Draft Comparative Effectiveness Review

Number xx

PCA3 Testing for the Diagnosis and Management of Prostate Cancer

Prepared for:
Agency for Healthcare Research and Quality
U.S. Department of Health and Human Services
540 Gaither Road
Rockville, MD 20850
www.ahrq.gov

Contract No. xxx-xx-xxxx

Prepared by:

Investigators:
First and Last Names, X.X.
First and Last Names, X.X.

AHRQ Publication No. xx-EHCxxx
<Month Year>

This report is based on research conducted by an Evidence-based Practice Center under contract to the Agency for Healthcare Research and Quality (AHRQ), Rockville, MD (Contract No. xxx-xxxx-xxxxx-I). The findings and conclusions in this document are those of the author(s), who are responsible for its content, and do not necessarily represent the views of AHRQ. No statement in this report should be construed as an official position of AHRQ or of the U.S. Department of Health and Human Services.

The information in this report is intended to help clinicians, employers, policymakers, and others make informed decisions about the provision of health care services. This report is intended as a reference and not as a substitute for clinical judgment.

This report may be used, in whole or in part, as the basis for the development of clinical practice guidelines and other quality enhancement tools, or as a basis for reimbursement and coverage policies. AHRQ or U.S. Department of Health and Human Services endorsement of such derivative products or actions may not be stated or implied.

This document is in the public domain and may be used and reprinted without special permission. Citation of the source is appreciated.

None of the investigators have any affiliations or financial involvement that conflicts with the material presented in this report.

Suggested citation: Author I, Author I. Comparative Effectiveness of <Title>. Comparative Effectiveness Review No. xx. (Prepared by the <Name> Evidence-based Practice Center under Contract No. xxx-xx-xxxx.) AHRQ Publication No. xx-EHCxxx. Rockville, MD: Agency for Healthcare Research and Quality. <Month Year>. Available at: www.effectivehealthcare.ahrq.gov/reports/final.cfm.

Preface

The Agency for Healthcare Research and Quality (AHRQ) conducts the Effective Health Care Program as part of its mission to organize knowledge and make it available to inform decisions about health care. As part of the Medicare Prescription Drug, Improvement, and Modernization Act of 2003, Congress directed AHRQ to conduct and support research on the comparative outcomes, clinical effectiveness, and appropriateness of pharmaceuticals, devices, and health care services to meet the needs of Medicare, Medicaid, and the Children's Health Insurance Program (CHIP).

AHRQ has an established network of Evidence-based Practice Centers (EPCs) that produce Evidence Reports/Technology Assessments to assist public- and private-sector organizations in their efforts to improve the quality of health care. The EPCs now lend their expertise to the Effective Health Care Program by conducting comparative effectiveness reviews (CERs) of medications, devices, and other relevant interventions, including strategies for how these items and services can best be organized, managed, and delivered.

Systematic reviews are the building blocks underlying evidence-based practice; they focus attention on the strength and limits of evidence from research studies about the effectiveness and safety of a clinical intervention. In the context of developing recommendations for practice, systematic reviews are useful because they define the strengths and limits of the evidence, clarifying whether assertions about the value of the intervention are based on strong evidence from clinical studies. For more information about systematic reviews, see http://www.effectivehealthcare.ahrq.gov/reference/purpose.cfm

AHRQ expects that CERs will be helpful to health plans, providers, purchasers, government programs, and the health care system as a whole. In addition, AHRQ is committed to presenting information in different formats so that consumers who make decisions about their own and their family's health can benefit from the evidence.

Transparency and stakeholder input from are essential to the Effective Health Care Program. Please visit the Web site (http://www.effectivehealthcare.ahrq.gov) to see draft research questions and reports or to join an e-mail list to learn about new program products and opportunities for input. Comparative Effectiveness Reviews will be updated regularly.

We welcome comments on this CER. They may be sent by mail to the Task Order Officer named below at: Agency for Healthcare Research and Quality, 540 Gaither Road, Rockville, MD 20850, or by e-mail to epc@ahrq.hhs.gov.

Carolyn M. Clancy, M.D.
Director
Agency for Healthcare Research and Quality

Jean Slutsky, P.A., M.S.P.H.
Director, Center for Outcomes and Evidence
Agency for Healthcare Research and Quality

Stephanie Chang, M.D., M.P.H.
Director
Evidence-based Practice Program
Center for Outcomes and Evidence
Agency for Healthcare Research and Quality

Supriya Janakiraman, M.D., M.P.H.
Task Order Officer
Center for Outcomes and Evidence
Agency for Healthcare Research and Quality

Acknowledgments

The authors gratefully acknowledge the following individuals for their contributions to this project: <Acknowledgments>.

Key Informants

<Name>
<Place>
<City>, <ST>

Technical Expert Panel

<Name>
<Place>
<City>, <ST>

<Name>
<Place>
<City>, <ST>

Peer Reviewers

<Name>
<Place>
<City>, <ST>

<Name>
<Place>
<City>, <ST>

PCA3 Testing for the Diagnosis and Management of Prostate Cancer

Structured Abstract

Objectives: Perform a comparative effectiveness review examining the use of PCA3 in improving biopsy or rebiopsy decisions in patients identified at risk for prostate cancer, or in improving decisionmaking about treatment choices (e.g., active surveillance versus aggressive therapy) in patients with prostate cancer positive biopsies. Comparators include total PSA elevations, free PSA, PSA density, PSA velocity, externally validated nomograms, complexed PSA and multivariate models.

Data Sources: MEDLINE® was searched from January 1, 1990 to August 8, 2011, EMBASE® was searched from January 1, 1990 to August 15, 2011, and, the Cochrane Database of Systematic Reviews was searched with no date restriction. A grey literature search included databases with regulatory information, clinical trial registries, abstracts and conference papers, grants and federally funded research, and information from manufacturers.

Review Methods: Inclusion criteria required PCA3 and at least one comparator to be measured in the same cohort in one of the three clinical settings: at-risk men considering first biopsy; at-risk men considering a repeat biopsy; and men with prostate cancer making treatment decisions. Data were extracted by one reviewer and audited by a second. Analyses were matched by comparing with-in study differences. Modeling was used to smooth consensus ROC curves and to associate cut-off levels with test performance. Diagnostic accuracy studies were assessed for quality using QUADAS. Strengths of evidence were judged insufficient, low, moderate, and high according to the AHRQ *Test Review Guide*.

Results: After further exclusion of six studies with strong likelihood of duplicate data, 17 studies provided data that could be used to address diagnostic accuracy (Key Question [KQ] 1 and 2); 11 studies addressed treatment decisionmaking (KQ3). All studies were of poor quality. Comparison of PCA3 to total PSA (tPSA) had the most available studies (17) but was subject to both partial verification and sampling biases. Both were accounted for in the analyses. We observed that PCA3 is more discriminatory for detecting cancers (i.e., at any sensitivity, the specificity is higher, or at any specificity, the sensitivity is higher) than tPSA elevations, but the strength of evidence is low. One subanalysis indicated that this finding holds for both initial and repeat biopsies. For all other diagnostic accuracy comparisons, and for all intermediate and long-term health outcomes, strength of evidence was insufficient. For treatment decisionmaking, in all comparisons for intermediate and long-term health outcomes, the strength of evidence was found to be insufficient.

Conclusions: For diagnostic accuracy, there was a low strength of evidence that PCA3 had better diagnostic accuracy than tPSA elevations, but insufficient evidence that this led to improved intermediate or long-term health outcomes. For all other settings, comparators and outcomes, there was insufficient evidence.

Contents

Introduction **1**
Background 1
Burden of Illness 1
Screening for Prostate Cancer 1
Diagnosis and Management of Prostate Cancer 2
Development of a New Biomarker: PCA3 3
Selected PCA3 Comparators 4
Total PSA (tPSA) 4
Free PSA (fPSA) 4
PSA Density 5
PSA Velocity and Doubling Time 5
Complexed PSA 5
Externally Validated Nomograms 6
Criteria for Distinguishing Low-Risk/Indolent and High-Risk/Aggressive Prostate Cancer 6
Key Questions 6
Analytic Frameworks 7
PICOTS Framework 10
Methods **13**
Topic Development and Refinement 13
Literature Search Strategy 13
Study Selection 14
Search Strategies for Grey Literature 15
Data Extraction and Management 15
Data Elements 16
Evidence Tables 16
Individual Study Quality Assessment 16
Definition of Ratings for Individual Studies and Reviews 16
Measuring Outcomes of Interest 17
Data Synthesis 18
Grading the Body of Evidence for Each Key Question 19
Assessment of Applicability 19
Peer Review and Public Commentary 19
Results **20**
Literature Search 20
Grey Literature Search 27
Description of Included Studies for KQ1/KQ2 “Combined” 31

Evaluation of PCA3 and Other Comparators to Identify Intermediate and Long-Term Outcomes Related to Prostate Cancer 32
Comparator: Total Serum PSA 32
Study Characteristics 35
PCA3 and tPSA: Area Under the Curve 36
PCA3 and tPSA: Reported Medians and Standard Deviations 40
PCA3 and tPSA: Performance at a PCA3 Cut-Off Score of 35 42
PCA3 and tPSA: ROC Curves - Sensitivity / Specificity 43
PCA3 and tPSA: Regression Analysis 46
Key Question 3 – Testing for PCA3 and Comparators to Identify Patients with Low Risk/Indolent Cancer Who May Be Candidates for Active Surveillance 51
Description of Included Studies 51
Discussion 57
Key Findings and Strength of Evidence 57
Key Question 1: Testing PCA3 and Comparators to Identify Prostate Cancer in Men Having an Initial Biopsy 57
Key Question 2: Testing PCA3 and Comparators to Identify Prostate Cancer in Men Having a Repeat Biopsy 57
Potential to Combine Studies Addressing Key Questions 1 and 2 57
Key Question 3: Testing PCA3 and Comparators Among Men with a Positive Prostate Biopsy to Identify Low Risk/Indolent Patients Who May Be Candidates for Active Surveillance 58
Findings in Relationship to What is Already Known 59
Systematic Review of PCA3 59
Applicability 59
Key Questions 1 and 2: Biopsy Decisionmaking 59
Population and Settings of Care 59
Interventions 60
Comparisons 60
Outcomes 61
Timeframes 61
Key Question 3: Management of Biopsy Positive Patients 61
Implications for Clinical and Policy Decisionmaking 62
Limitations of the Comparative Effectiveness Review Process 62
Limitations of the Evidence Base 63
Key Questions 1 and 2 63
Key Question 3 63

Research Gaps 64
Overview 64
Gaps in Knowledge for Key Questions 1 and 2 64
Gaps in Knowledge for KQ3 68
References 70
Abbreviations and Acronyms 76

Tables

Table 1. General descriptions of the included studies in matched populations 21
Table 2. Reported criteria[a] for inclusion of subjects into studies and specific exclusion criteria . 22
Table 3. Characteristics of populations studied 24
Table 4. PCA3 / DD3 assay characteristics 25
Table 5. Data available for the comparison of PCA3 scores and the five comparators, when analysis is restricted to studies reporting results only in men having initial biopsies 27
Table 6. Data available for the comparison of PCA3 scores and the five comparators, when analysis is restricted to studies reporting results only in men having a repeat biopsy 28
Table 7. Data available for the comparison of PCA3 scores and the five comparators, when analysis is restricted to studies reporting results with/without a previous prostate biopsy 29
Table 8. A comparison of PCA3 and tPSA elevations to identify men with prostate cancer with either the false positive rate, or the sensitivity held constant 34
Table 9. Summary results for the five analytic comparisons of PCA3 versus tPSA in matched populations of men having prostate biopsies 35
Table 10. Comparing PCA3 levels and extent of tPSA elevations in matched studies via AUC analysis to correctly diagnose prostate cancer, as defined by a positive needle biopsy 37
Table 11. Comparison of PCA3 and tPSA differences in central estimates in men with positive and negative prostate biopsy results, after accounting for study-specific variability in measurements 40
Table 12. A comparison of PCA3 and tPSA in identifying a positive prostate biopsy among matched studies: difference in sensitivities at the fixed specific associated with the commonly used PCA3 score cut-off of 35 42
Table 13. Matched studies of PCA3 and tPSA testing performance to identify positive biopsy men 44
Table 14. Comparison of modeled univariate continuous odds ratios (OR) for PCA3 and tPSA in matched studies 47
Table 15. Measures of independence of PCA3 and tPSA in identifying men having a positive biopsy finding in matched studies 47
Table 16. Key Questions 1 and 2 “Combined” Analyses: Summary Studies Available and Strength of Evidence for the Six PCA3 Comparators versus the four Outcomes of Interest 50
Table 17. Characteristics of biopsy studies 54
Table 18. Characteristics of prostatectomy only studies 55
Table 19. Comparators investigated along with PCA3 levels in studies addressing KQ3 56

Figures

Figure 1. Analytic framework for PCA3 as a diagnostic indicator for biopsy or rebiopsy in patients with elevated PSA and/or abnormal digital rectal examination (Key Questions 1 and 2).9
Figure 2. Analytic framework for PCA3 used to distinguish indolent versus aggressive prostate cancer (KQ3)....9
Figure 3. PRISMA diagram for identified published studies....20
Figure 4. PRISMA diagram for identified grey literature....26
Figure 5. A scatterplot showing the relationship between the proportion of men with an initial biopsy versus the difference in the area under the curves (AUC) for PCA3 scores and tPSA elevations....30
Figure 6. A scatterplot showing the relationship between the proportion of men with an initial biopsy versus the difference in sensitivity at a specificity of 50%....31
Figure 7. Examining the relationship between effect size (AUC difference) and an estimate of precision....38
Figure 8. The relationship between AUC difference and average AUC for PCA3 and tPSA....39
Figure 9. Distributions of PCA3 and tPSA in men with positive and negative prostate biopsies from four studies that reported such information in the same cohort....41
Figure 10. The performance of PCA3 score to identify subsequent positive prostate cancer biopsies from seven studies....43
Figure 11. Summary ROC curves for PCA3 and tPSA based on studies applying the tests to the same population....45

Appendixes

Appendix A. Search Strategies
Appendix B. Data Elements
Appendix C. Search Strategy for the Grey Literature
Appendix D. Articles Excluded at Full-Text Level
Appendix E. Quality Assessment Criteria and Category Definitions for Nonrandomized Comparative Intervention Studies
Appendix F. Quality Assessment of Studies Addressing KQ1 / KQ2 and KQ3
Appendix G. DistillerSR Abstract and Title Screening Form
Appendix H. DistillerSR Full-Text Article Screening Form
Appendix I. DistillerSR Data Extraction Forms
Appendix J. Detailed Description of PCA3 and Total PSA Interpretive Analysis
Appendix K. Summary of the Remaining Key Questions 1 & 2: Combined Analyses

Executive Summary

Background

Cancer of the prostate is the second most common cancer and the second leading cause of cancer deaths in men in the U.S.[1] Most patients have indolent tumors, and may live for years with no or minimal effects, ultimately dying of other causes.[2] The lifetime risk of being diagnosed with prostate cancer is 16 percent, but lifetime risk of dying from the disease is only 3 percent.[3] However, some patients have aggressive tumors that spread beyond the prostate, resulting in significant morbidity and death.

The rationale for prostate cancer screening using serum total prostate specific antigen (tPSA) levels[4, 5] was early detection of prostate tumors leading to timely intervention and reduced prevalence of disease. However, consensus on who to test, when, and the most effective clinical threshold has not been reached. A recent evidence review found no reduction in prostate cancer-specific mortality based on tPSA screening, and the test's low specificity has led to harms (e.g., false positives, unnecessary biopsy, overdiagnosis/treatment of low risk cancers). [6-8]

In 1999, researchers reported that the prostate cancer antigen 3 gene (*PCA3*; also known as *DD3*), was highly overexpressed in prostate cancer relative to normal prostate or benign prostatic hyperplasia tissue.[9] Subsequently, noninvasive PCA3 tests on messenger RNA from urine were developed. Two proposed intended uses of PCA3 and comparator tests were to inform decisionmaking about biopsy or rebiopsy of men with elevated tPSA and/or other risk factors. The third was to inform decisions about treatment (e.g., active surveillance, prostatectomy, radiotherapy) by classifying disease in men with positive biopsies as low risk (indolent) or high risk (aggressive). The U.S. Food and Drug Administration recently approved a PCA3 assay for use in men with one or more previous negative biopsies to inform decisionmaking about repeat biopsy.

Scope and Key Questions

For prostate cancer detection, biomarker comparators are tPSA, free PSA, PSA density, PSA velocity or doubling time, complexed PSA and externally validated risk-assessment calculators or nomograms. For risk classification, PCA3 comparators in a prognostic workup include Gleason score, tumor volume, risk factors, biochemical markers and clinical/pathological staging. The Key Questions (KQ) relate to the three proposed scenarios described above:

1. In patients with elevated PSA and/or an abnormal DRE who are candidates for initial prostate biopsy, what is the comparative effectiveness of PCA3 testing as a replacement for, or supplement to, standard tests, including diagnostic accuracy (clinical validity) for prostate cancer, intermediate outcomes (e.g., improved decision making about biopsy), and long-term health outcomes (clinical utility), including mortality/morbidity, quality of life and potential harms?
2. In patients with elevated PSA and/or an abnormal DRE who are candidates for repeat prostate biopsy, what is the comparative effectiveness of PCA3 testing as a replacement for, or supplement to, standard tests, including diagnostic accuracy (clinical validity) for prostate cancer, intermediate outcomes (e.g., improved decisionmaking about biopsy), and long-term health outcomes (clinical utility), including mortality/morbidity, quality of life and potential harms?

3. In patients with a positive biopsy for prostate cancer who are being evaluated to distinguish between indolent and aggressive disease, what is the effectiveness of using PCA3 testing alone, or in combination with the standard prognostic workup (e.g., tumor volume, Gleason score, clinical staging) or monitoring tests (e.g., PSA, PSA velocity), with regard to diagnostic accuracy (clinical validity) for aggressive (high risk) prostate cancer, intermediate outcomes (e.g., improved decisionmaking about prognosis and triage for active surveillance and/or aggressive treatment) and long-term health outcomes (clinical utility), including mortality/morbidity, quality of life, and potential harms?

Corresponding analytic frameworks are presented in the full report.

Methods

Literature Search Strategy

Searches were run on MEDLINE®, EMBASE® and the Cochrane Central Register of Controlled Trials for the time frame January 1, 1990 to August 15, 2011. The grey literature searches included regulatory information, clinical trial registries, conference papers, and selected web sites. Included studies were in English, reported primary data, addressed key questions and fulfilled the criteria for: 1) study design (matched study); 2) study subjects/populations (at risk men or men with a positive biopsy); 3) study interventions (biomarker testing, biopsy, risk classification); 4) study comparators; 5) intermediate (diagnostic accuracy, impact on decisionmaking, harms) and long-term (e.g., mortality, morbidity, function, quality of life, harms) outcomes.

