

# National Institute for Health and Care Excellence

## Diagnostics consultation document

### Diagnosing prostate cancer: PROGENSA PCA3 assay and Prostate Health Index

The National Institute for Health and Care Excellence (NICE) is producing guidance on using the PROGENSA PCA3 assay and the Prostate Health Index in the NHS in England. The Diagnostics Advisory Committee has considered the evidence submitted and the views of expert advisers.

**This document has been prepared for public consultation.** It summarises the evidence and views that have been considered, and sets out the draft recommendations made by the Committee. NICE invites comments from registered stakeholders, healthcare professionals and the public. This document should be read along with the evidence base (the diagnostics assessment report), which is available from <http://guidance.nice.org.uk/DT/InDevelopment>.

The Advisory Committee is interested in receiving comments on the following:

- Has all of the relevant evidence been taken into account?
- Are the summaries of clinical and cost effectiveness reasonable interpretations of the evidence?
- Are the provisional recommendations sound, and a suitable basis for guidance to the NHS?

#### Equality issues

NICE is committed to promoting equality of opportunity, eliminating unlawful discrimination and fostering good relations between people with particular protected characteristics and others. Please let us know if you think that the preliminary recommendations may need changing in order to meet these aims. In particular, please tell us if the preliminary recommendations:

- could have a different impact on people protected by the equality legislation than on the wider population, for example by making it more difficult in practice for a specific group to access the technology
- could have any adverse impact on people with a particular disability or disabilities.

Please provide any relevant information or data you have regarding such impacts and how they could be reduced.

**Note that this document is not NICE's final guidance on the PROGENSA PCA3 assay and the Prostate Health Index (PHI). The recommendations in section 1 may change after consultation.**

After consultation, the Committee will meet again to consider the evidence, this document and comments from the consultation. After considering these comments, the Committee will prepare its final recommendations, which will be the basis for NICE's guidance on the use of the technology in the NHS in England.

For further details, see the 'Diagnostics Assessment Programme process guide' (available at [www.nice.org.uk/aboutnice/howwework/developingnicediagnostictechnologiestheguidance](http://www.nice.org.uk/aboutnice/howwework/developingnicediagnostictechnologiestheguidance)).

**Key dates:**

Closing date for comments: 12 January 2015

Second Diagnostics Advisory Committee meeting: 21 January 2015

## **1 Provisional recommendation**

1.1 The PROGENSA PCA3 assay and the Prostate Health Index are not recommended for use in people having investigations for suspected prostate cancer, who have had a negative or inconclusive transrectal ultrasound prostate biopsy.

## **2 The technologies**

2.1 Two CE-marked technologies, the PROGENSA PCA3 assay and the Prostate Health Index, were identified during scoping as being relevant to this assessment. Additional details of these technologies are provided in section 4 of the guidance.

### **3 Clinical need and practice**

#### ***The problem addressed***

- 3.1 The PROGENSA PCA3 assay (PCA3 assay) and the Prostate Health Index (PHI) are in vitro diagnostic tests that are intended for use in people having investigations for suspected prostate cancer and who have had a negative or inconclusive transrectal ultrasound prostate biopsy. The tests detect specific biomarkers, PCA3 and PSA, that, when present at high levels, can suggest the presence of cancer. Both tests are intended to be used together with a review of risk factors, such as raised prostate specific antigen (PSA) levels and digital rectal examination findings, to help determine the need for a second biopsy to rule out prostate cancer.
- 3.2 Detection rates of prostate cancer are around 14–25% for first biopsy, and it is estimated that a significant proportion of people may get a negative or inconclusive result and need further investigations, including a second biopsy, to confirm the absence of prostate cancer. Prostate biopsies can detect clinically insignificant prostate cancer which may lead to unnecessary invasive treatments. Biopsy procedures are invasive, commonly associated with minor complications such as haemospermia, haematuria and rectal bleeding and are unpleasant for patients having them. In rare cases, biopsies can lead to major complications, such as sepsis, prostatitis, fever, urinary retention, epididymitis, and rectal bleeding for longer than two days (NHS Prostate Cancer Risk Management Programme 2010). The use of the PCA3 assay or the PHI may avoid second biopsies and associated complications by indicating which patients have a decreased likelihood of a positive biopsy result and therefore, are unlikely to have prostate cancer.
- 3.3 The purpose of this assessment is to evaluate the clinical and cost-effectiveness of using the PCA3 assay or the PHI in conjunction

with clinical assessment and other investigations to determine if people having investigations for prostate cancer need a second biopsy.

### ***The condition***

- 3.4 Prostate cancer is the most common cancer in males in the UK, and represents 26% of all male cancers in England and Wales in 2010. It is estimated that there are around 40,000 new cases of prostate cancer diagnosed in the UK every year. Prostate cancer is more likely to affect older people and most cases are diagnosed in people over 50 years of age.
- 3.5 Diagnosing prostate cancer often involves invasive testing such as a biopsy of the prostate gland. In 2012-13, it is estimated that there were 17,284 outpatient attendances in England associated with a rectal needle biopsy of the prostate and 1353 with perineal biopsy of the prostate (Hospital Episode Statistics [HES]). These figures are likely to substantially underestimate the number of biopsies performed because HES data only record inpatient episodes and day cases, and prostate biopsies are commonly performed as an outpatient procedure. Based on the number of people diagnosed with prostate cancer every year, anecdotal evidence suggests that the number of biopsies performed every year is more likely to be in the region of 80,000.

### ***The diagnostic and care pathways***

#### **Diagnosis**

- 3.6 The process for diagnosing and treating prostate cancer is described in the NICE clinical guideline on the diagnosis and treatment of prostate cancer. The guideline recommends the following before performing a biopsy:

- 3.7 To help people decide whether to have a prostate biopsy, discuss with them their prostate-specific antigen (PSA) level, digital rectal examination findings (including an estimate of prostate size) and comorbidities, together with their risk factors (including increasing age and black African-Caribbean family origin) and any history of a previous negative prostate biopsy. Do not automatically offer a prostate biopsy on the basis of serum PSA level alone.
- 3.8 Give people and their partners or carers information, support and adequate time to decide whether or not they wish to undergo prostate biopsy. Include an explanation of the risks (including the increased chance of having to live with the diagnosis of clinically insignificant prostate cancer) and benefits of prostate biopsy.
- 3.9 It is also recommended in [NICE clinical guideline 175](#) that prostate biopsies should be carried out following the procedure recommended by the Prostate Cancer Risk Management Programme (2006), Undertaking a transrectal ultrasound guided biopsy of the prostate. This recommends 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 to 12 core pattern that samples the midlobe peripheral zone and the lateral peripheral zone of the prostate only’. Transrectal ultrasound biopsy is usually carried out under local anaesthetic and involves using thin needles to take around 10 to 12 small pieces of tissue from the prostate.
- 3.10 For people who have a negative first prostate biopsy, [NICE clinical guideline 175](#) recommends that a core member of the urological cancer multi disciplinary team (MDT) should review the risk factors of all people who have had a negative first prostate biopsy, and discuss with the person that there is still a risk that prostate cancer

is present and the risk is slightly higher if any of the following risk factors are present:

- the biopsy showed high-grade prostatic intra-epithelial neoplasia (HGPIN)
- the biopsy showed atypical small acinar proliferation
- abnormal digital rectal examination.

3.11 [NICE clinical guideline 175](#) also recommends that multiparametric MRI (using T2 and diffusion-weighted imaging) be considered for people with a negative transrectal ultrasound 10–12 core biopsy to determine whether another biopsy is needed. A repeat biopsy should not be offered if the multiparametric MRI is negative, unless any of the risk factors (listed above) are present. In current NHS practice, a multiparametric MRI may not be carried out until 6 to 12 weeks after the transrectal ultrasound biopsy because any haemorrhage after the biopsy can cause artifacts in the images and this may reduce the diagnostic accuracy of the prostate multiparametric MRI.

3.12 A second biopsy may be taken using a template biopsy. A template biopsy uses a template grid, either with a cross sectional MRI (where available) or uses transrectal ultrasound imaging with transperineal sampling of the prostate under general anaesthetic. Usually, around 25 to 40 samples of the prostate are taken during a template biopsy.

3.13 Another type of second biopsy is the 'saturation' biopsy, which involves more than 20 cores being taken from the prostate. A saturation biopsy may be carried out transrectally or using a transperineal approach. The transperineal approach is generally carried out as a stereotactic template-guided procedure under general anaesthesia. The saturation approach leads to improved sampling of the anterior zones of the gland, which may be under-

sampled in a transrectal ultrasound biopsy and which may lead to cancer cells being missed.

