

# Recommendations from the EGAPP Working Group: does PCA3 testing for the diagnosis and management of prostate cancer improve patient health outcomes?

Evaluation of Genomic Applications in Practice and Prevention (EGAPP) Working Group\*

This recommendation statement is a product of the independent EGAPP Working Group. Although the Centers for Disease Control and Prevention (CDC) provides support to the EGAPP Working Group, including staff support in the preparation of this document, recommendations made by the EGAPP Working Group should not be construed as official positions of the CDC or the US Department of Health and Human Services.

**Summary of recommendations:** The Evaluation of Genomic Applications in Practice and Prevention (EGAPP) Working Group found insufficient evidence to recommend prostate cancer antigen 3 (PCA3) testing to inform decisions for when to rebiopsy previously biopsy-negative patients for prostate cancer or to inform decisions to conduct initial biopsies for prostate cancer in at-risk men (e.g., previous elevated prostate-specific antigen test or suspicious digital rectal examination).

The EGAPP Working Group found insufficient evidence to recommend PCA3 testing in men with cancer-positive biopsies to determine if the disease is indolent or aggressive in order to develop an optimal treatment plan.

Based on the available evidence, the overall certainty of clinical validity to predict the diagnosis of prostate cancer using PCA3 is deemed “low.” The EGAPP Working Group discourages clinical use for diagnosis unless further evidence supports improved clinical validity.

Based on the available evidence, the overall certainty of net health benefit is deemed “low.” The EGAPP Working Group discourages clinical use unless further evidence supports improved clinical outcomes.

**Rationale:** It has been suggested that PCA3 testing in the general male population might lead to earlier diagnosis and management changes (e.g., earlier detection and earlier initiation or higher rates of medical interventions) that improve outcomes. EGAPP Working Group found no direct evidence to support this possibility, so we sought indirect evidence aimed at documenting the extent to which PCA3 testing alters prostate cancer diagnosis or management, alone and in combination with traditional clinical management factors, and the extent to which this testing improves health outcomes.

**Analytic validity:** Assay-related evidence was deemed *adequate* for the PROGENSA PCA3 assay approved by the US Food and Drug Administration, available from Gen-Probe. Very few studies were available that investigated preanalytical effects, analytical performance, and diagnostic accuracy of other quantitative assays for PCA3.

**Clinical validity:** Evidence on clinical validity was rated *inadequate* to derive any conclusions about performance of PCA3 testing to inform decisions for when to rebiopsy previously biopsy-negative patients for prostate cancer, or to inform decisions to conduct initial biopsies for prostate cancer in at-risk men (e.g., previous elevated prostate-specific antigen test or suspicious digital rectal examination). Furthermore, there was little evidence to derive any conclusions about performance of PCA3 testing in men with cancer-positive biopsies to determine if the disease is indolent or aggressive in order to develop an optimal treatment plan.

**Clinical utility:** No studies were available to provide direct evidence on the balance of benefits and harms related to PCA3 testing for diagnosis and management in the general male population. Evidence for other populations (e.g., high risk) was not evaluated in the review.

**Contextual issues:** Early diagnosis of prostate cancer is central to minimizing morbidity and mortality. Prevention of prostate cancer mortality is a public health priority. Improvements in outcomes associated with PCA3 testing could have important impacts.

*Genet Med* advance online publication 26 September 2013

**Key Words:** diagnosis; genetic testing; management; PCA3; prostate cancer; recommendation

\*EGAPP Working Group: Chair: Ned Calonge, MD, MPH (Colorado Trust); Vice Chair: Roger D. Klein, MD, JD (Department of Molecular Pathology Cleveland Clinic Foundation); Members: Jonathan S. Berg, MD, PhD (Department of Genetics, UNC School of Medicine); Doug Campos-Outcalt, MD, MPA (Department of Family-Community Medicine, University of Arizona College of Medicine); Benjamin Djulbegovic, MD, PhD (Co-Director, Clinical & Translational Science Institute and Director, Center for Evidence-Based Medicine and Research, University of South Florida); Theodore Ganiats, MD (Department of Family and Preventive Medicine, University of California, San Diego); A. Cecile J.W. Janssens, PhD (Rollins School of Public Health, Emory University); Kenneth Offit, MD, MPH (Clinical Genetics Service, Memorial Sloan-Kettering Cancer Center); Stephen G. Pauker, MD, MACP, FACC, ABMH (Division of Clinical Decision Making, Informatics and Telemedicine, Department of Medicine, Tufts Medical Center); Margaret Piper, PhD, MPH (Blue Cross/Blue Shield Association Technology Evaluation Center); Carolyn Sue Richards, PhD, FACMG (Oregon Health & Science University); Ora L. Strickland, PhD, DSc (Hon.), RN, FAAN (Florida International University); Sean Tunis MD, MSc (Center for Medical Technology Policy); Marc S. Williams, MD (Genomic Medicine Institute, Geisinger Health System); and Doris Zallen, PhD (Department of Science and Technology in Society, Virginia Tech). Correspondence: (egappinfo@egappreviews.org)

Submitted 19 June 2013; accepted 7 August 2013; advance online publication 26 September 2013. doi:10.1038/gim.2013.141

## CLINICAL CONSIDERATIONS

### Definitions used by the Evaluation of Genomic Applications in Practice and Prevention Working Group

- Analytic validity refers to a test's ability to accurately and reliably measure the genotype or analyte of interest.
- Clinical validity defines the ability of the test to accurately and reliably identify or predict the intermediate or final outcomes of interest. This is usually reported as clinical sensitivity and specificity.
- Clinical utility defines the balance of benefits and harms associated with using the test in practice, including improvement in measurable clinical outcomes and added value in clinical management and decision making compared with not using the test.
- Credibility refers to the likelihood that an association exists after some evidence has been accumulated.

