.

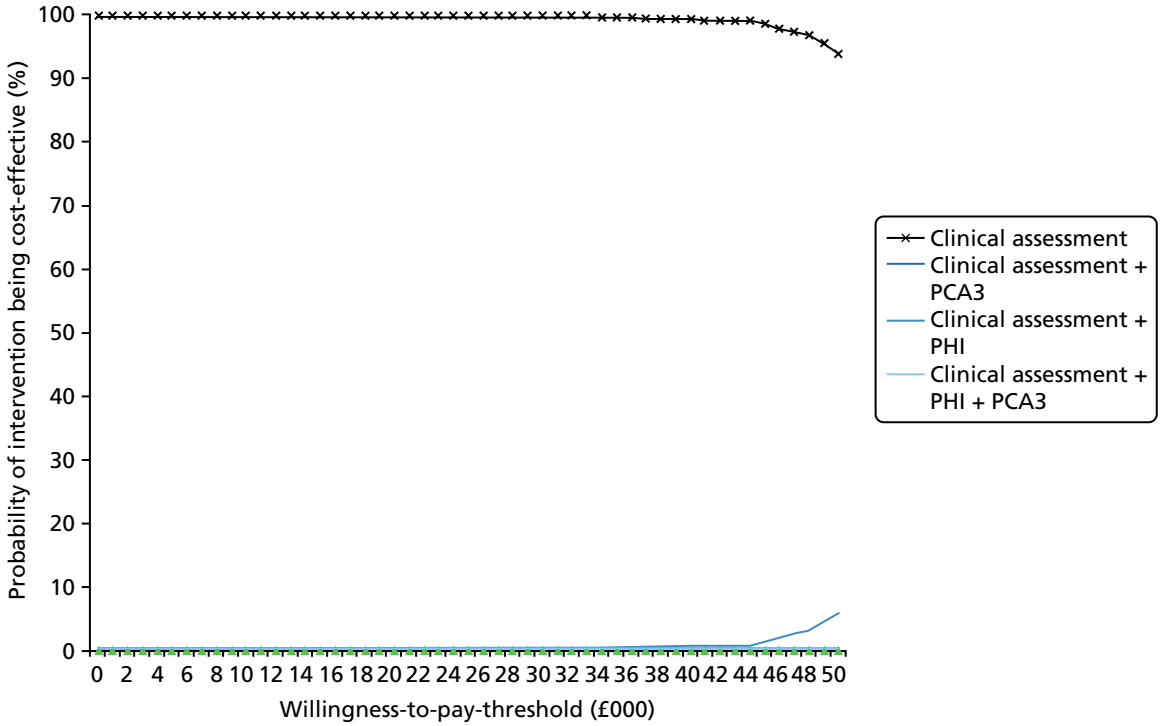

FIGURE 9 Cost-effectiveness acceptability curve (Scattoni *et al.*:[102] 80% sensitivity).

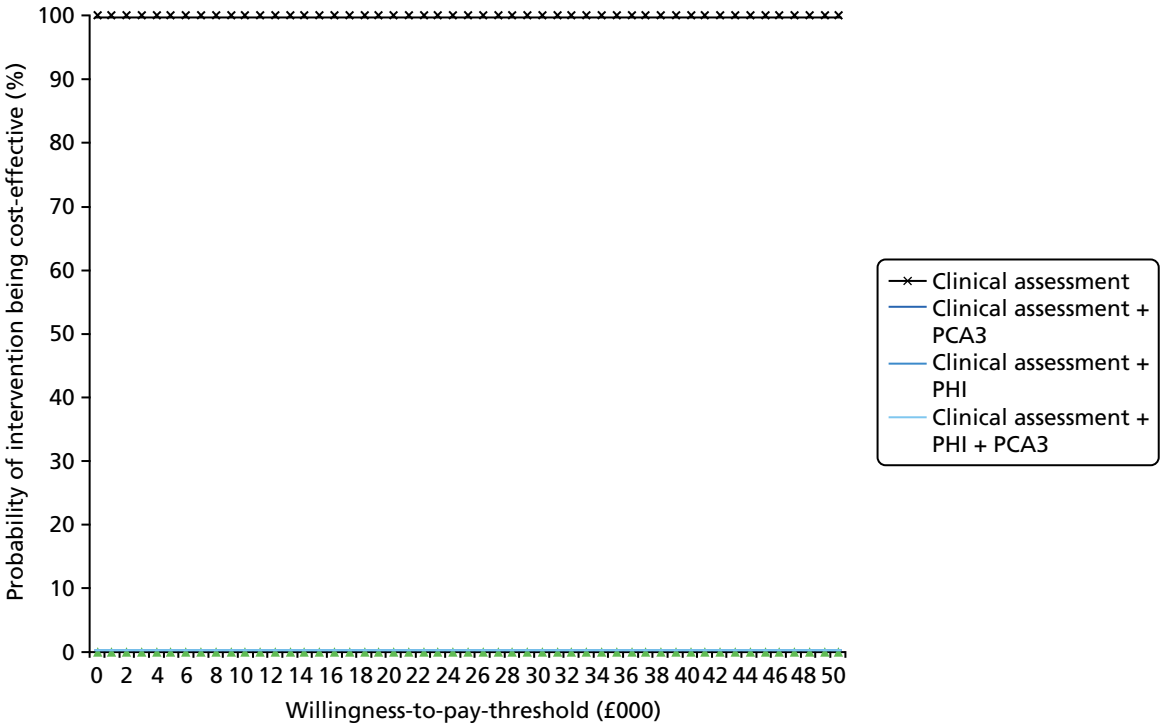

FIGURE 10 Cost-effectiveness acceptability curve (Scattoni *et al.*:[102] 90% sensitivity).

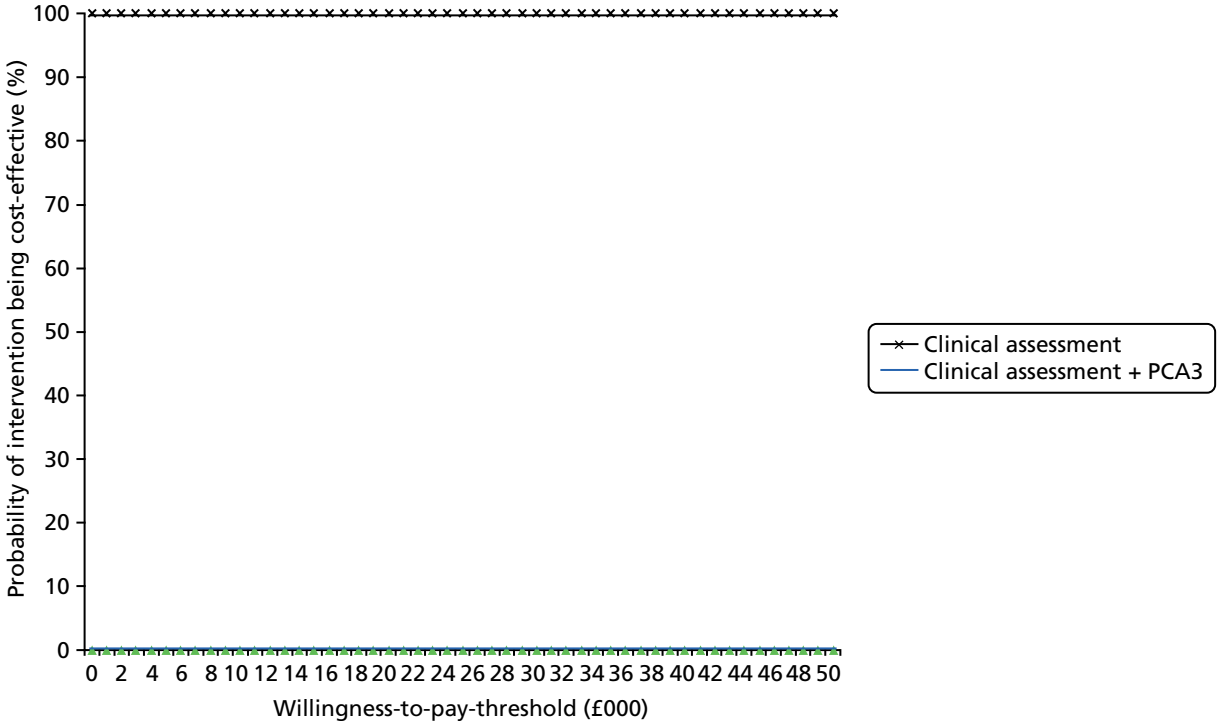

FIGURE 11 Cost-effectiveness acceptability curve (Gittelman *et al.*:[45] 90% sensitivity).

Summary of scenario analyses, deterministic sensitivity analyses and probabilistic sensitivity analyses

Other than the case in which the PSA monitoring strategy employed in the CG175[11] MRI model is used, the incremental cost-effectiveness ratios that were generated to test model uncertainty were all above £20,000 per QALY gained. The probabilistic sensitivity analyses, in particular, confirm that alternative testing strategies using any test in addition to clinical assessment are not cost-effective, although it should be noted that QALY loss associated with a biopsy was not varied in the probabilistic analyses.

Changing the desired sensitivity level does alter the efficiency frontier. However, the frontier does not change such that the inclusion of PCA3 or phi has a favourable incremental cost-effectiveness ratio. The change in frontier occurs because at different sensitivity levels the specificity of phi also changes. At 80% sensitivity, phi adds nothing to sensitivity if added into a strategy of clinical assessment and mpMRI and is therefore dominated by the strategy that excludes it. However, at 90% and 95% sensitivity, the inclusion of phi does improve specificity slightly when added to clinical assessment and mpMRI, and so is on the efficiency frontier, albeit at a high incremental cost-effectiveness ratio.

Discussion of the External Assessment Group model results

The de novo economic model, both in the base case and across an extensive range of scenarios and sensitivity analyses, shows that neither the PCA3 assay nor the phi is likely to be cost-effective when identifying patients for second biopsy over clinical assessment alone or over clinical assessment + MRI.

The only time that one of the tests appears to be cost-effective is when the PSA monitoring strategy used in the CG175[11] MRI model is employed. This approach leads to phi testing having an incremental cost-effectiveness ratio below £20,000 per QALY gained; the EAG cautions that this is a somewhat misleading finding. The testing strategy in that scenario, as stated in the methodology, favours the strategies that have lower specificity. Thus, as phi testing has the lowest specificity (90% sensitivity in the base case),

it generates the most QALYs. This finding, unless the PSA monitoring strategy used in the CG175[11] MRI model does accurately reflect routine clinical practice, should not be given any weight.

The base-case results show that the biggest reduction in the number of second biopsies performed comes from the use of mpMRI. In combination with PCA3 score, a further 2% reduction in second biopsies can be achieved. As, at 90% sensitivity, the PCA3 assay and the phi have a lower specificity than clinical assessment alone, model results show that their use leads to more biopsies being undertaken than when clinical assessment alone is employed. The lower specificity of these two tests can be interpreted as meaning that, to achieve 90% sensitivity, thresholds for the tests have to be set very low.

The use of the PCA3 assay or phi would appear to create more uncertainty in the decision-making process than use of clinical assessment alone with the result that, even though no more patients in total would actually have cancer, more patients would be identified as potentially having cancer than if clinicians had simply relied on their assessment of other clinical parameters.

Several caveats need to be considered when interpreting the cost-effectiveness results generated by the EAG model.

The modelling provides strong evidence that if mpMRI is undertaken then adding the PCA3 assay or phi to the diagnostic strategy is not cost-effective. The results from the model also suggest that mpMRI is not cost-effective compared with clinical assessment alone. However, in the analysis, the EAG assumed that mpMRI did not alter the sensitivity of the biopsy itself as the study[99] from which the data were taken blinded the clinicians taking the biopsy to the mpMRI results. This is, however, a conservative assumption, and there is evidence in CG175[11] that this is not the case and that biopsy is more accurate after mpMRI and as such the cost-effectiveness of mpMRI has probably been underestimated by the EAG model.

In any case, the focus of this assessment and the EAG model is on the PCA3 assay and the phi and whether or not these are clinically effective and/or cost-effective in combination with or without mpMRI rather than on whether or not mpMRI alone should be used. The EAG model found that adding PCA3 score or phi into a testing strategy with mpMRI was highly cost-ineffective. This finding will not change if the model has underestimated the sensitivity of mpMRI.

The mpMRI modelling undertaken in the CG175[11] and Mowatt *et al.*[14] attempted to address a different decision problem regarding the use of different forms of mpMRI, including T2-MRI, to inform the location of a second biopsy rather than to inform whether or not a second biopsy should be undertaken. Therefore, although much of the biopsy cost and complication information contained within these models are transferable to the EAG model presented here, the results are not directly relevant to the decision problem addressed by the EAG model.

It is noted that CG175[11] and Mowatt *et al.*[14] reported conflicting findings on the cost-effectiveness of MRI to inform biopsy; CG175[11] reported that mpMRI was not cost-effective, whereas Mowatt *et al.*[14] reported that there was a favourable incremental cost-effectiveness ratio for T2-MRI over saturation TRUS biopsy. Both studies[14,49] reported that more evidence was required on the effectiveness of MRI and that the level of available evidence may influence the results found.

The EAG analysis is built on the assumption that the same level of sensitivity can be achieved for all strategies with the only difference being the number of biopsies that need to be performed for this to be achieved. While allowing simple comparison between strategies, this assumption may be difficult to achieve in clinical practice. However, the clinical validity evidence shows that the clinician's decision regarding whether or not a patient is referred for a second biopsy is unlikely to be made simply on the results of a single test but rather on the test result in combination with assessment of a range of other patient characteristics and biological parameters.

The published clinical validity studies also show that the sensitivity and specificity of clinical assessment can vary quite markedly between studies, presumably because of the parameters that were incorporated into the assessment. This variation may influence the sensitivity and specificity of individual tests in combination with clinical assessment between studies; it also shows that different sensitivity–specificity combinations can be achieved depending on how the clinical assessment is undertaken. This may have implications for the EAG cost-effectiveness results.

The base-case results in the EAG model are reliant on data from the Porpiglia *et al.* study.[99] The EAG had no quality concerns about how the statistical analysis in the study[99] had been performed; however, the clinical assessment used in that study involved only DREs and age. This approach does not include assessment of PSA level and may not reflect clinical practice in the NHS or, indeed, anywhere outside the study setting.

Given these limitations, the EAG considered that the fairest way to compare the testing strategies was to analyse the individual tests from the perspective of which testing strategy required the fewest biopsies to identify a given percentage of cancers and then to explore how the different clinical assessments undertaken in different studies affect the results through scenario analyses. The results of the scenario analyses clearly show that different clinical effectiveness evidence did not affect the conclusions suggested by the base case.

Related to this point is the level of sensitivity that the EAG chose to incorporate into the model. As desired sensitivity rises, the specificity of the PCA3 assay and the phi falls. At 90% sensitivity and higher, as previously stated, there is evidence that the use of the PCA3 assay or phi can actually reduce specificity compared with clinical assessment alone. However, at lower sensitivities, the specificities of the PCA3 assay and the phi are higher than clinical assessment. It may be that, at sensitivities below 80%, either test may be cost-effective. However, the EAG was not able to explore this, as there was insufficient evidence to incorporate such scenarios into the model.

Varying the number of biopsies, the cost of a biopsy and utility loss from a biopsy for each testing strategy will affect cost-effectiveness results. Although there is evidence, which has been used in the model, on the number and cost of prostate biopsy, the EAG did not find a utility value in the literature specifically associated with prostate biopsy and so a disutility value associated with breast biopsy was used. This could have implications for findings if, as a result of future research, the disutility from prostate biopsy is demonstrated to be more severe and/or longer lasting than disutility from a breast biopsy.

There may also be disutilities resulting from the stress arising from the testing strategy itself, such as waiting for MRI or the patient being told that they have a probability of having cancer but are not being recommended for a second biopsy. With no evidence on the magnitude of such disutilities, these factors have not been included in the analysis.

For men who are more likely to experience adverse events because of biopsy or who suffer from marked anxiety about having a second biopsy, it is possible that utility gains and averted costs from avoided biopsy may be higher than for the average man used in the EAG analysis. In this case, mpMRI may well be cost-effective. When mpMRI is not available, then, unless a lower test sensitivity of 80% is thought desirable for a patient and his clinician, the potential QALY gain would have to be significant for either the PCA3 assay or the phi to be cost-effective.

It should be noted that sensitivity analysis was used to explore the impact of a significantly larger disutility associated with a biopsy and that this did not change the EAG conclusions regarding cost-effectiveness.

A targeted literature review failed to identify evidence (in a form that could be used in the model) on the costs associated with sepsis or antibiotic resistance resulting from a biopsy and so, in line with the CG175[11] MRI model, the impact of this complication was not included in any analyses. However, sensitivity analysis shows that even if complication costs had been underestimated by 100% it would make minimal difference to results. The exclusion of sepsis as a potential complication is, therefore, deemed unlikely to change the conclusions that can be drawn from the model results.

There was limited information in the literature to describe the characteristics of PSA monitoring strategies currently employed in clinical practice. The cost-effectiveness results of the scenarios that were explored suggest that, unless PSA monitoring is akin to that used in the CG175[11] MRI model, the PSA monitoring strategy employed is unlikely to influence results.

One area that could impact on the cost-effectiveness results that could not be explored because of a lack of evidence was whether or not the cancers identified under different strategies differed in levels of aggressiveness. If the PCA3 assay or phi has higher sensitivity for detecting aggressive cancer than clinical assessment alone, it may be that these tests are cost-effective options. Unfortunately, the studies that provided data that could be included in the model did not report this type of result.

Chapter 4 Discussion

Statement of principal findings

Ten documents[48,50,51,57,58,71–75] were included in the EAG review of analytical validity. The EAG concluded that the analytical validity of the PCA3 and p2PSA assays had been comprehensively documented. The EAG identified some important issues relating to the precision of PCA3 assay measurements. Issues were also highlighted in relation to the use of the p2PSA assay, namely sample handling and the thermal stability of samples.

The review of clinical validity data included results from 15 study populations from 17 publications.[45,46,85,86,89–92,94,96,97,99,102–106] More clinical validity data were available to assess the impact of adding the PCA3 assay to clinical assessment than were available to assess the impact of adding phi to clinical assessment. Although the addition of the PCA3 assay and the phi to clinical assessment improved measures of overall diagnostic accuracy, there was no consistent evidence of an improvement in derived sensitivity or derived specificity. The EAG concluded that it was not possible to identify a threshold for the PCA3 score or phi result for clinical use. Similarly, when MRI is carried out alongside clinical assessment, there is no evidence of any benefit associated with the addition either the PCA3 assay or the phi.

The EAG did not identify any studies for inclusion in the review of cost-effectiveness. The results from EAG base-case analyses involving either the PCA3 assay or the phi are unambiguous. The threshold below which NICE generally considers an intervention to be cost-effective (an incremental cost-effectiveness ratio of £20,000–30,000 per QALY gained) is clearly exceeded in all analyses (the lowest incremental cost-effectiveness ratio is over £2M per QALY gained). The probabilistic sensitivity analysis results, the deterministic sensitivity analysis results and results from the scenario analyses demonstrate that this finding is robust to variations in the magnitude of key parameters.

Comparison of results with other published studies

No systematic reviews of clinically relevant comparisons describing the addition of the PCA3 assay or the phi to clinical assessment, with or without MRI, were identified. One previous review[110] of the use of the PCA3 assay in a repeat biopsy population concluded that it was effective in improving accuracy, but the review did not consider its use in combination with other diagnostic tests. A major review[66,67] of the PCA3 assay compared with PSA in a combined population of initial and repeat biopsies concluded that the PCA3 assay improved accuracy but the strength of the evidence reviewed was considered to be low. The reviews by Bradley *et al.*[66,67] did not consider the use of PCA3 score in combination with PSA or clinical variables. No systematic reviews of phi in a repeat biopsy population were identified.

No relevant cost-effectiveness studies or reviews which included the PCA3 assay or phi were identified.

Strengths and limitations of the assessment

Strengths of analysis

The review of analytical validity has highlighted some important issues concerning the precision of PCA3 assay measurements and the requirements for storage and stability samples for phi.

From a clinical perspective, the key strength of this review is the restriction to those studies reporting the incremental effect of the addition of the PCA3 assay or the phi in combination with existing tests, scans and clinical judgement, in the diagnosis of prostate cancer in men who are suspected of having malignant disease and in whom the results of an initial prostate biopsy were negative or equivocal. This restriction was introduced as the issue of importance to clinical decision-makers is the impact of adding the PCA3 assay or

phi to tests currently carried out in routine clinical practice, rather than the theoretical efficacy as reflected in any assessment of the use of the PCA3 assay or phi as stand-alone diagnostic tests. Other authors have noted this important issue[125,126] and it is expected that future studies will focus on assessing the most clinically relevant comparisons, that is an approach which will involve considering combinations of tests.

The clinical validity review has reported results for a wide range of outcome measures from 10 different clinical comparisons. The EAG has made best use of all of the available published data and highlighted the comparisons that are most likely to be clinically relevant to clinicians working in the NHS in England and Wales.

A key strength of the EAG economic evaluation is that the de novo model provides a flexible framework that allows the comparison of many different diagnostic strategies. It is based on the best available clinical validity evidence (identified through the systematic review) and captures the trade-off between high upfront costs of diagnostic tests and the reduction in subsequent biopsies that they may offer. The model design captures all of the main factors that are relevant to the decision problem. It is user friendly and calculations are transparent. Furthermore, the model can easily be updated to incorporate new clinical validity evidence as it becomes available.

Limitations of the analysis

Predominance of one study

Of the 10 clinically relevant comparisons described in the 17 papers,[45,46,85,86,89–92,94,96,97,99,102–106] data from the study by Porpiglia *et al.*[99] are used in nine comparisons. Data from Scattoni *et al.*[102] are used in four comparisons and the remaining 13 populations provide data for a single comparison. Clearly, relying heavily on the study by Porpiglia *et al.*[99] is a limitation of both the clinical validity review and the results of the economic modelling undertaken by the EAG. The EAG acknowledges that clinical assessment in this study is not representative of clinical practice in the NHS and that MRI results were not considered before the biopsy because of the design of the study protocol. However, the EAG considers the Porpiglia *et al.* study[99] to be the only published study that reports data that could be used to inform the de novo economic model.

Clinical relevance of reported outcome measures

Only one of the included studies[105] reported sensitivity and specificity estimates and the EAG considers this to be a substantial weakness in the available data. Logistic regression models offer the potential to study multiple tests. However, although outcome measures such as AUC and multivariate diagnostic ORs may show improved accuracy, they lack clinical relevance as it is not clear whether improvements in overall accuracy are because of improved sensitivity or improved specificity.[127] Derived sensitivity and derived specificity values offer clearer advice but it is not possible to identify individual threshold values of the risk factors included in the model which are associated with a particular achieved level of sensitivity or specificity.

Decision curve analysis results were reported in several publications[90,92,97,99,102,106] included in the review of clinical validity. Neither the PCA3 assay nor phi when added to clinical assessment (or clinical assessment plus MRI) showed increased net benefit below a threshold risk of approximately 15% (i.e. if the assessed risk from the model is below 15% then the addition of a new test does help decision-making).

The EAG noted the following limitations in the use of decision curve methodology as used in the included studies:

- None of the reviewed studies used the option of adding a variable for the harm associated with a test (i.e. complications).
- The reviewed studies weighted the benefit of diagnosed cases as 1 but it is not clear whether or not this approach is appropriate when considering the identification of clinically insignificant cancers.
- The method does not consider the harm arising from missing cancers.

Clinical assessment

The process of clinical decision-making is difficult to capture, standardise and evaluate within a study population. In the reviewed studies, descriptions of clinical assessment varied widely. When definitions of clinical assessment are unclear or are very different across studies, it may not be clinically meaningful to compare results. In two studies[85,97] previously validated nomograms were used to reflect clinical assessment. Another study[105] used a clinical decision algorithm that had been developed in conjunction with 12 European urologists. The EAG considers that this type of decision tool may be the best representation of clinical assessment in the included studies.

The EAG notes that the inclusion of PSA in logistic regression models used to assess the efficacy of phi is inconsistent and gives rise to concerns about the validity of the model results as the phi result already includes a measure of PSA.

Use of different thresholds across the studies

The manufacturer of the PCA3 assay proposed a threshold value of 25 to differentiate between the presence and absence of prostate cancer. However, results using this threshold value were reported in only one study.[45] Other studies used 35,[86] 39[89] and 50[89] or used the PCA3 score as a continuous variable.[46,86,90,97,99,102]

The manufacturer of phi proposed using three groups: low risk of cancer (score of 0 to 20.9), moderate risk of cancer (score of 21 to 39.9) and high risk of cancer (score of 40 and above). However, the four studies[90,97,100,103] that used phi used the phi test results as a continuous variable.