- 3.14 A third option that can be used for a second biopsy is the targeted approach which uses MRI to map, target and track biopsy sites. Like the saturation approach, it aims to improve sampling of the anterior zones of the gland. The use of this approach relies on the availability of radiologists with relevant expertise and experience, and access to MRI.
- 3.15 For patients who have an initial negative transrectal ultrasound biopsy, with no indication of risk from other risk factors, the usual care is PSA surveillance, with repeat PSA testing taking place every 3 months.

## **4 The diagnostic tests**

### ***The interventions***

#### **The PROGENSA PCA3 assay**

- 4.1 The PROGENSA PCA3 assay (Hologic GenProbe) is an in vitro nucleic acid amplification test and is intended for the quantitative determination of prostate cancer antigen 3 (PCA3) ribonucleic acid (RNA) in urine. A digital rectal examination is performed, which releases prostate cells and RNA into the urinary tract, which are collected in a urine sample. Once collected, 2.5 ml of the sample is added to a transport tube containing a urine transport medium that triggers the breakdown of any remaining prostate cells and stabilises the RNA.
- 4.2 The PCA3 assay incorporates 2 nucleic acid amplification tests: 1 test for detecting PCA3 messenger RNA (mRNA) and 1 test for detecting prostate specific antigen (PSA) mRNA. PCA3 mRNA is

highly overexpressed in prostate cancer tissue cells compared with adjacent benign tissue, whereas PSA gene expression is relatively constant in normal prostate cells. By combining the detection of both genes in 1 assay, a PCA3 score based on the ratio of PCA3 mRNA to PSA mRNA can be generated. The PCA3 score can then be used to aid the risk stratification of people being considered for repeat biopsies. Higher PCA3 scores are associated with a higher probability of a positive biopsy. For the purposes of this assessment the threshold PCA3 score used to indicate the likelihood of a positive biopsy was 25 or above (as indicated in the company's information for use). This assay analyses the levels of PSA mRNA found in the urine following a digital rectal examination. So the PSA values reported by this assay are not the same as those that would be reported using the standard PSA test for prostate cancer, since the standard test detects the levels of PSA protein in the serum of a blood sample.

- 4.3 The PCA3 assay can be used with the Hologic Gen-Probe Direct Tube Sampling 400, 800 and 1600 molecular laboratory systems. The PCA3 assay is not compatible with other analysers. Each PCA3 assay kit is suitable for 2 x100 reactions and includes reagents, controls and calibrators for both the PCA3 and PSA reactions.
- 4.4 The instructions for use document states that the PCA3 assay should not be used for patients who are taking medication known to affect serum PSA levels such as finasteride, dutasteride and leuprorelin. The effect of these medications on PCA3 gene expression has not yet been evaluated. Certain therapeutic and diagnostic procedures such as prostatectomy, radiation, prostate biopsy, and others may affect the viability of prostatic tissue and subsequently impact the PCA3 score. The effect of these procedures on assay performance has not yet been evaluated.

Samples for PCA3 testing should be collected when the clinician believes prostate tissue has recovered from these medications and procedures.

### **The Prostate Health Index (PHI)**

- 4.5 The Prostate Health Index (PHI) (Beckman Coulter) is an in vitro diagnostic multivariate index assay that combines the results of 3 quantitative blood serum immunoassays (Access Hybritech PSA, fPSA and p2PSA) for different types of PSA into a single numerical result, the PHI. These assays can be carried out on the same blood sample without special sample handling or preparation. Therefore, the PHI can be calculated in a routine blood sciences laboratory using Beckman Coulter analysers with the PHI algorithm incorporated in the software.
- 4.6 PHI is calculated using the equation:  $(p2PSA/free\ prostate\ specific\ antigen) \times \sqrt{total\ PSA}$
- 4.7 The company reports that PHI is validated to perform equivalently with the traditional Beckman Coulter Hybritech calibration and the Beckman Coulter WHO calibration for both the Access Hybritech PSA and free PSA assays. The Beckman Coulter PHI is not intended to be calculated using PSA, or free PSA results, from any other company's assay and the PHI assay is only compatible with Beckman Coulter Access instruments (Access2, Dxl600, Dxl800, DxC600i, DxC680i, DxC800i, DxC880i).
- 4.8 The PHI is designed to detect prostate cancer in people aged 50 years and older with total PSA between 2–10 nanograms/ml and with digital rectal examination findings that are not suspicious for cancer. The PHI can be used to categorise patients into low, moderate and high probabilities of prostate cancer being found on biopsy. The company suggests that a score of 0–20.9 indicates low risk (8.4%) of cancer; 21–39.9 indicates moderate risk (21%) and

greater than 40 indicates high risk (44%). The company reports that the PHI demonstrates more than 3 times the specificity at 90% clinical sensitivity than PSA alone.

- 4.9 In the NHS, it is likely that a second biopsy would be carried out for people with a PHI score of 21 and above, that is, those classified as at moderate and high risk of cancer. For people with a PHI score of less than 21, it is likely they would have their condition monitored over time by PSA testing rather than having a second biopsy, although this is dependent on other risk factors.
- 4.10 Information provided by the company states that the effect of medication for benign prostate hyperplasia, and specifically the 5 alpha reductase inhibitors on the level of p2PSA is not known. As a result, PHI results cannot be interpreted in patients taking 5 alpha reductase inhibitors medication and PHI testing should not be offered to these people.
- 4.11 In addition the company has stated that stability studies showed that the p2PSA assay is not stable on coagulated blood. When left on clotted samples at room temperature, the p2PSA concentration increases significantly after 3 hours. Therefore it is important that the serum sample is prepared within this time frame.

### ***The comparator: clinical assessment or clinical assessment plus MRI***

- 4.12 In this assessment, 2 comparators were used:
- Clinical assessment was used to decide if a repeat prostate biopsy should be conducted. This was based on clinical judgement and previous findings such as age, prostate size, PSA levels, close male relative (brother or father) with prostate cancer, the presence of atypical small acinar proliferation, high-

grade prostatic intra-epithelial neoplasia or abnormal digital rectal examination.

- Clinical assessment (based on the above) plus the results of a multiparametric MRI.

## **5 Outcomes**

The Diagnostics Advisory Committee (section 9) considered evidence from a number of sources (section 10).

### ***How outcomes were assessed***

- 5.1 The External Assessment Group conducted 3 systematic reviews of the evidence on clinical effectiveness for the 2 tests, the PROGENSA PCA3 assay (PCA3 assay) and the prostate health index (PHI) when used in people who are having investigations for suspected prostate cancer, who have had a negative or inconclusive prostate biopsy. These covered analytical validity, clinical validity and clinical utility. Studies were considered for inclusion based on criteria developed for each systematic review as defined in the assessment protocol.
- 5.2 In total, 37 studies met the inclusion criteria and were included in this assessment; 6 studies reported the analytical validity and 31 studies reported the clinical validity of the tests. No studies reporting clinical utility of the tests were identified. Critical appraisal of the studies was carried out using the Tuetsch checklist or the QUADAS-2 tool.

### ***Clinical Effectiveness***

#### **Evidence on analytical validity**

- 5.3 Three studies reported the analytical validity of the PCA3 assay and 3 other studies used the p2PSA assay. (This assay is part of the PHI). All the studies using the PCA3 assay and 1 study using

the p2PSA assay were carried out in the USA. The remaining 2 studies using the p2PSA assay were carried out in Germany. The External Assessment Group also reviewed analytical validity data from the Food and Drug Administration (the Summary of Safety and Effectiveness Data [SSED] report), and information submitted by the company.

### PCA3 assay

5.4 Sokoll et al. (2008), the SSED report and company information reported the analytical sensitivity of the test. The analytical specificity was also reported in the company information. (Table 1)

Table 1 Analytical sensitivity of PCA3 assay

	Methods		LoB Copies/ml	LoD Copies/ml	LoQ Copies/ml
Sokoll 2008 <sup>46</sup>	LoD: lowest measureable concentration of controls LoB: 95 percentile of zero Calibrator LoQ: < 130% recovery and CV <35%	PCA3	176	259	259
		PSA	831	2338	2338
SSED 2012 <sup>48</sup>	4 blank female urine and 4 female urine spiked to Calibrator 2 concentrations LoD=LoB + 1.65 SD	PCA3	90	239	239
		PSA	254	3338	3338
Pack insert <sup>49</sup>	Diluted in-vitro transcripts. LoQ assessed according to CLSI EP17-A	PCA3	NR	80 c/mL	Calibrator 2~ 750
		PSA	NR	1438 c/mL	Calibrator 2 ~ 7500

LoB=limit of blank; LoD=limit of detection; LoQ=limit of quantitation; CV=coefficient of variation; SD=standard deviation; NR=not reported

5.5 The accuracy of the PCA3 assay was reported in Groskopf et al. (2006) and Sokoll et al. (2008), as well as in the SSED report and the company's information. As no gold standard (best available) method is available, accuracy was calculated by the percentage recovery of measured prostate cancer antigen 3 (PCA3) or prostate specific antigen (PSA) ribonucleic acid (RNA) (copies/ml).