### Patient population under consideration

These recommendations apply to three clinical patient populations:

1. Rebiopsy in men with previously negative prostate biopsies
2. Initial biopsies for prostate cancer in at-risk men (e.g., previously elevated prostate-specific antigen (PSA) test or suspicious digital rectal examination (DRE)) or
3. Men with cancer-positive biopsies to determine if the disease is indolent or aggressive in order to develop an optimal treatment plan.

### Considerations for practice

Prostate cancer antigen 3 (PCA3) tests have become available directly through physicians, who should routinely consider well-established clinical procedures for diagnosis and management of prostate cancer, in addition to any genetic testing.

## BACKGROUND AND CLINICAL CONTEXT FOR THE RECOMMENDATION

Prostate cancer is the second most common cancer in men, accounting for 217,000 new cases and 32,000 deaths per year in the United States.<sup>1</sup> The development and prognosis of the disease is unpredictable. Most patients have indolent tumors and may live for years with occult disease or slowly progressive disease, ultimately dying of other causes, but others have aggressive tumors that spread beyond the prostate, resulting in significant discomfort and death.

The paramount diagnostic challenge in dealing with prostate cancer is deciding which patients to biopsy and when. The most pressing challenge in managing clinically localized disease is distinguishing between men who have aggressive disease and need aggressive therapy and men who have indolent disease and can be safely managed by active surveillance. Screening programs that use the PSA test have been in place since the late 1980s and have sparked interest and controversy. These programs are based on the premise that PSA testing can lead to early detection of prostate cancer and that effective treatments

can be initiated to improve clinical outcomes. However, testing total PSA (tPSA) levels to decide which patients should undergo a biopsy has been found to lead to high rates of false-negative and false-positive results. In men with false-negative results, cancer may be missed. Men with false-positive results may undergo unwarranted biopsies that yield negative results. These men may experience unnecessary anxiety, discomfort, and occasionally significant procedural complications such as infection or hemorrhage.

Similar problems with disease misclassification may be observed in PSA-positive patients with cancer-positive biopsies. When tPSA testing identifies patients with cancer-positive biopsies, it can lead to overtreatment (use of aggressive therapy in patients with indolent disease) or undertreatment (use of active surveillance in patients with aggressive disease). Despite the publication of thousands of articles on PSA and prostate cancer screening, the value of early intervention remains unclear.<sup>2,3</sup>

The *PCA3* gene, formerly known as *DD3*, was first identified in 1999.<sup>2</sup> *PCA3* is a non-protein-coding messenger RNA (mRNA) that is highly overexpressed in prostate cancer tissue compared with normal prostate tissue or benign prostatic hyperplasia. In 2003, the strong association between *PCA3* mRNA levels and prostate cancer led to the development of a urinary assay to measure this analyte to aid in cancer detection.<sup>3</sup> Currently, only one *PCA3* mRNA-based test has been approved by the US Food and Drug Administration (FDA). The PROGENSA *PCA3* assay is available from Gen-Probe and provides a score based on the ratio of *PCA3* to PSA mRNAs, the latter of which is used for normalization. The FDA approval letter specifies that the test's intended use is for men 50 years of age or older who have had a previous negative biopsy (and no finding of atypical small acinar proliferation) and are being considered for a repeat biopsy. However, several reference laboratories offer *PCA3* testing as laboratory-developed tests, tests developed by and used at a single laboratory testing site, with two proposed clinical utilities: (i) to inform decisions about when to biopsy or rebiopsy patients versus when to wait; and (ii) to determine in patients with cancer-positive biopsies whether the disease is indolent or aggressive so that an optimal treatment plan can be developed.<sup>4</sup>

There are no universally accepted standards about whether and when to biopsy or rebiopsy. These decisions typically depend on consideration of a variety of clinical (e.g., age, family history, race) and laboratory factors.<sup>5</sup> Recently, attention has been directed toward creating algorithms or nomograms that combine multiple clinical and laboratory features into risk scores to help in clinical decision making.<sup>5-8</sup> Nomograms are risk assessment tools that combine multiple clinical and laboratory risk factors to inform clinical decision making about biopsy, risk classification, and/or treatment options. Although there is wide variation in the manner in which the algorithms and nomograms have been developed and validated, a recent systematic review suggested that these tools tend to provide more accurate diagnostic predictions for cancer-positive biopsies than the use of PSA testing or other factors alone.<sup>6</sup>

Recent reports described the use of PCA3 testing to identify patients with aggressive versus indolent prostate cancer.<sup>9-21</sup> Results of studies have been mixed. If the PCA3 test correlates with disease prognosis, it could be a valuable tool in identifying patients who are better treated with expectant management (e.g., aggressive follow-up of the tumor without radical treatment intervention) versus those better treated with curative therapy (e.g., surgery or radiation therapy).<sup>22</sup>

The burden of prostate cancer and the efforts to properly diagnose and treat the disease are substantial. Having a tool with enhanced diagnostic specificity and/or sensitivity has the potential, at least partially, to reduce the uncertainty that plagues this decision-making process. However, use of this test without a systematic review of current evidence has the potential to create harms rather than benefits to health-care outcomes.

In summary, the upregulation (overexpression) of PCA3 mRNA expression in prostate cancer tissue provided a rationale for detecting a small number of cancer cells within the background of a large number of normal or benign prostatic hyperplasia cells.<sup>23,24</sup> Three potential intended uses for PCA3 have been proposed: (i) to inform decisions about when to biopsy at-risk patients and when to wait; (ii) to inform decisions about when to rebiopsy at-risk patients and when to wait (the claim currently approved by FDA for the PROGENSA PCA3 test); and (iii) to determine in patients with positive biopsies whether the disease is indolent or aggressive, so that an optimal treatment plan can be developed.