Title/abstract and full-article review involved one reviewer to read and determine eligibility and a second reviewer to audit a subset of abstracts (and all marked uncertain) and all full articles; discrepancies were resolved by discussion or a third reviewer when needed. Data were extracted by a single reviewer, then fully audited by a second senior reviewer. Disagreements were resolved through review team discussion.

Quality Assessment of Individual Studies

In adherence with the Methods Guide,[10] grading the methodological quality of individual comparative studies was performed based on study design-specific criteria.[11, 12] The quality of diagnostic accuracy studies was assessed using the Quality Assessment of Diagnostic Accuracy Studies (QUADAS) tool.[13] The QUADAS ratings were summarized into general quality classes of good, fair, and poor.[14] In all cases, quality of individual studies and the overall body of evidence was assessed by two independent reviewers. Discordant decisions were resolved through discussion or third-party adjudication.

Data Synthesis

For KQ1 and KQ2, PCA3 scores were compared to all comparators for which published data were available. Analyses included sensitivity, specificity (or the false positive rate equal to 1-specificity), and positive and negative predictive values. All comparators were studied when possible using one or more of the following approaches for PCA3 and one or more comparators:
1) Differences in area under the curve (AUC), including direction and magnitude of differences.
2) Reported medians and standard deviations in positive and negative biopsy populations

(reported as z scores), including direction and strength of effect. 3) Performance at a PCA3 cut-off score of 35 (sensitivity and specificity). 4) Receiver Operator Characteristic (ROC) Curves (sensitivity and specificity), to evaluate fixed specificities and compare corresponding sensitivities. 5) Regression analysis using regression coefficients and associated relative odds ratios. Based on the limited studies addressing KQ3, we anticipate a qualitative analysis.

Applicability

Applicability of the results presented in this review was assessed in a systematic manner using the PICOT framework (Population, Intervention, Comparison, Outcome, Timing).

Results

Detailed description of analyses with tables and figures are included in the full report.

Results of Literature Searches

Inclusion criteria restrict the analyses to matched studies that provide data on PCA3 and at least one other comparator in the same patient population. Population matching was preserved by computing differences between PCA3 test results and comparator test results within each study, and comparing these differences across studies. Searches identified 1,449 citations, of which 163 underwent full text review and 34 were included. Grey literature search identified one additional study.

Key Question 1: Initial biopsy

Two matched studies addressed diagnostic accuracy for KQ1[15, 16] with comparisons of PCA3 to tPSA[15, 16], fPSA[15] and PSA density.[16] Studies differed in enrollment criteria, and strength of evidence was rated Insufficient. No studies addressed other comparators or outcomes; all other comparisons were graded Insufficient.

Key Question 2: Repeat Biopsy

Three matched studies addressed diagnostic accuracy for KQ2[17-19], with comparisons of PCA3 to tPSA[17-19], fPSA[18, 19] and externally validated nomograms[17]. With only three studies, strength of evidence was rated Insufficient. No studies addressed other comparators or outcomes; all other comparisons were graded Insufficient.

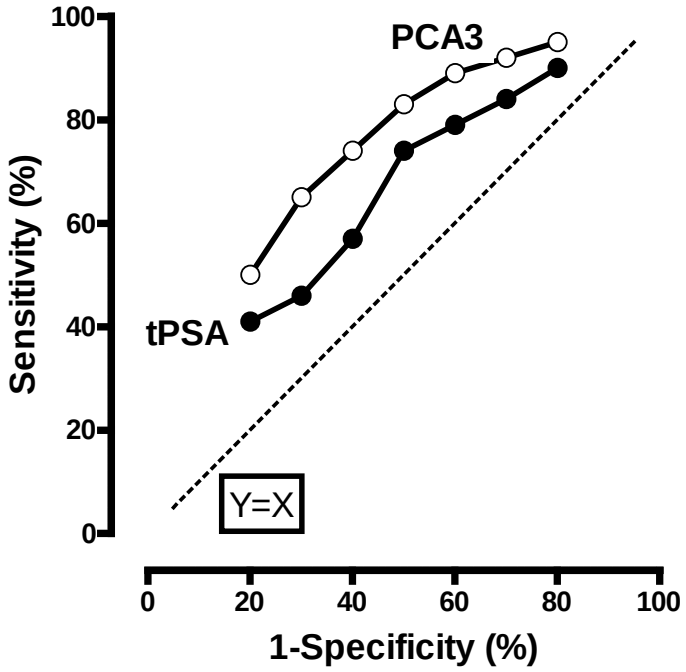
Key Questions 1 and 2: Initial and Repeat Biopsies

Only five studies qualified for KQ1 or KQ2.[15-19] Another 12 provided suitable data, but their populations were a mixture of initial and repeat biopsies.[20-31] Given the lack of specific data for these Key Questions, we examined all 17 studies to determine suitability for a combined analysis (Table ES-2). Two key findings emerged for PCA3 with tPSA. Eleven studies reported AUC and the proportion of men having an initial biopsy; no relationship was found (linear regression, p=0.97, test of slope). Ten studies provided ROC curves for both PCA3 and tPSA. The sensitivity, at a specificity of 50 percent, was not related to the proportion of men with an initial biopsy (p=0.79). No similar analyses can be made for any other comparator for diagnostic accuracy. This was considered sufficient to proceed with a combined analysis for KQ1/KQ2, without the biopsy history restriction.

Total PSA (tPSA) elevations and diagnostic accuracy for combined KQ1 / KQ2

Subsets of 17 studies provided sufficient data to compare the diagnostic accuracy of PCA3 with tPSA elevations, using the five described analyses (Table ES2). Two important biases were identified. Modeling was used to account for the potential impact of verification bias. A sampling bias was accounted for by stratification. Figure ES 1 shows the consensus observed ROC curves for PCA3 and tPSA. Table ES-1 shows the modeled ability of PCA3 and tPSA to identify prostate cancer among men at increased risk. Both Figure ES-1 and Table ES-1 indicate that PCA3 is associated with higher sensitivity at any given specificity than tPSA elevations, and higher specificity at any given sensitivity. This combined approach made it possible to reliably compare PCA3 and tPSA measurements for diagnostic accuracy. Quality of all individual studies was poor. For all other comparators (Table ES 2), and all other outcomes, studies were too few or not available, and strengths of evidence for all were considered insufficient.

Figure ES 1. Observed Consensus Receiver Operator Characteristic (ROC) curves for PCA3 scores and tPSA elevations



The open circles (solid line) indicates the consensus observed performance of PCA3 scores, while the filled circles (solid line) indicated the matched consensus observed tPSA performance. The dashed line indicated where the sensitivity equals 1-specificity, indicating a test with no predictive ability.

Table ES 1. A comparison of PCA3 and tPSA measurements to identify men with prostate cancer with either the false-positive rate or the sensitivity held constant

PCA3 Scores				tPSA elevations	
1-Specificity (FPR)		Sensitivity (DR)		Sensitivity (DR)	Improvement in PCa Detection
80%		95.8%		90.7%	5.1%
70%		92.0%		84.3%	7.7%
60%		87.2%		77.6%	9.6%
50%		81.1%		68.8%	14.0%
40%		73.7%		59.7%	14.0%
30%		63.8%		48.4%	15.4%
20%		51.6%		36.4%	15.2%
PCA3 Scores				tPSA elevations	
Sensitivity (DR)		1-Specificity (FPR)		1-Specificity (FPR)	Reduction in Biopsies
95%		77.7%		88.2%	10.5%
90%		65.6%		78.2%	12.6%
85%		56.3%		71.5%	15.2%
80%		48.5%		63.6%	15.1%
70%		36.2%		51.2%	15.0%
60%		26.6%		40.3%	13.7%
50%		18.9%		31.2%	12.3%

PCA3 and comparators for other intermediate and long term outcomes

Key Question 3: Testing PCA3 and comparators to identify men with low risk/indolent cancer who may benefit from active surveillance

One[32] of 11 studies[15, 21, 23, 32-39] reported[15, 21, 23, 32-39] on four to six year followup of the outcome biochemical recurrence, based on identification of micrometastases through measurement of tPSA and PCA3 in lymph node extracts. Another[37] reported two year followup of progression from active surveillance to treatment in men with prostate cancer, based on results of yearly biopsy. Quality of all studies was poor. Strength of evidence was insufficient due to the inability to compare or combine the two studies on different sample types, using different parameters (sensitivity / specificity, mean/median biomarker levels) and addressing different outcomes. No studies were identified for any other comparators and any other outcomes, and strength of evidence was insufficient.

Discussion

Key Findings and Strength of Evidence

Strength of evidence was insufficient for KQ3, and for all comparators and outcomes for KQ1 / KQ2 except the comparison of PCA3 and tPSA for the outcome of diagnostic accuracy (Figure ES 1, Table ES 1, Table ES 2). Among men at risk, PCA3 was more discriminatory for detecting prostate cancer than tPSA elevations. The finding that the relative performance of PCA3 versus tPSA elevations is not dependent on biopsy history is a new observation that could impact future studies. The quality of all studies was poor. The strength of evidence was considered low. An important consideration in this conclusion was the indirectness of evidence for identifying positive biopsy status. We made the underlying assumption that not all positive

biopsies are equal. For example, identifying a positive biopsy associated with a high Gleason score may be considered to be clinically more valuable than one with a low Gleason. None of the included studies provided a two-way cross tabulation of PCA3 and tPSA positive and negative test results among biopsy positive patients. Of most interest would be the clinical finding for the cases in the off-diagonal (when one test is positive and the other negative).

For KQ3, the literature review revealed few relevant matched studies and a lack of clinical followup after patients were placed into risk categories (low risk/indolent, high risk/aggressive) that were defined by the results of PCA3 and other biomarker and pathological tests. In nine of 11 studies, a reference clinical endpoint (or validated surrogate) was lacking. The quality of all studies was low. Strength of evidence was Insufficient. It is likely that more time will be needed for studies to assess the diagnostic accuracy of predicting long-term outcomes for patients based on categorization as having low risk or high risk disease.[40]

Applicability

The populations studied in the included articles were largely drawn from academic medical centers where patients with elevated tPSA results and/or other risk factors (e.g., positive DRE, family history, race) often seek, or are referred for, specialty care. Performance of PCA3 and comparators in a broader spectrum of health care settings may differ from that described in this review. It is not yet clear how PCA3, alone or in combination with other biomarkers/risk factors would be integrated into diagnostic or management pathways. The level of acceptance by physicians (and consumers) may well be impacted by Food and Drug Administration approval of a test kit to inform decisions about repeat biopsy in men with previous negative biopsies. While there was evidence that PCA3 performed better than tPSA with regard to diagnostic accuracy as a secondary test for men with increased risk, it is important to note that neither PCA3 nor tPSA have high performance. Combination of biomarkers and other risk information may be needed to improve overall performance in predicting prostate cancer at biopsy, or informing treatment based on risk classification. The intermediate outcome of diagnostic accuracy is key, as improvement could directly impact the number of biopsies performed in men without prostate cancer and the number of men with prostate cancer who are missed. It is also important to understand other potential harms, as well as the impact of the information on decisionmaking. The effect of even a great test is limited if uptake is low. Longer term outcomes are more challenging, due to the difficulty of following patients and collecting the necessary information.

Research Gaps

With the exception of analyses that include PCA3 and tPSA for the intermediate outcome of diagnostic accuracy, evidence was insufficient to answer the key questions. These questions, therefore, articulate remaining gaps in evidence. Other gaps in knowledge include:

- How much improvement in diagnostic accuracy is needed for any new test to impact biopsy decisionmaking.
- The potential of adding PCA3 alone or with other biomarkers to change decisionmaking in practice.
- How PCA3 compares to the two more frequently used add on tests (free PSA, PSA velocity) that have appeared in guidance documents.

- The need for matched studies not derived from “convenience” populations (e.g., biopsy referral centers), and more data on how key demographic factors (family history, race) impact on the performance of PCA3 and comparators.
- The need for outcome studies to determine how well PCA3 and other comparators used to categorize risk as low/indolent or high/aggressive predict the behavior of tumors over time.
- A range of methodological and statistical questions relating to modeling, assessing impact of verification bias, identifying most effective cutoffs for tests based on ROC analysis, and designs for future studies.

Conclusions

For diagnostic accuracy, there was a low strength of evidence that PCA3 had better diagnostic accuracy than tPSA elevations, but insufficient evidence that this led to improved intermediate or long-term health outcomes. In men at risk for prostate cancer based on elevated serum tPSA levels and/or suspicious digital rectal exam or other risk factor (e.g., family history), PCA3 was found to be more discriminatory for predicting prostate cancer at biopsy than tPSA elevations (i.e., at any sensitivity, the specificity is higher, or at any specificity, the sensitivity is higher). The finding that the relative performance of PCA3 versus tPSA elevations is not dependent on biopsy history (i.e., initial biopsy or repeat biopsy after one or more negative biopsies) is a new observation that allowed studies on KQ1 and KQ2 to be combined for analysis. Strength of evidence was insufficient for all other comparators and all other outcomes of interest in KQ1 / KQ2. Nine of 11 studies addressing KQ3 lacked a defined reference clinical endpoint (or validated surrogate), and the other two addressed different outcomes. Strength of evidence was insufficient. There is insufficient evidence for all other comparators and for all other outcomes of interest in KQ3. With one exception, these three questions continue to identify important gaps in knowledge, with other gaps identified in the review. Current uncertainty about the utility of tPSA screening for prostate cancer makes understanding followup tests (e.g., PCA3, other biomarkers, and algorithms) for assessing risk before biopsy and before treatment particularly important.

Table ES2. PCA3 vs. Comparators – Analyses and Strength of Evidence for the Intermediate Outcome of Diagnostic Accuracy

Comparator	tPSA[1]	fPSA[1]	PSAD[1]	EVN[1]	Multivariate Models Including tPSA[1]	cPSA[1]	tPSA DT and tPSA Velocity[1]	Comparators in Active Surveillance[2]
Risk of Bias	Moderate	Moderate	Moderate	Moderate	---	---	---	---
Consistency	Consistent, with 17 studies	Consistent, with 4 studies	Consistent, with 2 studies	Consistent, with 2 studies	Consistency unknown (1 study)	---	---	Consistency unknown (2 studies)[3]
Directness	Indirect	Indirect	Indirect	Indirect	Indirect	---	---	Indirect
Precision	Precise	Imprecise	Imprecise	Imprecise	Imprecise	---	---	---
Dose-Response Relationship	Present	---	---	---	---	---	---	---
Strength of Association	Weak	Weak	Weak	Weak	---	---	---	---
Overall Strength of Evidence (GRADE)	**Low**	**Insufficient**	**Insufficient**	**Insufficient**	**Insufficient**	**Insufficient**	**Insufficient**	**Insufficient**
KQ 1 and KQ2 **Area Under the Curve**	n=15[15-17, 19-30]	n=2 [19, 18]	n=2 [15, 26]	n=2 [17, 28]	0	0	0	---
Reported Mean and SD	n=4 [17, 25, 26, 28]	n=1 [28]	n=1 [26]	n=2 [17, 28]	0	0	0	---
Performance at a PCA3 cut-off score of 35	n=7 [15, 18, 20, 22, 25, 26, 31]	n=1[15]	n=1 [15]	0	0	0	0	---
ROC Curves- Sensitivity/ Specificity	n=10[15, 16, 18-20, 22, 25, 26, 28, 30]	n=3 [19, 18, 15]	n=2 [15, 26]	n=2[26, 28]	0	0	0	---
Regression Analysis	n=2 [15, 28]	n=2 [28, 15]	0	0	n=1 [15]	0	0	---
KQ3 **Progression and survival**	---	---	---	---	---	---	---	n=2[32, 37]

[1] Corresponds to KQ1 and KQ2; tPSA, total prostate specific antigen (PSA); fPSA, free PSA; PSAD, PSA density; EVN, externally validated nomograms; tPSA DT, PSA doubling time
[2] Corresponds to KQ3, for which comparators include biomarkers (e.g., tPSA, fPSA, PSA velocity), EVN/risk assessment tools and pathological tumor markers (e.g., Gleason score, staging, prostate volume). [3] Two studies had different sample types, different populations, and different surrogate/intermediate clinical outcome measures.

References

1. Jemal A, Siegel R, Xu J, et al. Cancer statistics, 2010. CA Cancer J Clin. 2010 Sep-Oct;60(5):277-300. PMID: 20610543.
2. Freedland SJ. Screening, risk assessment, and the approach to therapy in patients with prostate cancer. Cancer. 2011 Mar 15;117(6):1123-35. PMID: 20960523.
3. Yin M, Bastacky S, Chandran U, et al. Prevalence of incidental prostate cancer in the general population: a study of healthy organ donors. J Urol. 2008 Mar;179(3):892-5; discussion 95. PMID: 18207193.
4. Sutcliffe P, Hummel S, Simpson E, Young T, Rees A, Wilkinson A, Hamdy F, Clarke N, Staffurth J. Use of classical and novel biomarkers as prognostic risk factors for localised prostate cancer: a systematic review. Health Technol Assess. 2009 13(1-260. PMID:
5. Tosoian J, Loeb S. PSA and beyond: the past, present, and future of investigative biomarkers for prostate cancer. TheScientificWorldJournal. 2010 10(1919-31. PMID: 20890581.
6. Vlaeminck-Guillem V, Ruffion A, Andre J, et al. Urinary prostate cancer 3 test: toward the age of reason? Urology. 2010 Feb;75(2):447-53. PMID: 19586654.
7. Lin DW. Beyond PSA: utility of novel tumor markers in the setting of elevated PSA. Urol Oncol. 2009 May-Jun;27(3):315-21. PMID: 19414121.
8. Lin K, Lipsitz R, Miller T, et al. Benefits and harms of prostate-specific antigen screening for prostate cancer: an evidence update for the U.S. Preventive Services Task Force. Ann Intern Med. 2008 Aug 5;149(3):192-9. PMID: 18678846.
9. Bussemakers MJ, van Bokhoven A, Verhaegh GW, et al. DD3: a new prostate-specific gene, highly overexpressed in prostate cancer. Cancer Res. 1999 Dec 1;59(23):5975-9. PMID: 10606244.
10. Methods Guide for Effectiveness and Comparative Effectiveness Reviews. Rockville, MD: Agency for Healthcare Research and Quality; August 2011. AHRQ Publication No. 10(11)-EHC063-EF. Chapters available at: www.effectivehealthcare.ahrq.gov.
11. U.S. Preventive Services Task Force Procedure Manual. Rockville, MD: Agency for Healthcare Research and Quality, July 2008. AHRQ Publication No. 08-051180-EF.
12. Higgins JPT, Deeks JJ, Altman DG on behalf of the Cochrane Statistical Methods Group. Special topics in statistics. In: Cochrane Handbook for Systematic Reviews of Interventions. J. P. T. Higgins and S. Green (Eds). West Sussex, England:Wiley-Blackwell, pp. 481-529, 2010.
13. 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 May;63(5):513-23. PMID: 19595577.
14. Methods Guide for Medical Test Reviews. Rockville, MD: Agency for Healthcare Research and Quality; Draft Posted November 2010. http://www.effectivehealthcare.ahrq.gov/tasks/sites/ehc/assets/File/methods_guide_for_medical_tests.pdf. February 8, 2012.