Recovery varied from 90%– 118% for PCA3 and from 85%–121% for PSA RNA copies/ml.

- 5.6 Within-laboratory variation was reported in 4 studies as well as the SSED report and company information. In these six reports, the within-laboratory total coefficient of variation ranged from 4% to 27% for PCA3 and from 7% to 18.7% for PSA. In the SSED (2012) report and company information, the coefficient of variation of the PCA3 score ranged from 12% to 28%.
- 5.7 One study (Sokoll et al. 2008) and the SSED report (2012) reported both the within-laboratory and between-laboratory total coefficients of variation. These ranged from 5.9% to 17.2% for PCA3 and 10.1% to 11.3 % for PSA (Sokoll et al. 2008). The SSED report contained within-laboratory and between-laboratory total coefficient of variation for the PCA3 score; this ranged from 12.3% to 25.0%. Most variation appeared to occur within-laboratory; between-laboratory results contributed little additional variation.
- 5.8 The External Assessment Group considered there was uncertainty in interpreting PCA3 scores near the test cut-off point (25). For example, with a coefficient of variation of 25%, a sample with a true PCA3 score of 25 would give a result between 19 and 31 in two thirds of tests, with a third of tests lying outside this range.

### **p2PSA assay and the PHI assay**

- 5.9 Sokoll et al. (2012), Stephan et al. (2009) and the SSED report reported the analytic sensitivity of the PHI test. Sokoll et al. reported the limit of blank of p2PSA as 0.5 picograms/ml and a limit of detection of p2PSA as 0.7 picograms/ml. Stephan et al. only reported limit of detection data, while the Sokoll et al. included this, together with limit of blank and limit of quantitation data. The analytic specificity was also reported in the company information and the SSED report. (Table 2)

Table 2 Analytical sensitivity p2PSA assay

	Methods	p2PSA		
		LoB pg/mL	LoD pg/mL	LoQ pg/mL
SSED 2012 <sup>56</sup>	LoB: 95 percentile of zero analyte LoD: LOB+ 1.65 SD [SD from patient serum LoQ- dilutions of calibrators from LoD to 7x LoD] LoQ: concentration with CV20% from quadratic model	0.50	0.69	3.23
Sokoll 2012 <sup>72</sup>	Methods as for SSED. <i>Appears to be same study results as in SSED</i>	0.50	0.70	3.23
Stephan 2009 <sup>73</sup>	LoD: repeat measurement of zero calibrator + 2SD	Not reported	2.27	Not reported

LoB=limit of blank; LoD=limit of detection; LoQ=limit of quantitation; CV=coefficient of variation; SD=standard deviation; NR=not reported

- 5.10 The accuracy of the p2PSA/PHI assay was reported by Sokoll et al. (2012), Stephan et al. (2009) and the SSED report. As no gold standard is available, accuracy was calculated by the percentage recovery of measured p2PSA picogram/ml in male serum samples containing different known amounts of purified p2PSA. Recovery of this p2PSA reference standard ranged from 90% to 103%.
- 5.11 Within-laboratory variation was reported by Sokoll et al. (2012), Stephan et al. (2009) and in the SSED report. The SSED report also included data on between-laboratory variation. Both studies and the SSED report included data on the coefficient of variation for the p2PSA assay, but only the SSED report included data on the coefficient of variation for the PHI. The within-laboratory total coefficient of variation ranged from 3% to 13% for p2PSA and from 8.5% to 12% for the PHI assay. The within-laboratory and between-laboratory total coefficient of variation ranged from 5.54% to 9.4% for p2PSA and from 4.9% to 7.3% for the PHI assay.

## Evidence on clinical validity

- 5.12 The use of the PCA3 assay and the PHI was considered in 3 possible diagnostic pathways:
- The PCA3 score and the PHI alongside clinical assessment to inform the decision to perform a second biopsy.
  - The PCA3 score and the PHI alongside clinical assessment to inform the decision to perform a multiparametric MRI scan before second biopsy. If the multiparametric MRI is positive, a second biopsy would be performed.
  - The PCA3 score and the PHI alongside clinical assessment to inform the decision to perform a second biopsy in people who have had a negative multiparametric MRI scan.
- 5.13 Comparisons between the performance of the intervention tests (the PCA3 assay and the PHI) and the comparators (clinical assessment and clinical assessment with multiparametric MRI) were made using either data from studies carried out in the same study population (within-study or direct comparisons) or using studies where intervention and comparator tests were carried out in different populations (between-study or indirect comparisons).
- 5.14 There were 25 publications identified that met the inclusion criteria for the within-study comparisons. Of these, 21 reported within-study comparisons between clinical assessment with the PCA3 assay compared with a comparator diagnostic strategy, or clinical assessment with PHI compared with a comparator diagnostic strategy. The remaining 4 publications reported univariate assessments of the PCA3 assay or the PHI compared with univariate PSA which only provided limited data.
- 5.15 Twenty one publications reported within-study comparisons; clinical assessment versus clinical assessment with PCA3 or the PHI and clinical assessment with MRI and the PCA3 assay or the PHI. Of

these, 17 publications, were included in the review. (The remaining 4 publications, reporting data from the European Cohort study (3) and from Italy (1) did not present additional study results.) The majority of reviewed publications were cohort studies with only 1 randomised controlled trial identified.

- 5.16 Of the 17 studies included in the review, 9 were conducted in Italy, 3 in Europe, 3 in the USA and 2 were described as international. No studies were conducted in the UK.
- 5.17 There were 6 papers reporting 5 systematic reviews and meta-analyses identified that considered between-study comparisons and met the inclusion criteria. None of these reviews considered clinically relevant comparisons.
- 5.18 Critical appraisal of the identified studies reporting relevant within-study comparisons was carried out using the QUADAS-2 tool. The main potential sources of bias in the included studies related to patient selection and a lack of reported details on the intervention tests, comparators and biopsies. No meta-analyses were carried out because of the heterogeneity in the included studies.
- 5.19 Reported outcomes included a range of measures of diagnostic performance, most frequently from multivariate logistic regression models using receiver operating characteristic curves, area under the curve statistics, multivariate odds ratios and derived sensitivity and specificity values. Only 1 study, Tombal et al. (2013), presented independent sensitivity and specificity estimates.
- 5.20 There were 6 studies that 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 person or clinician will opt for a biopsy.

5.21 The External Assessment Group considered the 4 most clinically relevant comparisons for the NHS to be:

- clinical assessment alone compared with clinical assessment plus the PCA3 assay
- clinical assessment alone compared with clinical assessment plus the PHI
- clinical assessment plus MRI compared with clinical assessment plus MRI plus the PCA3 assay
- clinical assessment plus MRI compared with clinical assessment plus MRI plus the PHI.

**The PCA3 assay: clinical assessment compared with clinical assessment plus the PCA3 assay**

5.22 There were 8 area under the curve results reported from 6 study populations for the comparison of clinical assessment compared with clinical assessment with the PCA3; 1 study reported the results from 2 models; 1 study used the PCA3 score as a continuous variable and 1 study employed a threshold value of 35. The studies showed an increase in discrimination of between 1% to 19% when the PCA3 score was added to the clinical assessment, either as continuous or binary variable.

5.23 In addition, 2 studies reported area under the curve results only for models of clinical assessment with the PCA3, assay. These results were similar to the area under the curve results reported in other studies: Goode (2013) reported an area under the curve of 0.61 for a multivariate logistic regression model; and Perdona et al. (2011) reported an area under the curve of 0.74 for the Chun nomogram and an area under the curve of 0.74 for the Prostate Cancer

Prevention Trial nomogram. (The Chun nomogram and Prostate Prevention Trial nomograms are risk predicting systems for prostate cancer).

- 5.24 There were 5 studies that reported 7 multivariate odds ratios for clinical assessment with the PCA3 assay. Of these, 4 studies reported statistically significant results (odds ratios (OR) > 1, with confidence intervals that did not include 1). One study had an odds ratio above 1, but this was not statistically significant. Haese et al. (2008) reported that the multivariate odds ratio for the PCA3 score was significant (p=0.006), but did not report the effect size. These results are consistent with the area under the curve results and indicate that the addition of the PCA3 score to the clinical assessment increases discrimination. Pepe et al. (2013) reported derived sensitivity and specificity for various risk thresholds in a logistic regression model. At a 25% risk threshold, the Prostate Cancer Prevention Trial nomogram and the Prostate Cancer Prevention Trial nomogram with the PCA3 assay; both had 100% sensitivity but low specificity (1% and 8% respectively). Using a 40% risk threshold, the Prostate Cancer Prevention Trial nomogram alone had 75% sensitivity and 26% specificity. Prostate Cancer Prevention Trial with the PCA3 assay had 85.8% sensitivity and 25.0% specificity. This study population comprised white men with abnormal digital rectal examination findings and no family history of prostate cancer; the diagnostic power of the Prostate Cancer Prevention Trial nomogram was therefore likely to be reduced.
- 5.25 There were 2 studies that 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.(2014) showed a 9.5% decrease but Ankerst et al. (2008) showed a 2.4% increase in discrimination.