## Descriptions of tests and intended use claims

Several laboratories that offer testing for diagnosis and management of prostate cancer were identified. However, the Gen-Probe PROGENSA PCA3 assay is the only FDA-approved test. This test is intended and approved for use in men 50 years of age or older who have had a previous negative biopsy (and no finding of atypical small acinar proliferation) and are being considered for a repeat biopsy.

## REVIEW OF SCIENTIFIC EVIDENCE

This recommendation statement summarizes the supporting scientific evidence from a complete evidence review performed by the Agency for Healthcare Research and Quality,<sup>25</sup> which was used by the Evaluation of Genomic Applications in Practice and Prevention (EGAPP) Working Group (EWG) to support recommendations regarding the use of PCA3 testing for diagnosis and management of prostate cancer.

## Methods

EGAPP is an initiative developed by the Office of Public Health Genomics at the Centers for Disease Control and Prevention to support a rigorous, evidence-based process for evaluating genetic tests and other genomic applications that are in transition from research to clinical and public health practice in the United States.<sup>26</sup> The EWG-commissioned evidence review was contracted by the National Office of Public Health Genomics through an interagency agreement with the

Agency for Healthcare Research and Quality and conducted by the Blue Cross Blue Shield Association's Technology Evaluation Center. The Technical Expert Panel included Peter Albertsen, MD; Todd Alonzo, PhD; William Dotson, PhD; Peter Gann, MD, ScD; Roger D Klein, MD, JD; Stephen Spann, MD, MBA; and Thomas Trikalinos, MD, PhD.

This final EWG recommendation statement was formulated based on magnitude of effect, certainty of evidence, and consideration of contextual factors as outlined in the EWG methodology publication.<sup>26</sup> This publication outlines specific methods for evaluating the hierarchies of data sources and study designs for the components of evaluation, criteria for assessing quality of individual studies, and grading the quality of evidence for the individual components of the chain of evidence.<sup>26</sup>

## Technology description

In general, genotyping methods have involved *discrimination* of alleles by primer extension, hybridization, ligation, or enzymatic cleavage and *detection* using fluorescence, mass spectrometry, gel electrophoresis, or chemiluminescence. Mistaken alleles, allelic dropout (i.e., amplification of only one of two alleles in a heterozygous individual), and other genotyping errors can result from a number of causes. These include interaction with flanking DNA sequences, low quality/quantity of the DNA in samples, laboratory problems related to reagents/protocols/equipment, and human error (e.g., sample mislabeling or contamination and mistakes in data entry and interpretation). Less is known about causes of genotyping errors in newer technologies (e.g., multiplex assays, chips, and single-nucleotide polymorphism arrays) used in routine clinical practice and their potential impact on patient results. In this assay, target capture is used to isolate the mRNAs of PCA3 and the normalizing transcript PSA, which are then amplified using transcription-mediated amplification of PCA3 and detected by hybridization protection assay (complementary chemiluminescent-labeled nucleic acid probes). RNA is significantly more labile than DNA. A potential source of inaccurate results in such assays is (i) the differential degradation of mRNA of the measured analyte and the housekeeping gene or (ii) the poor quality of the mRNA templates. In addition, inaccuracies can result from suboptimal selection of the normalization control, for example, a control with variable expression levels. The PROGENSA PCA3 assay is intended to be performed on first-catch urine following a DRE. The assay should be performed by properly trained personnel, and the manufacturer's instructions should be followed.

## Analytic validity

For this recommendation, analytic validity can be defined in terms of the ability of the test to correctly identify the copy number of PCA3 mRNA or to correctly calculate PCA3 score. Because the gene that encodes PSA, *KLK3*, was not overexpressed in prostate cancer tissue,<sup>27</sup> studies commonly chose the PSA mRNA as the "housekeeping" gene against which PCA3 mRNA results were normalized. In most assays, the ratio of

PCA3 mRNA copies per milliliter to PSA mRNA copies per milliliter is multiplied by 1,000 to provide a PCA3 “score.”<sup>24,28</sup>

Sokoll et al.<sup>29</sup> reported the first multicenter study of PCA3 analytical performance in 2008 using the Gen-Probe assay and concluded that the assay performs well and is insensitive to pre-analytical factors. On 17 February 2012, Gen-Probe reported that they had received FDA approval for the PROGENSA PCA3 assay. With FDA approval, it is reasonable to anticipate that analytical validity is adequate for this particular test. PCA3 testing is also offered by reference laboratories as laboratory-developed tests, which are regulated under the Clinical Laboratory Improvement Amendments.

**Analytic validity conclusions.** There is convincing evidence that the test accurately identifies the mRNA PCA3 quantity or score in the specific populations described in the FDA approval or in laboratory-developed tests.

### Clinical validity

There are three scenarios investigated in the evidence review.<sup>25</sup> The first two relate to the use of the urine PCA3 test and other biomarker tests to predict detection of prostate tumor at biopsy or rebiopsy of at-risk men based primarily on elevated tPSA and/or suspicious DRE. The third key question concerns the use of the urine PCA3 test and other biomarker tests and pathological markers to classify the patient as being at low or high risk.

The comparative effectiveness review investigated the use of PCA3 testing in comparison with six serum biomarkers to predict risk of prostate cancer among men already identified as being at risk (KQ1 and KQ2); and in comparison with serum biomarkers, other risk factors (e.g., family history, age), and pathological tumor markers (e.g., Gleason score, staging) to identify men at high risk (i.e., aggressive) and low risk (i.e., indolent) prostate cancers. The first key question (KQ1) focused on predicting prostate cancer in men having an initial biopsy, whereas KQ2 focused on predicting prostate cancer in men with at least one previous negative biopsy. KQ3 focused on men with a positive prostate biopsy to inform decisions about management and treatment options (i.e., active surveillance versus treatment).