It was difficult for the EAG to draw conclusions from the limited data available, as the included studies used a range of threshold values for the PCA3 assay and none of the studies used the phi tests results recommended by the manufacturer.

Confidence intervals and statistical significance of clinical validity results

Many of the reported results for the clinical validity outcomes do not include either standard errors or CIs. It has, therefore, not been possible for the EAG to assess whether or not the differences between groups were statistically significant. Values for derived sensitivity and derived specificity reported in Porpiglia *et al.*[99] were similar for several models that involved different combinations of diagnostic tests. This may have been because there were small numbers of participants above or below the required threshold associated with a given level of sensitivity and specificity.

Lack of generalisable clinical validity data to inform the economic model

As is the case with all economic models, the results are limited by the generalisability of the available evidence data used to populate the model. In the study[99] used to inform the base-case analysis in the EAG model, the analysis undertaken is appropriate; however, there are differences in clinical practice between the study[99] and the NHS in England and Wales. To explore the impact that these differences may have had on the incremental cost-effectiveness ratios, data from other studies with alternative clinical assessments were modelled. The EAG is confident that using alternative assumptions did not change the model findings regarding the probable cost-effectiveness of adding the PCA3 assay or phi into a testing strategy.

Limited incorporation of utility values

Although the model attempted to capture all the important clinical and cost events, it was not possible to capture and/or value all the key factors that might influence cost-effectiveness. The main area where information is lacking is in relation to utility decrements associated with prostate biopsies. It was necessary to use a proxy value, based on the findings from a study[117] that focused on breast cancer biopsy, to represent pain and short-term complications. Any utility decrements associated with anxiety prior to a biopsy were omitted from the model because of lack of information. Inclusion of specific utility values would require a study that assessed utility across the testing process and would need to take into account anxiety not just from a second biopsy or waiting for mpMRI, but also from any change in anxiety associated with the change in risk information that different testing strategies may offer patients.

Uncertainties

Owing to the lack of published literature, the EAG assessment was unable to address three important clinical issues outlined in the final scope,[49] namely detection of clinically insignificant cancer, optimal order of the tests and the effect of using different forms of reference standard (biopsy). Other relevant uncertainties are also discussed.

Issues identified in the final scope

Detection of clinically insignificant cancers

The management of men who have been diagnosed with prostate cancer varies depending on the grade and extent of the cancer at diagnosis. Clinically insignificant cancers are monitored with active surveillance or watchful waiting. Many clinicians are concerned that, rather than improve health, increased diagnosis of clinically insignificant cancers may lead to an increase in morbidity because of anxiety.

One aim of the clinical validity review was to assess the ability of the PCA3 assay and the phi to improve the detection of more aggressive cancers. A lack of evidence meant that it was not possible for the EAG to draw any conclusions about the impact of the PCA3 assay or phi on the detection of clinically insignificant cancers.

Evidence for the relationship between the aggressiveness of tumours detected at prostate biopsy and the PCA3 assay or the phi has been reported in previous reviews.[128,129] These reviews[128,129] were not restricted to studies of repeat biopsy populations and do not consider the intervention test results used in combination with other diagnostic tests. Filella *et al.*[128] in a narrative review, highlighted inconsistencies in the evidence linking higher PCA3 scores to various markers of tumour aggressiveness. Wang *et al.*,[129] in a meta-analysis of four studies, found that the AUC of phi for discriminating cancers with a Gleason score of above or below 7 was 0.67 (95% CI 0.57 to 0.77), with a sensitivity of 90% (95% CI 87% to 92%) and a specificity of 17% (95% CI 14% to 19%). The authors of the review[129] comment on the need for further research.

If the PCA3 assay or phi has higher sensitivity for detecting aggressive cancer than clinical assessment alone (or clinical assessment plus MRI), the use of these tests may be cost-effective. Unfortunately, the studies that provided data that could be included in the EAG model did not report this type of result.

Order of tests

In the included studies, the results of tests were often presented as outputs from logistic regression models with all tests entered into one model. This approach meant that it was not possible to determine from the available data whether or not carrying out the diagnostic tests in one order was better than carrying out the tests in a different order. For example, it was not possible to determine whether or not diagnostic accuracy was improved if the PCA3 assay (or phi) was carried out before or after MRI. However, in clinical practice the order of the tests is important and has substantial cost and benefit implication, as the costs of MRI are higher than either the PCA3 assay or the phi and the order of the tests may result in different sensitivity and specificity estimates. There is no clinical, and therefore no economic, evidence on using the PCA3 assay or the phi to indicate whether or not mpMRI should be performed before a second biopsy. However, the economic model results provide evidence that if mpMRI is performed then the added information from the PCA3 assay or phi is minimal and incorporation of either into a testing strategy that will include mpMRI will, therefore, not be cost-effective.

Effect of different type of reference standard (prostate biopsy)

It has been shown that, in a given population, biopsy schemes that take a large number of cores spread widely across the prostate, such as saturation schemes, result in a higher prevalence of detected cancer than schemes that involve taking only a few cores.[1] An important clinical point is, therefore, to question if any incremental gain associated with the addition of the PCA3 assay or the phi to clinical assessment would vary depending on the biopsy scheme used to confirm the presence or absence of cancer. As discussed in *Chapter 1*, *Potential sources of bias*, any advantage gained by adding the PCA3 assay, or the phi, to clinical assessment could be reduced if a more extensive biopsy scheme were used. The EAG planned to assess this issue in the review of clinical validity by comparing results in studies which used different types of reference standard; however, the details of reference standards used in studies were poorly reported. An added complication was the fact that the number of biopsy cores taken often differed between patients within a single study. Where details were provided, 10- or 12-core biopsies were the most common.

Other relevant uncertainties

Extent to which the model reflects NHS clinical practice

The EAG economic model addresses clear questions:

> *Given a desired cancer detection rate for the target population, what proportion of the population would need a second biopsy and what proportion of these second biopsies would be necessary?*

However, the extent to which these questions reflect clinical practice is unclear. It is hypothesised that clinicians are more likely to think in terms of individual patients rather than in terms of desired cancer detection rates for the whole population of men suspected of having prostate cancer.

The actual achieved sensitivity and specificity of incorporating the PCA3 assay and the phi into a diagnostic strategy are unknown. The clinical evidence available does not address all of the factors that influence diagnostic practice, such as patient preferences for second biopsies given previous biopsy experience and increased/decreased anxiety levels resulting from the findings of additional tests that place patients in different risk categories. Ideally the cost-effectiveness model would be populated using 'real-world' data. In the absence of real-world data the EAG model has been constructed in such a way as to allow the tests to be compared fairly but at arbitrary levels of sensitivity.

Prostate-specific antigen monitoring strategy after a negative or equivocal biopsy

A key area of uncertainty is related to the best representation of the PSA monitoring strategy within the economic model. No published information was found by the EAG that described NHS monitoring practice in England and Wales. As a consequence, this parameter was varied in the sensitivity analyses and the incremental cost-effectiveness ratio per QALY gained fell below £20,000 only if the PSA monitoring regime employed in the CG175[11] MRI model was used. In the PSA monitoring strategy described in CG175,[11] all men with negative or equivocal results from an initial biopsy go on to receive at least one further biopsy and up to 6 years of PSA monitoring. When used in the EAG model, this assumption would mean that the optimal strategy would be to immediately carry out a further biopsy on everyone shown to be negative or equivocal on the initial biopsy and undertake no PSA monitoring.

Unclear clinical priorities

Improvements in diagnostic test accuracy are often a balance between a gain in sensitivity at the expense of lower specificity, or vice versa. Clinical priorities determine whether sensitivity or specificity is the most important outcome in any particular diagnostic situation. To understand the clinical implications of the findings of the clinical validity review, and to inform the design of the economic model, the EAG surveyed a convenience sample of five clinicians. The clinicians were asked whether, for men undergoing repeat prostatic biopsy, sensitivity or specificity was the most important and whether it was possible to identify a minimum level of sensitivity or specificity that that a test should achieve. Disparate views were expressed, with some clinicians favouring high sensitivity so that all cancers were identified, while others expressed a desire for a test that only identified the more aggressive cancers. No minimum level of sensitivity or specificity was suggested. The EAG, therefore, took the pragmatic approach of using sensitivity estimates of 85%, 90% and 95% in the economic model.

Unclear target population

The precise target patient population for the new tests is also unclear. As discussed in *Chapter 2*, *Quality assessment*, men with a negative result following an initial biopsy can be categorised into three groups: those with clear risk factors for a repeat biopsy (at high assessed risk), those with no remaining risk factors (at low assessed risk) and those where clinicians are uncertain (intermediate assessed risk). Most of the eligible studies included only men who had been referred for a repeat biopsy and had, therefore, presumably been assessed as at high or intermediate risk. It is not clear whether or not clinicians would wish to use the PCA3 assay and/or phi in all men, including those currently assessed as at low risk.

False-negative results

The impact of a FN result at repeat biopsy on the length of time to final diagnosis (and the impact that that delay might have on disease progression) is also an issue. However, recent data suggest that risk reductions associated with radical treatment for low-risk patients (and even moderate-risk patients) may be small and insignificant.[130] If this is the case, it might undermine the cost-effectiveness of strategies that increase cancer detection rates and costs over standard practice, unless those strategies are able to discriminate by grade of tumour. Furthermore, there appears to be no published information on the rate of FPs and overtreatment, although the EAG modelling approach means that these would be the same for all testing strategies and therefore should not impact on results.

Equality and diversity

The incidence of aggressive prostate cancer is greater in people with obesity, which can lead to the positive predictive value of a DRE being higher; DRE can be more difficult to perform in people with obesity.[131] The economic results rely on the results from clinical studies in which a DRE is part of a clinical assessment. It may be that cost-effectiveness results for the PCA3 assay and the phi differ depending on whether or not clinical assessment includes a DRE; this should be considered against the possibility that for some people a DRE may not be possible or may be more difficult to undertake.

Chapter 5 Conclusions

The main findings of the EAG assessment of using the PCA3 assay and the phi in combination with existing tests, scans and clinical judgement in the diagnosis of prostate cancer in men who are suspected of having malignant disease and in whom the results of an initial prostate biopsy were negative or equivocal, are presented in *Table 65*.

Implications for service provision

Several findings from the analytical validity review may affect the successful implementation of the assays in the NHS.

The PROGENSA prostate cancer antigen 3 assay

The patient must undergo a DRE before giving a urine sample for the PCA3 assay and the voided urine sample needs to be transferred to specialist transport tubes within 4 hours. It is likely that these requirements will pose little challenge within a secondary care setting; however, implementation within a primary care setting may require some staff training.

The published precision estimates for the PCA3 assay raise concerns about the interpretation and use of the PCA3 score in clinical practice for detecting men with prostate cancer.

TABLE 65 Summary table of conclusions

Clinical comparison	EAG clinical conclusions	Base-case EAG economic results
Clinical assessment vs. clinical assessment + PCA3	The implications of adding the PCA3 assay to clinical assessment are not clear and it is not possible to identify a single-threshold value for use in a clinical setting	Clinical assessment dominates clinical assessment + PCA3; clinical assessment costs less and generates more QALYs than clinical assessment + PCA3
Clinical assessment vs. clinical assessment + phi	The implications of adding phi to clinical assessment are not clear and it is not possible to identify threshold values for use in a clinical setting	Clinical assessment dominates clinical assessment + phi; clinical assessment costs less and generates more QALYs than clinical assessment + phi
Clinical assessment + MRI vs. clinical assessment + MRI + PCA3	The addition of the PCA3 assay to clinical assessment + MRI does not have a noticeable impact on discrimination	Clinical assessment + MRI costs less but is less effective than clinical assessment + MRI + PCA3; the incremental cost-effectiveness ratio per QALY gained for clinical assessment + MRI + PCA3 is £5,418,366 compared with clinical assessment + MRI
Clinical assessment + MRI vs. clinical assessment + MRI + phi	The addition of phi to clinical assessment + MRI does not have a noticeable impact on discrimination	Clinical assessment + MRI costs less but is less effective than clinical assessment + MRI + phi; the incremental cost-effectiveness ratio per QALY gained for clinical assessment + MRI + phi is £2,500,530 compared with clinical assessment + MRI

Prostate Health Index

The analytical review highlighted concerns about sample handling. Blood samples for the p2PSA assay need to centrifuged and the serum separated within 3 hours. The rationale for this 3-hour limit is unclear, but the current recommendation of 3 hours may pose challenges to implementing the test throughout the NHS. It is not clear if blood samples taken in a primary care setting could be routinely transported to a laboratory and processed as required within 3 hours.

Suggested research priorities

The clinical validity review has been limited by a lack of data directly addressing the dilemmas that clinicians and patients face when deciding whether or not to continue investigations after the results of an initial prostate biopsy are negative or equivocal. Longitudinal end-to-end studies following men from initial investigation through to diagnosis and treatment of prostate cancer are required. Ideally, these studies would be RCTs with men allocated to different diagnostic test pathways after an initial negative or equivocal biopsy. A RCT design would be required to address the following issues:

1. Minimisation of sampling and verification bias. By recruiting all men with negative results following an initial biopsy into a trial population, the contribution of the intervention tests to diagnosis can be assessed in men with all levels of risk and so the role of sampling and verification bias will be minimised.
2. Standardisation of clinical assessment. Within RCT protocols, the measurement of risk factors such as age, a DRE and family history should be standardised and this will enable results from different studies to be compared.
3. QoL. It would be beneficial to include measurement of health-related QoL into future RCTs assessing the accuracy of alternative approaches to diagnosis.

The EAG is aware that it may be many years before any reliable data are available from RCTs. Descriptive data from observational cohorts following men over several years from initial referral onwards could address some unanswered issues including:

1. Patient-reported outcomes. Available studies focus on clinical validity outcomes and do not report the morbidity associated with biopsy, either in the short or long term. These studies should also include men who do not receive a repeat biopsy, as the impact of continued monitoring and uncertainty in this group is not known. In particular, the disutility associated with undergoing a biopsy should be captured. It would also be helpful to gain an understanding of the level of anxiety and depression that results from waiting for mpMRI or biopsy as well as that resulting from receiving equivocal results from following these procedures.
2. Number of repeat biopsies. Longitudinal observational studies would also document how many men required multiple (more than two) biopsies in order to establish or exclude the presence of prostatic cancer.

A recent paper[78] raised the possibility that PCA3 scores may vary with genotype. Further research may be required on genotype variation in PCA3 scores and the implications that this variation has for setting appropriate PCA3 score thresholds to indicate increased risk of prostate cancer.

Acknowledgements

Dr Isabel Syndikus, Consultant Clinical Oncologist, The Clatterbridge Cancer Centre NHS Foundation Trust, Wirral, UK.

Dr Graham Scotland, Senior Research Fellow, University of Aberdeen, Aberdeen, UK.

Professor Chris Hyde, Professor of Public Health and Clinical Epidemiology, University of Exeter, Exeter, UK.

Dr Ana Alfirevic, Senior Lecturer, The Wolfson Centre for Personalised Medicine, Department of Molecular and Clinical Pharmacology, University of Liverpool, Liverpool, UK.

Contributions of authors

Amanda Nicholson: project lead and review of clinical evidence.

James Mahon: development of the de novo economic model.

Angela Boland: support of review process (clinical and economics).

Sophie Beale: support of review process (clinical and economics).

Kerry Dwan: clinical quality assessment, data extraction and statistical advisor.

Nigel Fleeman: review of analytical validity.

Juliet Hockenhull: literature selection and data extraction.

Yenal Dundar: literature searching.

Data sharing statement

All available data can be obtained from the corresponding author.

References

1. Eichler K, Wilby J, Hempel S, Myers L, Kleijnen J. *Diagnostic Value of Systematic Prostate Biopsy Methods in the Investigation for Prostate Cancer: A Systematic Review*. York: Centre for Reviews and Dissemination; 2005.
2. McNeal JE, Redwine EA, Freiha FS, Stamey TA. Zonal distribution of prostatic adenocarcinoma. Correlation with histologic pattern and direction of spread. *Am J Surg Pathol* 1988;**12**:897–906. http://dx.doi.org/10.1097/00000478-198812000-00001
3. McNeal JE. Normal histology of the prostate. *Am J Surg Pathol* 1988;**12**:619–33. http://dx.doi.org/10.1097/00000478-198808000-00003
4. American Cancer Society. *How is Prostate Cancer Staged?* 2013. URL: www.cancer.org/cancer/prostatecancer/detailedguide/prostate-cancer-staging (accessed March 2014).
5. Heidenreich A, Abrahamsson PA, Artibani W, Catto J, Montorsi F, Van Poppel H, *et al.* Early detection of prostate cancer: European Association of Urology recommendation. *Eur Urol* 2013;**64**:347–54. http://dx.doi.org/10.1016/j.eururo.2013.06.051
6. Prostate Cancer UK. *Prostate Cancer Treatment*. 2014. URL: http://prostatecanceruk.org/information/prostate-cancer/treatment (accessed 15 September 2014).
7. Heinlein CA, Chang C. Androgen receptor in prostate cancer. *Endocr Rev* 2004;**25**:276–308. http://dx.doi.org/10.1210/er.2002-0032
8. Lam JS, Leppert JT, Vemulapalli SN, Shvarts O, Belldegrun AS. Secondary hormonal therapy for advanced prostate cancer. *J Urol* 2006;**175**:27–34. http://dx.doi.org/10.1016/S0022-5347(05)00034-0
9. Cancer Research UK. *Prostate Cancer Incidence Statistics.* 2014. URL: www.cancerresearchuk.org/cancer-info/cancerstats/types/prostate/incidence/ (accessed 15 September 2014).
10. Cancer Research UK. *Prostate Cancer Survival Statistics*. 2014. URL: www.cancerresearchuk.org/cancer-info/cancerstats/types/prostate/survival/ (accessed 15 September 2014).
11. National Collaborating Centre for Cancer. *Prostate Cancer: Diagnosis and Treatment*. Cardiff: National Collaborating Centre for Cancer; 2014.
12. Cancer Research UK. *Prostate Cancer Mortality Statistics*. 2014. URL: www.cancerresearchuk.org/cancer-info/cancerstats/types/prostate/mortality/ (accessed 15 September 2014).
13. Glaser A, Fraser L, Corner J, Feltbower R, Morris E, Hartwell G, *et al.* Patient reported outcomes of cancer survivors in England 1–5 years after diagnosis: a cross-setional survey. *BMJ Open* 2013;**3**:e002317. http://dx.doi.org/10.1136/bmjopen-2012-002317
14. Mowatt G, Scotland G, Boachie C, Cruickshank M, Ford JA, Fraser C, *et al.* The diagnostic accuracy and cost-effectiveness of magnetic resonance spectroscopy and enhanced magnetic resonance imaging techniques in aiding the localisation of prostate abnormalities for biopsy: a systematic review and economic evaluation. *Health Technol Assess* 2013;**17**(20). http://dx.doi.org/10.3310/hta17200
15. Zaytoun OM, Jones JS. Prostate cancer detection after a negative prostate biopsy: lessons learnt in the Cleveland Clinic experience. *Int J Urol* 2011;**18**:557–68. http://dx.doi.org/10.1111/j.1442-2042.2011.02798.x
16. Kirby R, Fitzpatrick JM. Optimising repeat prostate biopsy decisions and procedures. *BJU Int* 2011;**109**:1750–4. http://dx.doi.org/10.1111/j.1464-410X.2011.10809.x

17. Scattoni V, Raber M, Capitanio U, Abdollah F, Roscigno M, Angiolilli D, *et al.* The optimal rebiopsy prostatic scheme depends on patient clinical characteristics: results of a recursive partitioning analysis based on a 24-core systematic scheme. *Eur Urol* 2011;**60**:834–41. http://dx.doi.org/10.1016/j.eururo.2011.07.036

18. Department of Health (DH). *National Schedule of Reference Costs: The Main Schedule*. London: DH; 2013. URL: www.gov.uk/government/publications/nhs-reference-costs-2012-to-2013 (accessed 15 September 2013).

19. Burford DC, Kirby M, Austoker J. *Prostate Cancer Risk Management Programme. Information for Primary Care; PSA Testing in Asymptomatic Men*. Sheffield: NHS Cancer Screening Programmes; 2010.

20. NHS Cancer Screening Programmes. *Prostate Cancer Risk Management Programme. Undertaking a Transrectal Ultrasound Guided Biopsy of the Prostate*. Sheffield: NHS Cancer Screening Programmes; 2006.

21. Norberg M, Egevad L, Holmberg L, Sparen P, Norlen BJ, Busch C. The sextant protocol for ultrasound-guided core biopsies of the prostate underestimates the presence of cancer. *Urology* 1997;**50**:562–6. http://dx.doi.org/10.1016/S0090-4295(97)00306-3

22. Djavan B, Waldert M, Zlotta A, Dobronski P, Seitz C, Remzi M, *et al.* Safety and morbidity of first and repeat transrectal ultrasound guided prostate needle biopsies: results of a prospective European prostate cancer detection study. *J Urol* 2001;**166**:856–60. http://dx.doi.org/10.1016/S0022-5347(05)65851-X

23. Lujan M, Paez A, Santonja C, Pascual T, Fernandez I, Berenguer A. Prostate cancer detection and tumor characteristics in men with multiple biopsy sessions. *Prostate Cancer Prostatic Dis* 2004;**7**:238–42. http://dx.doi.org/10.1038/sj.pcan.4500730

24. Mian BM, Naya Y, Okihara K, Vakar-Lopez F, Troncoso P, Babaian RJ. Predictors of cancer in repeat extended multisite prostate biopsy in men with previous negative extended multisite biopsy. *Urology* 2002;**60**:836–40. http://dx.doi.org/10.1016/S0090-4295(02)01950-7

25. Shah RB. Current perspectives on the Gleason grading of prostate cancer. *Arch Pathol Lab Med* 2009;**133**:1810–16. http://dx.doi.org/10.1043/1543-2165-133.11.1810

26. Epstein JI. Gleason score 2–4 adenocarcinoma of the prostate on needle biopsy: a diagnosis that should not be made. *Am J Surg Pathol* 2000;**24**:477–8. http://dx.doi.org/10.1097/00000478-200004000-00001

27. University of Liverpool. *Towards a Consensus Protocol on Prostate Biopsies: Indications, Techniques and Assessment*. Royal College of Pathologists, London, 6 June 2003.

28. Epstein JI, Allsbrook WC Jr, Amin MB, Egevad LL. The 2005 International Society of Urological Pathology (ISUP) Consensus Conference on Gleason Grading of Prostatic Carcinoma. *Am J Surg Pathol* 2005;**29**:1228–42. http://dx.doi.org/10.1097/01.pas.0000173646.99337.b1

29. Albertsen PC, Hanley JA, Fine J. 20-year outcomes following conservative management of clinically localized prostate cancer. *JAMA* 2005;**293**:2095–101. http://dx.doi.org/10.1001/jama.293.17.2095

30. Stamey TA, Freiha FS, McNeal JE, Redwine EA, Whittemore AS, Schmid HP. Localized prostate cancer. Relationship of tumor volume to clinical significance for treatment of prostate cancer. *Cancer* 1993;**71**:933–8. http://dx.doi.org/10.1002/1097-0142(19930201)71:3+<933::AID-CNCR2820711408>3.0.CO;2-L

31. Wolters T, Roobol MJ, van Leeuwen PJ, van den Bergh RCN, Hoedemaeker RF, van Leenders GJLH, *et al.* A critical analysis of the tumor volume threshold for clinically insignificant prostate cancer using a data set of a randomized screening trial. *J Urol* 2011;**185**:121–5. http://dx.doi.org/10.1016/j.juro.2010.08.082

32. Rodrigues G, Warde P, Pickles T, Crook J, Brundage M, Souhami L, *et al.* Pre-treatment risk stratification of prostate cancer patients: a critical review. *Can Urol Assoc J* 2012;**6**:121–7. http://dx.doi.org/10.5489/cuaj.11085

33. Shaw GL, Thomas BC, Dawson SN, Srivastava G, Vowler SL, Gnanapragasam VJ, *et al.* Identification of pathologically insignificant prostate cancer is not accurate in unscreened men. *Br J Cancer* 2014;**110**:2405–11. http://dx.doi.org/10.1038/bjc.2014.192

34. D'Amico AV, Whittington R, Malkowicz S, Schultz D, Blank K, Broderick GA, *et al.* Biochemical outcome after radical prostatectomy, external beam radiation therapy, or interstitial radiation therapy for clinically localized prostate cancer. *JAMA* 1998;**280**:969–74. http://dx.doi.org/10.1001/jama.280.11.969

35. Benson MC, Whang IS, Pantuck A, Ring K, Kaplan SA, Olsson CA, *et al.* Prostate specific antigen density: a means of distinguishing benign prostatic hypertrophy and prostate cancer. *J Urol* 1992;**147**:815–16.