5.26 There were 3 studies that reported derived specificity for sensitivity set at 80%, 90% or 95%. The results were conflicting. When sensitivity is set at 80% or 90%, Scattoni et al. (2013) showed that derived specificity decreases when the PCA3 score is added to clinical assessment. Gittleman et al. (2013) reported that at a set sensitivity of 80% the addition of the PCA3 score to clinical assessment increased derived specificity from 18.9% to 41.5%. Porpiglia et al. (2014) reported that adding the PCA3 score to clinical assessment increased derived specificity when sensitivity is set at 80% and 95% and reduced derived specificity when sensitivity is set at 90%.

5.27 There were 3 studies that presented decision curve analyses comparing net benefit for clinical assessment and for clinical assessment with the PCA3 assay. Busetto et al. (2013) and Porpiglia et al. (2014) reported that there is no benefit gained from adding the PCA3 score to clinical assessment between a 10% and 20% threshold probability. Net benefit was greater in Busetto et al. (2013) between 25% to 50% thresholds and in Porpiglia et al. (2014), between 20% and 35% thresholds. In Scattoni et al. (2013) net benefit was reduced when the PCA3 score was added to the clinical assessment between 10% and 40% threshold probabilities. At 40%, the curves reversed, with increased net benefit from 50% to 90% threshold probabilities.

#### **Clinical assessment plus MRI compared with clinical assessment plus MRI plus the PCA3 assay**

5.28 There were 2 studies that investigated the addition of the PCA3 score to a diagnostic model that included clinical assessment and MRI. Adding the PCA3 score to clinical assessment and MRI had very little effect on the size of the reported area under the curve.

Porpiglia et al. (2014) showed a 1% decrease in area under the curve and Busetto et al. (2013) reported a 3% increase in area under the curve. However, the diagnostic models with clinical assessment and MRI already had very high estimates of area under the curve, so adding the PCA3 score was unlikely to result in substantial changes.

- 5.29 Multivariate odds ratios for clinical assessment with MRI compared with clinical assessment with MRI and PCA3 assay were reported in 1 study. In the model containing both MRI and the PCA3 score, the odds ratio for MRI was much larger (94.55 [95% confidence interval (CI) 32.14 to 346.54]) than that for the PCA3 score (1.85 [95% CI 0.26 to 9.90]). In this model, the odds ratio for the PCA3 score was not statistically significant.
- 5.30 At 80% and 95% specificity, Porpiglia et al. (2014) reported no change in derived sensitivity for clinical assessment with MRI compared with clinical assessment using MRI and the PCA3 assay. At 90% specificity, derived sensitivity increased by 0.3%.
- 5.31 Porpiglia et al. (2014) reported minimal changes in derived specificity for clinical assessment with MRI compared with clinical assessment with MRI and the PCA3 assay; at 80% and 90% sensitivity, derived specificity increased by 0% and 0.8% respectively. At 95% sensitivity Porpiglia et al. (2014) reported a decrease in derived specificity (-5.9%) when the PCA3 score was added to clinical assessment with MRI.
- 5.32 Decision curve analysis results for 2 studies demonstrated that the addition of the PCA3 score does not improve diagnostic accuracy when added to the clinical assessment with MRI between threshold probabilities of 10% to 50%.

### **Between-study comparisons**

5.33 There were 2 reviews that assessed the clinical validity of using the PCA3 score to predict prostate cancer. Luo et al. (2014) considered a repeat biopsy population and included studies without a comparator. Luo et al. (2014) concluded that using 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%. A review by Bradley et al. (2014), which restricted inclusion to studies that compared the PCA3 score with a comparator of either clinical assessment or PSA, concluded that although the PCA3 score appeared to have better discrimination in detecting prostate cancer than total PSA, the strength of evidence was low.

### **PHI: clinical assessment compared with clinical assessment plus PHI**

5.34 There were 4 studies that reported area under the curve for the comparisons of clinical assessment compared with clinical assessment plus PHI. All studies showed an increase in discrimination of between 2–10% when the PHI was added to the clinical assessment model as a continuous variable.

5.35 There were 2 studies that reported multivariate odds ratios for the PHI. Both studies presented statistically significant results indicating that an increase in PHI score was associated with an increased probability of finding cancer by biopsy. These results are consistent with the area under the curve results and indicate that adding the PHI to the clinical assessment model increases discrimination.

5.36 There was 1 study that reported derived sensitivity values at 80%, 90% and 95% specificity. Adding the PHI to clinical assessment was associated with either a 2% increase at 90% and 95% specificity, or a 5.7% decrease in derived sensitivity at 80%

specificity. There were 2 other studies that reported derived specificity for 80% and 90% sensitivity. Scattoni et al. (2013) showed that adding the PHI to clinical assessment increased derived specificity at 80% sensitivity by 17% and, at 90% sensitivity by 2%. Porpiglia et al. (2014) showed that adding the PHI to clinical assessment was associated with a 2.5% and a 10.2% decrease in derived specificity at 80% and 90% sensitivity respectively and, a 0.9% increase in derived specificity at 95% sensitivity.

- 5.37 There were 3 studies that presented decision curve analyses comparing net benefit for clinical assessment and clinical assessment with the PHI. Lazzeri et al. (2012) showed that net benefit was greater for the clinical assessment model from threshold probabilities 20- 25%. Clinical assessment with the PHI had a greater net benefit between threshold probabilities of 25-40%. Scattoni et al. (2013) showed increased net benefit for the clinical assessment with the PHI model between threshold probabilities from 10- 50%. Porpiglia et al. (2014) demonstrated that estimates of net benefit for both models were similar between threshold probabilities of 10-70%.

#### **Clinical assessment plus MRI compared with clinical assessment plus MRI plus the PHI**

- 5.38 Porpiglia et al. (2014) showed that adding the PHI to a model comprised of clinical assessment with MRI had no effect on the size of the area under the curve.
- 5.39 Multivariate odds ratios for clinical assessment with MRI and the PHI compared with clinical assessment with MRI were reported in 1 study. In the model containing both MRI and the PHI, the odds ratios for MRI were larger (103.45; 95% CI 34.49 to 387.45) than the odds ratio for the PHI (0.76; 95% CI 0.17 to 4.40). The odds ratio for the PHI was not statistically significant.

- 5.40 Porpiglia et al. (2014) reported minimal changes in derived sensitivity at set specificity levels for clinical assessment with MRI and the PHI compared to clinical assessment with MRI. At 80%, 90% and 95% specificity, derived sensitivity increased by 0%, 0.3% and 0% respectively.
- 5.41 Porpiglia et al. (2014) reported minimal change in derived specificity for the addition of the PHI to clinical assessment with MRI. At 80%, 90% and 95% sensitivity, derived specificity increased by 0%, 0.8% and 0.9% respectively. Adding the PHI to diagnostic models incorporating clinical assessment with MRI had a negligible effect on derived specificity.
- 5.42 The decision curve analysis graphs in the study by Porpiglia et al. (2014) demonstrated that the PHI does not improve diagnostic accuracy when added to clinical assessment with MRI between threshold probabilities of 10– 60%.

#### **Identification of more aggressive cancers**

- 5.43 One of the parameters used to evaluate the prognosis of people with prostate cancer is the Gleason score. The score is based on the microscopic appearance of cancer, with a higher score ( $\geq 7$ ) indicating a more aggressive cancer, a worse prognosis for the patient.
- 5.44 There were 7 studies that reported diagnostic accuracy results for the PCA3 assay for detecting more aggressive cancers. In 6 studies, the authors used univariate analyses and showed the ability of the PCA3 score to predict a Gleason score of 7 or greater. One study reported how using the PCA3 score in combination with clinical assessment contributed to the prediction of more aggressive cancers.