The review compared PCA3 testing with multiple comparators through the selection of matched studies (i.e., paired studies). These are defined as studies in which both PCA3 and the comparator marker were measured in the same individuals, in the relevant clinical setting. The outcomes of interest were both intermediate (e.g., diagnostic accuracy, decision making) and long term (morbidity/mortality related to prostate cancer).

**Initial biopsy.** Among the 17 included studies, only two reported results in populations where all men had had initial biopsies as opposed to repeat biopsies.<sup>9,30</sup> Both studies reported data on tPSA, and one reported on free PSA and PSA density.<sup>9</sup> It was not possible to evaluate consistency (similar between-study results) using only the two studies with populations of men having an initial prostate biopsy. In addition, any estimates of effect size would, necessarily, be imprecise. This resulted in

assigned grades of “insufficient” for all six comparisons of PCA3 with tPSA, free PSA, PSA velocity, PSA density, complexed PSA, and externally validated nomograms. Furthermore, strength of evidence<sup>31</sup> was considered insufficient to derive any conclusions about relative performance or about the combination of PCA3 and one or more of the comparators. This included all intermediate and long-term outcomes of interest.

**Repeat biopsy.** Among the 17 included studies, only three<sup>32-34</sup> reported results in populations where all men had had a repeat biopsy. All three studies reported on tPSA, two on free PSA,<sup>33,34</sup> and one on externally validated nomograms.<sup>32</sup> It was not possible to evaluate consistency (similar between-study results) using only the three studies with populations of men having repeat prostate biopsies. In addition, one of the studies<sup>33</sup> restricted recruitment to men with tPSA in the “grey” zone. Estimates of effect size would, necessarily, be imprecise. Therefore, the authors assigned grades of “insufficient” for all six comparisons of PCA3 with tPSA, free PSA, PSA velocity, PSA density, complexed PSA, and externally validated nomograms.

These three studies also provided results from both PCA3 testing and a comparator marker in populations of men at risk for prostate cancer who had one or more previous negative biopsies. Strength of evidence was considered insufficient to derive any conclusions about relative performance or about the combination of PCA3 and one or more of the comparators. This included all intermediate and long-term outcomes of interest.

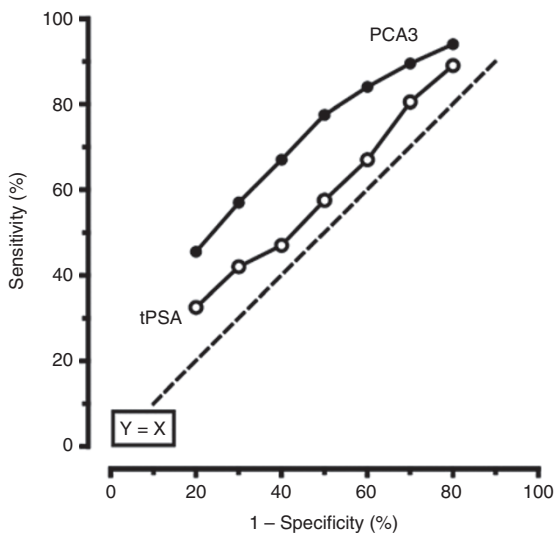
**Combined initial and repeat biopsy.** Five studies<sup>9,30,32-34</sup> exclusively studied men having an initial biopsy (KQ1) and men having repeat biopsy (KQ2). However, 12 additional studies included matched results of PCA3 and the comparators but enrolled men with both initial and repeat biopsies or they did not report biopsy history.<sup>10,11,35-44</sup> Most often, the results from these studies were not stratified by biopsy history.

Considering the inadequate strength of evidence found for the previous individual analyses of KQ1 and KQ2, the reviewers examined whether data from all 17 studies may be suitable for a combined analysis. Before performing this combined analysis, however, it is necessary to provide evidence that the biopsy status is not an important covariate that could bias the findings. The most common comparator is tPSA, and the most common analysis, by far, is the area under the receiver-operating characteristics curve (AUC). Fifteen of the 17 studies reported AUC results for both PCA3 and tPSA. Of these, 11 also provided the proportion of study subjects with no previous prostate biopsies. A regression analysis of AUC difference (PCA3 – tPSA) versus proportion of men with an initial biopsy would provide evidence regarding suitability of the combined analysis. The slope (–0.002) was nonsignificant ( $P = 0.97$ , test of slope), indicating that there was no significant relationship between the biopsy status and AUC difference. Ten studies reported the receiver operating characteristic (ROC) curves for both PCA3 and tPSA. Based on the linear regression (slope = 0.03,  $P = 0.79$ , test of slope), there again appeared to be no association

between the biopsy history and the performance of PCA3 and tPSA. Together, these two analyses provide evidence that combining results from studies of initial biopsies, repeat biopsies, and mixtures of initial and repeat biopsies will not affect the comparison of PCA3 with tPSA elevations.

On the basis of results from 15 studies, PCA3 was more discriminatory for detecting prostate cancer than level of tPSA elevations among men identified as being at risk (Figure 1). At any set clinical sensitivity, the clinical specificity of PCA3 testing is higher than that of tPSA (see Table A of the Executive Summary of the Agency for Healthcare Research and Quality Evidence Report<sup>25</sup>). Conversely, at any set clinical specificity, the clinical sensitivity of PCA3 was higher than that of tPSA (see Table B of the Executive Summary of the Agency for Healthcare Research and Quality Evidence Report<sup>25</sup>). These two biomarkers appeared to be independent in the detection of prostate cancer. The strength of evidence for diagnostic accuracy was low, mainly due to the poor quality rating of all studies and presence of verification bias.

**Clinical validity conclusions.** There is not convincing evidence that PCA3 testing in initial, repeat, or combined initial and repeat biopsy patients can be used to inform decisions about when to rebiopsy previously biopsy-negative patients for prostate cancer or to inform decisions to conduct initial biopsies for prostate cancer in at-risk men (e.g., previous elevated PSA test or suspicious DRE) primarily based on low strength of evidence.