36. Thanigasalam R, Mancuso P, Tsao K, Rashid P. Prostate-specific antigen velocity (PSAV): a practical role for PSA? *ANZ J Surg* 2009;**79**:703–6. http://dx.doi.org/10.1111/j.1445-2197.2009.05055.x

37. Patel AR, Klein EA. Risk factors for prostate cancer. *Nat Clin Pract Urol* 2009;**6**:87–95. http://dx.doi.org/10.1038/ncpuro1290

38. SWOP – The Prostate Cancer Research Foundation, Rotterdam. *Risk Calculators*. URL: www.prostatecancer-riskcalculator.com/seven-prostate-cancer-risk-calculators (accessed April 2014).

39. van Vugt HA, Roobol MJ, Kranse R, Maattanen L, Finne P, Hugosson J, *et al.* Prediction of prostate cancer in unscreened men: external validation of a risk calculator. *Eur J Cancer* 2011;**47**:903–9. http://dx.doi.org/10.1016/j.ejca.2010.11.012

40. Thompson IM, Ankerst DP, Chi C, Goodman PJ, Tangen CM, Lucia MS, *et al.* Assessing prostate cancer risk: results from the Prostate Cancer Prevention Trial. *J Natl Cancer Inst* 2006;**98**:529–34. http://dx.doi.org/10.1093/jnci/djj131

41. Karakiewicz PI, Benayoun S, Kattan MW, Perrotte P, Valiquette L, Scardino PT, *et al.* Development and validation of a nomogram predicting the outcome of prostate biopsy based on patient age, digital rectal examination and serum prostate specific antigen. *J Urol* 2005;**173**:1930–4. http://dx.doi.org/10.1097/01.ju.0000158039.94467.5d

42. University of California San Francisco. *Multi-Parametric Prostate Cancer Staging Exam*. San Francisco, CA: University of California; 2014. URL: www.radiology.ucsf.edu/patient-care/services/prostate-exam-study (accessed 4 September 2014).

43. Le JD, Huang J, Marks LS. Targeted prostate biopsy: value of multiparametric magnetic resonance imaging in detection of localized cancer. *Asian J Androl* 2014;**16**:522–9. http://dx.doi.org/10.4103/1008-682X.122864

44. Marks L, Young S, Natarajan S. MRI–ultrasound fusion for guidance of targeted prostate biopsy. *Curr Opin Urol* 2013;**23**:43–50 http://dx.doi.org/10.1097/MOU.0b013e32835ad3ee

45. Gittelman MC, Hertzman B, Bailen J, Williams T, Koziol I, Henderson RJ, *et al.* PCA3 molecular urine test as a predictor of repeat prostate biopsy outcome in men with previous negative biopsies: a prospective multicenter clinical study. *J Urol* 2013;**190**:64–9. http://dx.doi.org/10.1016/j.juro.2013.02.018

46. Haese A, de la Taille A, van Poppel H, Marberger M, Stenzl A, Mulders PF, *et al.* Clinical utility of the PCA3 urine assay in European men scheduled for repeat biopsy. *Eur Urol* 2008;**54**:1081–8. http://dx.doi.org/10.1016/j.eururo.2008.06.071

47. Pepe P, Fraggetta F, Galia A, Skonieczny G, Aragona F. PCA3 score and prostate cancer diagnosis at repeated saturation biopsy. Which cut-off: 20 or 35? *Int Braz J Urol* 2012;**38**:489–95. http://dx.doi.org/10.1590/S1677-55382012000400008

48. Sokoll LJ, Ellis W, Lange P, Noteboom J, Elliott DJ, Deras IL, *et al.* A multicenter evaluation of the PCA3 molecular urine test: pre-analytical effects, analytical performance, and diagnostic accuracy. *Clin Chim Acta* 2008;**389**:1–6. http://dx.doi.org/10.1016/j.cca.2007.11.003

49. National Institute for Health and Care Excellence. *Diagnosis and Monitoring of Prostate Cancer: Progensa PCA3 Assay and Prostate Health Index (PHI): Final Scope*. London: National Institute for Health and Care Excellence; 2014.

50. Food and Drug Administration. *Summary of Safety and Effectiveness Data. PMA P100033: PROGENSA PCA3 Assay*. Silverspring, MD: Food and Drug Administration; 2012.

51. Gen-Probe. *Pack Insert for PROGENSA PCA3 Assay*. 501377EN Rev. D2012.

52. Food and Drug Administration. *PMA Approval. PROGENSA PCA3 Assay*. 2012. URL: www.accessdata.fda.gov/cdrh_docs/pdf10/p100033a.pdf (accessed October 2014).

53. Lee R, Localio AR, Armstrong K, Malkowicz SB, Schwartz JS, Free PSA Study Group. A meta-analysis of the performance characteristics of the free prostate-specific antigen test. *Urology* 2006;**67**:762–8. http://dx.doi.org/10.1016/j.urology.2005.10.052

54. Roddam AW, Duffy MJ, Hamdy FC, Ward AM, Patnick J, Price CP, *et al.* Use of prostate-specific antigen (PSA) isoforms for the detection of prostate cancer in men with a PSA level of 2–10 ng/ml: systematic review and meta-analysis. *Eur Urol* 2005;**48**:386–99; discussion 98–9. http://dx.doi.org/10.1016/j.eururo.2005.04.015

55. 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**:921–7. http://dx.doi.org/10.1016/j.eururo.2010.02.003

56. Le BV, Griffin CR, Loeb S, Carvalhal GF, Kan D, 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**:1355–9. http://dx.doi.org/10.1016/j.juro.2009.12.056

57. Beckman Coulter. *Draft Directional Insert. Access Immunoassay Systems. Hybritech p2PSA*. 2011. URL: www.accessdata.fda.gov/cdrh_docs/pdf9/P090026c.pdf (accessed 21 August 2014).

58. Food and Drug Administration. *Summary of Safety and Effectiveness Data. PMA P090026. Quantitative test for determination of [–2]proPSA levels*. Silverspring, MD: Food and Drug Administration; 2012.

59. Vignati G, Giovanelli L. Standardization of PSA measures: a reappraisal and an experience with WHO calibration of Beckman Coulter Access Hybritech total and free PSA. *Int J Biol Markers* 2007;**22**:295–301.

60. Food and Drug Administration. *PMA Approval. Access Hybritech p2PSA on Access Immunoassay Systems*. 2012. URL: www.accessdata.fda.gov/cdrh_docs/pdf9/p090026a.pdf (accessed October 2014).

61. Centre for Reviews and Dissemination. *Systematic reviews: CRD's Guidance on Undertaking Reviews in Health Care*. York: Centre for Reviews and Dissemination; 2009. URL: http://srdta.cochrane.org/ (accessed July 2015).

62. National Institute for Health and Care Excellence. *Diagnostics Assessment Programme Manual*. London: National Institute for Health and Care Excellence; 2011.

63. Macaskill P, Gatsonis C, Deeks JJ, Harbord RM, Takwoingi Y. Analysing and Presenting Results. In Deeks JJ, Bossuyt PM, Gatsonis C, editors. *Cochrane Handbook for Systematic Reviews of Diagnostic test Accuracy (Version 10)*. Cochrane; 2010.

64. Jonas DE, Wilt TJ, Taylor BC, Wilkins TM, Matchar DB. Chapter 11: challenges in and principles for conducting systematic reviews of genetic tests used as predictive indicators. *J Gen Intern Med* 2012;**27**(Suppl. 1):S83–93. http://dx.doi.org/10.1007/s11606-011-1898-z

65. Teutsch SM, Bradley LA, Palomaki GE, Haddow JE, Piper M, Calonge N, *et al.* The Evaluation of Genomic Applications in Practice and Prevention (EGAPP) Initiative: methods of the EGAPP Working Group. *Genet Med* 2009;**11**:3–14. http://dx.doi.org/10.1097/GIM.0b013e318184137c

66. Bradley LA, Palomaki GE, Gutman S, Samson D, Aronson N. *PCA3 Testing for the Diagnosis and Management of Prostate Cancer*. Comparative Effectiveness Reviews No. 98. Rockville, MD: Agency for Healthcare Research and Quality (US). Report No. 13-EHC030-EF; 2013.

67. Bradley LA, Palomaki GE, Gutman S, Samson D, Aronson N. Comparative effectiveness review: prostate cancer antigen 3 testing for the diagnosis and management of prostate cancer. *J Urol* 2013;**190**:389–98. http://dx.doi.org/10.1016/j.juro.2013.02.005

68. Bruzzese D, Mazzarella C, Ferro M, Perdona S, Chiodini P, Perruolo G, *et al.* Prostate health index vs. percent free prostate-specific antigen for prostate cancer detection in men with 'gray' prostate-specific antigen levels at first biopsy: systematic review and meta-analysis. *Transl Res* 2014;**14**:441–51. http://dx.doi.org/10.1016/j.trsl.2014.06.006

69. Filella X, Gimenez N. Evaluation of [–2] proPSA and Prostate Health Index (phi) for the detection of prostate cancer: a systematic review and meta-analysis. *Clin Chem Lab Med* 2013;**51**:729–39. http://dx.doi.org/10.1515/cclm-2012-0410

70. Zhang ZX, Yang J, Zhang CZ, Li KA, Quan QM, Wang XF, *et al.* The value of magnetic resonance imaging in the detection of prostate cancer in patients with previous negative biopsies and elevated prostate-specific antigen levels: a meta-analysis. *Acad Radiol* 2014;**21**:578–89. http://dx.doi.org/10.1016/j.acra.2014.01.004

71. Groskopf J, Aubin SM, Deras IL, Blase A, Bodrug S, Clark C, *et al.* APTIMA PCA3 molecular urine test: development of a method to aid in the diagnosis of prostate cancer. *Clin Chem* 2006;**52**:1089–95. http://dx.doi.org/10.1373/clinchem.2005.063289

72. Shappell SB, Fulmer J, Arguello D, Wright BS, Oppenheimer JR, Putzi MJ. PCA3 urine mRNA testing for prostate carcinoma: patterns of use by community urologists and assay performance in reference laboratory setting. *Urology* 2009;**73**:363–8. http://dx.doi.org/10.1016/j.urology.2008.08.459

73. Semjonow A, Kopke T, Eltze E, Pepping-Schefers B, Burgel H, Darte C. Pre-analytical in-vitro stability of –2 proPSA in blood and serum. *Clin Biochem* 2010;**43**:926–8. http://dx.doi.org/10.1016/j.clinbiochem.2010.04.062

74. Sokoll LJ, Chan DW, Klee GG, Roberts WL, van Schaik RH, Arockiasamy DA, *et al.* Multi-center analytical performance evaluation of the Access Hybritech p2PSA immunoassay. *Clin Chim Acta* 2012;**413**:1279–83. http://dx.doi.org/10.1016/j.cca.2012.04.015

75. Stephan C, Kahrs AM, Cammann H, Lein M, Schrader M, Deger S, *et al.* A –2 proPSA-based artificial neural network significantly improves differentiation between prostate cancer and benign prostatic diseases. *INC* 2009;**69**:198–207. http://dx.doi.org/10.1002/pros.20872

76. Clinicaltrials.gov. *NCT01441687. Comparing the Reliability of Expressed Prostatic Secretion (EPS) and Post Massage Urine (PMU) for the Prediction of Prostate Cancer Biopsy Outcome*. URL: http://clinicaltrials.gov/show/NCT01441687 (accessed 30 September 2014).

77. World Health Organization. EUCTR2010-024649-61-NL. *Pilot Study: Performance of the Progensa PCA3 Test in Post-oxytocin Urine Specimens*. WHO: Geneva. URL: http://apps.who.int/trialsearch/Trial2.aspx?TrialID=EUCTR2010-024649-61-NL (accessed 30 September 2014).

78. Chen Z, Sun J, Kim ST, Groskopf J, Feng J, Isaacs WB, *et al.* Genome-wide association study identifies genetic determinants of urine PCA3 levels in men. *Neoplasia* 2013;**15**:448–53. http://dx.doi.org/10.1593/neo.122144

79. Kote-Jarai Z, Leongamornlert D, Tymrakiewicz M, Field H, Guy M, Al Olama AA, *et al.* Mutation analysis of the MSMB gene in familial prostate cancer. *BJC* 2010;**102**:414–18. http://dx.doi.org/10.1038/sj.bjc.6605485

80. Lou H, Yeager M, Li H, Bosquet JG, Hayes RB, Orr N, *et al.* Fine mapping and functional analysis of a common variant in MSMB on chromosome 10q11.2 associated with prostate cancer susceptibility. *Proc Natl Acad Sci USA* 2009;**106**:7933–8. http://dx.doi.org/10.1073/pnas.0902104106

81. Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, *et al.* QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med* 2011;**155**:529–36. http://dx.doi.org/10.7326/0003-4819-155-8-201110180-00009

82. Reitsma JB, Rutjes AWS, Whiting P, Vlassov VV, Leeflang MMG, Deeks JJ. Assessing Methodological Quality. In Deeks JJ, Bossuyt PM, Gatsonis C, editors. *Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy, Version 100*. The Cochrane Collaboration; 2009.

83. Whiting PF, Westwood ME, Rutjes AW, Reitsma JB, Bossuyt PN, Kleijnen J. Evaluation of QUADAS, a tool for the quality assessment of diagnostic accuracy studies. *BMC Med Res Methodol* 2006;**6**:9. http://dx.doi.org/10.1186/1471-2288-6-9

84. ClinicalTrials.gov. *NCT01632930. Medical Economics of Urinary PCA3 Test for Prostate Cancer Diagnosis*. URL: http://clinicaltrials.gov/show/NCT01632930 (accessed 30 September 2014).

85. Ankerst DP, Groskopf J, Day JR, Blase A, Rittenhouse H, Pollock BH, *et al.* Predicting prostate cancer risk through incorporation of prostate cancer gene 3. [Erratum published in *J Urol* 2009;**181**:1507.] *J Urol* 2008;**180**:1303–8; discussion 8. http://dx.doi.org/10.1016/j.juro.2008.06.038

86. Aubin SM, Reid J, Sarno MJ, Blase A, Aussie J, Rittenhouse H, *et al.* PCA3 molecular urine test for predicting repeat prostate biopsy outcome in populations at risk: validation in the placebo arm of the dutasteride REDUCE trial. *J Urol* 2010;**184**:1947–52. http://dx.doi.org/10.1016/j.juro.2010.06.098

87. Auprich M, Augustin H, Budaus L, Kluth L, Mannweiler S, Shariat SF, *et al.* A comparative performance analysis of total prostate-specific antigen, percentage free prostate-specific antigen, prostate-specific antigen velocity and urinary prostate cancer gene 3 in the first, second and third repeat prostate biopsy. *BJU Int* 2012;**109**:1627–35. http://dx.doi.org/10.1111/j.1464-410X.2011.10584.x

88. Auprich M, Haese A, Walz J, Pummer K, de la Taille A, Graefen M, *et al.* External validation of urinary PCA3-based nomograms to individually predict prostate biopsy outcome. *Eur Urol* 2010;**58**:727–32. http://dx.doi.org/10.1016/j.eururo.2010.06.038

89. Bollito E, De Luca S, Cicilano M, Passera R, Grande S, Maccagnano C, *et al.* Prostate Cancer Gene 3 urine assay cutoff in diagnosis of prostate cancer a validation study on an italian patient population undergoing first and repeat biopsy. *Anal Quant Cytol Histol* 2012;**34**:96–104.

90. Busetto GM, De Berardinis E, Sciarra A, Panebianco V, Giovannone R, Rosato S, *et al.* Prostate Cancer Gene 3 and multiparametric magnetic resonance can reduce unnecessary biopsies: decision curve analysis to evaluate predictive models. *Urology* 2013;**82**:1355–62. http://dx.doi.org/10.1016/j.urology.2013.06.078

91. Goode RR, Marshall SJ, Duff M, Chevli E, Chevli KK. Use of PCA3 in detecting prostate cancer in initial and repeat prostate biopsy patients. *INC* 2013;**73**:48–53. http://dx.doi.org/10.1002/pros.22538

92. Lazzeri M, Briganti A, Scattoni V, Lughezzani G, Larcher A, Gadda GM, *et al.* Serum Index Test percent –2 proPSA and Prostate Health Index are more accurate than prostate specific antigen and percent fPSA in predicting a positive repeat prostate biopsy. *J Urol* 2012;**188**:1137–43. http://dx.doi.org/10.1016/j.juro.2012.06.017

93. Marks LS, Fradet Y, Deras IL, Blase A, Mathis J, Aubin SM, *et al.* PCA3 molecular urine assay for prostate cancer in men undergoing repeat biopsy. *Urology* 2007;**69**:532–5. http://dx.doi.org/10.1016/j.urology.2006.12.014

94. Panebianco V, Sciarra A, De Berardinis E, Busetto GM, Lisi D, Buonocore V, *et al.* PCA3 urinary test versus 1H-MRSI and DCEMR in the detection of prostate cancer foci in patients with biochemical alterations. *Anticancer Res* 2011;**31**:1399–405.

95. Pepe P, Aragona F. PCA3 score vs. PSA free/total accuracy in prostate cancer diagnosis at repeat saturation biopsy. *Anticancer Res* 2011;**31**:4445–9.

96. Pepe P, Aragona F. Prostate cancer detection rate at repeat saturation biopsy: PCPT risk calculator versus PCA3 score versus case-finding protocol. *Can J Urol* 2013;**20**:6620–4.

97. Perdonà S, Cavadas V, Di Lorenzo G, Damiano R, Chiappetta G, Del Prete P, *et al.* Prostate cancer detection in the 'grey area' of prostate-specific antigen below 10 ng/ml: head-to-head comparison of the updated PCPT calculator and Chun's nomogram, two risk estimators incorporating prostate cancer antigen 3. *Eur Urol* 2011;**59**:81–7. http://dx.doi.org/10.1016/j.eururo.2010.09.036

98. Ploussard G, Haese A, Van Poppel H, Marberger M, Stenzl A, Mulders PF, *et al.* The prostate cancer gene 3 (PCA3) urine test in men with previous negative biopsies: does free-to-total prostate-specific antigen ratio influence the performance of the PCA3 score in predicting positive biopsies? *BJU Int* 2010;**106**:1143–7. http://dx.doi.org/10.1111/j.1464-410X.2010.09286.x

99. Porpiglia F, Russo F, Manfredi M, Mele F, Fiori C, Bollito E, *et al.* The roles of multiparametric magnetic resonance imaging, PCA3 and Prostate Health Index – which is the best predictor of prostate cancer after a negative biopsy? *J Urol* 2014;**192**:60–6. http://dx.doi.org/10.1016/j.juro.2014.01.030

100. Ramos CG, Valdevenito R, Vergara I, Anabalon P, Sanchez C, Fulla J. PCA3 sensitivity and specificity for prostate cancer detection in patients with abnormal PSA and/or suspicious digital rectal examination. First Latin American experience. *Urologic Oncol-Semin ORI* 2013;**31**:1522–6. http://dx.doi.org/10.1016/j.urolonc.2012.05.002

101. Remzi M, Haese A, van Poppel H, de la Taille A, Stenzl A, Hennenlotter J, *et al.* Follow-up of men with an elevated PCA3 score and a negative biopsy: does an elevated PCA3 score indeed predict the presence of prostate cancer? *BJU Intern* 2010;**106**:1138–42. http://dx.doi.org/10.1111/j.1464-410X.2010.09330.x

102. Scattoni V, Lazzeri M, Lughezzani G, De Luca S, Passera R, Bollito E, *et al.* Head-to-head comparison of Prostate Health Index and urinary PCA3 for predicting cancer at initial or repeat biopsy. *J Urol* 2013;**190**:496–501. http://dx.doi.org/10.1016/j.juro.2013.02.3184

103. Sciarra A, Panebianco V, Cattarino S, Busetto GM, Berardinis E, Ciccariello M, *et al.* Multiparametric magnetic resonance imaging of the prostate can improve the predictive value of the urinary prostate cancer antigen 3 test in patients with elevated prostate-specific antigen levels and a previous negative biopsy. *BJU Intern* 2012;**110**:1661–5. http://dx.doi.org/10.1111/j.1464-410X.2012.11146.x

104. Stephan C, Vincendeau S, Houlgatte A, Cammann H, Jung K, Semjonow A. Multicenter evaluation of [–2]proprostate-specific antigen and the prostate health index for detecting prostate cancer. *Clin Chem* 2013:**59**:306–14. http://dx.doi.org/10.1373/clinchem.2012.195784

105. Tombal B, Andriole GL, Taille A, Gontero P, Haese A, Remzi M, *et al.* Clinical judgment versus biomarker prostate cancer gene 3: which is best when determining the need for repeat prostate biopsy? *Urology* 2013;**81**:998–1004. http://dx.doi.org/10.1016/j.urology.2012.11.069

106. Wu AK, Reese AC, Cooperberg MR, Sadetsky N, Shinohara K. Utility of PCA3 in patients undergoing repeat biopsy for prostate cancer. *Prostate Cancer Prostatic Dis* 2012;**15**:100–5. http://dx.doi.org/10.1038/pcan.2011.52

107. 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**:280–8. http://dx.doi.org/10.1373/clinchem.2012.195560

108. Tombal B, Ameye F, de la Taille A, de Reijke T, Gontero P, Haese A, *et al.* Biopsy and treatment decisions in the initial management of prostate cancer and the role of PCA3; a systematic analysis of expert opinion. *World J Urol* 2012;**30**:251–6. http://dx.doi.org/10.1007/s00345-011-0721-0

109. Chun FK, de la Taille A, van Poppel H, Marberger M, Stenzi A, Mulders PF, *et al.* Prostate cancer gene 3 (PCA3): development and internal validation of a novel biopsy nomogram. *Eur Urol* 2009;**56**:659–68. http://dx.doi.org/10.1016/j.eururo.2009.03.029

110. Luo Y, Gou X, Huang P, Mou C. The PCA3 test for guiding repeat biopsy of prostate cancer and its cut-off score: a systematic review and meta-analysis. *Asian J Androl* 2014;**16**:487–92. http://dx.doi.org/10.4103/1008-682X.125390

111. Liverpool Reviews and Implementation Group. *Clinical and Cost Effectiveness of the PROGENSA PCA3 Assay and the Prostate Health Index in the Diagnosis of Prostate Cancer: A Systematic Review and Economic Evaluation [Protocol].* Liverpool: Liverpool Reviews and Implementation Group; 2014.

112. Drummond M, Jefferson T. Guidelines for authors and peer reviewers of economic submissions to the BMJ. The BMJ economic evaluation working party. *BMJ* 1996;**313**:275–83. http://dx.doi.org/10.1136/bmj.313.7052.275

113. Heijnsdijk E, Huang J, Denham D, De Koning H. The cost-effectiveness of prostate cancer detection using Beckman Coulter Prostate Health Index. *Eur Urol Suppl* 2012;**11**:pe260. http://dx.doi.org/10.1016/S1569-9056(12)60257-7

114. Malavaud B, Cussenot O, Mottet N, Rozet F, Ruffion A, Smets L. Impact of adoption of a decision algorithm including PCA3 for repeat biopsy on the costs for prostate cancer diagnosis in France. *J Med Econ* 2013;**16**:358–63. http://dx.doi.org/10.3111/13696998.2012.757552

115. Nepple K, Strope S, Kibel A, Sandhu G, Weigand L, Kymes S. Cost-analysis of PCA3 versus PSA in the detection of prostate cancer in men with a prior negative biopsy. *J Urol* 2012;**187**:e55. http://dx.doi.org/10.1016/j.juro.2012.02.184

116. Nichol M, Wu J, Huang J, Denham D, Frencher S, Jacobsen S. Cost-effectiveness of prostate health index for prostate cancer detection. *BJU Int* 2011;**110**:353–62. http://dx.doi.org/10.1111/j.1464-410X.2011.10751.x

117. Heijnsdijk EA, Wever EM, Auvinen A, Hugosson J, Ciatto S, Nelen V, *et al.* Quality-of-life effects of prostate-specific antigen screening. *N Engl J Med* 2012;**367**:595–605. http://dx.doi.org/10.1056/NEJMoa1201637

118. National Institute for Health and Care Excellence. *Guide to the Methods of Technology Appraisal 2013*. London: NICE; 2013. URL: www.nice.org.uk/article/pmg9/resources/non-guidance-guide-to-the-methods-of-technology-appraisal-2013-pdf (accessed 10 September 2014).