- 5.45 There were 2 studies that reported median PCA3 scores for detected cancers with a Gleason score above or below 7. Both studies found that the PCA3 scores were higher in cancers with higher Gleason scores. In Haese et al. (2008), the median PCA3 scores were 28.1 for cancers with Gleason score of less than 7 and 45.3 for cancers with Gleason score of 7 and higher ( $p=0.04$ ). In Aubin (2008), the corresponding median PCA3 scores were 31.8 and 49.5 respectively ( $p=0.002$ ).
- 5.46 Busetto et al. (2013) reported a statistically significant association ( $p<0.001$ ,  $\chi^2=71.27$ ) between Gleason score and the PCA3 score. Haese et al. (2008) also reported significant differences in the median PCA3 scores for clinical stage T1c cancers (small cancers inside the prostate) compared with T2 (larger cancers, in both lobes of the prostate but still inside the gland) cancers (26.8 vs. 61.7,  $p=0.005$ ) and for relatively benign cancers (defined as clinical stage T1c, PSA density  $<0.15$ , Gleason score  $\leq 6$  and percentage of positive cores  $\leq 33\%$ ) versus significant cancers (21.4 vs. 42.1,  $p=0.006$ ).
- 5.47 Gittelman et al. (2013) reported the sensitivity, specificity and area under the curve for the PCA3 assay using a score of 25 as the threshold for the detection of all cancers, the detection of cancers with a Gleason score of 7 or higher and the detection of significant cancers (defined as clinical stage T2 or above, PSA density  $>0.15$ , Gleason score  $\geq 7$  and  $\geq 3$  cores positive for cancer). The area under the curve values reported were 0.707 for all cancers, 0.638 for cancers with a Gleason score of 7 or higher 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) for the 3 groups respectively. 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), respectively. There was no evidence that the sensitivity or specificity of the PCA3 assay varied between the groups.

- 5.48 Bollito et al. (2012) and Haese et al. (2008) report the numbers of cancers that would have been missed using PCA3 screening alone and would have had a Gleason score of 7 or higher. In Haese et al. (2008), using a PCA3 score of 20 as the threshold for detecting 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 or higher. Using a PCA3 score of 35 as the threshold for detecting cancer, 68 out of 128 cancers would have been missed and 27 of these 68 cancers would have had a Gleason score of 7 or higher. In Bollito et al. (2012), using a PCA3 score of 39 for the threshold for detecting cancer, 22 out of 281 cancers would have been missed and none of these would have had a Gleason score of 7 or higher. 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 or higher.
- 5.49 Tombal et al. (2013) reported how the use of the PCA3 score in combination with clinical assessment contributed to predicting more aggressive cancers and showed that the sensitivity and specificity of the PCA3 score to detect cancers with Gleason score of 7 or higher was better than its specificity and sensitivity to detect all cancers. This study presented independent sensitivity and specificity estimates for all cancers and demonstrated that adding the PCA3 score to best clinical judgment 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 or higher

(prevalence=5.4%), adding the PCA3 score increased sensitivity from 75% to 85% and specificity from 26% to 67% meaning that 6 more cancers would have been detected and 395 biopsies would have been avoided compared with clinical assessment alone.

- 5.50 Only 1 study considered the relationship between the PHI and the Gleason score. The authors found a significant correlation with increased PHI being associated with higher Gleason score (Spearman's rho 0.299,  $p=0.013$ ). It is not clear from the publication whether these findings are for all biopsies or repeat biopsies only.
- 5.51 Filella (2013), in a narrative review, highlighted inconsistencies in the evidence linking higher PCA3 scores to various markers of tumour aggressiveness. Wang (2014), in a meta-analysis of four studies, found that the area under the curve of the 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%). These 2 reviews did not meet the inclusion criteria for the systematic review because they were not restricted to studies of repeat biopsies and did not consider the intervention test used in combination with other diagnostic tests.

### ***Costs and cost effectiveness***

- 5.52 The External Assessment Group conducted a systematic review to identify existing economic analyses using the PCA3 assay and the PHI. The review also sought to identify potentially relevant evidence sources to inform parameter values for the de novo economic model developed by the External Assessment Group. The de novo economic model aimed to assess the cost-effectiveness of using the PCA3 assay and the PHI in the current prostate cancer diagnostic pathway.

- 5.53 The systematic review of existing economic analyses did not identify any published economic studies which met the inclusion criteria.
- 5.54 The External Assessment Group developed a de novo economic model designed to assess the cost effectiveness of using 2 different tests, the PCA3 assay and the PHI, in the diagnosis of prostate cancer in people having investigations for suspected prostate cancer, who have had a negative or inconclusive transrectal ultrasound prostate biopsy.
- 5.55 The following diagnostic strategies were included in the model:
- Clinical assessment
  - Clinical assessment plus the PCA3 assay
  - Clinical assessment plus the PHI
  - Clinical assessment plus the PCA3 assay plus the PHI
  - Clinical assessment plus multiparametric MRI
  - Clinical assessment plus multiparametric MRI plus the PCA3 assay
  - Clinical assessment plus multiparametric MRI plus the PHI
  - Clinical assessment plus multiparametric MRI plus the PCA3 assay plus the PHI
- 5.56 The de novo economic model uses the derived specificities for stated sensitivity levels. The use of derived specificity at stated sensitivity levels allows a comparison to be made between different testing strategies. Using this approach, the percentage of cancers detected is always the same regardless of the diagnostic strategy chosen but the number of biopsies needed 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.

5.57 As the percentage of detected underlying cancers is the same for all diagnostic strategies in the 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. There is some evidence that biopsy may be linked to increased mortality in the short term, but this has not been proven. The de novo model, therefore, considers the short-term impact of a biopsy on quality of life and associated complications.

#### **Model structure**

- 5.58 In the model structure, after 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.
- 5.59 As part of developing NICE's clinical guidance on the diagnosis and treatment of [prostate cancer](#), an economic model was produced, which explored the use of multiparametric MRI before a transrectal ultrasound guided prostate biopsy in people with suspected prostate cancer. It was 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 request one. This assumption was also made in the de novo economic model for this evaluation.
- 5.60 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 need a lifetime horizon. In the base case, the PSA monitoring strategy runs for 3 years so the time horizon is also 3 years. Both costs and benefits have been discounted at 3.5% per annum.

## Model inputs

- 5.61 The model was populated using data derived from the systematic clinical effectiveness review. Because it was not possible to carry out between-trial analysis and pool effectiveness data, the data in each study have been considered independently. Other focused reviews to inform key parameters were also used to populate the model, for example, baseline risks and routine sources of cost data.
- 5.62 Data on the resource use and costs associated with the different diagnostic strategies were informed by published literature, existing guidance, companies' prices, and other routine sources of unit cost data. Some costs were informed by expert opinion where suitable data from other sources were not available.
- 5.63 The utility values used in the model are the disutilities associated with a biopsy. No primary studies collecting disutility values specifically associated with prostate biopsy were identified. The baseline utility value for people having a prostate biopsy was taken from a study (Heijnsdijk et al. 2012) which reported a utility decrement of 0.1 that lasted 3 weeks following a biopsy. However, this utility value was taken from an earlier study that focused on breast cancer biopsy and the duration of decrement was an assumption based on clinical opinion. In the absence of any other evidence, this utility value has been incorporated into the base-case model as a quality adjusted life year (QALY) loss of 0.0058 from a prostate biopsy.
- 5.64 For the purposes of decision-making, the incremental cost-effectiveness ratios (ICERs) per QALY gained were considered. The model was executed with a hypothetical cohort of 1000 patients. The following assumptions were applied in the base case analyses :

- Patients with undiagnosed cancer, either with or without a second biopsy, will continue to have elevated PSA levels.
- 25% of people without cancer will continue to have a rising PSA level.
- At 1, 2 and 3 years respectively, 25%, 50% and 100% of patients with a rising PSA level will have a saturation biopsy.
- All patients who are recommended for a second biopsy choose to go ahead with the biopsy.
- All those for whom a second biopsy is not recommended do not request one.
- All patients whose second biopsy results are negative or inconclusive do not immediately have a third biopsy; instead they enter a PSA monitoring phase.
- The PSA monitoring strategy runs for 3 years so the time horizon in the model is also 3 years.
- Biopsy and its associated complications only have a short-term impact on quality of life.
- Biopsy is not linked with mortality.

5.65 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 2 groups: second biopsies recommended by the testing strategy, and biopsies carried out during PSA monitoring.

5.66 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 so the number of patients having a repeat biopsy during PSA testing is independent of the strategy chosen and is always the same.

- 5.67 The base case analysis compared the number of biopsies needed, the disutility (QALY loss) and costs of each of the diagnostic strategies. Adding the PHI or the PCA3 assay to clinical assessment produced a small increase in total biopsies compared to clinical assessment. This was accompanied by an increase in costs and a slight loss of utility compared to clinical assessment. Adding MRI to clinical assessment strategy reduced the number of biopsies needed from 1099 for clinical assessment to 520. This was a more costly strategy than clinical assessment, but resulted in a lower disutility. Strategies that included clinical assessment with MRI and either the PCA3 assay or the PHI were more costly and had only a slight effect on the number of QALYs lost.
- 5.68 The incremental analysis shows that all diagnostic strategies are dominated by clinical assessment, with the exceptions of clinical assessment plus MRI and clinical assessment plus MRI plus the PHI. The ICER for clinical assessment with MRI was £33,911 per QALY and £2,500,530 per QALY for clinical assessment with MRI and the PHI. Both strategies exceed the NICE maximum acceptable ICER of £30,000 per QALY gained.