15. de la Taille A, Irani J, Graefen M, et al. Clinical evaluation of the PCA3 assay in guiding initial biopsy decisions. J Urol. 2011 Jun;185(6):2119-25. PMID: 21496856.

16. Rigau M, Morote J, Mir MC, et al. PSGR and PCA3 as biomarkers for the detection of prostate cancer in urine. Prostate. 2010 Dec 1;70(16):1760-7. PMID: 20672322.

17. Ankerst DP, Groskopf J, Day JR, et al. Predicting prostate cancer risk through incorporation of prostate cancer gene 3. J Urol. 2008 Oct;180(4):1303-8; discussion 08. PMID: 18707724.

18. 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 Nov;184(5):1947-52. PMID: 20850153.

19. 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 Oct;106(8):1143-7. PMID: 20230386.

20. Adam A, Engelbrecht MJ, Bornman MS, et al. The role of the PCA3 assay in predicting prostate biopsy outcome in a South African setting. BJU Intl. 2011 Apr 20; PMID: 21507188.

21. Cao DL, Ye DW, Zhang HL, et al. 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 May 15;71(7):700-10. PMID: 20957673.

22. Deras IL, Aubin SM, Blase A, et al. PCA3: a molecular urine assay for predicting prostate biopsy outcome. J Urol. 2008 Apr;179(4):1587-92. PMID: 18295257.

23. Hessels D, van Gils MP, van Hooij O, et al. Predictive value of PCA3 in urinary sediments in determining clinico-pathological characteristics of prostate cancer. Prostate. 2010 Jan 1;70(1):10-6. PMID: 19708043.

24. 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 Jun;14(4):235-43. PMID: 19489685.

25. Nyberg M, Ulmert D, Lindgren A, et al. PCA3 as a diagnostic marker for prostate cancer: a validation study on a Swedish patient population. Scand J Urol Nephrol. 2010 Dec;44(6):378-83. PMID: 20961267.

26. 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 Mar;18(3):200-5. PMID: 21332814.

27. Ouyang B, Bracken B, Burke B, et al. 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 Jun;181(6):2508-13; discussion 13-4. PMID: 19371911.

28. Perdona 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. European urology. 2011 Jan;59(1):81-7. PMID: 20947244.

29. Roobol MJ, Schroder 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. European urology. 2010 Oct;58(4):475-81. PMID: 20637539.

30. Schilling D, Hennenlotter J, Munz M, et al. Interpretation of the prostate cancer gene 3 in reference to the individual clinical background: implications for daily practice. Urol Int. 2010 85(2):159-65. PMID: 20424427.

31. Wang R, Chinnaiyan AM, Dunn RL, et al. Rational approach to implementation of prostate cancer antigen 3 into clinical care. Cancer. 2009 Sep 1;115(17):3879-86. PMID: 19517474.

32. Kusuda Y, Miyake H, Kurahashi T, et al. 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 May 18; PMID: 21600799.

33. Auprich M, Chun FK, Ward JF, et al. Critical assessment of preoperative urinary prostate cancer antigen 3 on the accuracy of prostate cancer staging. Euro Urol. 2011 Jan;59(1):96-105. PMID: 20980098.

34. Liss MA, Santos R, Osann K, et al. PCA3 molecular urine assay for prostate cancer: association with pathologic features and impact of collection protocols. World J Urol. 2011 Oct;29(5):683-8. PMID: 21152924.

35. 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 May;179(5):1804-9; discussion 09-10. PMID: 18353398.

36. 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. European urology. 2011 Mar;59(3):422-9. PMID: 21156337.

37. 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 Feb;183(2):534-8. PMID: 20006883.

38. Vlaeminck-Guillem V, Devonec M, Colombel M, et al. Urinary PCA3 score predicts prostate cancer multifocality. J Urol. 2011 Apr;185(4):1234-9. PMID: 21334023.

39. Whitman EJ, Groskopf J, Ali A, et al. PCA3 score before radical prostatectomy predicts extracapsular extension and tumor volume. J Urol. 2008 Nov;180(5):1975-8; discussion 78-9. PMID: 18801539.

40. Cooperberg MR, Carroll PR, Klotz L. Active surveillance for prostate cancer: progress and promise. J Clin Oncol. 2011 29(3669-76. PMID:

Introduction

Background

Burden of Illness

Prostate cancer is the second most commonly diagnosed cancer and the second leading cause of cancer deaths in men in the U.S.[1] There are approximately 217,000 new cases per year in the U.S. and 32,000 prostate cancer-related deaths per year.[1] Risk is associated with increasing age, with more than 60 percent of cases occurring in men 70 years of age or older.[2] The disease is unpredictable, with the rate of tumor growth varying from very slow to moderately fast.[3] Most patients have indolent tumors, and may live for years with no or minimal effects, ultimately dying of other causes.[3] Although the lifetime risk of being diagnosed with prostate cancer is 16 percent, the lifetime risk of dying from the disease is only 3 percent.[4] Even more striking is the prevalence of occult disease. One third of men over the age of 60, and 50 percent of men over the age of 70, were found to have prostate cancer at autopsy.[5]

However, some patients have aggressive tumors that spread beyond the prostate, resulting in significant morbidity and death. Therefore, the key 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 for Prostate Cancer

Screening programs that use the total prostate-specific antigen (tPSA) test have been in place since the late 1980s, and have sparked interest and controversy. Production of tPSA is increased in men with prostate cancer, and elevation of tPSA can precede clinical disease by as much as 5 to 10 years.[6] Therefore, the rationale for tPSA screening is that it will lead to early diagnosis of prostate cancer, and early initiation of effective treatment that will improve clinical outcomes.

However, tPSA elevation is not specific to prostate cancer, and is also found in men with benign conditions (e.g., prostatic hyperplasia, prostatitis). In fact, tPSA screening has been found to lead to high rates of both false-negative and false-positive results. In men with false-negative results, cancer may be missed. Men with false-positive screening results may undergo one or more unwarranted biopsies with negative results. These men may experience unnecessary anxiety, discomfort, and significant procedural complications (e.g., infection, hemorrhage) in some cases.

It has been estimated that as many as 75 percent of eligible men in the U.S. have undergone at least one tPSA test.[7] However, a consensus approach for tPSA testing has not been reached, including who should be tested, when to test (and retest), and what is the most effective tPSA clinical threshold (e.g., 2.5, 4.0, 10.0 ng/mL). Several professional association (e.g., American Urological Association, American Cancer Society) guidelines support tPSA screening in asymptomatic men age 50 years of age or older, and in younger men in higher-risk populations.[8-10] They do not propose “routine” screening, but rather propose that men be provided with information on the potential risks and benefits of tPSA screening, and then make a decision about screening in consultation with their physician.

The incidence of prostate cancer decreased 2.4 percent per year between 2000 and 2006, and prostate cancer-related death rates have also decreased (39 percent, 1990-2006).[1] While some association with increased tPSA screening is likely, the effect size is currently unknown. However, it is known that the low specificity of the test has subjected many men to unnecessary prostate biopsy (false positives) and to overdiagnosis of indolent cancers with low probability of progressing.[11-13]

Diagnosis and Management of Prostate Cancer

Clinical action points for tPSA screening results range from 2.5 to 10 ng/mL. Criteria for immediate consideration of biopsy in screen positive men vary, but almost always include men with tPSA results >10 ng/mL, and may include men with tPSA results ≥4 to 10 ng/mL and/or suspicious DRE findings.[9, 14] Patients with elevated (>4 ng/mL) or intermediate (also referred to as "grey zone") tPSA results (e.g., 2.5 to 4) may be subject to biopsy based on results of subsequent testing with tPSA and/or other biomarkers (e.g., fPSA, PSA density, PSA velocity) and/or prebiopsy individualized risk assessment tools (e.g., PCPT Risk Calculator[15]) or nomograms.[3, 7, 16] However, the most effective approaches for use of biomarkers and risk assessment tools/nomograms remain controversial, and evidence of impact on decision making and subsequent short- and long-term improvement in clinical outcomes (e.g., function, morbidity, mortality) are lacking.[12, 13] Decisions about biopsy may also be impacted by other risk factors, comorbidities, or patient and physician preferences.

Biopsy

When needle biopsy is performed, pathologic examination of tissue cores includes identification of the presence or absence of cancer, the percent of the tissue core that is cancer and the Gleason score (an assessment of tissue differentiation).[2, 14, 17] Classification systems have been developed to designate biopsy-positive men as high risk or low risk relative to the likelihood of disease progression without treatment, in order to inform decisions about the option of active surveillance.[18, 19] Patients who are biopsy positive for prostate cancer also have a variety of treatment options. There is considerable interest in the identification of new biomarkers, or effective combinations of biomarkers and/or other risk factors, to better inform prebiopsy decisionmaking.[20]

Treatment Options

For patients with "low-risk cancer" (i.e., low grade, low volume tumor with characteristics believed to be associated with indolent or insignificant disease), one option is active surveillance.[17] Instead of working to eradicate the tumor, the patient defers treatment and begins ongoing surveillance that minimally includes serial tPSA and PSA velocity testing and repeated biopsies. If there is evidence of conversion to more aggressive disease (i.e., increasing tPSA levels or PSA velocity, upgrading in tumor stage or volume), surgery and other options can be considered.

For patients with "high-risk cancer" (i.e., high grade, high volume or multifocal tumor with characteristics believed to be associated with aggressive disease), or those unwilling to accept the risk associated with active surveillance, there are three well-established treatment options.[3, 17]

- Prostatectomy involves complete excision of the prostate. Prostate tissue from radical prostatectomy undergoes clinical staging and gross and histopathological examination.[2,

[14, 17, 21] Quantitative tools for risk stratification are available to estimate the likelihood of prostate cancer recurrence post-prostatectomy or other treatment.[15, 22-24]Complications of prostatectomy include infection, urinary incontinence and erectile dysfunction. The 10-year rate of prostate cancer-related death ranges from one to eight percent, depending on classification of the tumor as high or low risk.

- Interstitial brachytherapy is a choice for patients with low to moderate risk disease, and involves implanting high-dose radioactive seeds into the prostate. Adverse effects include lower urinary tract symptoms, obstructive or irritative prostatitis, rectal urgency, and later onset of erectile dysfunction. Ten year disease free survival is reported to be about 85 percent.[25]
- External beam radiotherapy is commonly used in combination with androgen therapy to treat high-risk disease. Adverse effects include irritative voiding symptoms, hemorrhagic cystitis or proctitis, and risk of a second malignancy of the bladder and rectum. Reported ten year disease free survival is about 88 percent.[26]
- Newer techniques for focal ablation of tumor, including cryotherapy and high-intensity focused ultrasound (HIFU), are currently being studied. These are generally considered investigational, with ongoing long-term followup on benefits and harms.

While biomarkers and risk assessment tools/nomograms are being used to identify men with high and low risk cancers, evidence of their effectiveness based on long-term clinical outcomes is needed.

Development of a New Biomarker: PCA3

In 1999, researchers reported identification of the differential display 3 gene (*DD3*), highly over-expressed in prostate cancer tissue but having little or no expression in normal prostate tissue or benign prostatic hyperplasia tissue.[27] Subsequently renamed the prostate cancer antigen 3 gene (PCA3), PCA3 is a noncoding mRNA mapped to chromosome 9q21-22.[28, 29] Since 2002, quantitative methods to measure PCA3 mRNA in urine samples have been developed and improved[28, 30] (see Table 4 in Results). Early investigators found that prostate manipulation led to a general release of mRNA. Since the gene that encodes PSA, *KLK3*, was not over-expressed in prostate cancer tissue,[31] PSA mRNA was chosen as the 'housekeeping' gene against which PCA3 mRNA results were normalized.[32, 33] In most assays, the ratio of PCA3 mRNA copies per mL and PSA mRNA copies per mL is multiplied by 1,000 to provide a PCA3 "score."[32, 33]

Stability of mRNA has been shown to be a source of variability for some tests, but use of a detergent based stabilization buffer[34] and shipping frozen samples have improved sample quality. Sokoll et al.[35] 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 February 17, 2012, Gen-Probe reported that they had received U.S. Food and Drug Administration (FDA) approval for the PROGENSA® PCA3 assay. FDA approval 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. A negative PROGENSA PCA3 assay is noted to be associated with decreased likelihood of a positive biopsy. However, the label specifies that "...the performance of the assay has not been established in men for whom a repeat biopsy was not already recommended." PCA3 testing is also offered by reference laboratories as laboratory-developed tests (i.e., tests

developed by and used at a single laboratory testing site). With FDA approval, it is reasonable to anticipate that this test will be more widely available throughout the U.S.

In summary, the upregulation 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 hypertrophy cells.[28, 32] Three potential intended uses for PCA3 have been proposed: 1) to inform decisions about when to biopsy patients at-risk and when to wait; 2) to inform decisions about when to rebiopsy patients at-risk and when to wait (the claim currently approved by FDA for the PROGENSA® PCA test); and 3) to determine in patients with positive biopsies whether the disease is indolent or aggressive, so that an optimal treatment plan can be developed.

Selected PCA3 Comparators

Total PSA (tPSA)

In 1989, Catalona et al.[36] initiated a multicenter population-based study examining the use of tPSA and DRE for prostate cancer screening. Based on the findings of enhanced early prostate cancer detection, the FDA approved PSA for prostate cancer screening in 1994, and defined the upper limit of normal as 4.0 ng/mL.[29] Numerous studies have followed, and the test is widely available as a screening and monitoring test. A subsequent meta-analysis of studies on PSA as a screening test showed that a significant number of prostate cancers would be missed using 4 ng/mL as a cutoff.[37] Concerns about the number of false-negative results led many urologists to use tPSA cutoffs of 2.5 to 3 ng/mL. This change increased the yield of positive cancers, but at a cost of many false-positive tPSA results that led to unnecessary biopsies with associated risk of clinical harms.[29]

In spite of its importance in health care, serum tPSA is not a standardized analyte, meaning that there is not a national requirement for quality specifications defined by either the Clinical Laboratory Improvement Act (CLIA) regulations or by the FDA. Assay-specific and site-specific differences in serum tPSA test performance in proficiency testing have been reported for more than 15 years.[38-40] In 1999, Klee et al. reported that +/-10 percent bias ranges for PSA correspond to -19.9 percent to +20.4 percent variation in patient classification.[41] Redesign of proficiency testing (PT) materials was reported to improve College of American Pathology survey outcomes,[38] and these specimens are likely to be the best available for monitoring testing in U.S. laboratories.[42] However, a 2006 review of PT data for tPSA found that available methods could be assumed correct less than 40 percent of the time at 6.5 ng/mL, and only 30 to 40 percent of the time at 19 ng/mL.[42] They concluded that test results were insufficiently reliable when applying uniform national cutoffs, and could cause many false-positive results.

In summary, for KQ1 and KQ2 (and in some instances for KQ3), nearly all study subjects with have tPSA levels of at least 2 to 4 ng/mL or higher (i.e., screen positive). Therefore, the actual comparator for PCA3 is not a positive or negative tPSA but, among those with positive results, the extent of the tPSA elevation. For example, risk of a prostate cancer being identified among men with levels between 4 and 6 ng/mL is lower than among men with tPSA levels greater than 10.

Free PSA (fPSA)

Serum PSA circulates in blood as different isoforms, mainly as unbound or free PSA (fPSA) and a form that complexes with serum antiproteases (complexed PSA or cPSA).[36] The ratio of

fPSA to tPSA levels is reported as percent fPSA. High levels of fPSA are associated with benign prostatic disease, while low levels are associated with cancer.[43] Percent fPSA increases with age and prostate volume, and decreases as total PSA increases.[29] In addition, fPSA is less stable than tPSA, and requires processing with 24 hours of collection.

One meta-analysis of fPSA studies[44] reported that, in the tPSA diagnostic grey zone of 2 to 10 ng/mL, addition of fPSA testing can reduce the number of unnecessary biopsies while maintaining a high cancer detection rate, but a second meta-analysis reported that fPSA is a useful addition to tPSA testing in only one part of the grey zone.[45] However, FDA has approved fPSA testing for men with tPSA levels in the 4-10 ng/mL range. The National Cancer Consortium Network includes fPSA in its diagnostic algorithm for early detection of prostate cancer, as part of decisionmaking about biopsy or rebiopsy.[9] They suggest that patients with grey zone tPSA levels and fPSA of 10 percent or less are candidates for biopsy or repeat biopsy; patients with fPSA greater than 25 percent be should be followed closely (e.g., DRE, tPSA, fPSA, PSA velocity), and that patients with intermediate values be informed of choices.[9]

PSA Density

Benson et al.[46] described the concept of PSA density (PSAD) as the ratio of tPSA concentration to prostate volume in cc. They identified differences in mean PSAD between men with prostate cancer and men with BPH. Subsequent studies suggested a modest improvement in diagnostic accuracy when PSAD was added to tPSA values, but more recent studies failed to confirm the value of PSAD. Consequently, its current use appears limited.[29, 47] One recent report has suggested this measurement can be used to predict clinical pathological features of disease.[48]

PSA Velocity and Doubling Time

PSA velocity and PSA doubling time are measures of longitudinal increases in tPSA. PSA velocity is defined as the rate of change of tPSA levels in a specified period, typically reported as ng/mL per year.[29, 49, 50] PSA doubling time is defined as the time it takes (e.g., months) for the tPSA level to increase by a factor of two. Both have been defined in many ways, with variability in number of PSA measures needed to calculate a dynamics metric, the time between measures, and the statistical method for estimating change.[49, 50]

National Comprehensive Cancer Network (NCCN) Guidelines for early detection of prostate recommend that PSA velocity be considered in both screening and followup.[9] However, one systematic review of pretreatment use of these biomarkers reported a slight increase in prediction of positive biopsy using PSA velocity, but found the studies subject to verification bias.[50] A second systematic review found limited evidence that addition of PSA velocity testing to a single tPSA measurement increases diagnostic accuracy for prostate cancer.[49] It has been proposed that tPSA dynamics may have a more effective role in monitoring of patients and/or differentiating aggressive or life-threatening prostate cancer.