119. Bill-Axelson A, Holmberg L, Ruutu M, Garmo H, Stark J, Busch C, *et al.* Radical prostatectomy versus watchful waiting in early prostate cancer. *N Engl J Med* 2011;**364**:1708–17. http://dx.doi.org/10.1056/NEJMoa1011967

120. Nam RK, Saskin R, Lee Y, Liu Y, Law C, Klotz LH. Increasing hospital admission rates for urological complications after transrectal utlrasound guided prostate biopsy. *J Urol* 2010;**183**:963–8. http://dx.doi.org/10.1016/j.juro.2009.11.043

121. Rosario DJ, Lane JA, Metcalfe C, Donovan JL, Doble A, Goodwin L, *et al.* Short term outcomes of prostate biopsy in men tested for cancer by prostate specific antigen: prospective evaluation within ProtecT study. *BMJ* 2012;**344**:d7894. http://dx.doi.org/10.1136/bmj.d7894

122. Curtis L. *Unit Costs of Health and Social Care 2013*. Canterbury: PSSRU, University of Kent; 2013.

123. de Haes JC, de Koning HJ, van Oortmarssen GJ, van Agt HM, de Bruyn AE, van Der Maas PJ. The impact of a breast cancer screening programme on quality-adjusted life-years. *Int J Cancer* 1991;**49**:538–44. http://dx.doi.org/10.1002/ijc.2910490411

124. Ara R, Wailoo AJ. *The Use of Health State Utility Values in Decision Models*. NICE DSU Technical Support Document. 2011.

125. Vickers AJ. Counterpoint: Prostate-specific antigen velocity is not of value for early detection of cancer. *J Natl Compr Canc Netw* 2013;**11**:286–90.

126. Vickers AJ, Savage C, O'Brien MF, Lilja H. Systematic review of pretreatment prostate-specific antigen velocity and doubling time as predictors for prostate cancer. *J Clin Oncol* 2009;**27**:398–403. http://dx.doi.org/10.1200/JCO.2008.18.1685

127. Vickers AJ, Cronin AM. Everything you always wanted to know about evaluating prediction models (but were too afraid to ask). *Urology* 2010;**76**:1298–301. http://dx.doi.org/10.1016/j.urology.2010.06.019

128. Filella X, Foj L, Mila M, Auge JM, Molina R, Jimenez W. PCA3 in the detection and management of early prostate cancer. *Tumour Biol* 2013;**34**:1337–47. http://dx.doi.org/10.1007/s13277-013-0739-6

129. 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. http://dx.doi.org/10.1038/srep05012

130. Wilt TJ. The Prostate Cancer Intervention Versus Observation Trial: VA/NCI/AHRQ Cooperative Studies Program #407 (PIVOT): design and baseline results of a randomized controlled trial comparing radical prostatectomy with watchful waiting for men with clinically localized prostate cancer. *J Natl Cancer Inst Monogr* 2012;**2012**:184–90. http://dx.doi.org/10.1093/jncimonographs/lgs041

131. Chu DI, De Nunzio C, Gerber L, Thomas JA 2nd, Calloway EE, Albisinni S, *et al.* Predictive value of digital rectal examination for prostate cancer detection is modified by obesity. *Prostate Cancer Prostatic Dis* 2011;**14**:346–53. http://dx.doi.org/10.1038/pcan.2011.31

132. Saah AJ, Hoover DR. 'Sensitivity' and 'specificity' reconsidered: the meaning of these terms in analytical and diagnostic settings. *Ann Intern Med* 1997;**126**:91–4. http://dx.doi.org/10.7326/0003-4819-126-1-199701010-00026

133. Betz JM, Brown PN, Roman MC. Accuracy, precision, and reliability of chemical measurements in natural products research. *Fitoterapia* 2011;**82**:44–52. http://dx.doi.org/10.1016/j.fitote.2010.09.011

134. Labnetwork. *Method Validation*. URL: www.labnetwork.org/en/chemical-lab-recent/111-method-validation (accessed 1 September 2014).

135. Hayen A, Macaskill P, Irwig L, Bossuyt P. Appropriate statistical methods are required to assess diagnostic tests for replacement, add-on, and triage. *J Clin Epidemiol* 2010;**63**:883–91. http://dx.doi.org/10.1016/j.jclinepi.2009.08.024

136. Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. *Med Decis Making* 2006;**26**:565–74. http://dx.doi.org/10.1177/0272989X06295361

137. Steyerberg EW, Vickers AJ. Decision curve analysis: a discussion. *Med Decis Making* 2008;**28**:146–9. http://dx.doi.org/10.1177/0272989X07312725

Appendix 1 Outcome measures

Analytical validity

Analytical validity can be subdivided into the following components:

- Pre-analytical variability refers to the extent to which factors such as sampling methods, transport, storage and temperature of the samples before they are analysed affect the results of the assay. Pre-analytical variables considered can also include age, ethnicity and genotype, which affect the normal ranges of the results.
- Analytical specificity refers to the ability of an assay to measure a particular substance, rather than others, in a sample.[132] It is tested by examining the crossover reaction with other substances and drugs.
- Analytical sensitivity represents the smallest amount of substance in a sample that can accurately be measured by an assay.[132] It is usually measured by:
  - LoQ which is the lowest amount of analyte in a sample that can be reliably detected and at which the total error meets the pre-specified requirement for accuracy.[48]
  - LoB which is the highest measurement that is likely to be observed for a blank sample.[48]
  - LoD which may be defined as LoB plus 1.65 SD.[74]
- Accuracy is a measure of the closeness of the experimental value to the actual amount of the substance in the matrix.[133] Accuracy often depends on what is used as the true value and whether or not there is a gold standard available. Accuracy is typically assessed by spiked recovery studies in which the amount of a target compound is determined as a percentage of the theoretical amount present in the matrix.
- Precision measures how close individual measurements of a sample are to each other.[133] Precision is often measured using the CV, which is the SD of the repeated measurement divided by the mean value expressed as a percentage. Precision is subdivided into various components:
  - Repeatability is a measure of the within-laboratory uncertainty. It can be divided into within- and between-run variability.
  - Intermediate precision is a measure of the ruggedness of the method, that is reliability when performed in different environments. Demonstration of intermediate precision requires that the method be run on multiple days by different analysts and on different instruments. Robustness is the capacity of a method to remain unaffected by small deliberate variations in method parameters. The robustness of a method is evaluated by varying method parameters.[134]

Clinical validity

In clinical validity studies the diagnostic accuracy of a new or intervention test is assessed against a reference standard. The reference standard is the best test available, that is the current preferred method of diagnosing a disease. In the case of prostate cancer the reference standard is a biopsy. All new tests need to be compared against the diagnostic accuracy of a biopsy.

Measures for assessing a single test against a reference standard

The classic presentation of the results of a clinical validity study is the so-called 2 × 2 table as shown in *Table 66*.

The number entered into cell 'a' is the number of patients for whom the new test correctly diagnoses prostate cancer (as determined by the reference standard, in this case a biopsy). For these people, the new test is positive as is the reference standard: these are the TPs.

TABLE 66 Example of a 2 × 2 table

Test result	Biopsy results (reference standard)	
	Prostate cancer	No prostate cancer
New test positive	a	b
New test negative	c	d

The number entered into cell 'b' is the number of patients for whom the new test is positive (i.e. indicates the presence of prostate cancer) but who do not, according to the reference standard (biopsy), have prostate cancer. The new test has incorrectly diagnosed prostate cancer: these are FPs.

The number entered into cell 'c' is the number of patients who are identified through the reference standard (biopsy) as having prostate cancer but for whom the new test gave negative results. The new test has incorrectly labelled the patient as having prostate cancer: these are FNs.

The number in cell 'd' is the number of patients who do not, according to the reference standard (biopsy), have prostate cancer and who are also shown by the new test to be free from disease: these are TNs.

The numbers displayed in a 2 × 2 table are used to generate other summary measures. These are set out in *Table 67*.

In an ideal world, a test would be 100% sensitive and 100% specific. However, in reality there is often a trade-off between the two, with tests that have high sensitivity also having low specificity and vice versa.

The use of a 2 × 2 tables requires that the test results are dichotomous, that is can be divided into two groups: test positive and test negative. If the actual test results are continuous variables, similar to PCA3 or phi scores, this means that a threshold (or a cut-off point) needs to be selected to divide the results into positive and negative groups.

Differences in means or medians

Another way of comparing the results of continuous variables is to compare the means or medians of test results between biopsy-positive and -negative men. The difference can be compared using analysis of variance.

Receiver operating characteristic curve

When an intervention test has a range of possible thresholds which could be used to divide results into test positive and test negative, the relationship between the threshold used and the performance of the test can be examined in a ROC curve. This is a graphical plot of the sensitivity (TP rate) against 1 – specificity or the FP rate for each threshold; examples of a ROC curve are shown in *Figure 12* with the associated distribution of the intervention tests in diseased and non-diseased populations. An ideal test would have a point in the top-left corner, with 100% specificity and 100% sensitivity.

TABLE 67 Summary measures derived from numbers in a 2 × 2 table

Term	Formula	Notes
Sensitivity	a/(a + c)	Proportion of those who actually have disease who are correctly identified with positive test results. TP rate. High sensitivity = few FNs
Specificity	d/(b + d)	Proportion of those who do not actually have the disease who are correctly identified with negative test results. 1 – FP rate. High specificity = few FPs
Positive predictive value	a/(a + b)	The proportion of those with positive test results who actually have the disease
Negative predictive value	d/(c + d)	The proportion of those with negative test results who do not have the disease

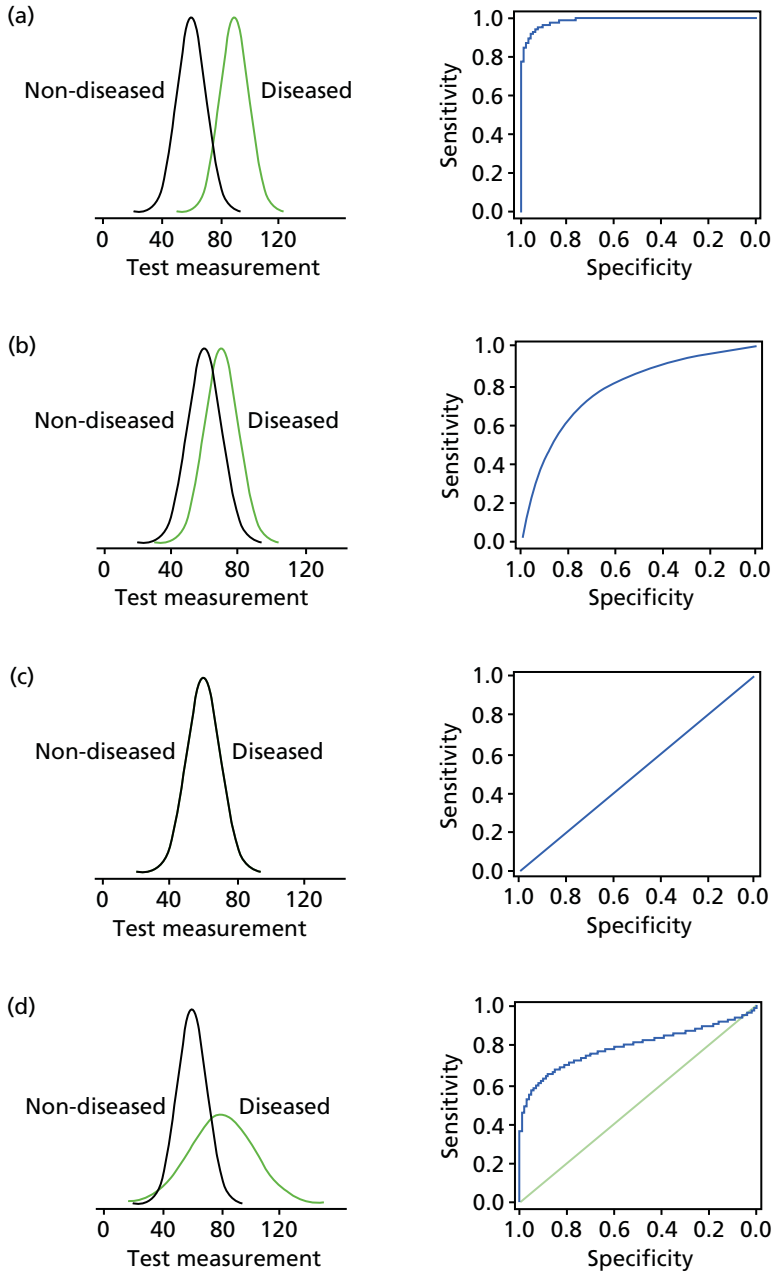

FIGURE 12 Examples of a ROC curve (image reproduced with permission from Chapter 10 of the Cochrane DTA handbook[63]). For explanatory text describing the ROC curves, please see Chapter 10 of the Cochrane handbook (p. 13).

The ROC curve can be used to assess the degree to which sensitivity changes at different levels of specificity or vice versa. Some studies report AUC as a proportion of the total area of the graph. This is a measure of the predictive accuracy or discrimination of the diagnostic test, that is the ability of the test to discriminate between those who have (or will develop) prostate cancer from those who do not have (or will not develop) prostate cancer. The AUC can also be expressed as the probability that someone with the disease will have a higher test result than someone without the disease. It is also referred to as the *c*-statistic. An AUC of 1.0 indicates a perfect test, and an AUC of 0.5 (the diagonal line) indicates that the test is no better than chance (i.e. 50% probability) in predicting whether or not the disease is present. An AUC of 0.5–0.7 is considered as poor discrimination, 0.7–0.8 acceptable discrimination, 0.8–0.9 excellent discrimination and above 0.9 exceptional discrimination.

Measures for assessing multiple tests against a reference standard

The measures discussed in *Measures for assessing a single test against a reference standard* can be used to compare a single intervention test with a reference standard. However, in clinical practice, test pathways often involve a series of tests used together. It is possible to combine 2 × 2 tables for a sequence of different tests if it is clear how the tests are used[135] (e.g. in parallel or in sequence), but results are rarely reported in this way. If the results of various tests are combined into an algorithm or decision tool within a study, data can be presented as a single test and analysed using sensitivity and specificity. However, when results are presented in this way, it can be unclear how each variable is used within the decision tool.

Most often, combinations of diagnostic tests are analysed using logistic regression models.

Logistic regression models

Logistic regression is a statistical method for analysing a data set in which there are one or more independent variables that determine an outcome. The outcome is measured with a dichotomous variable. A dichotomous variable is one with only two possible outcomes.

The goal of logistic regression is to find the best-fitting (yet biologically reasonable) model to describe the relationship between the dichotomous characteristic of interest (dependent variable = response or outcome variable) and a set of independent (predictor or explanatory) variables. In diagnostic logistic regression models, the outcome is the presence or absence of disease. In prostate cancer biopsy models, the outcome is presence or absence of prostate cancer and the independent variables are the intervention tests such as PCA3 score, age and/or PSA level. The independent variables may be used as dichotomous, continuous or categorical variables.

An OR is the outcome measure reported from the logistic regression model. An OR is a way of quantifying how strongly the outcome is associated with each of the variables used in the model, such as PSA level. The OR in a diagnostic logistic regression model (also called the diagnostic OR) is, for example, the odds that an individual with a 'raised PSA level' has prostate cancer relative to the odds that an individual without a 'raised PSA level' has prostate cancer. If the OR is greater than 1, then having 'prostate cancer' is considered to be 'associated' with having a 'raised PSA level', meaning that having a 'raised PSA level' raises (relative to not having a 'raised PSA level') the odds of having 'prostate cancer'. The OR demonstrates only an association between the two variables; causality has not been shown. When multiple variables are entered into a logistic regression model, the OR of each variable is adjusted to take account of the effects of other variables.

Receiver operating characteristics curves and derived sensitivity, and derived specificity from logistic regression models

The output from diagnostic logistic regression models can be used to generate ROC curves. These analyses rely on the predicted probability risk of having the outcome (in this case, prostate cancer) generated by the statistical model for each participant. By selecting a threshold probability risk of, say, 0.3, the participants can be classified as test positive or test negative depending on whether or not their predicted probability from the model is above or below 0.3. By varying the threshold predicted probability, ROC curves can be generated. The performance of different diagnostic models can be compared using the AUC. The AUC gives a measure of predictive accuracy, but is not very meaningful in clinical practice.

Receiver operating characteristics curves can also be used to derive sensitivity and specificity for alternative diagnostic models. For a set level, of for example 90% sensitivity, the specificity of various models can be calculated along with the associated threshold for predicted probability of a positive biopsy that has been used to generate these levels of sensitivity and specificity. However, the threshold predicted probability does not have relevance clinically and cannot be used to identify the threshold of an intervention test used in clinical practice above which patients should be recommended for biopsy.

Estimates of derived sensitivity and specificity from logistic regression models are more useful clinically than AUC estimates, as improvements in sensitivity or specificity can be described in terms of numbers of missed cancers or avoided biopsies. However, these derived sensitivity or specificity estimates are derived from ROC curves generated from logistic regression models and it is often not possible to associate the demonstrated improvement in sensitivity or specificity with the use of a particular threshold of the intervention test.

Decision curve analysis

Decision curve analysis is designed to present more clinically useful results when comparing diagnostic strategies.[136] The method calculates the net benefit of a diagnostic model by subtracting the harm of unnecessary biopsies from the benefit of diagnosed cases of prostate cancer. Unlike the conventional trade-off between sensitivity and specificity, in decision curve analysis there is an attempt to weight the relative harms and benefits using the threshold probability of cancer at which the patient or clinician will opt for a biopsy. For instance, when a clinician recommends a biopsy for any patient with a 10% or higher risk of cancer, which suggests that that the risk associated with an unnecessary biopsy is weighted less than the risk associated with an unnecessary biopsy when a 50% or higher risk of cancer is required before a biopsy is offered. The results are presented as graphs of net benefit over the range of probability risk stated to be clinically important, that is in which patients or clinicians might be uncertain whether or not to biopsy. This clinically important range of probability risk is typically from 10% to 40% for repeat prostatic biopsy. In the decision curve analysis graph, as well as displaying curves for each included model, there are two references lines shown: one for treating/biopsying no patients and one for biopsying all patients. The percentage reduction in biopsies for each diagnostic model compared with the biopsy-all strategy is another way of presenting the results. When interpreting results an emphasis is placed on whether or not the model adds any information to decision-making over the indicated range of probability.[137] The results do not present statistical significance tests and no methods of comparing or pooling the results across different studies are available.

Appendix 2 Literature search strategies

MEDLINE (via Ovid) and OLDMEDLINE (via Ovid)

Date range: 1946 to present with daily update.

Search name: PCA3_analytic.

Date run: 28 April 2014.

Search strategy

1. exp prostatic neoplasms/ (91,977)
2. (prostat* adj3 (cancer or carcinoma* or neoplasm* or malignant* or tumor* or tumour*)).tw. (84,448)
3. or/1-2 (104,576)
4. (Prostat* adj2 cancer* adj2 (antigen* or gene*) adj2 "3").tw. (107)
5. (PCA3 or PCA-3 or "PCA 3").tw. (275)
6. uPM3.tw. (7)
7. 7("differential display code 3 antigen" or DD3).tw. (80)
8. progensa.tw. (26)
9. or/4-8 (356)
10. prostate health index.tw. (30)
11. Beckman Coulter.tw. (515)
12. (proPSA or p2proPSA).tw. (70)
13. or/10-12 (582)
14. or/9,13 (928)
15. 3 and 14 (355)
16. exp animals/ not humans/ (3,930,803)
17. nonhuman/ not human/ (0)
18. or/16-17 (3,930,803)
19. 15 not 18 (354)
20. limit 19 to yr=2000-2014 (344)
21. Accuracy/ (106)
22. exp Diagnostic Errors/ (93,798)
23. exp "Sensitivity and Specificity"/ (411,853)
24. exp "reproducibility of results"/ (270,891)
25. analytic validity.mp. (47)
26. (repeatability or reproducibility).mp. (300,172)
27. or/21-26 (674,714)
28. 14 and 27 (410)
29. 28 not 19 (236)
30. or/9-10,12 (432)
31. 30 and 27 162)
32. 31 not 19 (4)

MEDLINE (via Ovid) and OLDMEDLINE (via Ovid)

Date range: 1946 to present with daily update.

Search name: PCA3_comparator.

Date run: 28 April 2014.

Search strategy

1. exp Magnetic Resonance Spectroscopy/ (177,430)
2. magnetic resonance imaging/ or exp diffusion magnetic resonance imaging/ (289,134)
3. magnetic resonance imag$.tw. (128,216)
4. magnetic resonance spectroscop*.tw. (15,525)
5. mrs.tw. (10,926)
6. (dynamic contrast enhanced adj3 (MRI or magnetic)).tw. (2026)
7. dce-mri.tw. (1185)
8. (diffusion weight$ adj3 (MRI or magnetic)).tw. (3682)
9. dw-mri.tw. (425)
10. ((multi-parametric or multiparametric or mp) adj (MRI or magnetic)).tw. (294)
11. or/1-10 (498,078)
12. exp Prostate/ah, pa, us [Anatomy & Histology, Pathology, Ultrasonography] (11,845)
13. (transrectal adj (biops* or ultrasound or ultrason*)).tw. (5135)
14. trus.tw. (1664)
15. exp Biopsy, Needle/ (52,622)
16. (biopsy or biopsies or pathol* or histopathol*).tw. (821,981)
17. or/12-16 (854,276)
18. exp Prostate-Specific Antigen/ (18,924)
19. psa.tw. (20,988)
20. prostat* specific antigen*.tw. (18,277)
21. or/18-20 (31,453)
22. exp nomograms/(1280)
23. nomogram*.tw. (4429)
24. (neural adj2 network).tw. (12,074)
25. or/22-24 (16,748)
26. exp prostatic neoplasms/ (91,977)
27. (prostat* adj3 (cancer or carcinoma* or neoplasm* or malignant* or tumor* or tumour*)).tw.(84,448)
28. or/26-27 (104,576)
29. or/11,17,21,25 (1,336,436)
30. 28 and 29 (38,357)
31. exp meta-analysis/ (47,236)
32. exp Meta-Analysis as Topic/ (13,686)
33. Meta-analys*.mp. or (meta adj analys*).ti,ab. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier] (74,826)
34. meta-regress*.mp. or (meta adj regress*).ti,ab. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier] (2040)
35. meta analysis.pt. (47,236)
36. systematic review.ti. (26,922)
37. or/31-36 (90,728)
38. 30 and 37 (262)

MEDLINE (via Ovid) and OLDMEDLINE (via Ovid)

Date range: 1946 to present with daily update.

Search name: PCA3_analytic.

Date run: 28 April 2014.

Search strategy

1. exp prostatic neoplasms/ (91,977)
2. (prostat* adj3 (cancer or carcinoma* or neoplasm* or malignant* or tumor* or tumour*)).tw. (84,448)
3. or/1-2 (104,576)
4. (Prostat* adj2 cancer* adj2 (antigen* or gene*) adj2 "3").tw. (107)
5. (PCA3 or PCA-3 or "PCA 3").tw. (275)
6. uPM3.tw. (7)
7. ("differential display code 3 antigen" or DD3).tw. (80)
8. progensa.tw. (26)
9. or/4-8 (356)
10. prostate health index.tw. (30)
11. Beckman Coulter.tw. (515)
12. (proPSA or p2proPSA).tw. (70)
13. or/10-12 (582)
14. or/9,13 (928)
15. 3 and 14 (355)
16. exp animals/ not humans/ (3,930,803)
17. nonhuman/ not human/ (0)
18. or/16-17 (3,930,803)
19. 15 not 18 (354)
20. limit 19 to yr=2000-2014 (344)
21. Accuracy/ (106)
22. exp Diagnostic Errors/ (93,798)
23. exp "Sensitivity and Specificity"/ (411,853)
24. exp "reproducibility of results"/ (270,891)
25. analytic validity.mp. (47)
26. (repeatability or reproducibility).mp. (300,172)
27. or/21-26 (674,714)
28. 14 and 27 (410)
29. 28 not 19 (236)
30. or/9-10,12 (432)
31. 30 and 27 (162)
32. 31 not 19 (4)

MEDLINE (via Ovid) and OLDMEDLINE (via Ovid)

Date range: 1946 to present with daily update.

Search name: PCA3_comparator.

Date run: 28 April 2014.