### ***Analysis of alternative scenarios***

- 5.69 The External Assessment Group undertook a number of scenario analyses and the scenarios which changed the number of biopsies are as follows:
- Different PSA monitoring assumptions
  - Varying the derived sensitivity
- 5.70 The scenario analysis in the External Assessment Group's model tested 2 different PSA monitoring strategies. In the PSA monitoring assumptions of Mowatt et al. (2013), PSA monitoring continued for

1 year. In the monitoring strategy used in NICE's clinical guideline on diagnosing and treating [prostate cancer](#), all people with a negative or inconclusive biopsy enter a PSA monitoring phase of up to 6 years during which time they may have second or subsequent biopsies; these reflect the "least costly" (Mowatt et al. 2013) and "most costly" (PSA monitoring strategies).

- 5.71 Under the PSA monitoring assumptions of [in NICE's clinical guidance on diagnosing and treating prostate cancer](#), the only strategy with an ICER below £30,000 per QALY was the addition of the PHI to clinical assessment (£15,898). However, the diagnostics assessment report cautions that there is limited evidence that this PSA monitoring strategy is widely used.

### ***Deterministic sensitivity analyses***

- 5.72 Deterministic sensitivity analysis tested the impact of using assumptions from other data sources. Assumptions from Scattoni et al. (2013) (80% and 90% sensitivity) and Gittleman et al. (2013) (90% sensitivity) were included in these analyses. Clinical assessment together with the PCA3 assay was the only non-dominated strategy, but the ICERs per QALY were £59,732 per QALY (80% sensitivity) and £963,964 per QALY (90%) (Scattoni et al. 2013) and £105,765 per QALY (90%) (Gittelman et al. 2013). Sensitivity analysis also included increasing the rate of complications to the upper level suggested in the literature, reducing the cost of the PHI by 50%, increasing the upper level of the QALY loss from biopsy to the upper limit in the literature, assuming that 50% of cancers are missed on second biopsy and varying the proportion of patients with negative second biopsies entering PSA monitoring. The costs of biopsy complications were also increased by 100%. The ICERs were stable to these changes; strategies involving the PHI or the PCA3 assay were dominated by clinical assessment, with the exception of clinical assessment

together with MRI and the PHI. However, the ICERs for this strategy ranged from £1,213,727 to £2,500,530 per QALY gained compared to clinical assessment alone.

### ***Probabilistic sensitivity analyses***

5.73 Probabilistic sensitivity analysis was carried out using: the base-case evidence and assumptions; and the alternative evidence sources and sensitivity rates. The cost effectiveness acceptability curve for the base-case analysis shows that the most cost-effective strategy, at £20,000 per QALY gained, is clinical assessment in 100% of model iterations. At a maximum acceptable ICER of £33,500 per QALY gained, approximately half of the iterations suggest that clinical assessment is the most cost-effective strategy. The remaining iterations suggest that clinical assessment with MRI is the most cost-effective strategy. At a maximum acceptable ICER of £37,000 per QALY gained, all iterations suggest that clinical assessment with MRI dominates (that is, is less expensive and more effective than) all other strategies.

## **6 Considerations**

6.1 The Diagnostics Advisory Committee considered the evidence available on the clinical and cost effectiveness of the PROGENSA PCA3 assay and the Prostate Health Index (PHI) in people who are having investigations for suspected prostate cancer who have had a negative or inconclusive transrectal ultrasound prostate biopsy. The Committee considered the evidence on the analytical validity of the PCA3 assay and the PHI. The Committee noted that 6 studies, together with data from the company's information and the Summary of Safety and Effectiveness Data [SSED] report, provided data on the analytical validity of the tests. The precision of the PCA3 assay at its lower threshold (25) was discussed, together

with the potential impact of the reported variation. The Committee noted that both tests were CE-marked and had approval from the US Food and Drug Administration. The Committee concluded that the analytical validity of the tests had been broadly established.

6.2 The Committee considered the particular sampling needs for the PCA3 assay and the PHI. The Committee noted that a urine sample must be placed in a special collecting tube within 4 hours of collecting it from a patient for the PCA3 assay to be effective and that a blood sample needs to be centrifuged within 3 hours of collection from the patient for the PHI. The Committee heard from clinical specialists that the accuracy of the tests could be reduced if these needs were not easily met in practice, for example if the tests were carried out in a primary care setting. The Committee heard that these needs were clearly detailed in the company's information, together with warnings about specific drug interactions that could reduce test accuracy. The Committee noted that healthcare professionals should read the company information to reduce the risk of inappropriate use of the tests and consequent false test results.

6.3 The Committee considered the generalisability of the evidence on clinical validity and noted that the majority were conducted in a European setting, but that no studies were conducted in the UK. The Committee heard from clinical specialists that studies conducted in a European setting were likely to have relevance for the NHS setting, but that the 2 studies conducted in the USA may be less relevant. This was because of differences, such as the earlier and more widespread use of prostate specific antigen (PSA) testing in the USA, together with differences in population prevalence. The Committee concluded that the majority of the reviewed studies would be relevant to a UK setting.

- 6.4 The Committee considered the quality of the studies on clinical validity. They heard from clinical specialists and noted from the diagnostic assessment report that studies showed heterogeneity of study populations, lack of consistent methodologies, variations in treatment pathways, together with differences in reported outcomes. For example, the type of clinical assessment used varied between studies and sometimes varied within individual studies. The Committee heard that it was not possible to combine study results or to conduct a meta-analysis of study outcomes. Studies therefore were presented and considered individually. The Committee concluded that the evidence was limited and therefore there is uncertainty in the clinical validity of the tests.
- 6.5 The Committee considered whether the 4 diagnostic strategies identified in the diagnostics assessment report were the most clinically relevant to an NHS setting: clinical assessment plus the PCA3 assay; clinical assessment plus the PHI; clinical assessment plus MRI plus the PCA3 assay; or clinical assessment plus the PHI. The Committee heard from clinical specialists that these strategies most closely reflected clinical practice. They also heard that the recent National Prostate Cancer Audit (2014) reported that on-site multiparametric MRI is currently available in 75% of NHS hospital trusts in England. The Committee considered the availability to be higher than previously thought, but noted that the availability of the technology did not necessarily correspond to the availability of trained staff to perform prostate multiparametric MRI. However, the Committee concluded that given the rapidly changing clinical practice for diagnosing and managing prostate cancer, all 4 diagnostic strategies were clinically relevant, particularly those including MRI.
- 6.6 The Committee considered the evidence on adding the PCA3 assay to clinical assessment. The Committee heard from the

External Assessment Group that, while there were some improvements in diagnostic accuracy reported in the reviewed studies, there was wide variation both within individual studies and between studies. This was reflected in the variation across different study outcomes such as the area under the curve, multivariate odds ratios, sensitivity and specificity and decision curve analyses. The Committee also heard from clinical specialists that, with the increasing availability of MRI, diagnostic strategies that exclude MRI would become less relevant. The Committee concluded that although some improvement in diagnostic accuracy was shown in the studies, there was too much variation in the results and consequently, too much uncertainty for the impact of using the PCA3 assay in clinical practice to be clear.

6.7 The Committee considered the evidence on adding the PCA3 assay to clinical assessment with MRI. The Committee noted that the type of outcome data reported in the reviewed studies was derived from logistic regression models and so did not allow conclusions to be drawn about the most appropriate sequence of tests in the diagnostic pathway. The Committee also noted that the evidence assessing the effect of adding the PCA3 assay to clinical assessment with MRI showed no substantial improvements in diagnostic accuracy. This was reflected in the very small changes in the area under the curve, non-significant changes in odds ratios and minimal changes in derived sensitivity and specificity. The Committee heard from clinical specialists that pathways including MRI more closely reflected clinical practice and clinical decision making. The Committee concluded that adding the PCA3 assay to clinical assessment together with MRI is unlikely to substantially improve diagnostic accuracy in clinical practice.

6.8 The Committee considered the evidence on adding the PHI to clinical assessment. The Committee noted that the diagnostic

assessment report indicated the effects of adding the PHI were mixed, with only slight increases in diagnostic discrimination reported for increased area under the curve (2–10%) and multivariate odds ratios. The Committee also heard from clinical specialists that the relatively small effects reported in the reviewed studies and the variability of the results meant they were unlikely to influence current diagnostic practice. The Committee concluded that although a small improvement in diagnostic accuracy was shown in the studies, this change was too small to substantially improve diagnostic performance in clinical practice. The Committee therefore concluded that adding the PHI to clinical assessment is unlikely to substantially impact clinical practice.