Complexed PSA

Another PSA isoform considered in early detection of prostate cancer is measurement of the PSA bound to serum antiproteases, termed complexed PSA (cPSA). A recent systematic review[44] concluded that cPSA and fPSA showed equivalent effects. The authors cautioned that a lack of detail on study methodology and the relatively small number of studies warranted caution in the interpretation of the findings.

Externally Validated Nomograms

Current decisionmaking about risk of prostate cancer and whether to biopsy or rebiopsy has not been standardized, but depends on consideration of a variety of clinical factors (e.g., age, family history, race) and laboratory test results.[16] Recently, attention has been directed at development of risk algorithms, nomograms or artificial neural networks that combine multiple clinical and laboratory risk factors to create a cumulative risk score that informs clinical decisionmaking.[7, 16, 22, 51] These risk assessment tools are intended to exploit the incremental value of running multiple tests, each with independent contributions to the estimation of the risks of biopsy or treatment outcomes for patients. Risk factors often included were tPSA level, age, ethnic background and family history, and other biomarkers that vary by study. A systematic review concluded that these tools produced improvements in area under the receiver operator characteristics (ROC) curve (AUC), when compared to tPSA levels alone, but noted that many nomograms were not externally validated. When external validation was taken into account, benefits of nomogram use were decreased, although still statistically significant.[7] Despite variation in development and validation, a recent systematic review suggested that these tools tend to provide more accurate diagnostic predictions for cancer-positive biopsies than the use of tPSA testing or other factors alone.

Criteria for Distinguishing Low-Risk/Indolent and High-Risk/Aggressive Prostate Cancer

Three criteria have been commonly recognized to identify candidates for active surveillance: low-grade disease, low-volume disease, and low tPSA levels. However, eligibility criteria vary from study to study and site to site.[52] The oldest and most established criteria for identifying patients with low risk disease are those proposed by Epstein.[19] Epstein criteria that define low-risk or indolent cancer include: PSA density less than 0.15 ng/mL; Gleason score of 6 or less; fewer than three biopsy cores containing prostate cancer; and 50 percent or less involvement of any core. Original and modified Epstein criteria are key inclusion measures for studies of active surveillance.[52]

Comparison of criteria for identifying low risk or indolent tumors are complicated by the development by multiple groups of web-based predictive and prognostic nomograms/risk assessment tools, with little or no comparison and standardization of underlying assumptions. Understanding the relative performance of measurements to identify candidates for active surveillance is complicated by the use of varying criteria to identify risk progression, including changes in tPSA levels, histological grade, extent of biopsy core involvement, and/or clinical stage.[52] Longitudinal studies are needed to determine long-term clinical outcomes of patients classified as low and high risk, who immediately entered treatment or initially chose active surveillance (possibly progressing to treatment).

Key Questions

This comparative effectiveness review addresses three key questions. 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 men at risk based primarily on elevated tPSA and/or suspicious DRE. A recognized problem in tPSA screening-based diagnosis of prostate cancer is the high rate of false-positive results that can lead to a relatively high number of unnecessary biopsies. The third key question concerns the use of the urine PCA3 test and other biomarker tests and pathological

markers to classify the patient as low or high risk. This review will not address the merits or limitations of prostate cancer screening.

The Key Questions relate to three proposed scenarios in which this testing may be used:

1. In patients with elevated PSA and/or an abnormal DRE who are candidates for initial prostate biopsy, what is the comparative effectiveness of PCA3 testing as a replacement for, or supplement to, standard tests, including diagnostic accuracy (clinical validity) for prostate cancer, intermediate outcomes (e.g., improved decision making about biopsy), and long-term health outcomes (clinical utility), including mortality/morbidity, quality of life and potential harms?
2. In patients with elevated PSA and/or an abnormal DRE who are candidates for repeat prostate biopsy (when all previous biopsies were negative), what is the comparative effectiveness of PCA3 testing as a replacement for, or supplement to, standard tests, including diagnostic accuracy (clinical validity) for prostate cancer, intermediate outcomes (e.g., improved decisionmaking about biopsy), and long-term health outcomes (clinical utility), including mortality/morbidity, quality of life and potential harms?
3. In patients with a positive biopsy for prostate cancer who are being evaluated to distinguish between indolent and aggressive disease, what is the effectiveness of using PCA3 testing alone, or in combination with the standard prognostic workup (e.g., tumor volume, Gleason score, clinical staging) or monitoring tests (e.g., PSA, PSA velocity), with regard to diagnostic accuracy (clinical validity) for aggressive (high risk) prostate cancer, intermediate outcomes (e.g., improved decisionmaking about prognosis and triage for active surveillance and/or aggressive treatment) and long-term health outcomes (clinical utility), including mortality/morbidity, quality of life, and potential harms?

The proposed Key Questions (KQs) were posted for public comment on the Effective Health Care Program Web site (www.effectivehealthcare.ahrq.gov) from May 4, 2011, to June 1, 2011. A total of eight comments were received. No respondent suggested a specific change in the questions, although several noted that data concerning the use of PCA3 testing were currently most compelling for decisionmaking about repeat biopsy in patients screened with a PSA test and a DRE. At least two comments were directed at the likelihood that our review would not be able to address the long-term outcomes of interest (e.g., mortality, morbidity, quality of life). One commentator addressed the value of PCA3 test results in multispecialty team decisionmaking and noted that this test should be evaluated in patients receiving treatment to aid decisions about management changes. Based on the public comments received, no changes were made to the KQs.

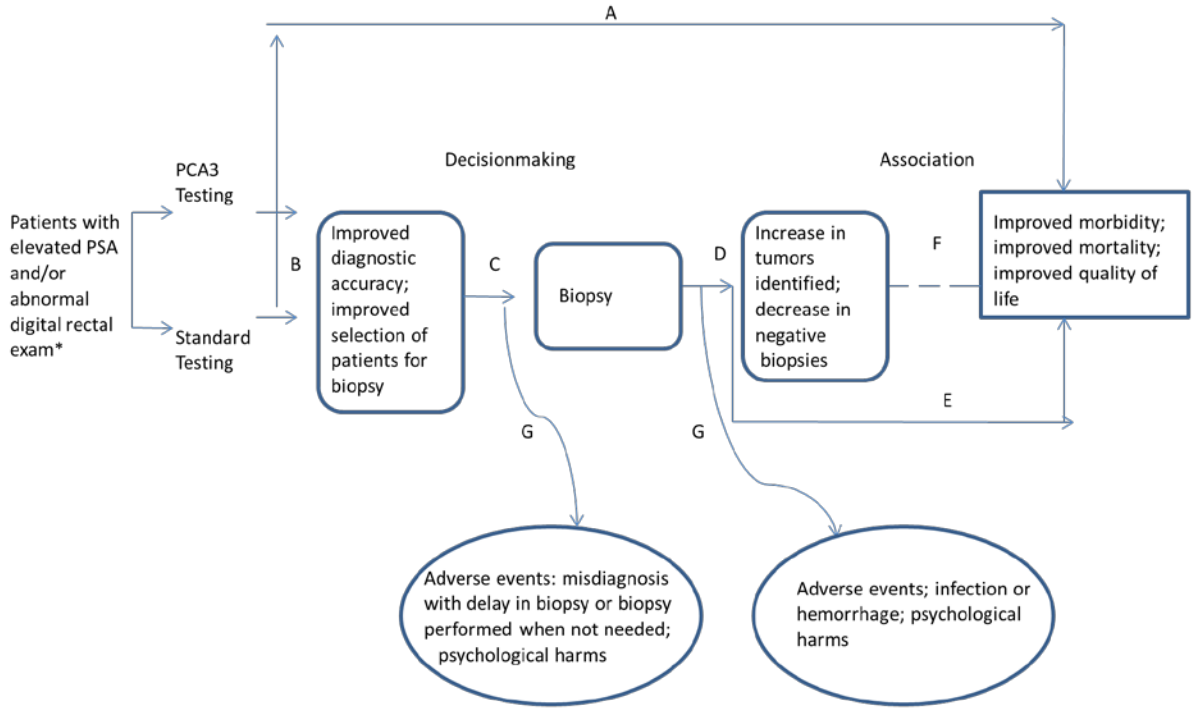
Analytic Frameworks

Two analytical frameworks (AFs) were developed for this review, one for Key Question (KQ) 1 and KQ2, and a second for KQ3. The first AF (Figure 1) addresses KQ1 and KQ2, and applies to men who are at risk for prostate cancer based on elevated tPSA results and/or abnormal DRE, and are having either an initial (KQ1) or repeat (KQ2) biopsy. The AF in Figure 1 depicts the comparative effectiveness of using PCA3 testing (alone or in combination with other biomarkers) and other standard tests (e.g., tPSA, fPSA) to predict intermediate and long-term health outcomes of interest. In Figure 1, direct evidence of the impact of testing on health

outcomes (e.g., mortality/ morbidity, quality of life) is shown by Link A. In the indirect chain of evidence, Link B addresses an intermediate outcome, the diagnostic accuracy (clinical validity) of the PCA3 test and its designated comparators in predicting positive biopsies. Link C addresses the impact of test results on the decision to proceed to an initial prostate biopsy, which, in turn, impacts other intermediate outcomes (Link D) and may affect health outcomes (Link E). Intermediate outcomes may have a direct association with long term health outcomes (Link F). Link G on the left addresses potential harms related to the effect testing has on the biopsy decision; Link G on the right focuses on clinical (e.g., bleeding, infection) and psychosocial (e.g., anxiety, quality of life) harms related to the biopsy procedure and/or repeated biopsy.

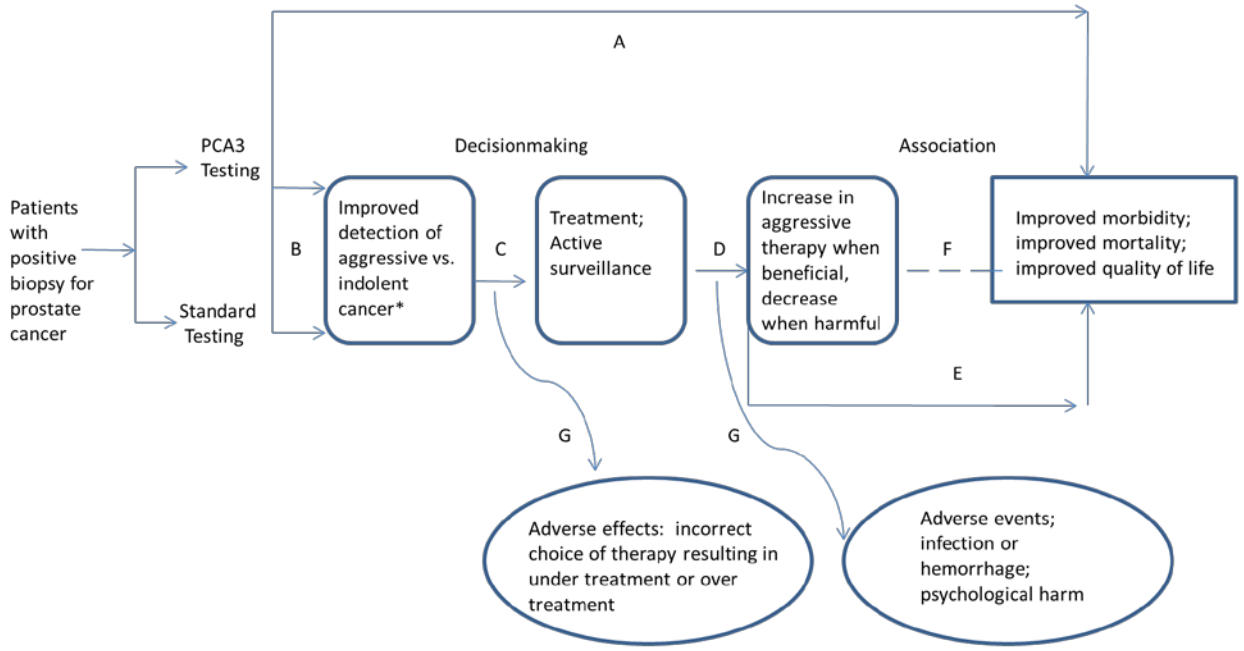
The second analytic framework (Figure 2) addresses KQ3, and applies to men who have had a biopsy result that is positive for prostate cancer. Figure 2 depicts the comparative effectiveness of using PCA3 testing (alone or in combination with other biomarkers) and other commonly used tests (e.g., Gleason score and other pathological markers, percent positive cores) on intermediate and long-term health outcomes of interest. Direct evidence of the impact of testing on health outcomes (e.g., mortality, morbidity, function, quality of life) is shown by Link A. In the indirect chain of evidence, Link B addresses the diagnostic accuracy (clinical validity) of the tests in categorizing tumors as aggressive and indolent. Link C addresses the impact of test results on decisionmaking related to prognosis and triage for active surveillance versus aggressive treatment. This link, in turn, impacts other intermediate outcomes (Link D) and indirectly affects health outcomes (Link E). Intermediate outcomes may have an association with health outcomes (Link F). Link G on the left addresses potential harms related to the effect testing has on treatment decisions; Link G on the right focuses on clinical (e.g., infection, hemorrhage) and psychosocial (e.g., anxiety, self-image, quality of life) harms related to the treatment.

Figure 1. Analytic framework for PCA3 as a diagnostic indicator for biopsy or rebiopsy in patients with elevated PSA and/or abnormal digital rectal examination (Key Questions 1 and 2)



*Patients may be evaluated for initial biopsy or rebiopsy after one or more negatives.

Figure 2. Analytic framework for PCA3 used to distinguish indolent versus aggressive prostate cancer (KQ3)



* Diagnostic accuracy

PICOTS Framework

The PICOTS (Population, Intervention, Comparator, Outcome, Timing, and Setting) for the Key Questions follow:

Population(s)

KQ1: Adult male patients who are candidates for initial prostate biopsy based on elevated prostate-specific antigen (tPSA) and/or abnormal digital rectal examination (DRE).

KQ2: Adult male patients with one or more previous negative prostate biopsies who are candidates for repeat biopsy based on elevated tPSA and/or abnormal DRE.

KQ3: Adult male patients with a positive prostate biopsy.

Interventions

- Testing for the prostate cancer antigen 3 gene (*PCA3*) mRNA alone or in conjunction with comparator tests.
- Prostate biopsy
- Prostatectomy

Comparators

- **KQs 1 and**
  - Total PSA
  - Percent free PSA
  - PSA velocity or doubling time
  - PSA density
  - Complexed PSA
  - Externally validated nomograms
- **KQ3**
  - Total PSA
  - Percent free PSA
  - PSA velocity or doubling time
  - Externally validated nomograms
  - Gleason score
  - Stage
  - Prostate volume
  - Epstein and other risk criteria
  - Other pathological markers

Outcomes

KQs 1 and 2

- *Long-term health outcomes*: Prostate cancer-related mortality, morbidity, function, quality of life (measured with validated instruments), and harms related to PCA3 testing and subsequent interventions (e.g., biopsy, surveillance, treatment).
- *Intermediate outcomes*: Diagnostic accuracy; impact on decisionmaking that leads to reduction in the number of unnecessary biopsies and increased identification of prostate tumors.

KQ3

- *Long-term outcomes*: Prostate cancer-related mortality, morbidity, function, quality of life (measured with validated instruments) and harms related to PCA3 testing and subsequent interventions (e.g., repeat biopsy, active surveillance, and treatment).
- *Intermediate outcomes*: Diagnostic accuracy; impact on decisionmaking that provides information on prognosis and informs treatment decisions.

Timing

- Any duration of followup will be evaluated.
- Timing of studies related to successive generations of PCA3 and PSA assays will be considered as part of quality assessment and as a potential source of heterogeneity.

Setting

- All settings.

Scope of the Review

Despite the large body of published literature on prostate cancer screening with tPSA, the value of early intervention remains unclear.[13] However, the burden of prostate cancer and the efforts to effectively diagnose and treat the disease are substantial. Consequently, there is continuing interest in identifying and validating biomarkers that can improve the clinical specificity and sensitivity of the prostate cancer diagnostic pathway (i.e., predicting prostate cancer in at risk men) or prognosis (i.e., classifying prostate tumors as high or low risk). Introduction of such a biomarker, alone or in combination with other biomarkers or risk factors, has the potential to reduce the current uncertainty in decisionmaking, which may lead to improved health outcomes with lower risk of associated harms. However, systematic review of the evidence supporting the diagnostic accuracy and utility of each intended use is needed to ensure an overall net balance of benefits over harms, and reduce the risk of introducing new unanticipated harms.

This is a comparative effectiveness review of testing with prostate cancer antigen 3 (PCA3), alone or in combination with other markers (comparators), in three proposed intended uses. The key questions addressed how PCA3 testing effectiveness compares to other markers: 1) alone or in combination with other risk factors (e.g., age, family history) or biomarkers in making decisions about which at risk patients to biopsy (KQ1); 2) alone or in combination with other risk factors (e.g., age, family history) or biomarkers in making decisions about which at risk patients to rebiopsy (KQ2); and 3) alone or in combination with other biomarkers, risk factors or pathological markers in biopsy positive patients, in making decisions about aggressive treatment (e.g., radical surgery or radiation therapy) or active surveillance (KQ3). KQ1 and KQ2 reflect important clinical decision points. However, they are narrowly focused on a specific subset of patients (i.e., only initial or only repeat biopsy), and studies often do not distinguish these two groups in their presentation of data. Therefore, it may be necessary to create a combined KQ1/KQ2 question if sufficient evidence exists that the biomarkers perform equally well in the two groups.

For prostate cancer prediction at biopsy or rebiopsy, the selected serum biomarker comparators were total prostate specific antigen (tPSA) elevations, free PSA (fPSA), PSA density, PSA velocity or doubling time, and complexed PSA. Externally validated nomograms (EVNs), a type of risk assessment tool, were also reviewed. For classification as high or low risk,

comparators included Gleason score, tumor stage, other pathological tumor markers, prostate volume, and biomarkers used for risk classification (e.g., tPSA, fPSA, PSA density) and monitoring disease progression (e.g., tPSA, PSA velocity). The selected outcomes of interest for all key questions included both intermediate (diagnostic accuracy, impact on decisionmaking, harms of biopsy or treatment) and long-term (mortality, morbidity, quality of life) outcomes.

Methods

Methodological practices followed in this review were derived from AHRQ *Methods Guide for Effectiveness and Comparative Effectiveness Reviews*[53] (hereafter *Methods Guide*) and AHRQ *Methods Guide for Medical Test Reviews* (hereafter AHRQ *Test Review Guide*).[54]

Topic Development and Refinement

Key questions were reviewed and refined as needed by the Evidence-based Practice Center (EPC) with input from Key Informants and the Technical Expert Panel (TEP) to ensure that the questions were specific and explicit about what information was being reviewed. In addition, for Comparative Effectiveness reviews, the key questions were posted for public comment and finalized by the EPC after review of the comments.