**Figure 1** Observed consensus receiver operating characteristic (ROC) curves for prostate cancer antigen 3 gene (PCA3) scores and total prostate-specific antigen (tPSA) elevations (reprinted from the Agency for Healthcare Research and Quality Evidence Report).<sup>25</sup> The open circles (solid line) indicate the consensus observed performance of PCA3 scores, while the filled circles (solid line) indicate the matched consensus observed tPSA performance. The dashed line indicates where the sensitivity equals 1 – specificity, indicating a test with no predictive ability. For each study, the sensitivities of PCA3 and tPSA at preselected false-positive (1 – specificity) rates (x axis) were estimated from the published ROC curves; median consensus sensitivities were derived for each (1 – specificity) rate (y axis).

## Clinical utility

Unlike the situation in KQ1 and KQ2, where positive or negative prostate biopsy results were the end points for diagnostic accuracy, KQ3 focuses on intermediate and long-term outcomes. Therefore, the reference standard must also be a longer-term clinical end point in order to investigate outcomes in the context of categorization of risk. These end points might include measures of progression, metastasis, and prostate cancer-related morbidity (e.g., function, quality of life) or mortality. Progression from active surveillance to treatment appears to be a commonly used intermediate marker of overall disease progression. tPSA level is commonly used as an indicator of risk for disease recurrence, but it is not a highly sensitive and specific marker of prostate cancer. Because longer-term clinical end points are not available, a surrogate (tPSA) has to be used in this situation.

Seven prospective cohort studies of men in active surveillance are currently ongoing.<sup>45</sup> One partially informative study described in the results was derived from one of these seven studies.<sup>18</sup> Additional follow-up time will be needed for assessment of progression-free survival, mortality, and other long-term outcomes.<sup>45</sup> Given the relatively recent advent of PCA3 testing and the longer follow-up time required, it is not surprising that no studies were identified that provided intermediate or long-term outcomes based on PCA3.

**Intermediate outcomes: diagnostic accuracy.** The extent of tPSA elevations were compared with PCA3 scores to determine their diagnostic accuracy to predict prostate biopsy results (cancer/no cancer). Measures included in the analyses are the AUC, sensitivity, specificity, and positive and negative predictive values. As a reminder, only studies in which the performance estimates for both comparators were made in the same individuals were included in the five analyses listed below.

- Area under the ROC curve (AUC). Fifteen studies<sup>9-11,30,32,34-43,46</sup> reported AUC estimates for tPSA and PCA3 in the same population. Overall, 13<sup>9-11,30,32,34,35,37,38,40-42,46</sup> of the 15 studies found that the AUC of PCA3 was higher than that of tPSA.
- Reported medians and SDs. Four studies<sup>32,35,39,40</sup> provided sufficient data for analysis, and none of these reported a logarithmic SD. These were estimated from interquartile ranges or ranges. The differences, reported as z scores, indicated that two studies of smaller populations found tPSA to be slightly better than PCA3 at separating populations of positive and negative prostate biopsies, whereas two larger studies found a larger difference in separation of these groups in favor of PCA3.
- Performance at a PCA3 cutoff score of 35. Seven studies<sup>9,33,36,37,39,40,43</sup> reported the sensitivity and specificity of PCA3 at a cutoff of 35. Six<sup>9,33,36,37,40,43</sup> of the seven studies reported a higher sensitivity in PCA3.
- Sensitivity/specificity of the ROC curves. Ten studies<sup>9,30,33-37,39,40,42</sup> provided an ROC curve or data representing

an ROC curve for both markers. At a set specificity of 50%, the corresponding sensitivity for PCA3 was equal to or higher than that for PSA in all 10 studies.

- Regression analysis. Only one study provided sufficient data to apply regression coefficients and create an odds ratio between the 25th and 75th centiles of the two distributions. A second study reported all but the interquartile range, which was estimated from the first study so that both data sets could be evaluated. These two studies<sup>9,35</sup> restricted recruitment patients in the “grey zone.” In both studies, the ratio of the odds ratios (PCA3/tPSA) was greater than 1 (1.38 and 1.97), and therefore, these calculations are likely to overestimate the relative superiority of PCA3 by underestimating tPSA performance.

No studies were identified that reported matched data for PCA3 and comparator results and also reported specific clinical outcomes of patients with tumors characterized as being at low risk and high risk, who

- Opted for active surveillance and never progressed to treatment
- Opted for active surveillance and progressed to treatment or
- Opted for immediate treatment.

The strength of evidence was insufficient.

**Intermediate outcome: biopsy reduction.** PCA3 testing has the potential to reduce unnecessary biopsies, while maintaining or increasing the detection of prostate cancer. Reducing unnecessary biopsies can avoid anxiety, improve decision making, and reduce adverse events related to this invasive procedure. However, no studies were identified that reported on any other intermediate outcome measures for PCA3 and any of the six comparators. The strength of evidence for all comparators for these intermediate outcomes was insufficient.

**Intermediate outcome: categorizing positive biopsies.** Identified studies investigated PCA3 and comparator tests in categorizing men with positive prostate biopsies into high risk (or aggressive) and low risk (or indolent) cancers. Variability was observed in the terminology and definitions of high risk/aggressive and low risk/indolent disease. These end points are not clinical outcomes or legitimate surrogates for clinical outcomes, but rather risk categories, defined by the prognostic markers, which have been proposed as a clinical guide to decision making about whether to proceed with treatment or active surveillance.

Eleven studies were identified that addressed PCA3 and other preoperative and pretreatment markers for characterizing tumors based on biopsy or prostatectomy results.<sup>10–12,14–18,20,21</sup> Three studies included data on biopsy results,<sup>9,10,18</sup> seven on prostatectomy results,<sup>12,14–17,20,21</sup> and one on both.<sup>11</sup> Two studies were conducted on subjects in an active surveillance program.<sup>14,18</sup> Only two

studies had a long-term outcome component and describe a clinical outcome (e.g., prostate cancer lymph node metastases, progression to treatment as defined by high Gleason score).<sup>14,18</sup> None of the other biomarkers or pathological markers used met the criteria for validated intermediate or surrogate outcomes.