6.9 The Committee considered the evidence on adding the PHI to clinical assessment with MRI and noted that adding the PHI had little or no impact on diagnostic discrimination as measured by area under the curve, odds ratios and derived sensitivity and specificity. The Committee noted the clinical view that such effects were unlikely to be clinically important. The Committee concluded that adding PHI to clinical assessment with MRI is unlikely to result in greater diagnostic certainty in clinical practice.

6.10 The Committee considered the overall results of adding the PCA3 assay or the PHI to current diagnostic pathways and heard from clinical specialists that the reported changes in diagnostic performance were neither large enough nor consistent enough to influence clinical diagnostic practice. The Committee noted that there were some improvements in diagnostic performance when the PCA3 assay or the PHI were added to clinical assessment but these were slight. In the diagnostic strategies involving MRI, adding the PCA3 assay or the PHI had no substantial impact on discrimination. The Committee concluded that the evidence did not

demonstrate an impact of using the PCA3 assay or the PHI on diagnostic performance .

- 6.11 The Committee considered the patient's perspective on being given a diagnosis of prostate cancer or possible prostate cancer. They heard from a lay member that living with, or awaiting a diagnosis of prostate cancer also affected families and carers and could give rise to much anxiety. They also heard that long-term monitoring could increase this anxiety. The Committee noted that patients valued improved diagnostic certainty, as it helped reduce anxiety and let them make better decisions. The Committee concluded that additional diagnostic information that could help distinguish between cancers that would not progress and more aggressive ones could reduce the need for long-term monitoring and would help patients make informed decisions together with their clinicians.
- 6.12 The Committee considered the patient's experience of prostate biopsy and heard that biopsy may be associated with discomfort and pain, together with side effects including bleeding, problems with catheterisation and possible infections. It also heard that patients experienced anxiety about biopsy, especially if they had previously had the procedure, and some declined further biopsies or requested general anaesthetic. The Committee noted that a reduction in the number of unnecessary prostate biopsies would be beneficial to patients, as it would avoid anxiety, discomfort and side effects. The Committee concluded that a diagnosis of prostate cancer has a substantial impact on patients and that greater diagnostic certainty and reducing the number of unnecessary biopsies offer substantial benefits to patients.
- 6.13 The Committee considered the importance of detecting more aggressive prostate cancers. It heard from clinical specialists that this was a complex issue, because of the heterogeneous nature of

prostate cancer. The Committee also heard that severity scoring schemes, such as Gleason score, can give variable results and consequently, it can be difficult to distinguish between relatively benign cancers and aggressive cancers. The Committee heard that this can lead to overtreatment and adverse effects in patients with relatively benign cancers or by contrast, undertreatment and misdiagnosis in patients with aggressive cancers. The Committee concluded that more accurately detecting aggressive prostate cancer is clinically important and could offer substantial benefits to patients.

6.14 The Committee considered the evidence in the diagnostic assessment report on the potential use of the PCA3 assay and the PHI in detecting more aggressive cancers and noted that it was inconclusive. Some studies reported a correlation between tumour aggressiveness as assessed at biopsy, and the PCA3 score or the PHI. However these reviews were not focused on the issue of repeat biopsy or the use of the test in conjunction with other diagnostic strategies. The Committee concluded that there was insufficient evidence to determine the clinical effectiveness of using the PCA3 assay or the PHI in detecting more aggressive cancers.

6.15 The Committee considered the cost-effectiveness analysis carried out by the External Assessment Group on the use of the PCA3 assay and the PHI. The Committee noted that no meta-analyses were conducted with the clinical data because of the heterogeneity between the studies, and accepted the reliance on data from a single study in the base-case analysis. The Committee also considered the assumptions and structure of the economic model and concluded that they were appropriate given the limitations in clinical evidence and the uncertainty in the values of sensitivity and specificity.

- 6.16 The Committee considered the results of the base-case analysis. They noted that the diagnostic assessment report indicated that at a set sensitivity level of 90%, all diagnostic strategies were dominated by clinical assessment (that is, clinical assessment was less expensive and more effective), apart from clinical assessment with MRI with an incremental cost-effectiveness ratio (ICER) of £33,911 per quality adjusted life year (QALY) gained, and clinical assessment with MRI and the PHI with an ICER of £2,500,530 per QALY gained. The Committee concluded that the ICERs for all diagnostic strategies involving either the PCA3 assay or the PHI were high and lay outside the range that NICE would normally consider as cost-effective. The tests were therefore unlikely to represent a cost-effective use of NHS resources.
- 6.17 The Committee considered the reported effectiveness of MRI. They heard from one of the companies that the study chosen to provide data for the cost-effectiveness model contained higher estimates of the effectiveness of MRI than other studies, and therefore may not reflect the accuracy of MRI in current practice in the UK. The Committee also heard from clinical specialists that there are varying effectiveness estimates reported for MRI and that the ongoing PROMIS study may provide a more robust estimate of effectiveness. The Committee noted the uncertainty in the precise effectiveness of MRI and concluded that the ICERs for the PCA3 assay and the PHI in combination with MRI may have been overestimated. However, the Committee noted that when the effectiveness of the 2 tests was compared to clinical assessment alone, the improvement in accuracy was small. Therefore, the Committee concluded that any change in the effectiveness of MRI was unlikely to substantially impact the cost effectiveness of the PCA3 assay or the PHI.

- 6.18 The Committee considered the uncertainty in the utility values used in the economic model. It noted that the disutility associated with potential complications of prostate biopsy, such as sepsis and patient anxiety was not included in the model. The Committee heard from clinical specialists and lay members that the model may not fully capture the impact of the disutility associated with a prostate biopsy and its potential complications. The Committee concluded that the disutility of a prostate biopsy was likely to be underestimated in the economic model.
- 6.19 The Committee considered the scenario analyses conducted by the External Assessment Group to assess the uncertainty in the cost-effectiveness analyses. It heard from the External Assessment Group that scenario analyses including varying the test sensitivity (80% or 95%) or the PSA monitoring strategy, or using sensitivity and specificity estimates from alternative data sources did not significantly impact on the ICERs, with the exception of 1 of the scenarios adopted for variation in PSA monitoring. Under the PSA monitoring assumptions of NICE's clinical guideline on diagnosing and treating [prostate cancer](#) clinical assessment with the PHI generated an ICER of £15,989 per QALY. The Committee heard from the External Assessment Group that this was because the assumptions in this guideline meant that all people with a negative or inconclusive biopsy would get PSA monitoring for up to 6 years. The Committee noted the uncertainties around current NHS monitoring practice and that it was unlikely that all men with a negative or inconclusive biopsy would be monitored for 6 years. Therefore, the Committee concluded that the cost of PSA monitoring was likely overestimated under the PSA monitoring assumptions.
- 6.20 The Committee considered the deterministic sensitivity analyses conducted by the External Assessment Group. The Committee

noted that varying the complication rate, the cost of the PHI, the QALY loss from biopsy and the costs of biopsy complications did not change the ICERs substantially. It also noted that ICERs were not substantially impacted when it was assumed that 50% of cancers were missed at second biopsy. The Committee concluded that the ICERs were robust to these changes.

6.21 The Committee considered the base case assumptions in the economic analyses and noted that various assumptions, including alternative evidence sources and sensitivity rates, were tested in probabilistic sensitivity analysis. The Committee noted that no testing strategy involving the PCA3 assay or the PHI were cost-effective in any model iteration below £30,000 per QALY gained compared to clinical assessment. Clinical assessment alone was the most cost-effective strategy in 100% of iterations below £20,000 per QALY gained. At around £33,500 per QALY gained, clinical assessment was the most cost effective strategy in approximately half the iterations and at a threshold of £37,000 per QALY gained, all iterations suggest clinical assessment with MRI dominates all other strategies.

6.22 The Committee heard that the External Assessment Group was unable to find utility data on the impact of prostate biopsy and therefore used breast biopsy utility values as a proxy. The Committee heard from clinical specialists that the 2 procedures would have very different impacts on patients and would be expected to have very different utility values. The committee heard from the specialist lay member about the anxiety and discomfort often associated with a prostate biopsy and noted that these may be substantially increased for a second biopsy. The Committee concluded that the breast biopsy utility was unlikely to fully capture the disutility associated with a prostate biopsy.