Key Informants are the end users of research, including patients and caregivers, practicing clinicians, relevant professional and consumer organizations, purchasers of health care, and others with experience in making health care decisions. Within the EPC program, the Key Informant role is to provide input into identifying the Key Questions for research that will inform health care decisions. The EPC solicits input from Key Informants when developing questions and an analytic framework for the systematic review or when identifying high priority research gaps and needed new research. The Key Informants selected to work on PCA3 included individuals with expertise in urology, pathology, laboratory medicine, internal medicine, family medicine, clinical trial design, as well as a patient advocate. The experts selected for the Technical Expert Panel to provide expertise and perspectives specific to the topic included individuals with expertise in urology, pathology, laboratory medicine, internal medicine, family medicine, clinical trial design and statistics. Technical Experts provided information to the EPC to identify literature search strategies and recommended approaches to specific issues as requested by the EPC.

Key Informants and Technical Experts are not involved in analyzing the evidence or writing the report and have not reviewed the report, except as given the opportunity to do so through the peer or public review mechanism. Key Informants and Technical Experts were required to disclose any financial conflicts of interest greater than $10,000, and any other relevant business or professional conflicts of interest. Because of their role as end-users or their unique clinical or content expertise, individuals were invited to serve as Key Informants or Technical Experts, and those who presented without potential conflicts were retained. The AHRQ Task Order Officer (TOO) and the EPC worked to balance, manage, or mitigate any potential conflicts of interest identified.

Literature Search Strategy

The research librarian, in collaboration with the review team, developed and implemented search strategies designed to identify articles relevant to each Key Question. Details on strategies with full search strings are presented in Appendix A. The search was limited to English-language articles or articles in other languages for which the journal provided an English translation. The rational for this decision is that this EPC's experience demonstrated that non-English references did not yield information of sufficiently high quality to justify the resources needed for translation. In addition, studies have demonstrated that excluding non-English language studies has little impact on effect size estimates or conclusions relative to the resources required.[55, 56] Systematic reviews/meta-analyses were identified through the MEDLINE® searches and grey

literature searches. Bibliographies of included studies were hand-searched to ensure complete identification of relevant articles. The time frame for the search was limited to literature published after January 1, 1990 based on FDA approval of the tPSA test for early detection of prostate cancer in 1993.

- MEDLINE® (January 1, 1990 to August 9, 2011)
- EMBASE® (January 1, 1990 to August 15, 2011)
- Cochrane Central Register of Controlled Trials (no date restriction)

Search results were stored in a project-specific EndNote9® database that was subsequently uploaded into DistillerSR (Evidence Partners Inc., Ottawa, Ontario, Canada), a web-based systematic review software application. Two independent reviewers used the DistillerSR software to determine study eligibility. Using selection criteria for screening abstracts, the two reviewers marked each abstract as: 1) yes (eligible for review of the full article); 2) no (ineligible for review); or 3) uncertain (review the full article to resolve eligibility). Reviewer discrepancies were resolved by discussion and consensus opinion; a third reviewer was consulted as needed. When abstracts were unavailable or unclear, full-text articles were obtained for review.

Using study selection criteria and the DistillerSR software, a single reviewer read each full-text article and determined eligibility of the study for data abstraction. A second reviewer audited a subset of articles, and reviewed all articles marked as uncertain. Discrepancies were resolved by discussion and consensus opinion; a third reviewer was consulted as needed. Key reasons for excluding studies were captured by DistillerSR and Excel® spreadsheet. Each paper retrieved in full-text, but excluded from the review, is listed in Appendix B with reasons for exclusion.

An updated search of the published literature was conducted upon submission of the draft report to determine if new information had been published since completion of the previous search. In addition, the Technical Expert Panel and individuals and organizations providing peer review were asked to inform the project team of any studies relevant to the KQs that were not included in the draft list of selected studies.

Study Selection

Studies were included if they fulfilled the following criteria:

- Study was a randomized controlled trial, a matched comparative study (e.g., prospective or retrospective cohort, diagnostic accuracy and case-control studies), or a systematic review of matched comparative studies. Note that systematic reviews of unmatched studies were initially retained in DistillerSR (but not extracted) based on potential usefulness in two areas: 1) providing references that might identify additional studies of PCA3; and 2) as sources of more broadly based unmatched data on performance characteristics of PCA3 and comparators (i.e., to compare to results based on smaller numbers of subjects in the primary matched studies of fPSA, to determine if the results are consistent or inconsistent).
- Study subjects were adult males with elevated total PSA tests and/or abnormal DRE who have not had a prostate biopsy or who have had one or more prostate biopsies (KQ1 and 2), OR adult male patients with prostate cancer positive biopsies (KQ3).
- Study intervention included testing for PCA3 and at least one designated pretreatment standard comparator test for prostate cancer, and a prostate biopsy (6 core minimum) or radical prostatectomy (KQ3 only).

- Study comparators for KQ1 and 2 were standard validated tests for prostate cancer that included tPSA, fPSA, PSA velocity, complexed PSA and externally validated nomograms/risk assessment programs; for KQ3, comparators included Gleason score, pathological staging, other pathological tumor characteristics, and tumor volume.
- Study outcomes included intermediate outcomes (e.g., diagnostic accuracy for prostate cancer, impact on biopsy decisionmaking and number of true and false positive biopsies), long-term outcomes (e.g., mortality, morbidity, function, quality of life) and potential harms (e.g., adverse effects of biopsy, misdiagnosis) (KQ1 and 2).

OR

- Study outcomes included the intermediate outcomes of diagnostic accuracy for tumor risk category (i.e., indolent/low risk, aggressive/high risk) and impact on decisionmaking about active surveillance versus aggressive treatment, as well as long-term outcomes (e.g., mortality, morbidity, function, quality of life) and potential harms (e.g., adverse effects of treatment, misdiagnosis) (KQ3).

Studies were excluded if they fulfilled at least one of the following criteria:

- Published in a non-English language for which the journal did not provide a translation
- Did not study prostate cancer
- Did not address one of the Key Questions
- Did not use a relevant study design
- Did not report primary data
- Did not report relevant outcomes

Search Strategies for Grey Literature

A systematic search of grey literature sources was undertaken to identify unpublished studies, or studies published in journals that are not indexed in major bibliographic citation database, in accordance with guidance from Effective Health Care Scientific Resource Center. The detailed search strategies and results can be found in Appendix C. Briefly, the searches included: regulatory information (i.e., FDA); clinical registries; abstracts and papers from professional annual meetings and conferences; organizations publishing guidance or review documents (e.g., National Guideline Clearinghouse, Cochrane, National Institute for Clinical Excellence); grants and federally funded research; and manufacturer web sites.

Search strategies were similar to those used in bibliographic databases, except for the following:

- Regulatory information: The FDA website was searched for PMA and 510(k) decision summary documents related to urine PCA3 mRNA assays.
- For clinical registries, NIH RePORTER, HSRPROJ, and AHRQ GOLD, searches were limited to completed studies only.
- For abstracts and conferences articles, studies published prior to 2009 were excluded.

Data Extraction and Management

The data elements from included studies were extracted using DistillerSR software into standard data formats and tables by one reviewer, and were subject to a full quality review for accuracy and completeness by a second reviewer. Data extraction question formats and tables were pilot-tested for completeness on a group of selected studies, and revised as necessary before

full data extraction began. Project staff met regularly to discuss the results at each phase, review studies that were difficult to classify and/or abstract, and to address any questions raised by team members.

Data Elements

Data elements extracted from the selected studies were defined in consultation with the TEP. A detailed list can be found in Appendix B, and the corresponding database fields in the DistillerSR Data Extraction Forms in Appendix I.

Evidence Tables

DistillerSR reports were created that contained content for specific evidence tables and downloaded into Excel® spreadsheets for editing. Final tables were formatted in Microsoft Word®. Primary reporting of DistillerSR data elements for each evidence table was done by one person; a second person reviewed articles and evidence tables for accuracy. Disagreements were resolved by discussion and, if necessary, by consultation with a third reviewer. When small differences occurred in quantitative estimates of data from published figures, the values were obtained by averaging the two reviewers' estimates.

Individual Study Quality Assessment

Definition of Ratings for Individual Studies and Reviews

In adherence with the Methods Guide[53], grading the methodological quality of individual comparative studies was performed based on study design-specific criteria. In all cases, quality of individual studies and the overall body of evidence was assessed by two independent reviewers. Discordant decisions were resolved through discussion or third-party adjudication. Quality assessments were summarized for each study and recorded in tables. Criteria for assessing quality of nonrandomized comparative intervention studies and quality rating definitions[57, 58] can be found in Appendix E. The quality of diagnostic accuracy studies was assessed using the Quality Assessment of Diagnostic Accuracy Studies (QUADAS) tool[59] included the following 14 questions:

1. Was the spectrum of patients representative of the patients who will receive the test in practice?
2. Were the selection criteria clearly described?
3. Is the reference standard likely to classify the target condition correctly?
4. Is the period between the reference standard and index test short enough to be reasonably sure that the target condition did not change between the two tests?
5. Did the whole sample or a random selection of the sample receive verification by using a reference standard of diagnosis?
6. Did patients receive the same reference standard regardless of the index test result?
7. Was the reference standard independent of the index test (i.e., the index test did not form part of the reference standard)?
8. Was the execution of the index test described in sufficient detail to permit replication of the test?
9. Was the execution of the reference standard described in sufficient detail to permit replication of the reference standard?

10. Were the index test results interpreted without knowledge of the results of the reference standard?
11. Were the reference standard results interpreted without knowledge of the results of the index test?
12. Were the same clinical data available when the test results were interpreted as would be available when the test is used in practice?
13. Were uninterpretable/intermediate test results reported?
14. Were withdrawals from the study explained?

For KQ1 and 2, the index test was PCA3 and the reference standard was biopsy. However, because selection of the screening positive populations was largely based on levels of tPSA, it was necessary to also consider QUADAS question 11 for tPSA to assess the potential for verification bias. This additional criterion was added at the end of the QUADAS questions (Table F-1, Appendix F), along with an entry to indicate whether verification bias was found (response, Yes or No). Because measurement of specific clinical outcomes was needed to assess diagnostic accuracy for KQ3, the additional criterion of clinical followup was added to the QUADAS questions (Table F-2, Appendix F).

The QUADAS ratings were summarized into general quality classes (from Table 5-4, Paper 5, AHRQ *Test Review Guide*[54]):

- **Good -** No major features that risk biased results.
- **Fair -** Susceptible to some bias, but flaws not sufficient to invalidate the results.
- **Poor -** Significant flaws that imply bias of various types which may invalidate the results.

Measuring Outcomes of Interest

There were several factors that supported the likelihood that most included studies would be focused on the predictive performance (e.g., clinical sensitivity and specificity, positive and negative predictive values for positive biopsy) of the PCA3 test. These factors included: the relatively short length of time that the PCA3 test, particularly the latest generation test, has been available; the comparative ease of conducting studies in which the end point is biopsy; and the length of time needed to collect long-term clinical outcomes related to the subsequent impact of interventions (e.g., active surveillance, aggressive treatment) related to the use of the test (as compared with no PCA3 testing or testing for other biomarkers). We expected that studies would provide a 2X2 table for PCA3 and other comparators, both for those subjects with positive biopsies and for those with negative biopsies (i.e., a matched analysis). In this way, one could evaluate not only the total performance of each test, but how the performance of the two tests varied in the population. For example, two tests could be shown to have equal sensitivity, but a matched analysis would indicate how often the two tests identified the same men with positive and negative biopsy results, and how often (and in what cases) they disagreed.

Two other intermediate outcomes for which data were sought were the impact of testing on physician and patient decisionmaking regarding biopsy versus surveillance, and on the potential harms related to the biopsy procedure (e.g., pain, bleeding, infection). Infectious complications appear to be increasing, with recent US and European reports indicating that 0.3 to 4.2 percent of men biopsied were admitted to the hospital for procedure-related infectious complications.[60, 61] Such data could be collected as followup to biopsy via records review or by conducting surveys of physicians and patients. Use of surveys requires particular attention to uptake rates and the reliability, validity and disease-specificity of survey instruments.

Long-term outcomes or study endpoints (e.g., 7-15 years) of interest include mortality and survival, morbidity and clinical and biochemical failure.[2] All-cause mortality at different time frames is reliable, but not a sensitive measure because it is dependent on age distribution and because most prostate cancer patients do not die of the disease. More sensitive measures are prostate cancer-specific 10 year survival or mortality if the cause of death is clear. Clinical failure may be measured as development of symptomatic disease, local disease progression or metastatic disease. Biochemical failure relates to increasing levels of total PSA (e.g., greater than 0.2 ng/mL) that may indicate disease recurrence. Morbidity also includes treatment-related adverse events (e.g., urinary incontinence, impotence, prostatitis) and other harms, as well as quality of life (QOL). Again, measuring QOL and the personal impact of symptoms related both to the cancer and to therapy requires the use of reliable and validated survey instruments. Minimally, assessment of QOL involves the use of a generic instrument to measure overall wellbeing, and a disease-specific instrument that focuses on specific symptoms and functions (e.g., incontinence, impotence).

Data Synthesis

After initial review of the extracted data from included studies, the analysis plan was finalized. Only matched studies were included and pair-wise relative performance of PCA3 scores versus comparator results summarized. Studies provided a wide variety of methods for comparing results, none of which were true matched analyses. For that reason, we chose to create the difference between the paired estimates and summarize these differences. Five separate analyses were designed:

1. A comparison of area under the ROC curve (or AUC);
1. Estimates of parameters defining the positive versus the negative biopsy populations;
2. Performance of PCA3 at a common cut-off score of 35;
3. Comparison of the ROC curves over a wide range of specificities/sensitivities; and
4. Results from logistic regression analysis.

As an example, consider a study reporting on a cohort of men age 50 or older who have a prostate biopsy and tPSA and PCA3 testing. For the first analysis, the AUC for tPSA was subtracted from the AUC for PCA3, resulting in the “difference of AUCs.” This comparison is an unbiased estimate of effect size differences. The next retrieved study is analyzed in the same way, and the two differences are then compared for consistency across studies. This is repeated for all relevant studies, and then repeated for each of the five analyses. The entire process is then repeated for each comparator.

Due to the small number of relevant matched studies for most comparators, heterogeneity of results could only be explored for the PCA3/tPSA comparison. This included stratification by studies including men with all elevations of tPSA versus those focusing on the “grey zone” of borderline tPSA elevations. The analysis of tPSA was complicated by the presence of partial verification bias in all of the studies. We relied on published results and in-house modeling in an attempt to account for this bias, as original data was not available to use published correction methods.[62, 63]

For KQ3, a qualitative analysis was considered likely. Identification of multiple matched studies in comparable populations, tested for PCA3 and one or more selected comparators, and reporting on the same intermediate or long-term clinical outcomes appeared to be unlikely.

Grading the Body of Evidence for Each Key Question

The strength of evidence for primary outcomes was graded by using the standard process of the Evidence-based Practice Centers as outlined in the *Methods Guide*.[53] The method is based on a system developed by the Grading of Recommendations Assessment, Development and Evaluation (GRADE) Working Group, and addresses four specific domains: risk of bias, consistency, directness, and precision of the evidence.[64] The evidence pertaining to key question were classified into four grade categories:

- **High -** High confidence that the evidence reflects the true effect. Further research is unlikely to change our confidence in the estimate of effect.
- **Moderate** – Moderate confidence that evidence reflects the true effect. Further research may change our confidence in the estimate of effect, or could change the estimate of effect.
- **Low -** Low confidence that the evidence reflects the true effect. Further research is likely to change our confidence in the estimate of effect and is likely to change the estimate.
- **Insufficient –** Evidence either is unavailable or does not permit a conclusion.
- Additional domains (e.g., strength of association, dose-response relationship, plausible confounding, publication bias) were assessed and reported if applicable based on the results of the evidence review.

The GRADE rating was determined by independent reviewers, and disagreements were resolved by consensus as necessary.

Assessment of Applicability

Applicability of the results presented in this review was assessed in a systematic manner using the PICOT framework (Population, Intervention, Comparison, Outcome, Timing). Assessment included both the design and execution of the studies, and their relevance with regard to target populations, interventions and outcomes of interest.

Peer Review and Public Commentary

Peer reviewers were invited to provide written comments on the draft report content based on their clinical and methodological expertise. Peer review comments on the preliminary draft of the report will be considered by the EPC in preparation of the final draft of the report. Peer reviewers have not participated in writing or editing of the final report or other products. The synthesis of the scientific literature presented in the final report will not necessarily represent the views of individual reviewers. The dispositions of the peer review comments will be documented and published three months after the publication of the evidence report.

Potential reviewers were required to disclose any financial conflicts of interest greater than $10,000, and any other relevant business or professional conflicts of interest. Peer reviewers who disclose potential business or professional conflicts of interest will be able to submit comments on draft reports through the public comment mechanism.

Results

Literature Search

Of the 1,449 citations identified through the literature searches, 1,415 were excluded at various stages of review. One additional study was identified through grey literature searches. No additional studies were identified from one identified systematic review on PCA3; this review was excluded from analyses as having no primary data. The 34 included articles reported the results of observational cohort studies with matched comparisons of PCA3 and other selected biomarkers. The PRISMA flow diagram (Figure 3) illustrates the review process for published studies, exclusions at each step and the selection results.

For KQ1 and KQ2, no randomized or comparative intervention trials were identified that included the use of PCA3 testing and reported long-term outcomes, or intermediate outcomes other than diagnostic accuracy. Of the 34 articles included, 25 were identified as addressing PCA3 and comparators in initial (KQ1) and/or repeat (KQ2) prostate biopsies.[20, 34, 51, 65-87] Six studies found to have duplicate data[22, 34, 66, 73, 84, 87] were excluded, and two studies[72, 83] did not have report data in a format usable for analyses. For KQ3, no randomized or comparative intervention trials were identified that included the use of PCA3 testing and reported intermediate or long-term outcomes. Two of 11 poor quality observational studies were identified that addressed two shorter-term health outcomes (i.e., biochemical recurrence and time to progression to treatment from active surveillance). Table 1 provides general descriptive information on the studies. Table 2 describes inclusion/exclusion criteria and Table 3 describes the populations studied. Table 4 provides key information on PCA3 testing. Table 17 provides information on biopsy results (KQ 1, 2, 3 as indicated in Table 17), and Table 18 and Table 19 provide information on prostatectomy results (KQ3 only).