Search strategy

1. exp Magnetic Resonance Spectroscopy/ (177,430)
2. magnetic resonance imaging/ or exp diffusion magnetic resonance imaging/ (289,134)
3. magnetic resonance imag$.tw. (128,216)
4. magnetic resonance spectroscop*.tw. (15,525)
5. mrs.tw. (10,926)
6. (dynamic contrast enhanced adj3 (MRI or magnetic)).tw. (2026)
7. dce-mri.tw. (1185)
8. (diffusion weight$ adj3 (MRI or magnetic)).tw. (3682)
9. dw-mri.tw. (425)
10. ((multi-parametric or multiparametric or mp) adj (MRI or magnetic)).tw. (294)
11. or/1-10 (498,078)
12. exp Prostate/ah, pa, us [Anatomy & Histology, Pathology, Ultrasonography] (11,845)
13. (transrectal adj (biops* or ultrasound or ultrason*)).tw. (5135)
14. trus.tw. (1664)
15. exp Biopsy, Needle/ (52,622)
16. (biopsy or biopsies or pathol* or histopathol*).tw. (821,981)
17. or/12-16 (854,276)
18. exp Prostate-Specific Antigen/ (18,924)
19. psa.tw. (20,988)
20. prostat* specific antigen*.tw. (18,277)
21. or/18-20 (31,453)
22. exp nomograms/ (1280)
23. nomogram*.tw. (4429)
24. (neural adj2 network).tw. (12,074)
25. or/22-24 (16,748)
26. exp prostatic neoplasms/ (91,977)
27. (prostat* adj3 (cancer or carcinoma* or neoplasm* or malignant* or tumor* or tumour*)).tw. (84,448)
28. or/26-27 (104,576)
29. or/11,17,21,25 (1,336,436)
30. 28 and 29 (38,357)
31. exp meta-analysis/ (47,236)
32. exp Meta-Analysis as Topic/ (13,686)
33. Meta-analys*.mp. or (meta adj analys*).ti,ab. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier] (74,826)
34. meta-regress*.mp. or (meta adj regress*).ti,ab. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier] (2040)
35. meta analysis.pt. (47,236)
36. systematic review.ti. (26,922)
37. or/31-36 (90,728)
38. 30 and 37 (262)

The Cochrane Library

Date range: start date to April 2014.

Search name: PCA3_studies.

Date run: 28 April 2014.

Search strategy

#1 MeSH descriptor: [Prostatic Neoplasms] explode all trees (3325)
#2 prostat* near/3 (cancer or carcinoma* or neoplasm* or malignant* or tumor* or tumour*) (5935)
#3 #1 or #2 (5935)
#4 Prostat* near/2 cancer* near/2 (antigen* or gene*) near/2 "3" (15)
#5 PCA3 or PCA-3 or "PCA 3" (26)
#6 uPM3 (0)
#7 "differential display code 3 antigen" or DD3 (3)
#8 progensa (6)
#9 [or #4-#8] (30)
#10 "prostate health index" (5)
#11 Beckman Coulter (20)
#12 proPSA or p2proPSA (3)
#13 [or #10-#12] (24)
#14 (#9 or #13) and #3 (28)
#15 #14 Publication Date from 2000 to 2014 (28)

EMBASE

Date range: 1980 to week 20 2014.

Date run: 19 May 2014.

Search strategy

1. exp prostate cancer/ (127,704)
2. (prostat* adj3 (cancer or carcinoma* or neoplasm* or malignant* or tumor* or tumour*)).tw. (119,035)
3. or/1-2 (153,396)
4. (prostate cancer* adj2 (antigen* or gene*) adj2 "3").tw. (204)
5. (PCA3 or PCA-3 or "PCA 3").tw. (600)
6. uPM3.tw. (11)
7. ("differential display code 3 antigen" or DD3).tw. (123)
8. progensa.tw. (81)
9. or/4-8 (732)
10. prostate health index.tw. (135)
11. Beckman Coulter.tw. (1719)
12. (proPSA or p2proPSA).tw. (182)
13. or/10-12 (1845)
14. 9 or 13 (2543)
15. 3 and 14 (794)

16. animal/ or animal experiment/ (3,209,456)
17. exp human/ or human experiment/ (14,662,346)
18. 16 not (16 and 17) (2,687,279)
19. 15 not 18 (793)
20. limit 19 to yr="2000 - 2014" (781)

EMBASE

Date range: 1974 to 16 May 2014.

Date run: 19 May 2014.

Search strategy

1. exp nuclear magnetic resonance spectroscopy/ (92,226)
2. exp nuclear magnetic resonance imaging/ or exp diffusion weighted imaging/ (528,862)
3. magnetic resonance imag*.tw. (163,725)
4. magnetic resonance spectroscop*.tw. (18,845)
5. mrs.tw. (17,870)
6. (dynamic contrast enhanced adj3 (MRI or magnetic)).tw. (2790)
7. dce-mri.tw. (1936)
8. (diffusion weight* adj3 (MRI or magnetic)).tw. (5296)
9. dw-mri.tw. (749)
10. ((multi-parametric or multiparametric or mp) adj (MRI or magnetic)).tw. (688)
11. or/1-10 (650,710)
12. exp prostate/ (38,897)
13. (transrectal adj (biops* or ultrasound or ultrason*)).tw. (7484)
14. trus.tw. (3448)
15. exp needle biopsy/ (34,761)
16. (biopsy or biopsies or pathol* or histopathol*).tw. (1,149,785)
17. or/12-16 (1,197,743)
18. exp prostate specific antigen/ (35,908)
19. psa.tw. (37,014)
20. prostat* specific antigen*.tw. (23,016)
21. or/18-20 (56,049)
22. exp nomogram/ (4169)
23. nomogram*.tw. (6493)
24. (neural adj2 network).tw. (17,045)
25. or/22-24 (24,162)
26. exp prostate tumor/ (154,040)
27. (prostat* adj3 (cancer or carcinoma* or neoplasm* or malignant* or tumor* or tumour*)).tw. (120,976)
28. 26 or 27 (164,596)
29. 11 or 17 or 21 or 25 (1,817,863)
30. 28 and 29 (68,602)
31. exp meta analysis/ (78,651)
32. exp Meta-Analysis as Topic/ (13,178)
33. Meta-analys*.mp. or (meta adj analys*).ti,ab. (118,781)
34. meta-regress*.mp. or (meta adj regress*).ti,ab. (2908)
35. systematic review*.ti,ab. (64,099)
36. or/31-35 (158,492)
37. 30 and 36 (674)

Appendix 3 Data extraction forms

TABLE 68 Data extraction form for analytical validity

Number	Item	Comment
Miscellaneous		
1	Publication type (e.g. full report, abstract, letter, unpublished)	
2	Other reports from same study population?	
3	Funding	
4	Conflict of interest?	
5	Notes	
Test and study population		
6	Test name (e.g. PCA3/phi/p2PSA)	
7	Details of test platform/methods evaluated	
8	Country and setting	
9	Number of participants	
10	Age of participants	
11	Ethnicity of participants	
12	% of participants with prostatic disease	
13	Pre-analytical variables studied (age, DRE, genetic or ethnicity)	
14	Number of centres/labs tested	
15	Number of samples tested	
16	Timing and locations of repeat assays	
17	Standard/control sample used	
18	Other notes about how test conducted and/or data collected (e.g. likely to reflect how samples collected in clinical practice?)	
Results		
19	Test failure rate	
20	Analytical sensitivity (e.g. LoB, LoD or LoQ)	
21	Analytical specificity (e.g. crossreactivity and carry over)	
22	Accuracy (e.g. comparison to a 'gold standard' reference test and recovery)	
23	Linearity and range	
24	Precision (reproducibility and %CV)	
25	Other	

continued

TABLE 68 Data extraction form for analytical validity (*continued*)

Number	Item	Comment
Quality assessment		
26	Adequate descriptions of test under evaluation (reports specific methods/platforms evaluated, quality assurance measures, e.g. control samples; see responses to 6 to 18 above)	
27	Comparison to a 'gold standard' reference test?	
28	Specimens represent routinely analysed clinical specimens in all aspects (e.g., collection, transport, processing; see response to 18 above)	
29	Relevant outcomes to assess analytical validity adequately addressed? (see responses to 19 to 25 above)	
30	Sample size/power calculations addressed?	

TABLE 69 Data extraction form for clinical validity studies within-study comparisons

Study details	Description/location in text
Date form completed (dd/mm/yyyy)	
Name/ID of person extracting data	
Record number, author, year (ID for this paper/abstract/report)	
Name of study	
Other reports from same study population	
Publication type (e.g. full report, abstract or letter)	
Funding/conflicts of interest	

Study design	Description/location in text
Aim of study	
Design (e.g. cohort, cross-sectional, case–control, randomised)	
Number of centres	
Country, type of hospital	
Method/s of recruitment of participants (e.g. consecutive, random sample, retrospective selection)	
Informed consent obtained	
Ethical approval needed/obtained for study	
Method of allocation to test pathway if not all participants received both tests	
Start date	
End date	
Total study duration	

Participants	Description/location in text
Total number in study	
Number with previous negative biopsy (use this number for results)	
Inclusion criteria include: • PSA level • Clinical findings such as a DRE • Imaging results • Clinician recommendation	
Exclusion criteria	
Age	
Race/ethnicity	
PSA mean	
Other (DRE +ve, family history, # previous biopsy, imaging abnormalities, HGPIN, etc.)	
Baseline imbalances if not all participants received both tests	
Details of first negative biopsy taken: • Route (transrectal or transperineal) • Ultrasound guided • Sextant, extended or template • Type and number of cores taken • Definition of negative results • Proportion with abnormal histopathology	

Intervention test group: repeat if needed	Description as stated in report/paper
Test name (e.g. PCA3/phi/p2PSA)	
Details of urine/blood sample collection	
Details of test platform used	
Details of a DRE and test collection	
Number (%) informative test	
Threshold values used	
Was threshold pre-specified in Methods?	
Timing of test in relation to initial biopsy	
Timing of test in relation to repeat biopsy	
Was assessor blinded to other study results?	

Comparator tests reported	Tick if included	Details
PSA	☐	
MRI	☐	
Nomograms	☐	
Clinical risk factors (e.g. age, a DRE, etc.) – please list	☐	
Other – please list	☐	

Comparator test: PSA	Description as stated in report/paper. Location in text
Test name (e.g. tPSA/fPSA etc.)	
Number of participants test collected from	
Threshold values used	
Was threshold pre-specified in methods?	
Timing of test in relation to initial biopsy	
Timing of test in relation to repeat biopsy	
Was assessor blinded to other study results?	
Comparator test: MRI	**Description as stated in report/paper. Location in text**
Type of MRI name (e.g. T2, DW, DCE-MRI)	
Details of MRI used	
Number of participants received MRI	
Number (%) informative results	
Who did assessment/interpretation of MRI?	
Definition of abnormality?	
Was definition pre-specified in methods?	
Was assessor blinded to other results?	
Timing of MRI in relation to initial biopsy	
Timing of MRI in relation to PCA3/phi	
Timing of MRI in relation to repeat biopsy	
Comparator test: nomograms/clinical risk factors	**Description as stated in report/paper. Location in text**
Number of participants with nomogram/clinical assessment results	
Name of nomograms used	
Reference for nomogram	
Data incorporated in nomogram or clinical risk factors used	
Threshold values used for nomogram	
Was threshold/abnormal definition pre-specified in methods?	
Timing of data collection in relation to initial biopsy	
Timing of data collection in relation to PCA3/phi	
Timing of data collection in relation to repeat biopsy	
Blinding of clinical assessment to other results?	

Reference standard	Description as stated in report/paper
Type of repeat biopsy: • transrectal or transperineal • sextant/extended or template	
(Include description of any differences in reference standard dependent on result of intervention/comparator tests)	
Number of cores taken	
Timing of biopsy	
Definition of positive biopsy. HGPIN/ASAP included?	
Number (%) positive	
Other end points reported	
(Gleason score, % cores positive)	
Histopathology procedures and expertise	
Use of ultrasound for guiding biopsy	
Use of MRI-targeting technology	
Results of intervention/comparator test known to: • person taking biopsy • person reporting pathology	
Please draw up a flow chart of number of participants completing study	

TABLE 70 Comparisons and results: all cancers (copy and repeat table if > four comparisons reported)

Reference standard	Test pathway 1	Test pathway 2	Test pathway 3	Test pathway 4
PCA3	☐	☐	☐	☐
Cut-off/continuous				
phi	☐	☐	☐	☐
Cut-off/continuous				
Clinical risk factors/ nomogram	☐	☐	☐	☐
Details				
PSA	☐	☐	☐	☐
Details				
MRI	☐	☐	☐	☐
Details				

continued

TABLE 70 Comparisons and results: all cancers (copy and repeat table if > four comparisons reported) (*continued*)

Results								
N								
Means/medians (SD/IQR), units	☐		☐		☐		☐	
2 × 2 table – TP, TN, FN, FP	☐	Bx+ Bx– Test+ Test–	☐	Bx+ Bx– Test+ Test–	☐	Bx+ Bx– Test+ Test–	☐	Bx+ Bx– Test+ Test–
Sensitivity, specificity, LRs		Sensitivity: Specificity:		Sensitivity: Specificity:		Sensitivity: Specificity:		Sensitivity: Specificity:
ROC curves – graph	☐		☐		☐		☐	
Area under curve (95% CI)	☐		☐		☐		☐	
Derived sensitivity and specificity from curves	☐		☐		☐		☐	
Univariate ORs	☐	PCA3 phi PSA clin MRI	☐	PCA3 phi PSA clin MRI	☐	PCA3 phi PSA clin MRI	☐	PCA3 phi PSA clin MRI
Multivariate ORs	☐	PCA3 phi PSA clin MRI	☐	PCA3 phi PSA clin MRI	☐	PCA3 phi PSA clin MRI	☐	PCA3 phi PSA clin MRI

Bx, biopsy; IQR, interquartile range; LR, likelihood ratio.

TABLE 71 Comparisons and results: high-grade cancers (copy and repeat table if > four comparisons reported)

Definition of high grade cancer				
Reference standard				
	Test pathway 1	**Test pathway 2**	**Test pathway 3**	**Test pathway 4**
PCA3	☐	☐	☐	☐
Cut-off/continuous				
phi	☐	☐	☐	☐
Cut-off/continuous				
Clinical risk factors/ nomogram	☐	☐	☐	☐
Details				
PSA	☐	☐	☐	☐
Details				
MRI	☐	☐	☐	☐
Details				
Results				
N				
Means/medians (SD/IQR), units	☐	☐	☐	☐
2 × 2 table – TP, TN, FN, FP	☐	☐	☐	☐
	Bx+ Bx–	Bx+ Bx–	Bx+ Bx–	Bx+ Bx–
	Test+	Test+	Test+	Test+
	Test–	Test–	Test–	Test–
Sensitivity, specificity, LRs	Sensitivity:	Sensitivity:	Sensitivity:	Sensitivity:
	Specificity:	Specificity:	Specificity:	Specificity:
ROC curves – graph	☐	☐	☐	☐
Area under curve (95% CI)	☐	☐	☐	☐
Derived sensitivity and specificity from curves	☐	☐	☐	☐
Univariate ORs	☐	☐	☐	☐
	PCA3	PCA3	PCA3	PCA3
	phi	phi	phi	phi
	PSA	PSA	PSA	PSA
	clin	clin	clin	clin
	MRI	MRI	MRI	MRI

continued

TABLE 71 Comparisons and results: high-grade cancers (copy and repeat table if > four comparisons reported) (*continued*)

Results								
Multivariate ORs	☐	PCA3	☐	PCA3	☐	PCA3	☐	PCA3
		phi		phi		phi		phi
		PSA		PSA		PSA		PSA
		clin		clin		clin		clin
		MRI		MRI		MRI		MRI

Bx, biopsy; clin, clinical; IQR, interquartile range; LR, likelihood ratio.

TABLE 72 Quality assessment: Quality Assessment of Diagnostic Accuracy Studies – version 2

Patient selection	
A. Risk of bias	***Risk assessed as low/high/unclear***
Was a consecutive or random sample of patients enrolled?	
Was a case–control design avoided?	
Did the study avoid inappropriate exclusions?	
Were men selected into study on basis of cancer risk such as on PSA range, a DRE, MRI, etc.	
Could the selection of patients have introduced bias?	
Comments:	
B. Concerns regarding applicability	***Concerns assessed as low/high/unclear***
Was risk of underlying risk of cancer in men in study population representative?	
Are there concerns that the included patients and setting do not match the review question?	
Comments:	
Intervention test	
A. Risk of bias	***Risk assessed as low/high/unclear***
Were the intervention test results interpreted without knowledge of the results of the reference standard?	
If a threshold was used, was it prespecified?	
Were the intervention test results interpreted without knowledge of the results of the comparator tests?	
Could the conduct or interpretation of the intervention test have introduced bias?	
Comments:	
B. Concerns regarding applicability	***Concerns assessed as low/high/unclear***
Are there concerns that the intervention test, its conduct or its interpretation differs from the review question?	
Comments:	

TABLE 72 Quality assessment: Quality Assessment of Diagnostic Accuracy Studies – version 2 (*continued*)

Comparator test – clinical and PSA	
A. Risk of bias	***Risk assessed as low/high/unclear***
Were the comparator test results interpreted without knowledge of the results of the reference standard?	
If a threshold was used, was it prespecified?	
Were the comparator test results interpreted without knowledge of the results of the comparator tests?	
Could the conduct or interpretation of the comparator test have introduced bias?	
Comments:	
B. Concerns regarding applicability	***Concerns assessed as low/high/unclear***
Are there concerns that the comparator test, its conduct, or interpretation differs from the review question?	
Comments:	
Comparator test – MRI	
A. Risk of bias	***Risk assessed as low/high/unclear***
Were the comparator test results interpreted without knowledge of the results of the reference standard?	
If a threshold was used, was it prespecified?	
Were the comparator test results interpreted without knowledge of the results of the comparator tests?	
Could the conduct or interpretation of the comparator test have introduced bias?	
Comments:	
B. Concerns regarding applicability	***Concerns assessed as low/high/unclear***
Are there concerns that the comparator test, its conduct or its interpretation differ from the review question?	
Comments:	
Reference standard	
A. Risk of bias	***Risk assessed as low/high/unclear***
Is the reference standard likely to correctly classify the target condition?	
Was the reference standard performed and results interpreted without knowledge of the results of the intervention tests?	
Was the reference standard performed and results interpreted without knowledge of the results of the comparator tests?	
Were the same number and pattern of cores taken in all participants?	
Could the reference standard, its conduct, or its interpretation have introduced bias?	
Comments:	
B. Concerns regarding applicability	***Concerns assessed as low/high/unclear***
Are there concerns that the target condition as defined by the reference standard does not match the question?	
Comments:	

continued

TABLE 72 Quality assessment: Quality Assessment of Diagnostic Accuracy Studies – version 2 (*continued*)

Flow and timing	
A. Risk of bias	***Risk assessed as low/high/unclear***
Was there an appropriate interval between intervention test and reference standard?	
Could the patient flow have introduced bias?	
Comments:	
Summary	
Key conclusions of study authors	
Comments of review authors	
Action/queries/further investigation needed	

Appendix 4 Study characteristics of included studies for analytical validity review

TABLE 73 Study characteristics of included studies for analytical validity review

Study	Samples used	Outcomes reported					
		Pre-analytical variables	Analytical sensitivity	Analytical specificity	Accuracy	Precision	Linearity and range
PCA3 studies							
FDA SSED report[50]	Controls based on in vitro transcripts, clinical samples	Temperature stability of 12 clinical urine samples before and after processing, reagents, urine transport kit	LoB, LoD, LoQ using control samples	Unspliced transcript, interfering substances, carry-over	Recovery of female urine spiked with in vitro transcripts	Within-laboratory repeatability of control samples and patient samples; between-laboratory reproducibility of control samples	Linearity using both control and clinical samples
Pack insert[51]	Controls based on in vitro transcripts, clinical samples	Temperature stability of 10 clinical urine samples before and after processing	LoD, LoQ using control samples	Unspliced transcript, urine from post-prostatectomy patients, tissue specificity, interfering substances	Recovery of female urine spiked with in vitro transcripts	Within-laboratory repeatability of control samples; between-laboratory reproducibility of control and pooled clinical samples	Linearity using both control material diluted in female urine and diluent
Groskopf 2006[71]	Controls based on transcripts in detergent solution, clinical samples or 'previously characterised pooled processed urine specimens'	Temperature stability of three clinical urine samples after processing	NR	Urine from post-prostatectomy patients, urine from female patients	Recovery of three control samples	Within-laboratory repeatability of three control and three pooled urine samples	NR
Sokoll 2008[48]	Clinical samples, control samples based on in vitro transcripts	Informative rate for 179 clinical samples taken with or without a DRE and varying the strokes/lobe	LoB, LoD, LoQ using control samples	NR	Recovery of three control samples in two different sites	Within-laboratory repeatability of three control; between-laboratory reproducibility of three control	NR
Shappell 2009[72]	Clinical samples	NR	NR	NR	NR	Between-laboratory reproducibility of 50 clinical samples	NR

Study	Samples used	Outcomes reported					
		Pre-analytical variables	Analytical sensitivity	Analytical specificity	Accuracy	Precision	Linearity and range
p2PSA or phi studies							
FDA SSED[58] Draft Pack insert[57]	Patient samples (unspecified source), control samples based on internal reference preparation of p2PSA	Temperature stability of samples (same as in Semjonow *et al.*[73]), thermal sensitivity of assay; stability of reagents, calibrator and controls	LoB, LoD, LoQ using zero analyte and calibration samples (same as in Sokoll *et al.*[74])	Interfering substances, cross-reactivity with other forms PSA (same as in Sokoll *et al.*[74]), carryover	Recovery of six spiked samples, (same as in Sokoll *et al.*[74])	p2PSA: within-laboratory repeatability of three control and six patient samples; between-laboratory reproducibility of three control and three patient samples (same as in Sokoll *et al.*[74]) phi score: within-laboratory repeatability of one control and four patient samples; between-laboratory reproducibility of 10 patient samples	Linearity of 12 unspecified samples. Hook effect examined
Stephan 2009[75]	Control materials, spiked patient serum, in-house serum pool	NR	LoD, based on zero calibrator	Cross-reactivity with other forms PSA	Recovery of six spiked samples	p2PSA: within-laboratory repeatability of four control or three control and one pooled clinical sample; inter-assay precision of sample from serum pool and control samples	Linearity of six spiked samples
Semjonow 2010[73]	22 clinical samples from volunteers	Temperature stability of: clotted samples at 21 °C; serum at 4 °C, 21 °C, –20 °C and –70 °C freeze–thaw cycles	NR	NR	NR	NR	NR
Sokoll 2012[74]	Control samples, patient samples	NR	LoB, LoD, LoQ using zero analyte	Interfering substances, cross-reactivity with other forms PSA	Recovery of six spiked samples	p2PSA: within-laboratory repeatability of three control and three patient samples. Reported separately for four different laboratories	Linearity of three serum samples. Hook effect examined

NR, not reported.