## ***Research considerations***

- 6.23 The Committee considered the value of developing research recommendations for the use of the PCA3 assay and the PHI. The Committee noted the existence of a number of ongoing trials and the rapid pace of developments in clinical practice for prostate cancer. The Committee considered the potential benefits of using the 2 tests in clinical practice and noted that the evidence showed variation in accuracy and, where improvements in accuracy were shown, the improvements were small. The Committee concluded that if the potential benefits of using the PCA3 assay and the PHI were realised, they were unlikely to be sufficiently large to offset the costs of the test and make a substantial difference to the number of people having a second biopsy unnecessarily. The Committee concluded not to recommend further research on the 2 tests.
- 6.24 The Committee also heard from clinical specialists that identifying aggressive prostate cancers and the development of individual risk-adjusted scoring systems were clinically important and research priorities. The Committee noted the views of clinical specialists and evidence from the diagnostics assessment report, which did not allow conclusions to be reached on the effectiveness of these tests in identifying more aggressive cancers. The Committee concluded that there was too much uncertainty in the evidence to determine the clinical effectiveness of these tests in distinguishing aggressive prostate cancers from relatively benign prostate cancers. The Committee also concluded that the development of diagnostics tests for distinguishing aggressive cancers is of clinical importance with potentially large patient benefits and further research is encouraged.
- 6.25 The Committee considered the lack of evidence on the specific disutility of prostate biopsy, and noted that the utility value for breast biopsy was unlikely to fully capture the disutility of prostate

biopsy and associated complications. The Committee also noted that the disutility for a second biopsy was likely to be greater than for the first biopsy because of an increased risk in complications for the second biopsy and an increase in patient anxiety associated with the experience of the first biopsy. The Committee encouraged further research on the utility of prostate biopsies.

## 7 Related NICE guidance

### Published

- [Prostate cancer: diagnosis and treatment](#). NICE clinical guideline 175 (2014)
- [Abiraterone for castration-resistant metastatic prostate cancer previously treated with a docetaxel-containing regimen](#). NICE technology appraisal 259 (2012)
- [Cabazitaxel for the second line treatment of hormone refractory, metastatic prostate cancer](#). NICE technology appraisals 255 (2012)
- [Focal therapy using high-intensity focused ultrasound \(HIFU\) for localised prostate cancer](#). NICE interventional procedure guideline 424 (2012)
- [Focal therapy using cryoablation for localised prostate cancer](#). NICE Interventional procedure guidance 423 (2012)
- [Transperineal template biopsy of the prostate](#). NICE interventional procedure guidance 364 (2010)
- Laparoscopic radical prostaectomy. NICE interventional procedure guidance 364 (2006)
- Docetaxel for the treatment of hormone-refractory metastatic prostate cancer. NICE technology appraisal 101 (2006)

### Under development

NICE is developing the following guidance (details available from [the NICE website](#)):

- Prostate cancer NICE quality standard. Publication expected June 2015
- Prostate cancer - intensity modulated radiotherapy. NICE Technology appraisal ) Publication date to be confirmed
- Prostate cancer (advanced, hormone dependent) - degarelix depot. NICE Technology appraisal. Publication expected June 2015
- Prostate cancer (hormone refractory) – atrasentan. NICE Technology appraisal. Publication date to be confirmed
- Prostate cancer (hormone relapsed, bone metastases) - radium-223 dichloride NICE Technology appraisal. Publication expected February 2015
- Prostate cancer (metastatic, hormone relapsed) - sipuleucel-T (1st line). NICE Technology appraisal. Publication expected February 2015
- Prostate cancer (metastatic, hormone relapsed, not treated with chemotherapy) - abiraterone acetate (with prednisolone). NICE Technology appraisal. Publication date to be confirmed
- Prostate cancer (metastatic, hormone-relapsed) – enzalutamide. NICE Technology appraisal Publication expected September 2015
- Prostate cancer (prevention) – dutasteride. NICE Technology appraisal. Publication date to be confirmed

## 8 Review

NICE updates the literature search at least every 3 years to ensure that relevant new evidence is identified. NICE will contact product sponsors and other stakeholders about issues that may affect the value of the diagnostic technology. NICE may review and update the guidance at any time if significant new evidence becomes available.

Adrian Newland

Chair, Diagnostics Advisory Committee

December 2014

## **9       Diagnostics Advisory Committee members and NICE project team**

### ***Diagnostics Advisory Committee***

The Diagnostics Advisory Committee is an independent Committee consisting of 22 standing members and additional specialist members. A list of the Committee members who participated in this assessment appears below.

#### **Standing Committee members**

##### **Professor Adrian Newland**

Chair, Diagnostics Advisory Committee

##### **Dr Mark Kroese**

Vice Chair, Diagnostics Advisory Committee and Consultant in Public Health Medicine, PHG Foundation and UK Genetic Testing Network

##### **Professor Ron Akehurst**

Professor in Health Economics, School of Health and Related Research, University of Sheffield

##### **Professor Paul Collinson**

Consultant Chemical Pathologist and Professor of Cardiovascular Biomarkers, St George's Hospital

##### **Dr Sue Crawford**

GP Principal, Chillington Health Centre

##### **Professor Ian A Cree**

Senior Clinical Advisor, National Institute for Health Research Evaluation Trials and Studies Coordinating Centre, University of Southampton

##### **Professor Erika Denton**

National Clinical Director for Diagnostics, NHS England, Honorary Professor of Radiology, University of East Anglia and Norfolk and Norwich University Hospitals NHS Foundation Trust

**Dr Steve Edwards**

Head of Health Technology Assessment, BMJ Evidence Centre

**Mr David Evans**

Lay Member

**Dr Simon Fleming**

Consultant in Clinical Biochemistry and Metabolic Medicine, Royal Cornwall Hospitals NHS Trust

**Mr John Hitchman**

Lay Member

**Professor Chris Hyde**

Professor of Public Health and Clinical Epidemiology, Peninsula Technology Assessment Group

**Mr Matthew Lowry**

Director of Finance and Infrastructure, Doncaster and Bassetlaw Hospitals NHS Foundation Trust

**Dr Michael Messenger**

Deputy Director and Scientific Manager National Institute for Health Research Diagnostic Evidence Cooperative, Leeds

**Dr Peter Naylor**

GP, Chair, Wirral Health Commissioning Consortium

**Dr Dermot Neely**

Consultant in Clinical Biochemistry and Metabolic Medicine, Newcastle upon Tyne Hospitals NHS Foundation Trust

**Dr Richard Nicholas**

Consultant Neurologist; Honorary Senior Lecturer, Frimley Health NHS Foundation Trust

**Dr Gail Norbury**

National Institute for Health and Care Excellence

Page 48 of 52

Diagnostics consultation document: Diagnosis and monitoring of prostate cancer: PROGENSA PCA3 assay and Prostate Health Index (PHI)

Issue date: December 2014

Consultant Clinical Scientist, Guy's Hospital

**Dr Diego Ossa**

Director of Market Access Europe, Novartis Molecular Diagnostics

**Professor Mark Sculpher**

Professor of Health Economics, Centre for Health Economics, University of York

**Dr Steve Thomas**

Consultant Vascular and Cardiac Radiologist, Sheffield Teaching Hospitals NHS Foundation Trust

**Mr Paul Weinberger**

Chief Executive Officer, DiaSolve Ltd

**Specialist Committee members**

**Mr Naeem Soomro**

Consultant Urologist, Newcastle upon Tyne Hospitals NHS Foundation Trust

**Mr Robert Mills**

Consultant Urologist, Norfolk and Norwich University Hospitals NHS Foundation Trust

**Dr Chris Parker**

Consultant Clinical Oncologist, Royal Marsden Hospital

**Mr Allan Higgin**

Lay Member

**Professor Craig Robson**

Professor of Molecular Urology, Northern Institute for Cancer Research

**Dr Iain Woodrow**

Clinical Biochemist, Mid Yorkshire Hospitals NHS Trust

**Dr Ashish Chandra**

Guy's & St.Thomas' NHS Foundation Trust

### ***NICE project team***

Each diagnostics assessment is assigned to a team consisting of a Technical Analyst (who acts as the topic lead), a Technical Adviser and a Project Manager.

#### **Bob Harbin**

Topic Lead

#### **Sarah Byron**

Technical Adviser

#### **Rob Fernley**

Project Manager

## **10 Sources of evidence considered by the Committee**

The diagnostics assessment report was prepared by Liverpool Reviews and Implementation Group.

- Nicholson A, Mahon J, Boland A, et al. The clinical and cost effectiveness of the PROGENSA PCA3 Assay and the Beckman Coulter Prostate Health Index (PHI) in the diagnosis of prostate cancer: a systematic review and economic evaluation. September 2014.

### ***Registered stakeholders***

The following organisations accepted the invitation to participate in this assessment as registered stakeholders. They were invited to attend the scoping workshop and to comment on the diagnostics assessment report.

#### **Manufacturer(s) of technologies included in the final scope:**

- Gen-Probe Lifesciences Ltd
- Beckman Coulter

#### **Professional groups and patient/carer groups:**

- Royal College of Pathologists
- Royal College of Physicians
- Royal College of Nursing
- Association for Clinical Biochemistry and Laboratory Medicine
- Prostate Cancer UK
- Prostate Cancer Support Organisation
- British In Vitro Diagnostics Association

#### **Others:**

- Department of Health
- Wirral University Teaching Hospital NHS Foundation Trust
- Healthcare Improvement Scotland

- NHS England
- Welsh Government
- NHS Cancer Screening Programmes