Figure 3. PRISMA diagram for identified published studies

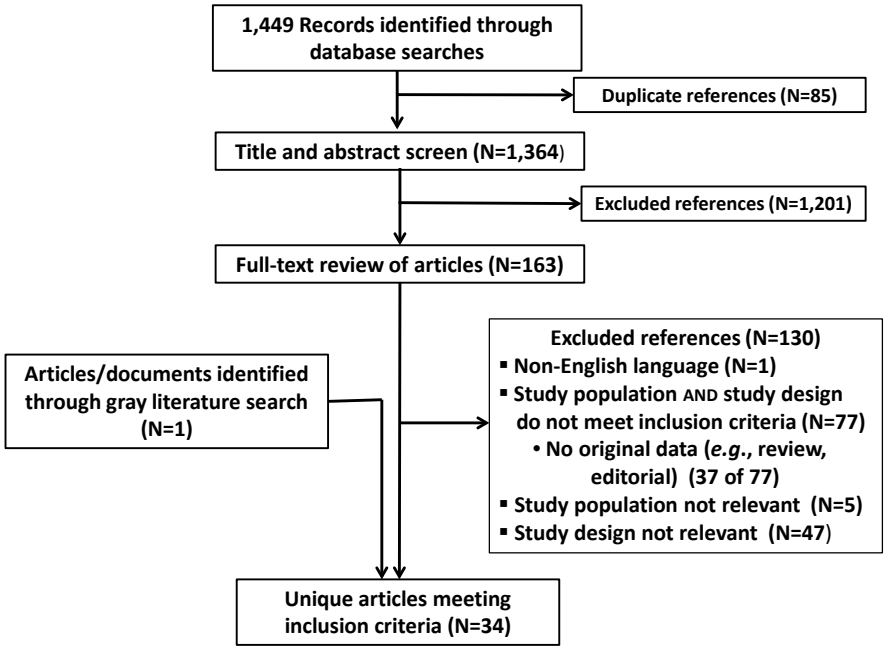


Table 1. General descriptions of the included studies in matched populations

Author[a], Year Country	Enrollment Period (Month/Year)	Reference Standard	N Biopsies Reported / Total N Biopsies	Unreported Biopsies Explained[b]	% Positive Biopsies	Key Question Addressed	Matched Analysis Reported	Potential COI Funding[c] / Disclosures[d]
Adam, 2011[65], South Africa	07/09 - 02/10	Biopsy	105 / 107	Yes	43	KQ1, 2	No	No / Yes
Ankerst, 2008[51] Europe, Canada, U.S.	NR	Biopsy	443 / 443	NA	28	KQ2	No	No / Yes
Aubin, 2010[66], International, U.S.	NR	Biopsy	1072 / 1140	Yes	18	KQ2	No	Yes / Yes
Cao, 2011[68], China	06/09 – 04/10	Biopsy	131 / 143	Yes	60	KQ1, 2, 3	No	No / NR
de la Taille, 2011[70], Europe	02/08 - 08/09	Biopsy	516 / 528	Yes	40	KQ1, 3	No	Yes / Yes
Deras, 2008[71], Canada, U.S.	04/04 – 05/06	Biopsy	557 / 570	Yes	36	KQ1, 2	No	NR / No
Hessels, 2010[74],The Netherlands	07/03 - 09/06	Biopsy	336 / 351	Yes	40	KQ1, 2, 3	No	No / NR
Mearini, 2009[75], Italy	NR	Biopsy	96 / 96	NA	73	KQ1, 2	No	No / No
Nyberg, 2010[76], Sweden	01/08 - 09/08	Biopsy	62 / 62	NA	29	KQ1, 2	No	No / No
Ochiai, 2011[77], Japan	05/07 - 05/08	Biopsy	102 / 105	Yes	36	KQ1, 2	No	Yes / No
Ouyang, 2009[78], US	NR	Biopsy	92/ 106	Yes	47	KQ1, 2	No	No / NR
Perdona, 2011,[20] Italy	10/08 - 10/09	Biopsy	218 / 218	NA	33	KQ1, 2	No	No / No
Ploussard, 2010[79], Europe	08/06 – 07/07	Biopsy	301 / 301	Yes	24	KQ2	No	NR / Yes
Rigau, 2010[80], Spain	NR	Biopsy	215 / 262	No	34	KQ1, 2	No	No / NR
Roobol, 2010[81], Europe	09/07 - 02/09	Biopsy	492 / 492	NA	17	KQ1, 2	No	NR / Yes
Schilling, 2010[88], Germany	01/08 - 06/08	Biopsy	103 / 104	Yes	56	KQ1, 2	Yes[e]	NR / No
Tosoian, 2010[89], U.S.	00/95 – 00/09	Biopsy surveillance	294 / 301	No	13	KQ3	No	Yes / NR
Wang, 2009[86], U.S.	09/06 – 12/07	Biopsy	187 / 192	Yes	46	KQ1, 2	No	No / Yes
Auprich, 2011[67] Germany, U.S., Austria	11/06 - 10/09	P	621 / 805	Yes	NA	KQ3	No	No / No
Kusuda, 2011[90], Japan	10/01 - 07/04	P	120 / 120	NA	NA	KQ3	No	NR / NR
Liss, 2011[91], U.S.	05/07 - 04/08	P	98 / 100	Yes	NA	KQ3	No	Yes / NR
Nakanishi, 2008[92], U.S.	06/05 - 05/06	P	96 / 96	NA	NA	KQ3	No	Yes / Yes
Ploussard, 2011[93], France	02/09 - 06/10	P	106 / 106	NA	NA	KQ3	No	No / No
Vlaeminck-Guillem, 2011[94], France	01/08 – 05/10	P	102 / 102	NA	NA	KQ3	No	NR / Yes
Whitman, 2008[95], U.S.	09/06 - 11/07	P	72 / 72	NA	NA	KQ3	No	Yes / Yes

NR=Not reported; NA=Not applicable, P=prostatectomy. [a] In alphabetical order for studies with a reference standard = biopsy (N=20), then alphabetical order for studies with a reference standard = prostatectomy (N=7). [b] 'Unreported biopsies' are biopsies completed without results reported. This column indicates whether the studies provided an explanation for missing biopsy results. [c] 'Yes' indicates that funding for the study was provided entirely or in part by a developer of a PCA3 assay used in the study; 'No' indicates funding from another source. [d] 'Yes' indicates that one or more authors disclosed a paid consultancy or other relationship with the developer of a PCA3 assay used in the study. [e] This one matched analysis was small (N=18).

Table 2. Reported criteria[a] for inclusion of subjects into studies and specific exclusion criteria

Author[b], Year	PSA Cutoff (ng/ml)	Abnormal DRE[c] (%)	Positive Biopsy[d]	Previous Negative Biopsy (%)	Positive Family History (%)	African-American	Other Risk Factors (%)	Specific Exclusion Criteria[e]
Adam[65]	All values	Yes	No	18	Yes	Yes	NR	BPH, indwelling catheters
Ankerst[51]	≥2.5	Yes (19)	No	100	5.9	NR	NR	
Aubin[66]	2.5-10 (<60y) 3-10 (≥60y)	No	No	100	No	NR	NR	PSA levels >10; meds that affect PSA levels; HGPIN, ASAP
Cao[68]	≥4	Yes	No	NR	NR	NR	NR	Meds that affect PSA levels
de la Taille[70]	2.5-10	Yes (19)	No	0	NR	NR	NR	PSA levels >10
Deras[71]	≥2.5	Yes (15)	No	50	Yes	Yes	Yes	
Hessels[74]	>3	Yes	No	Yes[6]	Yes	NR	NR	
Mearini[75]	>1	Yes	No	NR	NR	NR	NR	
Nyberg[76]	2.5	Yes	No	45	NR	NR	NR	
Ochiai[77]	≥ 2.5	Yes	No	19	NR	NR	NR	Meds that affect PSA levels
Ouyang[78]	≥ 4	NR	No	NR	NR	NR	NR	
Perdona[20]	<10	Yes (22)	No	38	No	No	NR	PSA levels > 10; meds that affect PSA levels; previous dx PCa
Ploussard[79]	2.5-10	No	No	100	NR	NR	NR	PSA levels > 10; , meds that affect PSA levels; initial biopsies
Rigau[80]	4	Yes (20)	No	0	NR	NR	NR	Meds that affect PSA levels
Roobol[96]	≥3.0	No	No	29	No	NR	NR	
Schilling,[88]	>4	Yes	No	14	No	No	NR	
Wang[86]	Yes	Yes	No	27	Yes (19)	NR	Yes (5.3)	History PCa

Table 2. Reported criteria[a] for inclusion of subjects into studies and specific exclusion criteria (continued)

Author[b], Year	PSA Cutoff (ng/ml)	Abnormal DRE[c]	Positive Biopsy[d]	Previous Negative Biopsy	Positive Family History	Race: Black	Other Risk Factors	Specific Exclusion Criteria[e]
Auprich[67]	NA	NA	Yes	NA	NA	NA	NA	Meds that affect PSA levels
Kusuda[90]	NA	NA	Yes	NA	NA	NA	NA	
Liss[91]	NA	NA	Yes	NA	NA	NA	NA	
Nakanishi[92]	NA	NA	Yes	NA	NA	NA	NA	Not age 40-70, PSA levels < 50; meds that affect PSA levels
Ploussard[93]	NA	NA	Yes	NA	NA	NA	NA	Not "low risk" = PSA ≤ 10; stage T1c-T2a; Gleason score 6
Tosoian[89]	NR	Yes	Yes	--	NR	NR	NR	Positive biopsy, not "low risk" based on Epstein criteria
Vlaeminck-Guillem[94]	NA	NA	Yes	NA	NA	NA	NA	Meds that affect PSA levels
Whitman[95]	NA	NA	Yes	NA	NA	NA	NA	Meds that affect PSA levels

NR = Not Reported; No = Not a criterion for inclusion; NA=Not applicable.

[a] Criteria represent the common risk factors for prostate cancer that are used to characterize subjects in the included studies.

[b] In alphabetical order for studies with a reference standard = biopsy (N=20), then alphabetical order for studies with a reference standard = prostatectomy (N=7).

[c] DRE = Digital Rectal Examination (DRE)

[d] 'Yes' answer indicates subjects having radical prostatectomy rather than biopsy; 'No' answer associated with study of biopsy patients.

[e] ASAP, atypical small acinar proliferation; BPH, benign prostatic hyperplasia; HGPIN, high-grade prostatic intraepithelial neoplasia.

Table 3. Characteristics of populations studied

Author, Year	Reference Standard	Age Distribution in years Mean (sd) / Median (range)	Race Distribution %	% Positive DRE	% Positive Family History	Co-Existing Pathology Name, %
Adam[65]	Biopsy	NR / 67 (35-89)	White 25.7; AA 68.6; Other 5.7	48.6	4.8	NR
Ankerst[51]	Biopsy	NR / 66 (11-83)	White 97.5; AA 2.0; Other 0.5	18.7	NR	NR
Aubin[66]	Biopsy	66.1 (6.0) / 66.1 (52.7-80)	NR	NR	13.8	NR
Cao[68]	Biopsy	NR	Asian (Chinese) 100	NR	NR	NR
de la Taille[70]	Biopsy	63 (7.6) / 63	NR	NR	NR	NR
Deras[71]	Biopsy	64 / 64 (32-89)	White 82.5; AA 5.3; Hispanic 2.3; Asian 0.4; Other 9.6	NR	NR	NR
Hessels[74]	Biopsy	63 / 64 (38-83)	NR	NR	NR	NR
Mearini[75]	Biopsy	NR	NR	NR	NR	BPH, 27
Nyberg[76]	Biopsy	NR / 63 (IQR 57-70)	NR	NR	NR	HGPIN, 5
Ochia[77]	Biopsy	NR / 66 (44-87)	Asian (Japanese) 100	NR	NR	NR
Ouyang[78]	Biopsy	NR	White 98; AA 2	NR	NR	NR
Perdona[20]	Biopsy	NR / 66 (60-72)	NR	NR	4.6	NR
Ploussard[79]	Biopsy	64.6 / 65.4 (43.3-83.4)	NR	NR	NR	ASAP. 4
Rigau[80]	Biopsy	Mean 65.7 / Range 44-85	NR	26.5	NR	NR
Roobol[96]	Biopsy	70.0 / 70.2 (63.6-77.5)	NR	13.1	NR	NR
Schilling[88]	Biopsy	NR / 67 (42-88)	NR	NR	NR	NR
Tosoian[89]	Biopsy surveillance	68.2 (6.2) / 68.2 (50.3-84.2)	White 91.8; AA 5.4; Other 2.8	NR	NR	NP
Wang[86]	Biopsy	62 (8.3) / Median NR (44-86)	White 91.5; AA 5.3; Other 3.2	16	18.7	HGPIN, 5.9 ASAP, 2.1
Auprich[67]	Prostatectomy	NR / 63 (44-79)	NR	23.9	NR	NR
Kusuda[90]	Prostatectomy	67.2 (6.5) / NR	NR	NR	NR	NR
Liss[91]	Prostatectomy	62.7 (7.2) / NR	NR	16	NR	NR
Nakanishi[92]	Prostatectomy	60 / 60 (45-70)	White 78.1; AA 15.6; Hispanic 6.3	NR	NR	NR
Ploussard[93]	Prostatectomy	62 / 62 (43-75)	NR	NR	NR	NR
Vlaeminck-Guillem[94]	Prostatectomy	62 (6) / 63 (47-72)	NR	NR	NR	NR
Whitman[95]	Prostatectomy	NR / 58 (42-73)	White 75; AA 25	NR	NR	NR

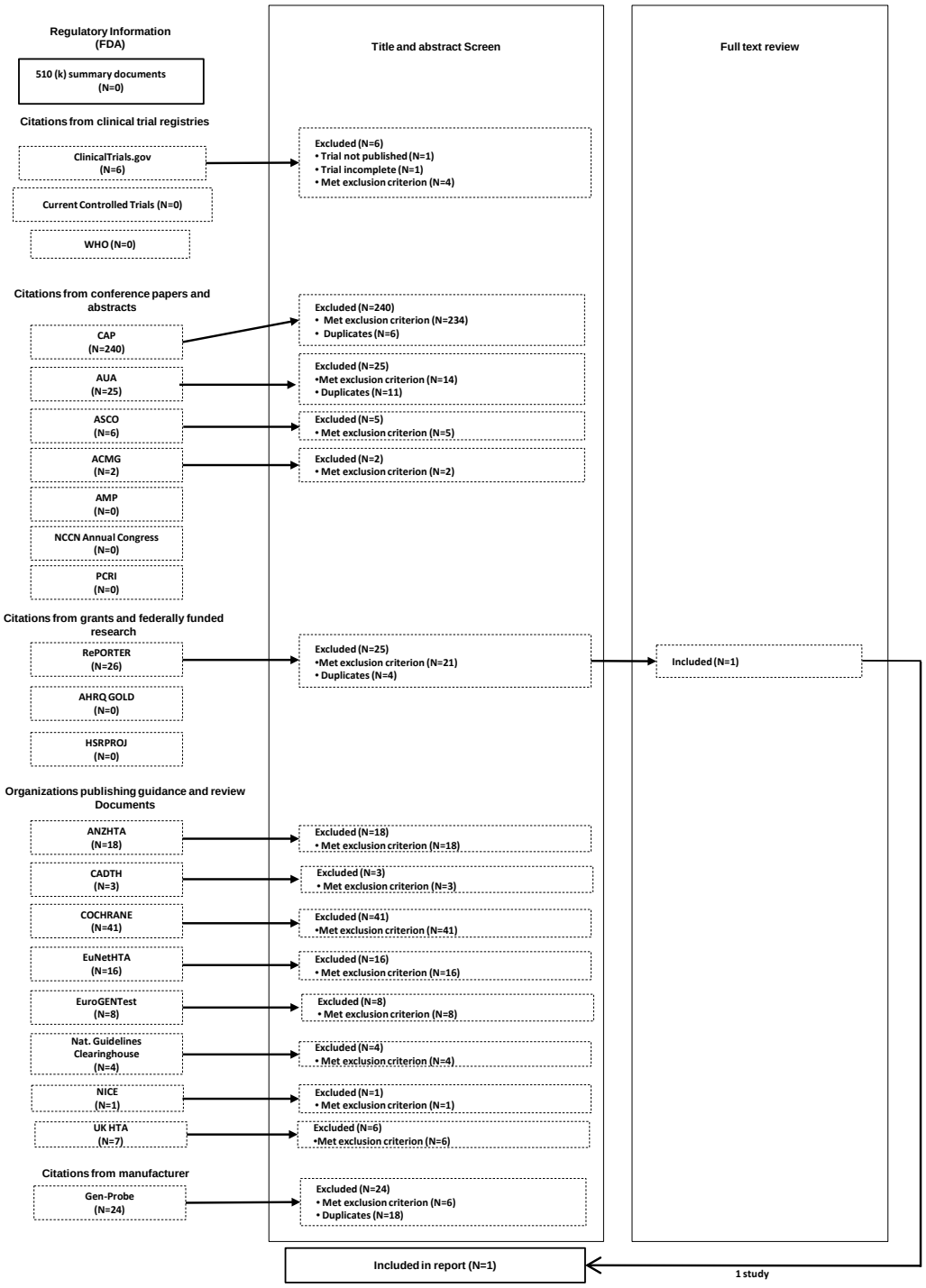
AA, African American; DRE, Digital Rectal Examination; ASAP, atypical small acinar proliferation; BPH, benign prostatic hyperplasia; HGPIN, high-grade prostatic intraepithelial neoplasia; NR – Not Reported

Table 4. PCA3 / DD3 assay characteristics

Author, Year	Attentive Massage Used	Specimen	Method Used	Assay Specified	Reporting Units	House-keeping Gene	Handling Temperatures Holding (C); Storage (C)	Informative Results %
Adam, 2011[65]	Yes	UU	TMA	Progensa	PCA3 Score	PSA	NR	98.1
Aubin, 2010[66]	Yes	UU	TMA	“Gen-Probe”	PCA3 Score	PSA	NR; -70	94
Auprich, 2011[67]	Yes	UU	TMA	Progensa	PCA3 Score	PSA	NR	100
de la Taille, 2011[70]	Yes	UU	TMA	Progensa	PCA3 Score	PSA	NR	99
Deras, 2008	Yes	UU	TMA	Progensa	PCA3 Score	PSA	NR; -70	100
Hessels, 2010[74]	Yes	US	TMA	Aptima	PCA3 Score	PSA	NR	95.8
Liss, 2011[91]	Yes	UU	TMA	NR	PCA3 Score	PSA	2-8; NR	100
Nakanishi, 2008	Yes	UU	TMA	“Gen-Probe”	PCA3 Score	PSA	2-8; -70	100
Nyberg, 2010[76]	Yes	UU	TMA	Progensa	PCA3 Score	PSA	NR	100
Ochiai, 2011[77]	Yes	UU	TMA	Aptima	PCA3 Score	PSA	NR; -70	100
Perdona, 2011[20]	Yes	UU	TMA	Progensa	PCA3 Score	PSA	NR	100
Ploussard, 2011[93]	Yes	UU	TMA	Progensa	PCA3 Score	PSA	NR	NR
Roobol, 2010[96]	Yes	UU	TMA	Progensa	PCA3 Score	PSA	NR	100
Schilling, 2010[88]	Yes	UU	TMA	“Gen-Probe”	PCA3 Score	PSA	NR	99
Tosoian, 2010[89]	Yes	UU	TMA	Aptima	PCA3 Score	PSA	NR; -80	97.7
Vlaeminck-Guillem, 2011[94]	Yes	UU	TMA	Progensa	PCA3 Score	PSA	NR	100
Wang, 2009[86]	Yes	UU	TMA	“Gen-Probe”	PCA3 Score	PSA	30; NR	97.4
Whitman, 2008	Yes	UU	TMA	“Gen-Probe”	PCA3 Score	PSA	NR	100
Fradet, 2004[72]	Yes	UU	NASBA	uPM3	Probability	PSA	NR; 2-8	86
Tinzl, 2004[83]	Yes	UU	NASBA	uPM3	PCA3 Score	PSA	4; -20	79
Cao, 2011[68]	Yes	US	QRT-PCR	NR	PCA3 Score	PSA	4; -20	92
Kusuda, 2011[90]	NR	LN	QRT-PCR	NR	Quantification	GADPH	NR	NR
Mearini, 2009[75]	Yes	US	QRT-PCR	NR	Quantification	Beta actin	NR -70	NR
Ouyang, 2009[78]	Yes	US	QRT-PCR	NR	PCA3 Score	GADPH	NR; -80	87
Rigau, 2010[80]	Yes	US	QRT-PCR	NR	PCA3 Score	PSA	NR; -80	82.1
Ankerst, 2008[51]	NR	NR	NR	NR	PCA3 Score	PSA	NR	NR
Ploussard, 2010[79]	NR	UU	NR	NR	PCA3 Score	PSA	NR	NR

LN, lymph node tissue; NASBA, nucleic acid sequence-based amplification; NR, not reported; QRTPCR, quantitative real time polymerase chain reaction; TMA, transcription-mediated amplification; UU, urine-unsedimented; US, urine-sedimented.