Appendix 5 Tables of excluded studies

TABLE 74 Details of selected ineligible primary studies with reasons for exclusion

PCA3 studies excluded	Reason for exclusion
Chevli KK, Duff MI, Walter P, Yu C, Capuder B, Elshafeli A, *et al.* Urinary PCA3 as a predictor for prostate cancer in a cohort of 3073 men undergoing initial prostate biopsy. *J Urol* 2014;**191**:1743–8	Initial biopsy population and abstract
Crawford ED, Rove KO, Trabulsi EJ, Qian J, Drewnowska KP, Kaminetsky JC, *et al.* Diagnostic performance of PCA3 to detect prostate cancer in men with increased prostate specific antigen: a prospective study of 1,962 cases. *J Urol* 2012;**188**:1726–31	Initial biopsy population
Day JR, Jones LA, Meyer SE, Hodge PN, Aussie J, Saltzstein DR, *et al.* Urinary PCA3 and TMPRSS2:ERG help predict biopsy outcome prior to initial prostate biopsy using a risk group analysis. 28th Annual EAU Congress, Milan, Italy, 15–19 March 2013. *Eur Urol Suppl* 2013;**12**:e1045	Initial biopsy population and abstract
de la Taille A, Irani J, Graefen M, Chun F, de Reijke T, *et al.* Clinical evaluation of the PCA3 assay in guiding initial biopsy decisions. *J Urol* 2011;**185**:2119–25	Initial biopsy population
Deras IL, Aubin SMJ, Blase A, Day JR, Koo S, Partin AW, *et al.* PCA3: a molecular urine assay for predicting prostate biopsy outcome. *J Urol* 2008;**179**:1587–92	PSA/PCA3 only. Single study
Kella N, Day JR, Jones LA, Meyer SE, Hodge PN, Aussie J, *et al.* Urinary PCA3 and TMPRSS2: ERG help predict biopsy outcome prior to initial prostate biopsy using a risk group analysis. Annual Meeting of the American Urological Association, San Diego, CA, 4–8 May 2013	Initial biopsy population and abstract
Ochiai A, Okihara K, Kamoi K, Iwata T, Kawauchi A, Miki T, *et al.* Prostate cancer gene 3 urine assay for prostate cancer in Japanese men undergoing prostate biopsy. *Int J Urol* 2011;**18**:200–5	Mixed population
Roobol MJ, Schröder FH, van Leeuwen P, Wolters T, van den Bergh RC, van Leenders GJ, *et al.* Performance of the prostate cancer antigen 3 (PCA3) gene and prostate-specific antigen in prescreened men: exploring the value of PCA3 for a first-line diagnostic test. *Eur Urol* 2010;**58**:475–81	Not all repeat biopsies
Wei J, Sanda M, Thompson I, Partin A, Feng Z, Sokoll L, *et al.* The NCI Early Detection Research Network (EDRN) Urinary PCA3 Validation Trial. Annual Meeting of the American Urological Association, Atlanta, Georgia, 19–23 May 2012	Abstract only
Aubin SMJ, Reid J, Sarno MJ, Blase A, Aussie J, Rittenhouse H, *et al.* PCA3 molecular urine test for predicting repeat prostate biopsy outcome in populations at risk: validation in the placebo arm of the dutasteride REDUCE trial. *J Urol* 2010;**184**:1947–52	Ineligible population
Deras IL, Aubin SMJ, Blase A, Day JR, Koo S, Partin AW, *et al.* PCA3: a molecular urine assay for predicting prostate biopsy outcome. *J Urol* 2008;**179**:1587–92	Mixed biopsy population
Tombal B, Ameye F, de la Taille A, de Reijke T, Gontero P, Haese A, *et al.* Biopsy and treatment decisions in the initial management of prostate cancer and the role of PCA3; a systematic analysis of expert opinion. *World J Urol* 2012;**30**:251–6	Expert opinion of PCA3 impact on repeat biopsy decision
Wei J, Sanda M, Thompson I, Partin A, Feng Z, Sokoll L, *et al.* The NCI Early Detection Research Network (EDRN) Urinary PCA3 Validation Trial. Annual Meeting of the American Urological Association Atlanta, Georgia, 19–23 May 2012	Abstract only
Auprich M, Chun FKH, Ward JF, Pummer K, Babaian R, Augustin H, *et al.* Critical assessment of pre-operative urinary prostate cancer antigen 3 on the accuracy of prostate cancer staging. *Eur Urol* 2011;**59**:96–105	Indolent and aggressive prostate cancers following diagnosis
Lin DW, Newcomb LF, Brown EC, Brooks JD, Carroll PR, Feng Z, *et al.* Urinary TMPRSS2: ERG and PCA3 in an active surveillance cohort: results from a baseline analysis in the Canary Prostate Active Surveillance Study. *Clin Cancer Res* 2013;**19**:2442–50	Indolent and aggressive prostate cancers following diagnosis
Nakanishi H, Groskopf J, Fritche HA, Bhadkamkar V, Blase A, Kumar SV, *et al.* PCA3 molecular urine assay correlates with prostate cancer tumor volume: implication in selecting candidates for active surveillance. *J Urol* 2008;**179**:1804–10	Unclear population

continued

TABLE 74 Details of selected ineligible primary studies with reasons for exclusion (*continued*)

PCA3 studies excluded	Reason for exclusion
Ploussard G, Durand X, Xylinas E, Moutereau S, Radulescu C, Forgue A, *et al.* PCA3 score accurately predicts tumor volume and might help in selecting prostate cancer patients for active surveillance. *Eur Urol* 2011;**59**:422–9	Indolent and aggressive prostate cancers following diagnosis
van Poppel H, Haese A, Graefen M, de la Taille A, Irani J, de Reijke T, *et al.* The relationship between Prostate CAncer gene 3 (PCA3) and prostate cancer significance. *BJU Int* 2011;**109**:360–6	Results not reported separately for repeat
Vlaeminck-Guillem V, Devonec M, Colombel M, Rodriguez-Lafrasse C, Decaussin-Petrucci M, Ruffion A. Urinary PCA3 Score predicts prostate cancer multifocality. *J Urol* 2011;**185**:1234–9	Indolent and aggressive prostate cancers following diagnosis
Whitman EJ, Groskopf J, Ali A, Chen Y, Blase A, Furusato B, *et al.* PCA3 score before radical prostatectomy predicts extracapsular extension and tumor volume. *J Urol* 2008;**180**:1975–9	Indolent and aggressive prostate cancers following diagnosis
Marks LS, Bostwick DG. Prostate cancer specificity of PCA3 gene testing: examples from clinical practice. *Rev Urol* 2008;**10**:175–81	Review without meta-analysis
Schilling D, de Reijke T, Tombal B, de la Taille A, Hennenlotter J, Stenzl A. The Prostate Cancer gene 3 assay: indications for use in clinical practice. *BJU Int* 2009;**105**:452–5	Non-systematic review
Schilling D, Hennenlotter J, Munz M, Bökeler U, Sievert KD, Stenzl A. Interpretation of the prostate cancer gene 3 in reference to the individual clinical background: implications for daily practice. *Urol Int* 2010;**85**:159–65	Unclear population and study design – only some biopsied
Wang R, Chinnaiyan AM, Dunn RL, Wojno KJ, Wei JT. Rational approach to implementation of prostate cancer antigen 3 into clinical care. *Cancer* 2009;**115**:3879–86	Repeat results not reported separately
phi studies excluded	**Reason for exclusion**
Rhodes T, Jacobson DJ, McGree MS, St Sauver JL, Sarma AV, Girman CJ, *et al.* Distribution and associations of [–2]proenzyme-prostate specific antigen in community dwelling black and white men. *J Urol* 2012;**187**:92–6	Ineligible design
Nichol MB, Wu J, An JJ, Huang J, Denham D, Frencher S, *et al.* Budget impact analysis of a new prostate cancer risk index for prostate cancer detection. *Prostate Cancer Prostatic Dis* 2011;**14**:253–61	Ineligible design
Eichholz A, McCarthy F, Nening D, Thomas K, Howlett T, Iqbal J, *et al.* Prostate Health Index (phi) as a novel biomarker in active surveillance of prostate cancer (PCa). *J Clin Oncol* 2014;**32**:81	Abstract only
Boegemann M, Vincendeau S, Stephan C, Houlgatte A, Krabbe LM, Blanchet J-S, *et al.* The effect of [–2]proPSA and prostate health index (phi) on the accuracy of the prediction of initial and repeat prostate biopsies compared to tPSA and percent fPSA in young men (age 65 or younger). *J Clin Oncol* 2014;**32**:Abstract 171	Abstract only
Lughezzani G, Lazzeri M, Haese A, McNicholas T, de la Tailee A, Buffi NM, *et al.* Multicenter European external validation of a prostate health index-based nomogram for predicting prostate cancer at extended biopsy. *Eur Urol* 2014;**66**:906–12	Initial biopsy population
Lippi G, Aloe R, Cervellin G. p2PSA but not total and free PSA increases after myocardial infarction: results of a preliminary investigation. *Int J Cardiol* 2011;**153**:119	Letter RE: MI
Guazzoni G, Lazzeri M, Buffi NM, Abrate A, Mistretta FA, Hurle R, *et al.* Preoperative prostate-specific antigen isoform p2PSA and its derivatives, percent p2PSA and prostate health index, predict pathologic outcomes in patients undergoing radical prostatectomy for prostate cancer. *Eur Urol* 2012;**61**:455–66	Men diagnosed with clinically localised PCa
Filella X, Gimenez N. Evaluation of [–2] proPSA and Prostate Health Index (phi) for the detection of prostate cancer: a systematic review and meta-analysis. *Clin Chem Lab Med* 2013;**51**:729–39	Not a repeat biopsy population
Lazzeri M, Haese A, Abrate A, de la Taille, Redorta JP, NcNicholas T, *et al.* Clinical performance of serum prostate-specific antigen isoform [–2]proPSA (p2PSA) and its derivatives, per cent p2PSA and the prostate health index (PHI), in men with a family history of prostate cancer: results from a multicentre European study, the PROMEtheuS project. *BJU Int* 2013;**112**:313–21	Not a repeat biopsy population

TABLE 74 Details of selected ineligible primary studies with reasons for exclusion (*continued*)

phi studies excluded	Reason for exclusion
Catalona WJ, Partin AW, Sanda MG, Wei JT, Klee GG, Bangma CH, *et al.* A multicenter study of [–2]pro-prostate specific antigen combined with prostate specific antigen and free prostate specific antigen for prostate cancer detection in the 2.0 to 10.0 ng/ml prostate specific antigen range. *J Urol* 2011;**185**:1650–5	Not a repeat biopsy population
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**:921–7	Not a repeat biopsy population
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**:305–11	Not a repeat biopsy population
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**:1274–8	Not repeat biopsy population
Heidegger I, Klocker H, Steiner E, Skradski V, Ladurner M, Pichler R, *et al.* [–2]proPSA is an early marker for prostate cancer aggressiveness. *Prostate Cancer Prostatic Dis* 2014;**17**:70–4	Not repeat biopsy population
Ferro M, Bruzzese D, Perdona S, Mazzarella C, Marino A, Sorrentino A, *et al.* Prostate Health Index (Phi) and Prostate Cancer Antigen 3 (PCA3) significantly improve prostate cancer detection at initial biopsy in a total PSA range of 2–10 ng/ml. *PLOS ONE* 2013;**8**:e67687	Not repeat biopsy population
Lazzeri M, Haese A, de la Taille A, Palou Rodorta 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**:986–94	Not repeat biopsy population
Perdona 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**:227–35	Not repeat biopsy population
Ng CF, Chiu PK, Lam NY, Lam HC, Lee KW, Hou SS, *et al.* 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**:711–17	Not repeat biopsy population
Isharwal S, Makarov DV, Sokoll LJ, Landis P, Marlow C, Epstien JI, *et al.* ProPSA and diagnostic biopsy tissue DNA content combination improves accuracy to predict need for prostate cancer treatment among men enrolled in an active surveillance program. *Urology* 2011;**77**:e761–6	Patients with low-risk cancer
Tosoian JJ, Loeb S, Feng Z, 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**:1131–6	Patients with low-risk cancer
Hirama H, Sugimoto M, Ito K, Shiraishi T, Kakehi Y. The impact of baseline [–2]proPSA-related indices on the prediction of pathological reclassification at 1 year during active surveillance for low-risk prostate cancer: the Japanese multicenter study cohort. *J Cancer Res Clin Oncol* 2014;**140**:257–63	Patients with low-risk cancer

MI, myocardial infarction; PCa, prostate cancer.

TABLE 75 Ineligible systematic reviews with reasons for exclusion

Reference	Reason for exclusion
Filella X, Gimenez N. Evaluation of [–2] proPSA and prostate health index (phi) for the detection of prostate cancer: a systematic review and meta-analysis. *Clin Chem Lab Med* 2013;**51**:729–39	Initial biopsies only
Wang W, Wang M, Wang L, Adams TS, Tian Y, Xu J. Diagnostic ability of percent p2PSA and prostate health index for aggressive prostate cancer: a meta-analysis. *Sci Rep* 2014;**4**:5012	Mixed or unclear biopsy population
Luo Y, Gou X, Huang P, Mou C. Prostate cancer antigen 3 test for prostate biopsy decision: a systematic review and meta analysis. *Chin Med J* 2014;**127**:1768–74	Mixed or unclear biopsy population
Bruzzese D, Mazzarella C, Ferro M, Perdona S, Chiodini P, Perruolo G, *et al.* Prostate health index vs. percent free prostate-specific antigen for prostate cancer detection in men with 'gray' prostate-specific antigen levels at first biopsy: systematic review and meta-analysis. *Transl Res* 2014;**164**:444–51	Initial biopsies only
Harvey P, Basuita A, Endersby D, Curtis B, Lacovidou A, Walker M. A systematic review of the diagnostic accuracy of prostate specific antigen. *BMC Urol* 2009;**9**:14	Unclear initial or repeat biopsy
Lawrentschuk N, Fleshner N. The role of magnetic resonance imaging in targeting prostate cancer in patients with previous negative biopsies and elevated prostate-specific antigen levels (Structured abstract). *BJU Int* 2009;**103**:730–3	No meta-analysis, no reference standard
Overduin CG, Futterer JJ, Barentsz JO. MRI-guided biopsy for prostate cancer detection: a systematic review of current clinical results. *Curr Urol Rep* 2013;**14**:209–13	No meta-analysis, no reference standard
de Rooij M, Hamoen EH, Futterer JJ, Barentsz JO, Rovers MM. Accuracy of multiparametric MRI for prostate cancer detection: a meta-analysis. *Am J Roentgenol* 2014;**202**:343–51	Unclear initial or repeat biopsy
Nelson AW, Harvey RC, Parker RA, Kastner C, Doble A, Gnanapragasam VJ. Repeat prostate biopsy strategies after initial negative biopsy: meta-regression comparing cancer detection of transperineal, transrectal saturation and MRI guided biopsy. *PLOS ONE* 2013;**8**:e57480	No meta-analysis, no reference standard

Appendix 6 Within-study comparisons reporting univariate prostate cancer antigen 3 or Prostate Health Index scores only

TABLE 76 Study characteristics of studies reporting univariate PCA3 or phi scores only

Study name	Inclusion/exclusion criteria	Repeat biopsies (type, number of positive/total sample (%)	Comparisons reported	Author conclusions
Marks 2007[93]	Consecutive men with serum PSA levels of 2.5 ng/ml or greater who had a history of at least one negative biopsy, documented by the study site investigator and who had been scheduled for a follow-up biopsy	12 cores peripheral; 60/226 (22.6%)	Univariate only. PCA3 continuously and PSA continuously	In men undergoing repeat prostate biopsy to rule out cancer, the urinary PCA3 score was superior to serum PSA determination for predicting the biopsy outcome. The high specificity and informative rate suggest that the PCA3 assay could have an important role in prostate cancer diagnosis
Auprich 2012[87]	Previously biopsy with 8 or 10 cores, aged ≤ 70 years, a suspicious DRE and/or persistently raised age-specific tPSA thresholds (2.5–6.5 ng/ml) and/or suspicious prior histology (ASAPs ≥ 2 cores affected by HGPIN), but no patient with a tPSA levels of > 50 ng/ml	12 or 24 TRUS; specific sampling of anterior/transition zone; 44/127 (34.6%). Note that the first/second/third repeat biopsy reported separately	Univariate only. PCA3, PSA, PSA velocity and %fPSA	The findings of the present study promote the concept that the number of previous repeat biopsy sessions strongly influences the performance characteristics of biopsy risk factors. tPSA was no significant risk factor in the entire analysis. By contrast, %fPSA performed best at second and third and higher repeat biopsies. PSAV's diagnostic potential was reserved to patients at second and third and higher repeat biopsies. Finally, PCA3 demonstrated the highest diagnostic accuracy and potential to reduce unnecessary biopsies at first repeat biopsy. However, this advantage dissipated at second and third and higher repeat biopsies

continued

TABLE 76 Study characteristics of studies reporting univariate PCA3 or phi scores only (*continued*)

Study name	Inclusion/exclusion criteria	Repeat biopsies (type, number of positive/total sample (%)	Comparisons reported	Author conclusions
Ramos 2013[100]	Indication of transrectal prostate biopsy, either for elevated PSA and/or a suspicious DRE	≥ 12 core TRUS, at least two cores per sextant; 9/15 (60%)	Univariate analysis reported a PCA3 score of > 35 and a PSA level of > 4	This is the first report in Latin America on the use of PCA3 in diagnosing PCa. Our results are comparable to those reported in other populations in the literature, demonstrating the reproducibility of the test. PCA3 score was highly specific and we specially recommend its use in patients with persistent elevated PSA and prior negative biopsies
Stephan 2013[107]	Men scheduled for prostate biopsy owing to a suspicious DRE, suspicious transrectal ultrasonography findings, or increased PSA concentration or PSA velocity. Study exclusion criteria included urinary infections, medications (androgen or 5-α-reductase inhibitors) or interventions that could alter PSA concentrations	10–22 cores; 40/110 (36.7%)	Univariate analysis of PCA3, phi and PSA %fPSA. Multivariate model with PCA3, phi and T2:ERG	PCA3 and phi were superior to the other evaluated parameters but their combination gave only moderate enhancements in diagnostic accuracy for PCa at first or repeat prostate biopsy

ERG, Evidence Review Group; PCa, prostate cancer; PSAV, PSA velocity.
It is important to note that the author conclusions are those presented in the papers and are not the EAG's conclusions.

Appendix 7 Full results of quality assessment exercise

The outputs in this appendix are from RevMan: Review Manager (RevMan) [Computer program]. Version 5.3. Copenhagen: The Nordic Cochrane Centre, The Cochrane Collaboration; 2014.

PCA3 versus MRI for Prostate in men with negative Bx 02-Oct-2014

Characteristics of studies

Characteristics of included studies

Bollito 2012

Patient Selection

A. Risk of Bias	
Patient Sampling	Consecutive cohort. Mixed initial and repeat biopsy. Repeat reported separately. Men receiving PCA3 test and referred for repeat biopsy based on persistent PSA elevation. Men with positive DRE and/or ASAP on initial biopsy excluded.A*ssumed this means all DRE normal.*

Was a consecutive or random sample of patients enrolled?	Yes
Was a case-control design avoided?	Yes
Did the study avoid inappropriate exclusions?	Yes
Were men selected into study on basis of cancer risk such as on PSA range, DRE MRI etc.	Yes
Could the selection of patients have introduced bias?	High risk

B. Concerns regarding applicability	
Patient characteristics and setting	3 centres Italian: Turin, Orbassano. Milan
Was risk of underlying risk of Cancer in men in study population representative?	No
Are there concerns that the included patients and setting do not match the review question?	High concern

Index Test

Index tests	PCA3 Clinical : age, PSA, %fPSA No mention of blinding.

Intervention test

A. Risk of Bias	
Were the index test results interpreted without knowledge of the results of the reference standard?	Unclear
If a threshold was used, was it pre-specified?	Yes
Were the comparator test results interpreted without knowledge of the results of the intervention test?	Unclear
Could the conduct or interpretation of the index test have introduced bias?	Low risk

B. Concerns regarding applicability	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	Low concern

PCA3 versus MRI for Prostate in men with negative Bx 02-Oct-2014

Comparator test - clinical & PSA

A. Risk of Bias	
Were the index test results interpreted without knowledge of the results of the reference standard?	Unclear
If a threshold was used, was it pre-specified?	
Were the comparator test results interpreted without knowledge of the results of the intervention test?	Unclear
Could the conduct or interpretation of the index test have introduced bias?	Unclear risk

B. Concerns regarding applicability	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	High concern

Reference Standard

A. Risk of Bias	
Target condition and reference standard(s)	14-18 peripheral and transition zone cores - taken by experienced urologist. All specimens evaluated by experienced pathologist with an interest in uropathology. No mention of blinding.
Is the reference standards likely to correctly classify the target condition?	No
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear
Were the same number & pattern of cores taken in all participants?	No
Was the reference standard performed & results interpreted without knowledge of the results of the comparator tests?	Unclear
Could the reference standard, its conduct, or its interpretation have introduced bias?	Unclear risk

B. Concerns regarding applicability	
Are there concerns that the target condition as defined by the reference standard does not match the question?	High concern

Flow and Timing

A. Risk of Bias	
Flow and timing	Biopsy after PCA3 assessment. Assume < 1 yr. 6 out of 515 excluded due to inconclusive biopsy result
Was there an appropriate interval between index test and reference standard?	Yes
Were all patients included in the analysis?	No
Could the patient flow have introduced bias?	Low risk

Review Manager 5.2

Notes

Notes	

Busetto 2013

Patient Selection

A. Risk of Bias	
Patient Sampling	Inclusion: a first random TRUS-guided prostate biopsy that was negative for PC or high-grade prostatic intraepithelial neoplasm and a PSA level of 4-10 ng/mL.
Was a consecutive or random sample of patients enrolled?	Yes
Was a case-control design avoided?	Yes
Did the study avoid inappropriate exclusions?	Yes
Were men selected into study on basis of cancer risk such as on PSA range, DRE MRI etc.	Yes
Could the selection of patients have introduced bias?	Unclear risk

B. Concerns regarding applicability	
Patient characteristics and setting	Italy? University hospital Rome. Prospective cohort.
Was risk of underlying risk of Cancer in men in study population representative?	Yes
Are there concerns that the included patients and setting do not match the review question?	Low concern

Index Test

Index tests	PCA3; Clinical: *Age, PSA and DRE;* multiparametric MRI with magnetic resonance spectroscopic imaging, diffusion-weight imaging, and dynamic contrast-enhanced imaging. No mention of blinding of results

Intervention test

A. Risk of Bias	
Were the index test results interpreted without knowledge of the results of the reference standard?	Unclear
If a threshold was used, was it pre-specified?	Yes
Were the comparator test results interpreted without knowledge of the results of the intervention test?	Unclear
Could the conduct or interpretation of the index test have introduced bias?	Low risk

B. Concerns regarding applicability	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	Low concern

PCA3 versus MRI for Prostate in men with negative Bx 02-Oct-2014

Comparator test - clinical & PSA

A. Risk of Bias	
Were the index test results interpreted without knowledge of the results of the reference standard?	Unclear
If a threshold was used, was it pre-specified?	Unclear
Were the comparator test results interpreted without knowledge of the results of the intervention test?	Unclear
Could the conduct or interpretation of the index test have introduced bias?	Unclear risk

B. Concerns regarding applicability	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	Low concern

Comparator test- MRI

A. Risk of Bias	
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes
If a threshold was used, was it pre-specified?	
Were the comparator test results interpreted without knowledge of the results of the intervention test?	Unclear
Could the conduct or interpretation of the index test have introduced bias?	Low risk

B. Concerns regarding applicability	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	Low concern

Reference Standard

A. Risk of Bias	
Target condition and reference standard(s)	The biopsy protocol was a 10-core. (2 cores from the basal portion, lateral and paramedial; 2 from the midgland, lateral and paramedial; and 1 from the apex, on each side of the gland). In those cases with areas described by MRSI, DWI, and DCEI as suspicious for PC, 2 additional TRUS-guided cores were taken from each site considered abnormal.
Is the reference standards likely to correctly classify the target condition?	No
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear
Were the same number & pattern of cores taken in all participants?	No
Was the reference standard performed & results interpreted without knowledge of the results of the comparator tests?	Yes

Could the reference standard, its conduct, or its interpretation have introduced bias?	High risk

B. Concerns regarding applicability	
Are there concerns that the target condition as defined by the reference standard does not match the question?	High concern

Flow and Timing

A. Risk of Bias	
Flow and timing	PCA3 test before biopsy. 171 consecutive patients in the study. Two patients (1.2%) were excluded from the analysis because of insufficient PSA messenger RNA to evaluate the PCA3 test. Another 2 patients (1.2%) were excluded because of the impossibility of performing mMRI, and 4 patients (2.3%) declined informed consent.
Was there an appropriate interval between index test and reference standard?	Yes
Were all patients included in the analysis?	No
Could the patient flow have introduced bias?	Low risk

Notes

Notes	

European cohort

Patient Selection

A. Risk of Bias	
Patient Sampling	Men with one or two previous negative prostate biopsies (>=6 cores performed at >=3 mo prior to enrolment) scheduled for repeat biopsy were enrolled. Prospective.
Was a consecutive or random sample of patients enrolled?	Unclear
Was a case-control design avoided?	Yes
Did the study avoid inappropriate exclusions?	Yes
Were men selected into study on basis of cancer risk such as on PSA range, DRE MRI etc.	Yes
Could the selection of patients have introduced bias?	Unclear risk

B. Concerns regarding applicability	
Patient characteristics and setting	6 European centres. Prospective cohort.
Was risk of underlying risk of Cancer in men in study population representative?	Yes
Are there concerns that the included patients and setting do not match the review question?	Low concern

PCA3 versus MRI for Prostate in men with negative Bx 02-Oct-2014

Index Test

Index tests	PCA3 Progensa Clinical: Total and fPSA,number of previous biopsies, age, prostate volume No mention of blinding

Intervention test

A. Risk of Bias	
Were the index test results interpreted without knowledge of the results of the reference standard?	Unclear
If a threshold was used, was it pre-specified?	
Were the comparator test results interpreted without knowledge of the results of the intervention test?	Unclear
Could the conduct or interpretation of the index test have introduced bias?	Low risk

B. Concerns regarding applicability	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	Low concern

Comparator test - clinical & PSA

A. Risk of Bias	
Were the index test results interpreted without knowledge of the results of the reference standard?	Unclear
If a threshold was used, was it pre-specified?	
Were the comparator test results interpreted without knowledge of the results of the intervention test?	Unclear
Could the conduct or interpretation of the index test have introduced bias?	Unclear risk

B. Concerns regarding applicability	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	Low concern

Reference Standard

A. Risk of Bias	
Target condition and reference standard(s)	At least 10 standardized periph zone. Bx taken by an experienced physician. The specimens were evaluated by the pathologist at each site.No mention of blinding.
Is the reference standards likely to correctly classify the target condition?	No
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear
Were the same number & pattern of cores taken in all participants?	Unclear

Review Manager 5.2

Was the reference standard performed & results interpreted without knowledge of the results of the comparator tests?	Unclear
Could the reference standard, its conduct, or its interpretation have introduced bias?	Unclear risk

B. Concerns regarding applicability	
Are there concerns that the target condition as defined by the reference standard does not match the question?	High concern

Flow and Timing

A. Risk of Bias	
Flow and timing	Biopsy Immediately after blood and urine samples taken.470 subjects, 467 urine samples adequate for PCA3, 463 had conclusive biopsy results.
Was there an appropriate interval between index test and reference standard?	Yes
Were all patients included in the analysis?	Yes
Could the patient flow have introduced bias?	Low risk

Notes

Notes	

Gittelman 2013

Patient Selection

A. Risk of Bias	
Patient Sampling	Participants were men without PCa with 1 or more previous negative prostate biopsy session who were recommended by their physician for repeat biopsy
Was a consecutive or random sample of patients enrolled?	Unclear
Was a case-control design avoided?	Yes
Did the study avoid inappropriate exclusions?	Yes
Were men selected into study on basis of cancer risk such as on PSA range, DRE MRI etc.	Yes
Could the selection of patients have introduced bias?	Unclear risk

B. Concerns regarding applicability	
Patient characteristics and setting	Geographically diverse, community based urology clinics, group health organizations and academic institutions in the United States. Prospective cohort
Was risk of underlying risk of Cancer in men in study population representative?	Yes

Are there concerns that the included patients and setting do not match the review question?	Low concern

Index Test

Index tests	PCA3 Progensa Assay. Laboratory personnel were blinded to subject clinical status, and sPSA and biopsy results. Clinical variables: age, DRE result, family history of PCa, race and number of previous negative biopsy sessions.