- Lymph node involvement in a prostate cancer patient is an indicator of poor clinical outcome. One study<sup>14</sup> attempted to identify “micrometastases,” based on identifying tumor cells within the lymph nodes that produce the prostate cancer markers tPSA and PCA3. The study followed 120 patients with localized prostate cancer for 4 to 6 years and used biochemical recurrence (any serum tPSA >0.2 ng/ml) as the surrogate outcome of interest. As expected, they found significantly decreased biochemical recurrence-free survival among the 11 subjects with histologically confirmed lymph node metastases, compared with 77 subjects with no lymph node involvement. Among the remaining 32 patients with biochemical recurrence, many were identified as having “micrometastases” based on either PSA or PCA3 (or both) reverse transcriptase–polymerase chain reaction testing. For those with positive tPSA, the incidence of biochemical recurrence was 73.5% (positive predictive value); the incidence of biochemical recurrence among tPSA negatives was 8.1% (negative predictive value was 92%). The sensitivity for PSA was 78% (95% CI: 61–89; 25/32); false-positive rate was 8.1% (95% CI: 5–19; 1 – 79/88). For those with positive PCA3 (DD3), the incidence of biochemical recurrence was 40.9% (positive predictive value); incidence of biochemical recurrence among PCA3 negatives was 23.5% (negative predictive value: 76.5%). The sensitivity for PCA3 was 28% (95% CI: 15–45; 9/32); false-positive rate was 15% (95% CI: 9–24; 1 – 75/88). Although this appears to indicate that PSA testing is more predictive, the use of PSA mRNA as the test and a rise in serum tPSA levels as the outcome suggests an important risk of bias.
- Based on no more than a 2-year follow-up of patients in an active surveillance program, the study reported PCA3 and tPSA results (mean, SD, median) for 38 of the 294 patients progressing to treatment based on yearly biopsy results.<sup>18</sup> Progression to treatment was recommended for “unfavorable findings,” defined as any Gleason pattern 4 or 5, >2 positive biopsy cores, or >50% involvement of any core with cancer (modified Epstein criteria). No difference in PCA3 and tPSA levels was observed between the 13% who progressed and those remaining in active surveillance ( $P = 0.13$ ). However, the authors state that only 140 of the 294 study subjects submitted a urine sample; furthermore, the authors did not report how many of these 140 men had an unfavorable result on biopsy. This study did not provide matched results for all subjects (partially matched).

Both studies were judged to be of poor quality, either because of missing follow-up to clinical end points, unclear data presentation, and/or inadequate blinding. The first study was partially

funded by Gen-Probe,<sup>18</sup> and the second<sup>14</sup> did not report on source of funding or conflicts of interest. No studies were identified that reported on other intermediate outcomes (e.g., diagnostic accuracy, decision making, harms) or long-term clinical outcomes (e.g., mortality/survival, morbidity, quality of life). The strength of evidence was insufficient.

*Intermediate outcomes: other.* No studies were identified that reported PCA3 and comparator results and intermediate outcome data (e.g., physician or patient surveys, chart review) on the degree to which PCA3 or comparator test results and categorization of risk as high or low affected decisions made with reference to selection of active surveillance versus aggressive treatment. The strength of evidence was insufficient.

*Intermediate outcomes: adverse events.* Studies have been conducted that document treatment-related clinical harms, such as incontinence, impotence, and prostatitis. On the basis of general studies on potential psychosocial harms of diagnostic testing, it is possible to generalize that patients facing treatments such as radical prostatectomy might also experience anxiety or perceive a reduction in quality of life. However, no studies were identified that reported PCA3 and comparator test results and intermediate outcome data (e.g., physician- or patient-reported adverse events, biochemical recurrence, progression to treatment) on the degree to which categorization of risk as high or low and choice of active surveillance or treatment related to the occurrence of adverse clinical events. The strength of evidence was insufficient.

*Long-term outcomes.* Data were missing or inadequate for comparison of PCA3 testing to the other selected biomarkers with reference to long-term outcomes, such as prostate cancer-related morbidity/mortality, function, and quality of life. The strength of evidence for all comparators for these long-term outcomes was insufficient.

*Clinical utility conclusions.* There is not convincing evidence that PCA3 testing offers improved intermediate or long-term outcomes.

## Contextual issues important to the recommendation

- The published literature on the use of PCA3 and comparators in the two intended uses described in KQ1 and KQ2 was found to be limited and of poor quality. However, the recent FDA approval of the Gen-Probe PCA3 test for the intended use addressed in KQ2 will raise awareness of this test and possibly accelerate its adoption into practice. Practice guidelines currently recommend that a decision to have tPSA testing should be based on discussion between the physician and patient on the balance of potential benefits and harms. An increase in the knowledge base on the comparative effectiveness of PCA3 and other biomarkers is needed to support more informed choices.

- The pros and cons of prostate cancer screening are affected by any diagnostic or demographic information that will help physicians and their patients at risk for prostate cancer to make better informed decisions about biopsy. In biopsy-positive men, the impact of additional diagnostic information on decisions regarding treatment options is of equal importance. However, to achieve potential improvement in outcomes, reliable information is needed on the diagnostic accuracy of a new test and its comparators for the outcomes of interest. Ultimately, direct or indirect evidence is needed to measure improvement in long-term health outcomes related to the use of the test and subsequent decision making.

*Cost-effectiveness.* This review did not include any economic analyses.