Figure 4. PRISMA diagram for identified grey literature



Grey Literature Search

The process for evaluation of grey literature search results is summarized in Figure 4. and Appendix C. Although Gen-Probe has announced FDA clearance for the Progensa® assay kit, the data were not available for analysis for the draft report. Two clinical trial registry citations were relevant to the review:

- Prostate Cancer Antigen 3 (PCA-3) Gene Project (NCT01177436) – This trial is still recruiting participants. Of interest is the use of three housekeeping genes for PCA3 testing in addition to *KLK3* (PSA): *ACTB* (beta-actin), *TUA* (Ka 1 tubulin), and *GAPDH* (glyceraldehyde-3-phosphate). Results may resolve remaining concerns about potential bias related to the use of PSA as the housekeeping gene.
- Clinical Evaluation of the Progensa® PCA3 Assay in Men With a Previous Negative Biopsy Result (NCT01024959) – This trial, conducted by GenProbe and completed in April, 2011, is likely to have provided data for the submission to FDA. However, no published article was identified that reported the results of this clinical trial.

Overall, the search of grey literature yielded one published study.

Key Question 1 – Testing PCA3 and Comparators to Identify Prostate Cancer in Men Having an Initial Biopsy

Among the 17 included studies, only two reported results in populations where all men were having initial biopsies (Table 5).[70, 80] Both studies reported data on tPSA, and one reported on fPSA and PSA density.[70] The actual data will be presented in later analyses (KQ1/KQ2 combined), but clearly there were too few data for reliable interpretations. The five matched analyses are outlined in Methods, and discussed in more detail in later sections (Table 7).

Table 5. Data available for the comparison of PCA3 scores and the five comparators, when analysis is restricted to studies reporting results only in men having initial biopsies

			Matched analyses (of five) available for each PCA3 comparator					
Author	N	Initial Biopsy	tPSA	fPSA	PSA velocity	PSA Density	cPSA	Validated Nomogram
de la Taille[70]	516	100%	A,C,D,E	C,D,E	-	A,C,D	-	-
Rigau[80]	21	100%	A,D	-	-	-	-	-

The dash ('-') indicates no data provided for that comparator. The letters 'A' through 'E' represent the analyses: A = AUC, B = mean/sd, C = PCA3 > 35, D = Sensitivity/Specificity and E = Regression.

Shaded rows indicate studies focusing on the 'grey zone' of tPSA (2.5-10 ng/mL) when enrolling patients.

Strength of Evidence: When data are restricted to the two studies reporting only on populations of men having an initial prostate biopsy, it is not possible to evaluate consistency (similar between-study results). In addition any estimates of effect size will, necessarily, be imprecise. This results in assigning grades of "insufficient" for all six comparisons of PCA3 with tPSA, fPSA, PSA velocity, PSA density, complexed PSA and externally validated nomograms, respectively.

Key Question 2 – Testing PCA3 and Comparators to Identify Prostate Cancer in Men Having a Repeat Biopsy

Among the 17 studies, only three[51, 66, 79] reported results in populations where all of the men were having a repeat biopsy (Table 6). All studies reported on tPSA, two on fPSA[66, 79]and one on externally validated nomograms.[51] The actual data will be presented in later analyses, but clearly there are too few data for reliable interpretations.

Table 6. Data available for the comparison of PCA3 scores and the five comparators, when analysis is restricted to studies reporting results only in men having a repeat biopsy

			Matched analyses (of five) available for each PCA3 comparator					
Author	N	Initial Biopsy	tPSA	fPSA	PSA velocity	PSA Density	cPSA	Validated Nomogram
Ankerst[51]	443	0%	A,B	-	-	-	-	A,B
Aubin[66]	1072	0%	C,D	A,D	-	-	-	-
Ploussard[79]	301	0%	A,D	A,D	-	-	-	-

A dash ('-') indicates no data provided for that comparator. The letters 'A' through 'E' represent the analyses: A = AUC, B = mean/sd, C = PCA3 > 35, D = Sensitivity/Specificity and E = Regression.

Shaded rows indicate studies focusing on the 'grey zone' of tPSA (2.5-10 ng/mL) when enrolling patients.

Strength of Evidence: When data are restricted to the three studies reporting only on populations of men having repeat prostate biopsy, it is not possible to evaluate consistency (similar between-study results). In addition, one of the studies[66] restricted recruitment to men with tPSA in the "grey" zone. Estimates of effect size will, necessarily, be imprecise. This results in assigning grades of "insufficient" for all six comparisons of PCA3 with tPSA, fPSA, PSA velocity, PSA density, complexed PSA and externally validated nomograms, respectively.

Potential to Combine Key Questions 1 and 2 – Testing PCA3 and Comparators to Identify Prostate Cancer in Men With or Without a Previous Negative Biopsy (Initial and Repeat)

Five studies[51, 66, 70, 79, 80] exclusively studied men having an initial biopsy (KQ1) and men having repeat biopsy (KQ2). These have been shown in the previous pages. However, 12 additional studies include matched results of PCA3 and the comparators, but enrolled men with both initial and repeat biopsies, or did not report biopsy history.[20, 65, 68, 71, 74-78, 82, 86, 97] Most often, the results from these studies were not stratified by biopsy history. At this point, one could ignore the data in these 12 studies, as they do not directly apply to either KQ1 or KQ2. Instead, based on the inadequate strength of evidence found for the prior individual analyses of KQ1 and KQ2, we decided to examine whether data from all 17 studies may be suitable for a combined analysis (Table 7). Prior to 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. An examination of Table 7 finds that the most common comparator is tPSA, and the most common analysis, by far, is the area under the curve (AUC), indicated by an "A."

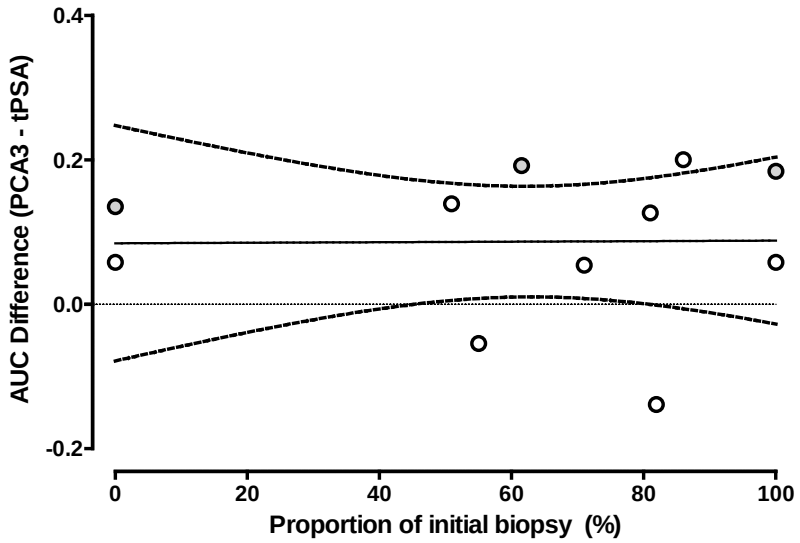
Table 7. Data available for the comparison of PCA3 scores and the five comparators, when analysis is restricted to studies reporting results with/without a previous prostate biopsy

			Matched analyses (of five) available for each PCA3 comparator					
Author	**N**	**Initial Biopsy**	**tPSA**	**fPSA**	**PSA velocity**	**PSA density**	**cPSA**	**Validated Nomogram**
Cao[68]	143	-	A	-	-	-	-	-
Hessels[74]	336	-	A	-	-	-	-	-
Mearini[75]	-	-	A	-	-	-	-	-
Ouyang[78]	92	-	A	-	-	-	-	-
Ankerst[51]	443	0%	A,B	-	-	-	-	A,B
Aubin[66]	1072	0%	C,D	A,D	-	-	-	-
Ploussard[79]	301	0%	A,D	A,D	-	-	-	-
Deras[71]	570	51%	A,C,D	-	-	-	-	-
Nyberg[76]	62	55%	A,B,C,D	-	-	-	-	-
Perdona[20]	218	61%	A,B,D,E	B,E	-	-	-	A,B,D
Roobol[96]	721	71%	A	-	-	-	-	-
Wang[86]	516	73%	C	-	-	-	-	-
Ochiai[77]	105	81%	A,B,C,D	-	-	A,B,D	-	D
Adam[65]	106	82%	A,C,D	-	-	-	-	-
Schilling[88]	32	86%	A,C	-	-	-	-	-
de la Taille[70]	516	100%	A,C,D,E	C,D,E	-	A,C,D	-	-
Rigau[80]	215	100%	A,D	-	-	-	-	-

A dash ('-') indicates no data provided for that comparator. The letters 'A' through 'E' represent the analyses: A = AUC, B = mean/sd, C = PCA3 > 35, D = Sensitivity/Specificity and E = Regression.
Shaded rows indicate studies focusing on the 'grey zone' of tPSA when enrolling patients.

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. Figure 5 shows this analysis. The raw data for this figure can be found in Table 10. The slope (-0.00227) is nonsignificant (p=0.97, test of slope), indicating that there is no significant relationship between the biopsy status and AUC difference.

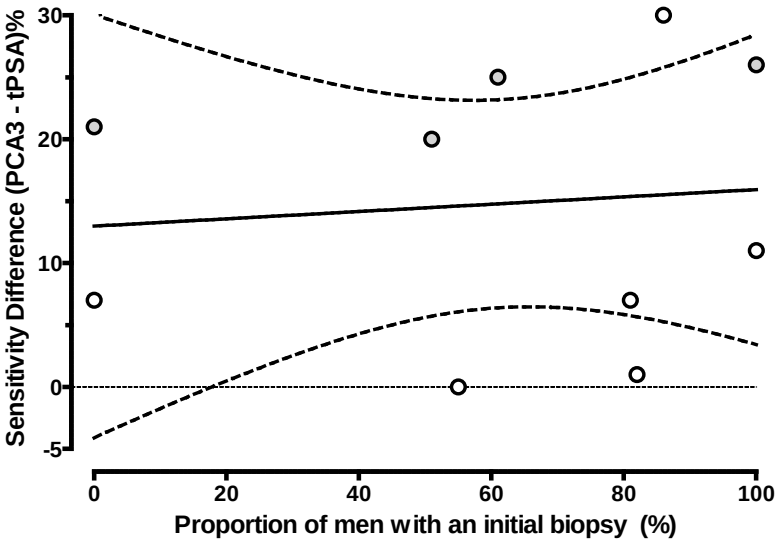
Figure 5. A scatterplot showing the relationship between the proportion of men with an initial biopsy versus the difference in the area under the curves (AUC) for PCA3 scores and tPSA elevations



The horizontal axis shows the proportion of men with an initial biopsy (range 0 to 100%). The vertical axis shows the AUC Difference; a value of 0 (dashed horizontal line) indicates the AUCs are equivalent. Each circle indicates the results from one included study. Results from three studies focusing on the grey zone are filled. There is no significant relationship, as indicated by the solid line and 95% confidence intervals.

Examining Table 7 also indicates that 10 studies reported the ROC curves for both PCA3 and tPSA. The raw data for this figure can be found in Table 13. Figure 6 again shows the proportion of men with an initial biopsy on the x-axis but now indicates the relative increase (or decrease) in the PCA3 – tPSA sensitivities, at a constant specificity of 50 percent. Based on linear regression (slope = 0.02956, p=0.79, test of slope) there again appears to be little or no association between the biopsy history and the relative performance of PCA3 and tPSA.

Figure 6. A scatterplot showing the relationship between the proportion of men with an initial biopsy versus the difference in sensitivity at a specificity of 50%



The horizontal axis shows the proportion of men with an initial biopsy (range 0 to 100%). The vertical axis shows the difference in sensitivity of detecting prostate cancer, when the specificity is held constant at 50%; a value of 0 (dashed horizontal line) indicates the specificities are equivalent. Each circle indicates the results from one included study. Results from three studies focusing on the grey zone are filled. There is no significant relationship, as indicated by the solid line and 95% confidence intervals.

Together, these two analyses (Figure 5, Figure 6) provide evidence that combining results from studies of initial biopsies, repeat biopsies, and mixtures of initial and repeat biopsies will not impact the comparison of PCA3 with tPSA elevations. No such analyses can be made for any of the other comparators, but given the lack of data for those comparisons, we have chosen to comprehensively list all potentially relevant results, regardless of the biopsy status of the enrolled men. The following sections provide the results of the combined analysis of Key Questions 1 and 2.

Description of Included Studies for KQ1/KQ2 "Combined"

The inclusion criteria restricted the analyses to matched studies. These were defined as studies that provided estimates of diagnostic test performance for PCA3 and at least one other comparator (e.g., total PSA elevations or free PSA) using the same patient population. Thus, a study of PCA3 alone, or a comparator alone, would not be included. In examining the included studies, it was clear that although the same population was used, the reports rarely applied a true matched analysis. However, the results were still considered as being "matched," due to the application of the test(s) to the same underlying population. We preserved this population matching by computing differences between PCA3 test results and comparator test results within each study. These matched differences could then be compared across studies. For example, one study of biopsied patients might report an AUC for PCA3 and then separately report an AUC for tPSA in the same population. The difference in the two would then be computed and compared to the difference in AUCs from other similarly matched studies. Although this restriction limited

the number of included publications, it was aimed at improving the consistency of results. For example, an analysis of unmatched studies might have provided sufficient information to stratify PCA3 performance by number of previous biopsies. Similar data could be obtained from the literature for tPSA. Comparing the results between these unmatched studies might have shown differences related to variations in study populations or design rather than the variable of interest.

All studies were judged to be of poor quality either because of verification bias, missing followup to clinical endpoints, or unclear data presentation or blinding. Almost a third of studies (N=8) were funded by GenProbe and almost half (N=12) indicated conflicts of interest by investigators (Table 1). All of these studies focused on determining the diagnostic accuracy of PCA3 testing using biopsy results as the reference or gold standard. No studies were identified that reported on intermediate outcomes other than diagnostic accuracy, or long-term clinical outcomes.

Evaluation of PCA3 and Other Comparators to Identify Intermediate and Long-Term Outcomes Related to Prostate Cancer

Comparator: Total Serum PSA

Study design was a crucial criterion for this comparison, because tPSA measurements are integral to men's decisionmaking regarding uptake of prostate biopsy after the finding of an initial tPSA elevation obtained at routine prostate cancer screening. Men were likely offered biopsy based on the extent of tPSA elevations, suspicious findings on a digital rectal exam (DRE), a combination of the two or, less commonly, other risk factors such as family history or race. This leads to only a subset of initially identified men having the "gold standard" test (biopsy) that defines one of the outcomes of interest – diagnostic accuracy. This association of test result with uptake of the diagnostic test is labeled verification bias. In this setting, verification bias will occur because men with higher tPSA elevation are more likely to undergo biopsy compared to men with lower levels. Thus, test sensitivity will be overestimated (as a higher proportion of cancers with lesser elevations will not be identified by biopsy). This bias will also underestimate specificity (or overestimate the false positive rate), because the larger number of men without cancer and negative results are not identified via biopsy). Appendix J contains a more complete description of verification bias, an example relevant to prostate cancer and tPSA elevations, a review of directly relevant literature, and an evaluation of what will, or will not be compromised in this comparison. Appendix J also contains a more complete description of the modeling used to overcome the major impact of verification bias and a more extensive comparison of PCA3 and tPSA test performance characteristics. These analyses indicate that the relative performance of tPSA elevations (sensitivity at a given specificity) is not influenced by verification bias, but the tPSA cut-off level at which this performance occurs cannot be directly observed. We employed a simple model to determine approximate tPSA cut-off levels in the presence of verification bias. This bias is unlikely to be an issue for the other comparators (e.g., fPSA, PSA density). This is because they are not routinely used in all men with a tPSA/DRE positive result and are, therefore, not strongly associated with biopsy uptake. A second known bias relates to a subset of studies that limit their reporting to men in the "grey zone" of tPSA measurements, usually defined as between 2.5 and 10 ng/mL.[20, 70, 79] These studies will underestimate the performance of tPSA compared to studies that include all men with elevated results. We accounted for this bias by stratifying results, when possible.

Intermediate Outcome: Diagnostic Accuracy

Key Points

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 sensitivity, specificity (or the false positive rate equal to 1-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 population were included in the five analyses listed below.