Intervention test

A. Risk of Bias	
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes
If a threshold was used, was it pre-specified?	
Were the comparator test results interpreted without knowledge of the results of the intervention test?	Yes
Could the conduct or interpretation of the index test have introduced bias?	Low risk

B. Concerns regarding applicability	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	Low concern

Comparator test - clinical & PSA

A. Risk of Bias	
Were the index test results interpreted without knowledge of the results of the reference standard?	Unclear
If a threshold was used, was it pre-specified?	
Were the comparator test results interpreted without knowledge of the results of the intervention test?	Unclear
Could the conduct or interpretation of the index test have introduced bias?	Unclear risk

B. Concerns regarding applicability	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	Low concern

Reference Standard

A. Risk of Bias	
Target condition and reference standard(s)	12 core or greater TRUS biopsy. Each specimen was evaluated at the site pathology facility according to institutional procedures. All pathologists were blinded to PCA3 assay and other test results
Is the reference standards likely to correctly classify the target condition?	No

Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear
Were the same number & pattern of cores taken in all participants?	No
Was the reference standard performed & results interpreted without knowledge of the results of the comparator tests?	Unclear
Could the reference standard, its conduct, or its interpretation have introduced bias?	Unclear risk

B. Concerns regarding applicability	
Are there concerns that the target condition as defined by the reference standard does not match the question?	High concern

Flow and Timing

A. Risk of Bias	
Flow and timing	Test order = blood, urine, biopsy – usually all within 24 hrs. Bx within 7 days of blood and urine within 7 days of blood samples. 6/474 excluded due to < 50 yrs. Not enough to affect results.
Was there an appropriate interval between index test and reference standard?	Yes
Were all patients included in the analysis?	No
Could the patient flow have introduced bias?	Low risk

Notes

Notes	

Goode 2013

Patient Selection

A. Risk of Bias	
Patient Sampling	Retrospective review of notes. Mixed biopsy population. Repeat reported separately. Men with no known personal history of prostate cancer who underwent a prostate biopsy because of an elevated PSA level, abnormal digital rectal exam (DRE), or abnormal previous prostate biopsy-prostatic intraepithelial neoplasia (PIN) or atypical small acinar proliferation (ASAP).
Was a consecutive or random sample of patients enrolled?	Unclear
Was a case-control design avoided?	Yes
Did the study avoid inappropriate exclusions?	Yes
Were men selected into study on basis of cancer risk such as on PSA range, DRE MRI etc.	Yes
Could the selection of patients have introduced bias?	Unclear risk

B. Concerns regarding applicability		
Patient characteristics and setting	1 centre. US	
Was risk of underlying risk of Cancer in men in study population representative?		Yes
Are there concerns that the included patients and setting do not match the review question?		Low concern

Index Test

Index tests	PCA3 Clinical: prostate volume, patient age, patient race, family history, and digital rectal exam status. **PSA not included.** The laboratories processing the blood and urine specimens and the pathologists examining the biopsy cores were unaware of the patients' clinical status.

Intervention test

A. Risk of Bias	
Were the index test results interpreted without knowledge of the results of the reference standard?	Unclear
If a threshold was used, was it pre-specified?	No
Were the comparator test results interpreted without knowledge of the results of the intervention test?	Unclear
Could the conduct or interpretation of the index test have introduced bias?	Low risk

B. Concerns regarding applicability	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	Low concern

Comparator test - clinical & PSA

A. Risk of Bias	
Were the index test results interpreted without knowledge of the results of the reference standard?	Unclear
If a threshold was used, was it pre-specified?	Unclear
Were the comparator test results interpreted without knowledge of the results of the intervention test?	Unclear
Could the conduct or interpretation of the index test have introduced bias?	Unclear risk

B. Concerns regarding applicability	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	High concern

Reference Standard

A. Risk of Bias	
Target condition and reference standard(s)	12 core TRUS. Pathologists examining the biopsy cores were unaware of the patients' clinical status.
Is the reference standards likely to correctly classify the target condition?	No
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear
Were the same number & pattern of cores taken in all participants?	Yes
Was the reference standard performed & results interpreted without knowledge of the results of the comparator tests?	Unclear
Could the reference standard, its conduct, or its interpretation have introduced bias?	Unclear risk

B. Concerns regarding applicability	
Are there concerns that the target condition as defined by the reference standard does not match the question?	High concern

Flow and Timing

A. Risk of Bias	
Flow and timing	Retrospective design. Unclear selection and timing.
Was there an appropriate interval between index test and reference standard?	Unclear
Were all patients included in the analysis?	Unclear
Could the patient flow have introduced bias?	High risk

Notes

Notes	

Lazzeri 2012

Patient Selection

A. Risk of Bias	
Patient Sampling	a negative first biopsy but persistent suspicion of PCa who were scheduled for repeat biopsy according to the European Association of Urology guidelines of increasing and/or persistently elevated PSA, suspicious DRE, atypical small acinar proliferation and high grade prostate intraepithelial neoplasia
Was a consecutive or random sample of patients enrolled?	Unclear
Was a case-control design avoided?	Yes
Did the study avoid inappropriate exclusions?	Yes
Were men selected into study on basis of cancer risk such as on PSA range, DRE MRI etc.	Yes

Could the selection of patients have introduced bias?	Unclear risk

B. Concerns regarding applicability	
Patient characteristics and setting	University hospital Milan. Prospective cohort
Was risk of underlying risk of Cancer in men in study population representative?	Yes
Are there concerns that the included patients and setting do not match the review question?	Low concern

Index Test

Index tests	phi Clinical: PSA, prostate volume, and DRE, %fPSA and PSA density No mention of blinding

Intervention test

A. Risk of Bias	
Were the index test results interpreted without knowledge of the results of the reference standard?	Unclear
If a threshold was used, was it pre-specified?	Unclear
Were the comparator test results interpreted without knowledge of the results of the intervention test?	Unclear
Could the conduct or interpretation of the index test have introduced bias?	Low risk

B. Concerns regarding applicability	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	Low concern

Comparator test - clinical & PSA

A. Risk of Bias	
Were the index test results interpreted without knowledge of the results of the reference standard?	Unclear
If a threshold was used, was it pre-specified?	Unclear
Were the comparator test results interpreted without knowledge of the results of the intervention test?	Unclear
Could the conduct or interpretation of the index test have introduced bias?	Unclear risk

B. Concerns regarding applicability	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	Low concern

Reference Standard

A. Risk of Bias	
Target condition and reference standard(s)	Ambulatory transrectal ultrasonography guided prostate biopsies according to a standardized institutional scheme to obtain the highest detection rate. 24 core saturation Bx. Range of cores 12-26. Specimens were processed and evaluated by a single experienced genitourinary pathologist. Pathologist blinded but not clear about person performing biopsy.
Is the reference standards likely to correctly classify the target condition?	No
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear
Were the same number & pattern of cores taken in all participants?	No
Was the reference standard performed & results interpreted without knowledge of the results of the comparator tests?	Unclear
Could the reference standard, its conduct, or its interpretation have introduced bias?	Unclear risk

B. Concerns regarding applicability	
Are there concerns that the target condition as defined by the reference standard does not match the question?	High concern

Flow and Timing

A. Risk of Bias	
Flow and timing	Blood sample was drawn at the time of repeat biopsy.8/230 samples not analyzed according to p2PSA product insert claimed stability informa. 22 analysed
Was there an appropriate interval between index test and reference standard?	Yes
Were all patients included in the analysis?	Yes
Could the patient flow have introduced bias?	Low risk

Notes

Notes	

Panebianco 2011

Patient Selection

A. Risk of Bias	
Patient Sampling	first random TRUS-guided prostate biopsy negative for prostate adenocarcinoma or high-grade prostate intraepithelial neoplasm; persistent elevated PSA levels (total PSA ≥ 4 ng/ml and <10 ng/ml) and negative digital rectal examination (DRE). *Assumed this means all DRE normal*.

Was a consecutive or random sample of patients enrolled?	Yes
Was a case-control design avoided?	Yes
Did the study avoid inappropriate exclusions?	Yes
Were men selected into study on basis of cancer risk such as on PSA range, DRE MRI etc.	Yes
Could the selection of patients have introduced bias?	High risk

B. Concerns regarding applicability		
Patient characteristics and setting	Italian hospital ? Rome. Prospective cohort	
Was risk of underlying risk of Cancer in men in study population representative?		No
Are there concerns that the included patients and setting do not match the review question?		High concern

Index Test

Index tests	PCA3 MRI

Intervention test

A. Risk of Bias	
Were the index test results interpreted without knowledge of the results of the reference standard?	Unclear
If a threshold was used, was it pre-specified?	Yes
Were the comparator test results interpreted without knowledge of the results of the intervention test?	Unclear
Could the conduct or interpretation of the index test have introduced bias?	Low risk

B. Concerns regarding applicability	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	Low concern

Comparator test- MRI

A. Risk of Bias	
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes
If a threshold was used, was it pre-specified?	Yes
Were the comparator test results interpreted without knowledge of the results of the intervention test?	Unclear
Could the conduct or interpretation of the index test have introduced bias?	Low risk

B. Concerns regarding applicability	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	Low concern

Reference Standard

A. Risk of Bias	
Target condition and reference standard(s)	10-core laterally directed (two cores from the basal portion lateral and paramedial, two from the midgland lateral and paramedial, and one from the apex, on each side of the gland for each patient, plus 3 additional biopsies from other areas suspicious for PCa at MRSI)
Is the reference standards likely to correctly classify the target condition?	No
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear
Were the same number & pattern of cores taken in all participants?	No
Was the reference standard performed & results interpreted without knowledge of the results of the comparator tests?	No
Could the reference standard, its conduct, or its interpretation have introduced bias?	High risk

B. Concerns regarding applicability	
Are there concerns that the target condition as defined by the reference standard does not match the question?	High concern

Flow and Timing

A. Risk of Bias	
Flow and timing	Tests before repeat biopsy. Assume < 1 yr. 41/43 participants had informative PCA3 results.
Was there an appropriate interval between index test and reference standard?	Yes
Were all patients included in the analysis?	Yes
Could the patient flow have introduced bias?	Low risk

Notes

Notes	

Pepe 2013

Patient Selection

PCA3 versus MRI for Prostate in men with negative Bx 02-Oct-2014

A. Risk of Bias	
Patient Sampling	All men had negative family history, abnormal DRE, PSA 4.1-10 or 2.6-4. All caucasian.
Was a consecutive or random sample of patients enrolled?	Unclear
Was a case-control design avoided?	Yes
Did the study avoid inappropriate exclusions?	No
Were men selected into study on basis of cancer risk such as on PSA range, DRE MRI etc.	Yes
Could the selection of patients have introduced bias?	High risk

B. Concerns regarding applicability	
Patient characteristics and setting	Italy ? catania Unclear whether prospective/retrospective cohort
Was risk of underlying risk of Cancer in men in study population representative?	No
Are there concerns that the included patients and setting do not match the review question?	High concern

Index Test

Index tests	PCA3 Clinical; PCPT nomogram -Age, race, PSA, DRE, family history, previous negative biopsy No mention of blinding

Intervention test

A. Risk of Bias	
Were the index test results interpreted without knowledge of the results of the reference standard?	Unclear
If a threshold was used, was it pre-specified?	Yes
Were the comparator test results interpreted without knowledge of the results of the intervention test?	Unclear
Could the conduct or interpretation of the index test have introduced bias?	Low risk

B. Concerns regarding applicability	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	Low concern

Comparator test - clinical & PSA

A. Risk of Bias	
Were the index test results interpreted without knowledge of the results of the reference standard?	Unclear
If a threshold was used, was it pre-specified?	Unclear

Review Manager 5.2

Were the comparator test results interpreted without knowledge of the results of the intervention test?	Unclear
Could the conduct or interpretation of the index test have introduced bias?	Unclear risk

B. Concerns regarding applicability	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	High concern

Reference Standard

A. Risk of Bias	
Target condition and reference standard(s)	Transperineally, saturation biopsy. At least 12 in the posterior zone of each lobe and 2-3 in the transition zone Median 30, range 24-38. No mention of blinding.
Is the reference standards likely to correctly classify the target condition?	No
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear
Were the same number & pattern of cores taken in all participants?	No
Was the reference standard performed & results interpreted without knowledge of the results of the comparator tests?	Unclear
Could the reference standard, its conduct, or its interpretation have introduced bias?	Unclear risk

B. Concerns regarding applicability	
Are there concerns that the target condition as defined by the reference standard does not match the question?	High concern

Flow and Timing

A. Risk of Bias	
Flow and timing	PCA3 test 3-10 days before biopsy. All patients had adequate PCA3.
Was there an appropriate interval between index test and reference standard?	Yes
Were all patients included in the analysis?	Yes
Could the patient flow have introduced bias?	Low risk

Notes

Notes	

Perdona 2011

Patient Selection

A. Risk of Bias	
Patient Sampling	Men referred for prostate biopsy because of abnormal PSA and/or suspicious DRE.Mixed and repeat but repeat reported separately. No PSA > 10 ng/mL.
Was a consecutive or random sample of patients enrolled?	Unclear
Was a case-control design avoided?	Yes
Did the study avoid inappropriate exclusions?	Yes
Were men selected into study on basis of cancer risk such as on PSA range, DRE MRI etc.	Yes
Could the selection of patients have introduced bias?	Unclear risk

B. Concerns regarding applicability	
Patient characteristics and setting	3 centre Italian study - Naples, Catanzaro. Prospective.
Was risk of underlying risk of Cancer in men in study population representative?	Yes
Are there concerns that the included patients and setting do not match the review question?	Low concern

Index Test

Index tests	PCA3 Chun: age, PSA, DRE, previous Bx, prostate volume. PCPT: race, age, PSA, fam Hx, DRE & previous Bx. Multivariate: AGE, transrectal ultrasound (TRUS) abnormalities, prostate volume, history of previous biopsy, family history of PCa No mention of blinding

Intervention test

A. Risk of Bias	
Were the index test results interpreted without knowledge of the results of the reference standard?	Unclear
If a threshold was used, was it pre-specified?	
Were the comparator test results interpreted without knowledge of the results of the intervention test?	Unclear
Could the conduct or interpretation of the index test have introduced bias?	Low risk

B. Concerns regarding applicability	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	Low concern

Comparator test - clinical & PSA

A. Risk of Bias	
Were the index test results interpreted without knowledge of the results of the reference standard?	Unclear
If a threshold was used, was it pre-specified?	Unclear
Were the comparator test results interpreted without knowledge of the results of the intervention test?	Unclear
Could the conduct or interpretation of the index test have introduced bias?	Unclear risk

B. Concerns regarding applicability	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	Low concern

Reference Standard

A. Risk of Bias	
Target condition and reference standard(s)	systematic, laterally directed,? transrectal ? >= 12-core, median 12 (IQR 12-16). Evaluated by an experienced pathologist at each site. No mention of blinding.
Is the reference standards likely to correctly classify the target condition?	No
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear
Were the same number & pattern of cores taken in all participants?	No
Was the reference standard performed & results interpreted without knowledge of the results of the comparator tests?	Unclear
Could the reference standard, its conduct, or its interpretation have introduced bias?	Unclear risk

B. Concerns regarding applicability	
Are there concerns that the target condition as defined by the reference standard does not match the question?	High concern

Flow and Timing

A. Risk of Bias	
Flow and timing	PCA3 test immediately before biopsy. 84 men with repeat Biopsy - no other details.
Was there an appropriate interval between index test and reference standard?	Yes
Were all patients included in the analysis?	Unclear
Could the patient flow have introduced bias?	Unclear risk

Notes

Notes	

Porpiglia 2014

Patient Selection

A. Risk of Bias	
Patient Sampling	Negative initial Bx – 12 cores. Persistently elevated PSA levels, **and/or** positive digital rectal examination
Was a consecutive or random sample of patients enrolled?	Unclear
Was a case-control design avoided?	Yes
Did the study avoid inappropriate exclusions?	Yes
Were men selected into study on basis of cancer risk such as on PSA range, DRE MRI etc.	Yes
Could the selection of patients have introduced bias?	Unclear risk

B. Concerns regarding applicability		
Patient characteristics and setting	Italy - Orbassano. Prospective	
Was risk of underlying risk of Cancer in men in study population representative?		Yes
Are there concerns that the included patients and setting do not match the review question?		Low concern

Index Test

Index tests	PCA3 & Phi Clinical: DRE, age, **NOT PSA** No mention of blinding for PCA3 / PSA lab personnel. mp-MRI: diffusion-weighted imaging (DWI) and dynamic contrast enhanced (DCE) MRI. The radiologist was blinded to the pathologist's biopsy reports and to the biomarker results (but ? knew clinical status).

Intervention test

A. Risk of Bias	
Were the index test results interpreted without knowledge of the results of the reference standard?	Unclear
If a threshold was used, was it pre-specified?	No
Were the comparator test results interpreted without knowledge of the results of the intervention test?	Unclear
Could the conduct or interpretation of the index test have introduced bias?	Low risk

B. Concerns regarding applicability	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	Low concern

Comparator test - clinical & PSA

A. Risk of Bias	
Were the index test results interpreted without knowledge of the results of the reference standard?	Unclear
If a threshold was used, was it pre-specified?	Unclear
Were the comparator test results interpreted without knowledge of the results of the intervention test?	Unclear
Could the conduct or interpretation of the index test have introduced bias?	Unclear risk

B. Concerns regarding applicability	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	High concern

Comparator test- MRI

A. Risk of Bias	
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes
If a threshold was used, was it pre-specified?	Yes
Were the comparator test results interpreted without knowledge of the results of the intervention test?	Yes
Could the conduct or interpretation of the index test have introduced bias?	Low risk

B. Concerns regarding applicability	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	Low concern

Reference Standard

A. Risk of Bias	
Target condition and reference standard(s)	Two dedicated urologists blinded to the mp-MRI reports and to the biomarkers results performed all RB. 18 or 24 core depending on prostate volume. No extra cores for MRI result. May have been affected by clinical findings but better controlled than many other studies. Histological examination was conducted by a dedicated uropathologist, who was blinded to the biomarkers and to the mp-MRI results, according to a standardised protocol.
Is the reference standards likely to correctly classify the target condition?	No

Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes
Were the same number & pattern of cores taken in all participants?	Yes
Was the reference standard performed & results interpreted without knowledge of the results of the comparator tests?	Yes
Could the reference standard, its conduct, or its interpretation have introduced bias?	Low risk

B. Concerns regarding applicability	
Are there concerns that the target condition as defined by the reference standard does not match the question?	High concern

Flow and Timing

A. Risk of Bias	
Flow and timing	PCA3, phi and MRI Prior to repeat biopsy- asume < 1 yr.4 /174 excluded due anterior Ca
Was there an appropriate interval between index test and reference standard?	Yes
Were all patients included in the analysis?	No
Could the patient flow have introduced bias?	Low risk

Notes

Notes	

REDUCE placebo

Patient Selection

A. Risk of Bias	
Patient Sampling	Cohort of patients from placebo arm of REDUCE trial. Followed for 4 years. Selection into this study depended on trial site being able to process urine sample for PCA3. Only scheduled biopsies used. Low risk population as "for cause" biopsies excluded.
Was a consecutive or random sample of patients enrolled?	Unclear
Was a case-control design avoided?	Yes
Did the study avoid inappropriate exclusions?	Yes
Were men selected into study on basis of cancer risk such as on PSA range, DRE MRI etc.	Yes
Could the selection of patients have introduced bias?	Unclear risk

B. Concerns regarding applicability	
Patient characteristics and setting	REDUCE trial. Multi centre international study. Prospective cohort within.
Was risk of underlying risk of Cancer in men in study population representative?	No
Are there concerns that the included patients and setting do not match the review question?	Low concern

Index Test

Index tests	PCA3 Progensa Assay. PCA3 Operators were blinded with respect to biopsy results and study arm (placebo vs dutasteride). Not quite clear how being used in algorithm . Clinical variables used; life expectancy, DRE findings, PSA level, prostate volume, number of previous negative PBxs

Intervention test

A. Risk of Bias	
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes
If a threshold was used, was it pre-specified?	
Were the comparator test results interpreted without knowledge of the results of the intervention test?	Unclear
Could the conduct or interpretation of the index test have introduced bias?	Low risk

B. Concerns regarding applicability	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	Low concern

Comparator test - clinical & PSA

A. Risk of Bias	
Were the index test results interpreted without knowledge of the results of the reference standard?	Unclear
If a threshold was used, was it pre-specified?	
Were the comparator test results interpreted without knowledge of the results of the intervention test?	Unclear
Could the conduct or interpretation of the index test have introduced bias?	Unclear risk

B. Concerns regarding applicability	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	Low concern

Reference Standard

A. Risk of Bias	
Target condition and reference standard(s)	10 core transrectal biopsies. Biopsies were read at the central pathology laboratory (CPL, which processed the majority, 94%, of biopsies).No mention of blinding.
Is the reference standards likely to correctly classify the target condition?	No
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear
Were the same number & pattern of cores taken in all participants?	Unclear
Was the reference standard performed & results interpreted without knowledge of the results of the comparator tests?	Unclear
Could the reference standard, its conduct, or its interpretation have introduced bias?	Unclear risk

B. Concerns regarding applicability	
Are there concerns that the target condition as defined by the reference standard does not match the question?	High concern

Flow and Timing

A. Risk of Bias	
Flow and timing	At time of repeat Biopsy?. Assume < 1 yr. 48/ 172 with informative PCA3 not included in model in Tombal due to missing covariates.
Was there an appropriate interval between index test and reference standard?	Yes
Were all patients included in the analysis?	No
Could the patient flow have introduced bias?	Low risk

Notes

Notes	

Scattoni 2013

Patient Selection

A. Risk of Bias	
Patient Sampling	Candidates for initial or repeat PBx at 2 tertiary care institutions.Indication for repeat Bx ASAP, plurifocal HGPIN , PSA 2-15 and/or positive DRE.
Was a consecutive or random sample of patients enrolled?	Yes
Was a case-control design avoided?	Yes
Did the study avoid inappropriate exclusions?	Yes

Were men selected into study on basis of cancer risk such as on PSA range, DRE MRI etc.	Yes
Could the selection of patients have introduced bias?	Unclear risk

B. Concerns regarding applicability		
Patient characteristics and setting	Consecutive, prospective cohort	
Was risk of underlying risk of Cancer in men in study population representative?		Yes
Are there concerns that the included patients and setting do not match the review question?		Low concern

Index Test

Index tests	PCA3, phi Clinical: age, DRE, volume, PSA, f/tPSA

Intervention test

A. Risk of Bias	
Were the index test results interpreted without knowledge of the results of the reference standard?	Unclear
If a threshold was used, was it pre-specified?	
Were the comparator test results interpreted without knowledge of the results of the intervention test?	Unclear
Could the conduct or interpretation of the index test have introduced bias?	Low risk

B. Concerns regarding applicability	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	Low concern

Comparator test - clinical & PSA

A. Risk of Bias	
Were the index test results interpreted without knowledge of the results of the reference standard?	Unclear
If a threshold was used, was it pre-specified?	Unclear
Were the comparator test results interpreted without knowledge of the results of the intervention test?	Unclear
Could the conduct or interpretation of the index test have introduced bias?	Unclear risk

B. Concerns regarding applicability	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	Low concern

Reference Standard

A. Risk of Bias	
Target condition and reference standard(s)	Ambulatory transrectal ultrasound guided PBx according to a standardized institutional saturation scheme. at least 14 to 24 biopsy cores. Mean 18.7 ± 3.2. No mention of blinding.
Is the reference standards likely to correctly classify the target condition?	No
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear
Were the same number & pattern of cores taken in all participants?	No
Was the reference standard performed & results interpreted without knowledge of the results of the comparator tests?	Unclear
Could the reference standard, its conduct, or its interpretation have introduced bias?	Unclear risk