*Research gaps.* The EWG agrees with important gaps in knowledge presented in the evidence review,<sup>25</sup> including the following:

- Does the addition of PCA3, either alone or in combination with other markers, change prostate cancer biopsy decision making for the patient or physician? Several studies (and the evidence review) provide evidence that PCA3 may improve individualized risk prediction among men with an initial positive tPSA and/or DRE. However, no information is available on whether the clinical use of PCA3 can be effectively used to change current practice.
- What improvement in diagnostic accuracy is needed for any new test (e.g., PCA3) to provide sufficient value to affect biopsy decision making? Was there clear guidance on how much improvement in diagnostic accuracy would be required to affect clinical protocols? How can the methods required to assess and accept/reject prospective markers be streamlined? The relative importance of other factors to be considered (e.g., convenience, cost) would also be useful.
- How does PCA3 compare with the two commonly used add-on tests of free PSA and tPSA velocity/doubling time? These comparisons have been singled out because both comparators have been recommended for clinical implementation (National Comprehensive Cancer Network guidelines), but their use has generated controversy rather than bringing consensus. Special attention should be paid to the relative performance of PCA3 versus these two comparators in the context of outcomes of decision making as a way to avoid further fracturing of protocols based on limited evidence.
- Is PCA3 affected by key demographic features known to change risk for prostate cancer (ethnicity, family history)? These features were not well reported in most studies. Their impact on performance of PCA3 (in addition to some of the comparators) is unknown but may be important.
- What is the population from which the convenience samples of biopsied men have been selected? Nearly all of the

“matched” studies were convenience samples gathered by centers performing prostate biopsies. These sites should be encouraged to gather information regarding the catchment population as a way to estimate the potential for partial verification bias.

- What should the gold standard be for defining intermediate outcomes for use in establishing the clinical validity of PCA3? Studies evaluating PCA3 as selection criteria for entering a program of active surveillance have focused on how well PCA3 compares with other selection criteria (tumor volume, tumor grade, clinical stage, Epstein criteria, etc.). These intermediate measures were not well described in most studies and vary considerably among studies.
- How can PCA3 alone or when integrated with one or more comparators be used to improve decision making about whether to choose active surveillance or aggressive treatment for biopsy-positive men? No studies have yet examined the impact of PCA3 on decision making compared with existing criteria such as the Epstein Criteria. There have been no outcome studies performed to determine how well PCA3 scores predict the behavior of a particular tumor over time.

### Recommendations of other groups

At the time of publication, the EWG is not aware of other organizations providing recommendations relating to specifically addressing the use of PCA3 testing to inform decisions about when to rebiopsy previously biopsy-negative patients for prostate cancer, to inform decisions about performing initial biopsies for prostate cancer in at-risk men (e.g., previous elevated PSA test or suspicious DRE), or to determine whether the disease is indolent or aggressive in men with prostate cancer.