B. Concerns regarding applicability	
Are there concerns that the target condition as defined by the reference standard does not match the question?	High concern

Flow and Timing

A. Risk of Bias	
Flow and timing	95 repeat patients. A blood sample was drawn at biopsy just before prostatic manipulations.Urine sample just before Bx.
Was there an appropriate interval between index test and reference standard?	Yes
Were all patients included in the analysis?	Yes
Could the patient flow have introduced bias?	Low risk

Notes

Notes	

Sciarra 2012

Patient Selection

A. Risk of Bias	
Patient Sampling	First negative prostate biopsy to cancer & HGPIN , persistent total PSA > 4 ng/mL and negative DRE. *Assumed this means all DRE normal.* Consecutive patients who were referred to the Department of Urology. Randomly assigned (1:1) to PCA3 only or PCA3 plus MRI before repeat biopsy.
Was a consecutive or random sample of patients enrolled?	Yes
Was a case-control design avoided?	Yes

Did the study avoid inappropriate exclusions?	Yes
Were men selected into study on basis of cancer risk such as on PSA range, DRE MRI etc.	Yes
Could the selection of patients have introduced bias?	High risk

B. Concerns regarding applicability	
Patient characteristics and setting	
Was risk of underlying risk of Cancer in men in study population representative?	No
Are there concerns that the included patients and setting do not match the review question?	High concern

Index Test

Index tests	PCA3 MRI

Intervention test

A. Risk of Bias	
Were the index test results interpreted without knowledge of the results of the reference standard?	Unclear
If a threshold was used, was it pre-specified?	Yes
Were the comparator test results interpreted without knowledge of the results of the intervention test?	Unclear
Could the conduct or interpretation of the index test have introduced bias?	Low risk

B. Concerns regarding applicability	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	Low concern

Comparator test- MRI

A. Risk of Bias	
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes
If a threshold was used, was it pre-specified?	No
Were the comparator test results interpreted without knowledge of the results of the intervention test?	Unclear
Could the conduct or interpretation of the index test have introduced bias?	Low risk

B. Concerns regarding applicability	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	Low concern

Reference Standard

A. Risk of Bias	
Target condition and reference standard(s)	All TRUS and biopsies were performed using an end-fire ultrasonographic transducer and biopsy gun with an 18-gauge needle (Esaote Technos MP with a C10-5 transducer. laterally directed 10-core. In cases with areas described by MRI as being suspicious for cancer, two additional cores were taken from each area that was labelled abnormal. All biopsies were performed in the department by a single physician (M.C.) with a long experience of this procedure. Histological assessments were carried out blind to the results of the MRI. ? blind to PCA3?
Is the reference standards likely to correctly classify the target condition?	No
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear
Were the same number & pattern of cores taken in all participants?	No
Was the reference standard performed & results interpreted without knowledge of the results of the comparator tests?	No
Could the reference standard, its conduct, or its interpretation have introduced bias?	High risk

B. Concerns regarding applicability	
Are there concerns that the target condition as defined by the reference standard does not match the question?	High concern

Flow and Timing

A. Risk of Bias	
Flow and timing	2nd biopsy within 90 days of 1st biopsy. Unclear timing of PCA3 test but at or after 1st biopsy. 180 cases with first negative random biopsy and persistent total PSA > 4 ng/ml. 12 indaequate PCA3 sample Baseline PCA3. 168 cases entered trial.
Was there an appropriate interval between index test and reference standard?	Yes
Were all patients included in the analysis?	Yes
Could the patient flow have introduced bias?	Low risk

Notes

Notes	

Stephan 2013

Patient Selection

A. Risk of Bias	
Patient Sampling	Unclear – described as both case-control and cohort. Patients enrolled prospectively and retrospectively. 1362 men; 681 patients (50%) were included for initial biopsy and 280 patients (21%) were scheduled for a repeated biopsy, and for the remaining 401 patients (29%) this information was missing.tPSA results between 1.6 and 8.0g/L (calibration against a WHO PSA reference material)
Was a consecutive or random sample of patients enrolled?	Unclear
Was a case-control design avoided?	Unclear
Did the study avoid inappropriate exclusions?	Unclear
Were men selected into study on basis of cancer risk such as on PSA range, DRE MRI etc.	Yes
Could the selection of patients have introduced bias?	High risk

B. Concerns regarding applicability	
Patient characteristics and setting	4 centres in Germany and France
Was risk of underlying risk of Cancer in men in study population representative?	Unclear
Are there concerns that the included patients and setting do not match the review question?	Unclear concern

Index Test

Index tests	phi - p2PSA Clinical: Age, prostate volume, DRE, tPSA, %fPSA. Participants and investigators were blinded to p2PSA results and the personnel involved in testing (p2PSA?) were blinded to patients' clinical information

Intervention test

A. Risk of Bias	
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes
If a threshold was used, was it pre-specified?	
Were the comparator test results interpreted without knowledge of the results of the intervention test?	Yes
Could the conduct or interpretation of the index test have introduced bias?	Low risk

B. Concerns regarding applicability	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	Low concern

Comparator test - clinical & PSA

A. Risk of Bias	
Were the index test results interpreted without knowledge of the results of the reference standard?	Unclear
If a threshold was used, was it pre-specified?	Unclear
Were the comparator test results interpreted without knowledge of the results of the intervention test?	Yes
Could the conduct or interpretation of the index test have introduced bias?	Unclear risk

B. Concerns regarding applicability	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	Low concern

Reference Standard

A. Risk of Bias	
Target condition and reference standard(s)	transrectal ultrasound (TRUS)-guided needle biopsy. 10 -22 cores. According to standard clinical practice routinely used at each site. No mention of blinding.
Is the reference standards likely to correctly classify the target condition?	No
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear
Were the same number & pattern of cores taken in all participants?	No
Was the reference standard performed & results interpreted without knowledge of the results of the comparator tests?	Unclear
Could the reference standard, its conduct, or its interpretation have introduced bias?	Unclear risk

B. Concerns regarding applicability	
Are there concerns that the target condition as defined by the reference standard does not match the question?	High concern

Flow and Timing

A. Risk of Bias	
Flow and timing	All blood samples were obtained before any manipulations involving the prostate and at least 3 weeks after a digital rectal examination (DRE). Patient flow unclear.
Was there an appropriate interval between index test and reference standard?	Yes
Were all patients included in the analysis?	Unclear
Could the patient flow have introduced bias?	Unclear risk

Notes

Notes	

Wu 2012

Patient Selection

A. Risk of Bias	
Patient Sampling	Consecutive retrospective study.Indications for repeat prostate biopsy were based on suspicious DRE, persistently elevated PSA, previous suspicious histology (such as high-grade prostatic intraepithelial neoplasia or atypical small acinar proliferation) and/or patient preference.
Was a consecutive or random sample of patients enrolled?	Yes
Was a case-control design avoided?	Yes
Did the study avoid inappropriate exclusions?	Yes
Were men selected into study on basis of cancer risk such as on PSA range, DRE MRI etc.	Yes
Could the selection of patients have introduced bias?	Unclear risk

B. Concerns regarding applicability		
Patient characteristics and setting	I centre. San Francisco, US	
Was risk of underlying risk of Cancer in men in study population representative?		Yes
Are there concerns that the included patients and setting do not match the review question?		Low concern

Index Test

Index tests	PCA3 Clinical : own nomogram PSA, PSAD, **TRUS** and DRE Chun nomogram: *Age,DRE,previous bx, vol* PSA, PSAD No mention of blinding.

Intervention test

A. Risk of Bias	
Were the index test results interpreted without knowledge of the results of the reference standard?	Unclear
If a threshold was used, was it pre-specified?	Yes
Were the comparator test results interpreted without knowledge of the results of the intervention test?	Unclear
Could the conduct or interpretation of the index test have introduced bias?	Low risk

B. Concerns regarding applicability	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	Low concern

Comparator test - clinical & PSA

A. Risk of Bias	
Were the index test results interpreted without knowledge of the results of the reference standard?	Unclear
If a threshold was used, was it pre-specified?	Unclear
Were the comparator test results interpreted without knowledge of the results of the intervention test?	Unclear
Could the conduct or interpretation of the index test have introduced bias?	Unclear risk

B. Concerns regarding applicability	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?	Unclear concern

Reference Standard

A. Risk of Bias	
Target condition and reference standard(s)	transrectal ultrasound (TRUS). >=12 (two cores from each sextant of the prostate are taken plus additional cores from suspicious areas by TRUS and/or anterior prostate cores). All performed by same clinician. No mention of blinding.
Is the reference standards likely to correctly classify the target condition?	No
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear
Were the same number & pattern of cores taken in all participants?	No
Was the reference standard performed & results interpreted without knowledge of the results of the comparator tests?	Unclear
Could the reference standard, its conduct, or its interpretation have introduced bias?	High risk

B. Concerns regarding applicability	
Are there concerns that the target condition as defined by the reference standard does not match the question?	High concern

Flow and Timing

A. Risk of Bias		
Flow and timing	103 out of 188 patients with full data included (54.7%). PCA3 before repeat Bx - time gap not given. Assume < 1yr	
Was there an appropriate interval between index test and reference standard?		Yes
Were all patients included in the analysis?		No
Could the patient flow have introduced bias?		High risk

Notes

Notes	

Footnotes

Characteristics of excluded studies

Footnotes

Characteristics of studies awaiting classification

Footnotes

Characteristics of ongoing studies

Footnotes

Summary of findings tables

Additional tables

Appendix 8 Graphs of decision curve analysis results

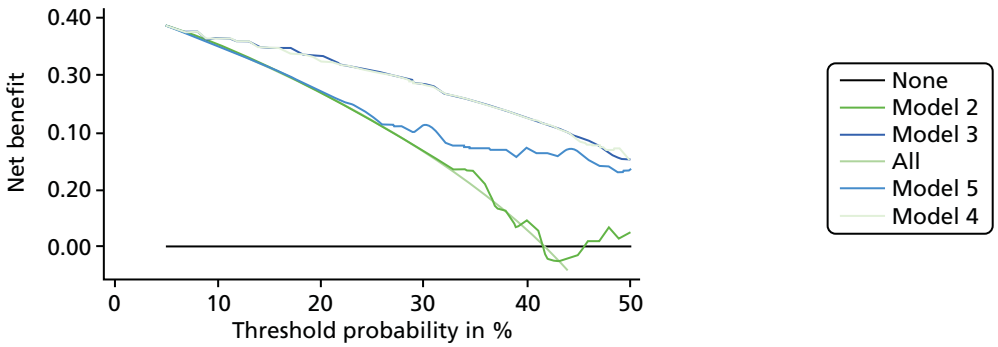

FIGURE 13 Decision curve analysis graph from Busetto *et al.*[90] The dark green line indicates the base clinical model (age, digital rectal examination, and prostate-specific antigen); the light blue line indicates the base clinical model plus prostate cancer gene 3 (PCA3) assay; the dark blue line indicates the base clinical model plus multiparametric magnetic resonance imaging (mMRI); and the light green line indicates the base clinical model plus PCA3 and mMRI. [Reprinted from *Urology*, vol. 82, Busetto GM, De Berardinis E, Sciarra A, Panebianco V, Giovannone R, Rosato S, D'Errigo P, Di Silverio F, Gentile V, Salciccia S, Prostate cancer gene 3 and multiparametric magnetic resonance can reduce unnecessary biopsies: decision curve analysis to evaluate predictive models, pp. 1355–1362, © 2013, with permission from Elsevier.]

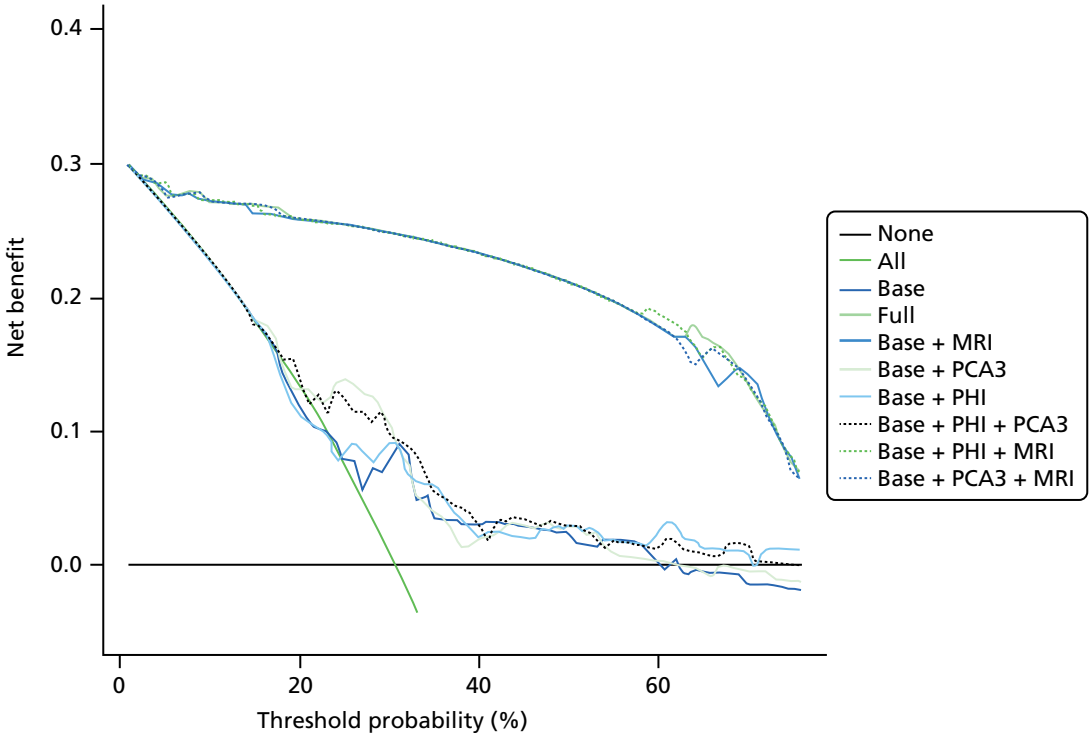

FIGURE 14 Decision curve analysis graph from Porpiglia *et al.*[99] DCAs of effect of various models on PCa detection. Threshold probability to undergo biopsy is reported vs. net benefit. Broken black line represents assumption that all patients will harbour PCa (biopsy all patients). Horizontal line represents assumption that no patients will harbour PCa (biopsy no patients). [Reprinted from *The Journal of Urology*, vol. 192, Porpiglia F, Russo F, Manfredi M, Mele F, Fiori C, Bollito E, Papotti M, Molineris I, Passera R, Regge D, The roles of multiparametric magnetic resonance imaging, PCA3 and prostate health index – which is the best predictor of prostate cancer after a negative biopsy?, pp. 60–6, © 2014, with permission from Elsevier.] DCA, decision curve analysis.

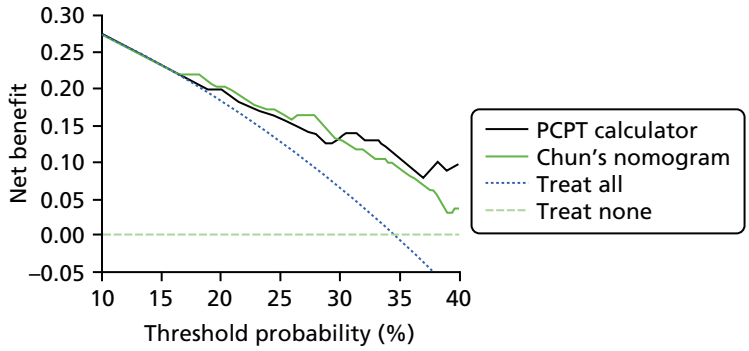

FIGURE 15 Decision curve analysis graph from Perdonà *et al.*[97] [Reprinted from *European Urology*, vol. 59, Perdonà S, Cavadas V, Di Lorenzo G, Damiano R, Chiappetta G, Del Prete P, Franco R, Azzarito G, Scala S, Arra C, De Sio M, Autorino R, Prostate cancer detection in the 'grey area' of prostate-specific antigen below 10 ng/ml: head-to-head comparison of the updated PCPT calculator and Chun's nomogram, two risk estimators incorporating prostate cancer antigen 3, pp. 81–7, © 2011, with permission from Elsevier.]

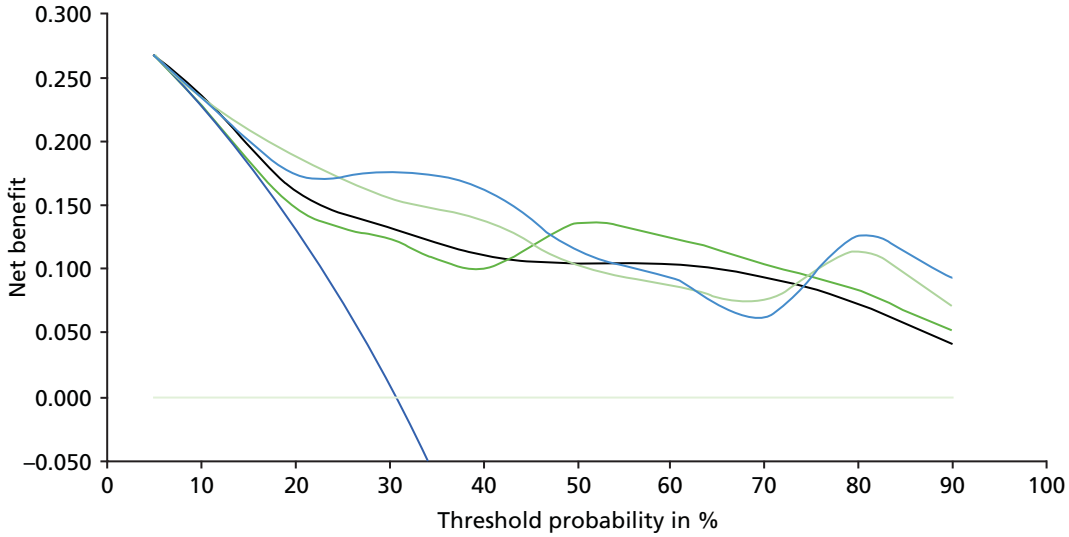

FIGURE 16 Decision curve analysis graph from Scattoni *et al.*[102] DCA of PCa diagnosis at biopsy shows net benefit of BMM 1 (black curve) consisting of age, PSA, per cent fPSA, DRE and prostate volume, and net benefit of adding PCA3 model 2 (dark green curve), PHI model 3 (light green curve) or PCA3 plus PHI model 4 (light blue curve) to BMM. Dark blue curve represents assumption that all patients will harbour PCa and be treated. Horizontal green curve represents assumption that no patient will harbour PCa and none will be treated. A, initial biopsy. B, repeat biopsy. [Reprinted from *The Journal of Urology*, vol. 190, Scattoni V, Lazzeri M, Lughezzani G, De Luca S, Passera R, Bollito E, Randone D, Abdollah F, Capitanio U, Larcher A, Lista G, Maria Gadda G, Bini V, Montorsi F, *et al.*, Head-to-head comparison of prostate health index and urinary PCA3 for predicting cancer at initial or repeat biopsy, pp. 496–501, © 2013, with permission from Elsevier.] BBM, base multivariate model; fPSA, free PSA.

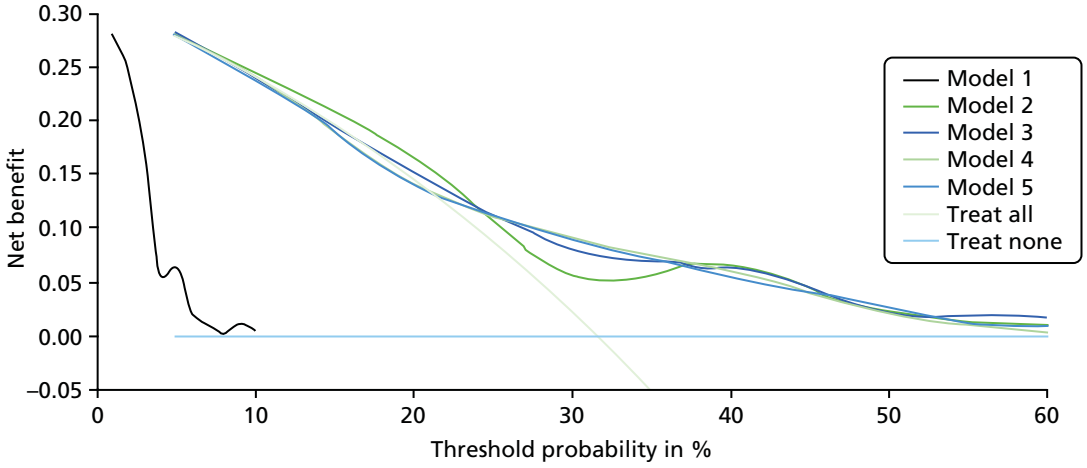

FIGURE 17 Decision curve analysis graph from Lazzeri *et al.*[92] DCA for models shown in Table 3. Models 3 (base plus %fPSA and p2PSA) 4 (base plus %fPSA and %p2PSA) and 5 (base plus %fPSA and phi) resulted in greater net benefit in PCa probability threshold range from 25% to 40% compared with model 1 (base) and model 2 (base plus %fPSA), which are perfectly superimposed. (Reprinted from *The Journal of Urology*, vol. 188, Lazzeri M, Briganti A, Scattoni V, Lughezzani G, Larcher A, Maria Gadda G, Lista G, Cestari A, Buffi N, Bini V, Freschi M, Rigatti P, Montorsi F, Guazzoni G, Serum Index Test %[-2]proPSA and Prostate Health Index are more accurate than prostate specific antigen and %fPSA in predicting a positive repeat prostate biopsy, pp. 1137–43, © 2012, with permission from Elsevier.) DCA, decision curve analysis.

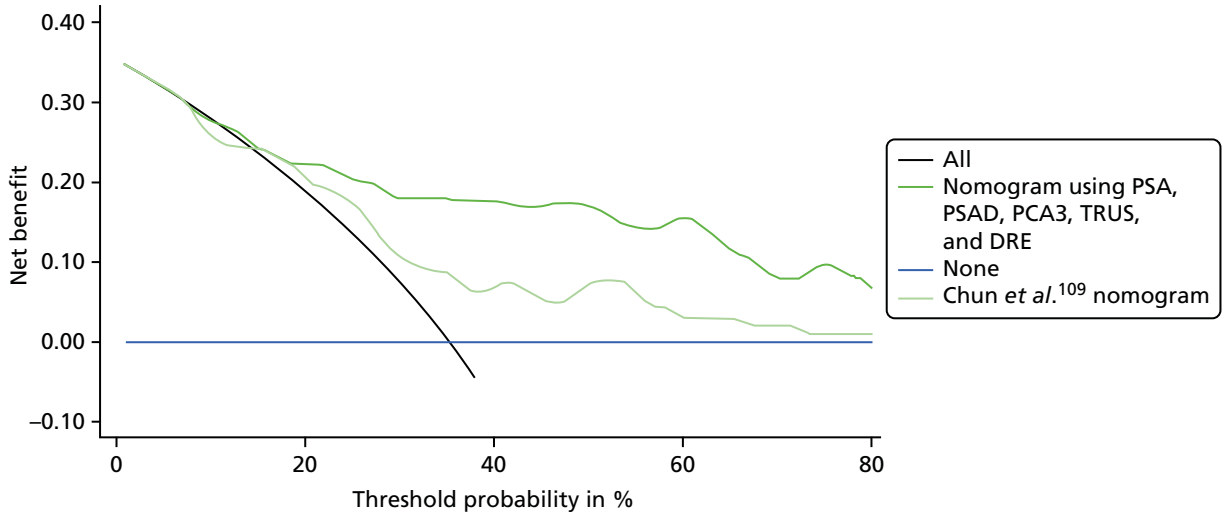

FIGURE 18 Decision curve analysis graph from Wu *et al.*[106] The 'all' line shows the net benefit if all patients were taken for repeat prostate biopsy. The 'none' line shows the net benefit if no patients were taken for repeat prostate biopsy. As the threshold probability increases, the net benefit for each nomogram decreases. DRE, digital rectal examination; PCA3, prostate cancer antigen 3; PSAD, PSA density; TRUS, transrectal ultrasound. (Adapted by permission from Macmillan Publishers Ltd: *Prostate Cancer and Prostatic Diseases* Wu AK, Reese AC, Cooperberg MR, Sadetsky N, Shinohara K. Utility of PCA3 in patients undergoing repeat biopsy for prostate cancer. *Prostate Cancer Prostatic Dis* 2012;**15**:100–5, © 2012.)

EME
HS&DR
HTA
PGfAR
PHR

Part of the NIHR Journals Library
www.journalslibrary.nihr.ac.uk

Published by the NIHR Journals Library