### REFERENCES

- Jemal A, Siegel R, Xu J, Ward E. Cancer statistics, 2010. *CA Cancer J Clin* 2010;60:277–300.
- Sutcliffe P, Hummel S, Simpson E, et al. Use of classical and novel biomarkers as prognostic risk factors for localised prostate cancer: a systematic review. *Health Technol Assess* 2009;13:iii, xi–xiii, 1–219.
- Freedland SJ. Screening, risk assessment, and the approach to therapy in patients with prostate cancer. *Cancer* 2011;117:1123–1135.
- Yin M, Bastacky S, Chandran U, Becich MJ, Dhir R. Prevalence of incidental prostate cancer in the general population: a study of healthy organ donors. *J Urol* 2008;179:892–5; discussion 895.
- National Cancer Institute. SEER Cancer Statistics Review, 1975–2005. [http://seer.cancer.gov/csr/1975\\_2005/](http://seer.cancer.gov/csr/1975_2005/) Accessed 19 June, 2013.
- Draisma G, Boer R, Otto SJ, et al. Lead times and overdiagnosis due to prostate-specific antigen screening: estimates from the European Randomized Study of Screening for Prostate Cancer. *J Natl Cancer Inst* 2003;95:868–878.
- Schröder F, Kattan MW. The comparability of models for predicting the risk of a positive prostate biopsy with prostate-specific antigen alone: a systematic review. *Eur Urol* 2008;54:274–290.
- Wolf AM, Wender RC, Etzioni RB, et al.; American Cancer Society Prostate Cancer Advisory Committee. American Cancer Society guideline for the early detection of prostate cancer: update 2010. *CA Cancer J Clin* 2010;60:70–98.
- de la Taille A, Irani J, Graefen M, et al. Clinical evaluation of the PCA3 assay in guiding initial biopsy decisions. *J Urol* 2011;185: 2119–2125.
- Cao DL, Ye DW, Zhang HL, Zhu Y, Wang YX, Yao XD. A multiplex model of combining gene-based, protein-based, and metabolite-based with positive and negative markers in urine for the early diagnosis of prostate cancer. *Prostate* 2011;71:700–710.
- Hessels D, van Gils MP, van Hooij O, et al. Predictive value of PCA3 in urinary sediments in determining clinicopathological characteristics of prostate cancer. *Prostate* 2010;70:10–16.
- Auprich M, Chun FK, Ward JF, et al. Critical assessment of preoperative urinary prostate cancer antigen 3 on the accuracy of prostate cancer staging. *Eur Urol* 2011;59:96–105.
- Durand X, Xylinas E, Radulescu C, et al. The value of urinary prostate cancer gene 3 (PCA3) scores in predicting pathological features at radical prostatectomy. *BJU Int* 2012;110:43–49.
- Kusuda Y, Miyake H, Kurahashi T, Fujisawa M. Assessment of optimal target genes for detecting micrometastases in pelvic lymph nodes in patients with prostate cancer undergoing radical prostatectomy by real-time reverse transcriptase-polymerase chain reaction. *Urol Oncol* 2011;31:615–621.
- Liss MA, Santos R, Osann K, Lau A, Ahlering TE, Ornstein DK. PCA3 molecular urine assay for prostate cancer: association with pathologic features and impact of collection protocols. *World J Urol* 2011;29:683–688.
- Nakanishi H, Groskopf J, Fritsche HA, et al. PCA3 molecular urine assay correlates with prostate cancer tumor volume: implication in selecting candidates for active surveillance. *J Urol* 2008;179:1804–1809; discussion 1809.
- Ploussard G, Durand X, Xylinas E, et al. Prostate cancer antigen 3 score accurately predicts tumour volume and might help in selecting prostate cancer patients for active surveillance. *Eur Urol* 2011;59:422–429.
- Tosoian JJ, Loeb S, Kettermann A, et al. Accuracy of PCA3 measurement in predicting short-term biopsy progression in an active surveillance program. *J Urol* 2010;183:534–538.
- van Poppel H, Haese A, Graefen M, et al. The relationship between Prostate Cancer gene 3 (PCA3) and prostate cancer significance. *BJU Int* 2012;109:360–366.
- 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;185:1234–1239.
- Whitman EJ, Groskopf J, Ali A, et al. PCA3 score before radical prostatectomy predicts extracapsular extension and tumor volume. *J Urol* 2008;180:1975–8; discussion 1978.
- National Comprehensive Cancer Network Guidelines—Prostate Cancer Early Detection, Version 1.2011. accessed 28 February 2013. <http://www.NCCN.org>.
- Durand X, Moutereau S, Xylinas E, de la Taille A. ProgenSA™ PCA3 test for prostate cancer. *Expert Rev Mol Diagn* 2011;11:137–144.
- Day JR, Jost M, Reynolds MA, Groskopf J, Rittenhouse H. PCA3: from basic molecular science to the clinical lab. *Cancer Lett* 2011;301:1–6.
- Bradley LA, Palomaki G, Gutman S, Samson DJ, Aronson N, et al. *PCA3 Testing for the Diagnosis and Management of Prostate Cancer*. Rockville, MD: Agency for Healthcare Research and Quality, 2013.
- Teutsch SM, Bradley LA, Palomaki GE, et al.; EGAPP Working Group. The Evaluation of Genomic Applications in Practice and Prevention (EGAPP) Initiative: methods of the EGAPP Working Group. *Genet Med* 2009;11:3–14.
- Moher D, Pham B, Klassen TP, et al. What contributions do languages other than English make on the results of meta-analyses? *J Clin Epidemiol* 2000;53:964–972.
- de la Taille A. ProgenSA PCA3 test for prostate cancer detection. *Expert Rev Mol Diagn* 2007;7:491–497.
- Sokoll LJ, Ellis VW, Lange P, 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.
- Rigau M, Morote J, Mir MC, et al. PSGR and PCA3 as biomarkers for the detection of prostate cancer in urine. *Prostate* 2010;70:1760–1767.
- Owens DK, Lohr KN, Atkins D, et al. AHRQ series paper 5: grading the strength of a body of evidence when comparing medical interventions—agency for healthcare research and quality and the effective health-care program. *J Clin Epidemiol* 2010;63:513–523.
- Ankerst DP, Groskopf J, Day JR, et al. Predicting prostate cancer risk through incorporation of prostate cancer gene 3. *J Urol* 2008;180:1303–1308; discussion 1308.
- Aubin SM, Reid J, Sarno MJ, 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;184:1947–1952.
- Ploussard G, Haese A, Van Poppel H, 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;106:1143–1147.



35. Perdonà S, Cavadas V, Di Lorenzo G, 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;59:81–87.
36. Adam A, Engelbrecht MJ, Bornman MS, Manda SO, Moshokoa E, Feilat RA. The role of the PCA3 assay in predicting prostate biopsy outcome in a South African setting. *BJU Int* 2011;108:1728–1733.
37. Deras IL, Aubin SM, Blase A, et al. PCA3: a molecular urine assay for predicting prostate biopsy outcome. *J Urol* 2008;179:1587–1592.
38. Mearini E, Antognelli C, Del Buono C, et al. The combination of urine DD3(PCA3) mRNA and PSA mRNA as molecular markers of prostate cancer. *Biomarkers* 2009;14:235–243.
39. Nyberg M, Ulmert D, Lindgren A, Lindström U, Abrahamsson PA, Bjartell A. PCA3 as a diagnostic marker for prostate cancer: a validation study on a Swedish patient population. *Scand J Urol Nephrol* 2010;44:378–383.
40. Ochiai A, Okihara K, Kamoi K, et al. Prostate cancer gene 3 urine assay for prostate cancer in Japanese men undergoing prostate biopsy. *Int J Urol* 2011;18:200–205.
41. Ouyang B, Bracken B, Burke B, Chung E, Liang J, Ho SM. A duplex quantitative polymerase chain reaction assay based on quantification of alpha-methylacyl-CoA racemase transcripts and prostate cancer antigen 3 in urine sediments improved diagnostic accuracy for prostate cancer. *J Urol* 2009;181:2508–2513; discussion 2513.
42. 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;85:159–165.
43. Wang R, Chinnaiyan AM, Dunn RL, Wojno KJ, Wei JT. Rational approach to implementation of prostate cancer antigen 3 into clinical care. *Cancer* 2009;115:3879–3886.
44. Roobol MJ, Schröder FH, van Leenders GL, et al. Performance of prostate cancer antigen 3 (PCA3) and prostate-specific antigen in Prescreened men: reproducibility and detection characteristics for prostate cancer patients with high PCA3 scores (= 100). *Eur Urol* 2010;58:893–899.
45. Cooperberg MR, Carroll PR, Klotz L. Active surveillance for prostate cancer: progress and promise. *J Clin Oncol* 2011;29:3669–3676.
46. Roobol MJ, Schröder FH, van Leeuwen P, 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;58:475–481.