- Area under the curve (AUC). Fifteen studies[20, 51, 65, 68, 70, 71, 74-80, 82, 86, 96] reported AUC estimates for tPSA and PCA3 in the same population and the difference of the two [AUC(PCA3) – AUC(tPSA)] was computed. Overall, 13[20, 51, 68, 70, 71, 74, 75, 77-80, 82, 96] of the 15 studies found a positive difference. The two[65, 76] studies finding tPSA elevations to have a greater AUC were among the four smallest studies. Removing the three studies[51, 65, 76] that restricted recruitment to the grey zone resulted in an AUC difference of 0.0750 in the remaining 12 studies.[20, 68, 70, 71, 74, 75, 77-80, 82, 96]
- Reported medians and standard deviations (SD). Four studies[20, 51, 76, 77] provided sufficient data for analysis, and none of these reported a logarithmic SD (most, if not all studies examining the distribution found both PCA3 and tPSA to be highly right skewed). These were estimated from inter-quartile ranges or ranges. The differences, reported as z-scores, indicated that two studies (both small) found tPSA to be slightly better than PCA3 at separating populations of positive and negative prostate biopsies, while two others found a larger difference in separation of these groups in favor of PCA3.
- Performance at a PCA3 cut-off score of 35. Seven studies[65, 66, 70, 71, 76, 77, 86] reported the sensitivity and specificity of PCA3 at this cut-off. We computed the difference in sensitivity (PCA3 – tPSA) when tPSA was held at the PCA3-related specificity. Six[65, 66, 70, 71, 77, 86] of the seven studies reported a positive difference (median 19.8 percent, range -23 to 31 percent) favoring PCA3.
- ROC Curves - Sensitivity / Specificity. Ten studies[20, 65, 66, 70, 71, 76, 77, 79, 80, 82] provide an ROC curve, or data representing an ROC curve for both markers. At a specificity of 50 percent, the difference in corresponding specificities (PCA3 – tPSA) was zero or positive for all 10. Removing the four studies that restricted recruitment to the grey zone resulted in a difference in sensitivity of 9 percent favoring PCA3 (range 0 to 39 percent).
- Regression analysis. Only one study provided sufficient data to apply the respective regression coefficients to create a relative odds ratio (OR) between the 25th and 75th centiles of the two distributions. A second study reported all but the inter-quartile range, and that was estimated from the first study so that both datasets could be evaluated. In both studies, the ratio of the ORs (PCA3 / tPSA) was greater than 1 (1.38 and 1.97). These two studies[20, 70] both restricted recruitment to the "grey zone" so these are likely to overestimate the relative superiority of PCA3 by underestimating tPSA performance.

Interpretation

The results of analyzing the literature regarding the matched analyses of PCA3 score versus extent of tPSA elevations is summarized in Table 8. A more complete description of how these data were computed is described fully in Appendix J. Table 8 compares the diagnostic accuracy of PCA3 scores and tPSA elevations to independently identify men who would have a positive

biopsy (prostate cancer). In the top half of Table 8, the false-positive rate (1-specificity) is held constant, while in the bottom half, the sensitivity (detection rate) is held constant. This display was chosen because an undetected cancer is not equivalent to an unnecessary prostate biopsy and, therefore, comparing a loss in sensitivity to a gain in specificity is difficult. In the last column is the difference between the two estimates (PCA3 – tPSA). When comparing the sensitivities (top half) this column contains the improvement in prostate cancer detection. When comparing the false positive rates, it contains the reduction in unnecessary biopsies occurring in men without prostate cancer. For example, assume that one would like to set test sensitivities to 85 percent. Using the row with 85 percent sensitivity (bottom half of Table 8), only 56 percent of men without cancer would be subject to biopsy with PCA3 testing (cut-off score of about 17). Using tPSA elevations, 71.5 percent of those men without cancer would be biopsied (cut-off of about 2.5 ng/mL). This means that using PCA3 instead of tPSA elevations, the same proportion of cancers might be detectable while performing 15 percent fewer biopsies. Additional tables at fixed PCA3 and tPSA cut-off levels, individual risks and positive and negative predictive values at several difference prostate cancer rates can be found in Appendix J.

Table 8. A comparison of PCA3 and tPSA elevations to identify men with prostate cancer with either the false positive rate, or the sensitivity held constant

	PCA3 Scores			tPSA Measurements		
1-Specificity (FPR)	**Cut-off (Score)**	**Sensitivity (DR)**		**Cut-off (ng/mL)**	**Sensitivity (DR)**	**Improvement in PCa Detection**
80%	9.2	95.8%		2.0	90.7%	5.1%
70%	12.3	92.0%		2.6	84.3%	7.7%
60%	15.8	87.2%		3.2	77.6%	9.6%
50%	20.0	81.1%		4.0	68.8%	14.0%
40%	25.1	73.7%		4.9	59.7%	14.0%
30%	32.5	63.8%		6.2	48.4%	15.4%
20%	43.4	51.6%		8.0	36.4%	15.2%
	PCA3 Scores			**tPSA Measurements**		
Sensitivity (DR)	**Cut-off (Score)**	**1-Specificity (FPR)**		**Cut-off (ng/mL)**	**1-Specificity (FPR)**	**Reduction in Biopsies**
95%	9.9	77.7%		1.5	88.2%	10.5%
90%	13.8	65.6%		2.1	78.2%	12.6%
85%	17.3	56.3%		2.5	71.5%	15.2%
80%	20.7	48.5%		3.0	63.6%	15.1%
70%	27.7	36.2%		3.9	51.2%	15.0%
60%	35.6	26.6%		4.9	40.3%	13.7%
50%	45.0	18.9%		6.0	31.2%	12.3%

DR: proportion of biopsy positive men with a value at or above the cut-off level ; FPR: proportion of biopsy negative men with a value at or above the cut-off level.

Study Characteristics

A total of 17 studies reported PCA3 and tPSA comparisons that could be used in one or more of the five matched analyses (Table 9). Three studies provided data for four analyses, three provided data for three, five provided data for two, and six provided data for only one of the analyses. Fourteen of the 17 studies (82 percent) were consistent in finding the matched PCA3 estimate better at identifying positive prostate biopsies than the corresponding tPSA estimate for all analyses performed using that study's data. In only three[51, 65, 76] of the 17 studies (18 percent), one or more analysis found the tPSA estimate to be higher than the matched PCA3 estimate (p=0.013, binomial distribution). These entries are highlighted in grey in Table 10. All study quality ratings were poor.

Of interest are the four studies that used an upper cut-off for tPSA elevations to define a "grey zone."[20, 66, 70, 79] When the tPSA is truncated, it will reduce the effectiveness of the marker to predict biopsy outcome. In Table 9, all four of these studies are associated with positive findings, with PCA3 being more discriminatory. If these four studies were to be removed from the analysis, 10 of 13 studies (77 percent) would find in favor of PCA3 compared to tPSA (p=0.09, binomial distribution).

Table 9. Summary results for the five analytic comparisons of PCA3 versus tPSA in matched populations of men having prostate biopsies

			Results for a matched comparison available[1]					PCA3 > tPSA
Author	**Year**	**Number**	**AUC**	**Mean/SD**	**PCA3>35**	**Sens/Spec**	**Reg**	**% (analyses)**
Wang[86]	2011	516	-	-	28.4%	-	-	100 (1)
Aubin[66]	2010	1072	-	-	12.9%	7%	-	100 (2)
Adam[65]	2011	105	-0.1389	-	3.1%	1%	-	67 (3)
Nyberg[76]	2010	62	-0.0543	-0.22	-23.3%	0%	-	0 (4)
Roobol[96]	2010	721	0.0540	-	-	-	-	100 (1)
Rigau[80]	2010	215	0.0580	-	-	39%	-	100 (2)
Ankerst[51]	2008	443	0.0580	-0.24	-	-	-	50 (2)
Hessels[74]	2010	336	0.0700	-	-	-	-	100 (1)
Ouyang[78]	2009	92	0.0800	-	-	-	-	100 (1)
Mearini[75]	2010	-	0.1180	-	-	-	-	100 (1)
Ochiai[77]	2011	105	0.1264	0.98	19.8%	11%	-	100 (4)
Ploussard[79]	2010	301	0.1350	-	-	21%	-	100 (2)
Deras[71]	2008	570	0.1390	-	21.4%	20%	-	100 (3)
Cao[68]	2011	143	0.1730	-	-	-	-	100 (1)
de la Taille[70]	2011	516	0.1840	-	31.0%	26%	1.38	100 (3)
Perdona[20]	2011	218	0.1920	0.38	-	25%	1.97	100 (4)
Schilling[82]	2011	32	0.2000	-	-	7%	-	100 (2)

[1]AUC = area under the curve for PCA3 minus the AUC of tPSA; Mean/DS = difference in separation between PCA3 scores and tPSA results, when expressed as z-scores; PCA3>35 = difference of the PCA3 minus the tPSA sensitivities at the specificity found for a PCA3 cut-off of 35; Sens/Spec = difference between PCA3 and tPSA sensitivity at a specificity of 50%; Reg = relative change in the PCA3 ORs (between the 25th and 75th centiles) and the corresponding tPSA ORs. The corresponding full analyses resulting in these summaries can be found on the following pages.

Shaded rows indicate studies in which tPSA outperformed PCA3 measurements in identify prostate cancer.

PCA3 and tPSA: Area Under the Curve

Fifteen studies[20, 51, 65, 68, 70, 71, 74-80, 82, 96] reported the diagnostic performance of PCA and extent of tPSA elevation among men with initially screen positive test results (elevated tPSA with or without positive DRE) to discriminate between positive and negative needle biopsy test results. These studies and related information are shown in Table 10, sorted by effect size. Two studies in which all individuals already had one negative biopsy[51, 79] were among the included studies in order to strengthen the potential bias by including men with previous biopsies. The studies are ordered by the difference between the matched AUC estimates of PCA3 minus tPSA (positive numbers indicate PCA3 performed better), ordered from smallest to largest effect size (negative differences indicate that tPSA had a higher AUC then PCA3).

All but two studies[65, 76] found the matched AUC point estimate for PCA3 higher than that for tPSA. Those two studies were among the three smallest reported with 62 and 105 enrollees, respectively. The largest effect size was reported by the smallest study of all[82] that provided matched results for only 32 men. The median AUC difference was 0.118 (range -0.139 to 0.200). Only seven[20, 51, 70, 75-77, 96] of the 15 studies reported the matched p-values comparing the two AUCs. Using these as a guide, at least the eight studies[20, 68, 70, 71, 75, 77, 79, 82] from row eight[75] to the end of the table are likely to have been statistically significant. No study reported a statistically significant lower performance for PCA3. The study reporting a difference of -0.139 did not report a p-value, but did provide the respective 95 percent confidence intervals on the PCA3 and tPSA AUC estimates. These overlapped, indicating the differences were not likely to be significant (0.705, 95 percent CI: 0.599 to 0.812; and 0.844, 95 percent CI: 0.765 to 0.910, respectively). None of the studies reported a confidence interval or standard deviation for the matched difference of the two AUCs. Although the AUCs for PCA3 and tPSA ranged widely (indicating relatively high heterogeneity), the variability of the differences seemed more consistent. This may be due to the requirement that only paired estimates of the AUCs be included in this analysis.

Three studies[65, 76, 51] employed an upper cut-off of 10 ng/mL essentially limiting their population to the so-called “grey zone.” In general, only 1 percent to 2 percent of biopsy negative men have tPSA levels over 10, while about 20 percent of biopsy positive men are in this range[98,99] Removing this subset from the overall population of men with positive tPSA/DRE will have the effect of reducing the ability of tPSA to predict positive prostate biopsies. Thus, one would expect these studies to show greater differences in favor of PCA3 than the remaining ones. These three studies focusing on the grey zone are highlighted in grey in Table 10. They are all near the bottom of the table, indicating that they do in fact find greater differences. The median difference in AUC for all of the included studies is 0.1180. If these three studies are removed, the AUC difference is reduced to 0.0750. Although not formally computed, the heterogeneity will also be expected to be reduced. One could argue that these three studies should be excluded, as they do not, technically, satisfy fully the inclusion criteria. However, they are included for two reasons. First, the performance in this subset is a closely related question that has clinical implications. Clinicians may want to act based solely on a very elevated tPSA, but use additional markers to evaluate the remaining “grey zone” patients. In addition, the analysis shows one source of heterogeneity that other meta-analyses may not have examined and in doing so helps demonstrate the validity of these analyses.

Table 10. Comparing PCA3 levels and extent of tPSA elevations in matched studies via AUC analysis to correctly diagnose prostate cancer, as defined by a positive needle biopsy

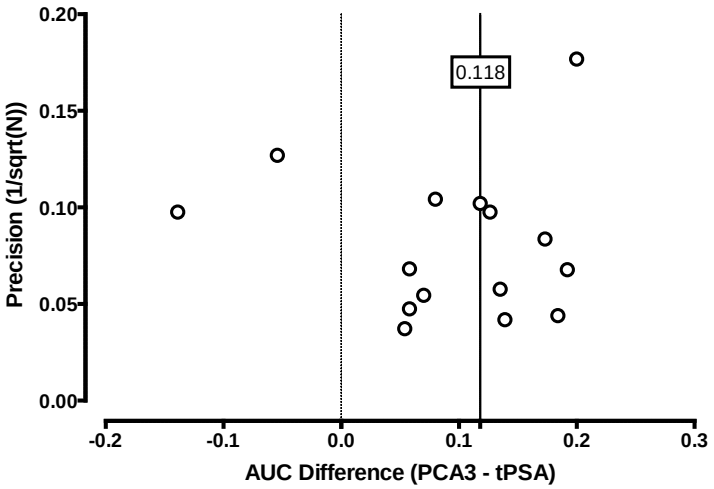
Author	Year	Number	Initial Bx	PCA3 AUC	tPSA AUC	Difference[1]	P-value[2]
Adam[65]	2011	105	82%	0.7054	0.8443	-0.1389	-
Nyberg[76]	2010	62	28%	0.7418	0.7961	-0.0543	0.07
Roobol[96]	2010	721	-	0.6350	0.5810	0.0540	0.14
Ankerst[51]	2008	443	0%	0.6650	0.6070	0.0580	>0.05
Rigau[80]	2010	215	-	0.6600	0.6020	0.0580	-
Hessels[74]	2010	336	-	0.7200	0.6500	0.0700	-
Ouyang[78]	2009	92	-	0.6700	0.5900	0.0800	-
Mearini[75]	2010	-	-	0.8140	0.6960	0.1180	<0.05
Ochiai[77]	2011	105	81%	0.8507	0.7243	0.1264	0.025
Ploussard[79]	2010	301	0%	0.6880	0.5530	0.1350	-
Deras[71]	2008	570	51%	0.6860	0.5470	0.1390	-
Cao[68]	2011	143	-	0.7390	0.5660	0.1730	-
de la Taille[70]	2011	516	100%	0.7610	0.5770	0.1840	<0.001
Perdona[20]	2011	218	61%	0.8280	0.6360	0.1920	<0.001
Schilling[82]	2011	32	56%	0.8100	0.6100	0.2000	-
Median (all data, N=15)						**0.1180**	
Median (only 'grey zone', N=3)						**0.1840**	
Median (excluding 'grey zone', N=12)						**0.0750**	

Bx = biopsy, AUC = area under the curve, '-' = no value reported
Shaded rows indicate studies focusing on the 'grey zone' of tPSA
1 All comparisons were between all subjects, regardless of repeat/initial biopsy status.
2 Reported p-value for the comparison of the two AUCs computed among the same set of men.
Shaded rows indicate studies focusing on the 'grey zone' of tPSA

An estimate of the potential for publication bias for this analysis can be generated under the assumption that the standard error of the AUC difference is proportional to the reciprocal of the square root of the number of enrolled men for each study.

Figure 7 shows a plot of the computed AUC difference (x-axis) versus its estimated precision (i.e., the reciprocal of the square root of the sample size [y-axis]). The solid vertical line shows the median difference of 0.118 while the dashed vertical line at 0.000 shows where the AUC would be equivalent. As predicted, the data fit a symmetric "inverted funnel" suggesting that at least some of the variability is due to the small samples sizes for several of the studies. The results seem far more consistent for the eight largest studies[20, 51, 70, 71, 74, 79, 80, 96] that provided matched results for 200 or more men.

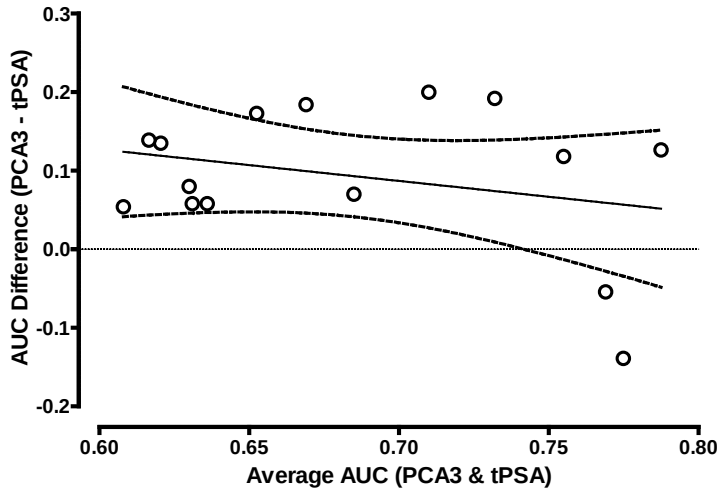
Figure 7. Examining the relationship between effect size (AUC difference) and an estimate of precision



In this analysis, precision is estimated by the reciprocal of the square root of N. The vertical dashed line indicates the points at which the two tests perform equally.

Figure 8 explores the relationship of the AUC difference, this time comparing the results against the average AUC (average of PCA3 and tPSA AUCs). The actual AUC for the markers may be indicative of extraneous factors (e.g., tPSA cut-off level, age of enrollees) that may be different between the 15 studies. Of interest, the two studies reporting the negative AUC differences were two[65, 76] of the three highest average AUCs. However, regression analysis shows no significant relationship between average AUC and the AUC difference (p=0.32). The median average AUC was 0.6690 (range 0.6080 to 0.7875).

Figure 8. The relationship between AUC difference and average AUC for PCA3 and tPSA



The solid line indicates the results of linear regression (AUC_Difference = -0.4032 * Avg AUC + 0.3692). The dashed lines show the 95% prediction limits. The slope is not significant (p=0.32).

Among the six of 15 studies that reported the racial/ethnic distribution in the study population,[51, 65, 68, 71, 77, 78] one was in an Asian (Japanese) population[77] and this group's AUC difference of 0.126 was slightly higher than the summary estimate of 0.118. Another Asian study (China) had an AUC difference estimate of 0.173. Only one study performed in South Africa reported on a population composed of a majority of black men (68.6 percent)[65], and the group's AUC difference of -0.140 was the lowest observed in all studies (Table 3). The five North American studies reporting a small black population (5.4,[89] 5.3,[71, 86] and 2 percent[51, 78]) had AUC differences near the consensus estimate.

Two manufactured kits (Aptima, Progensa), and one in-house methodology (quantitative RT PCR) were used to test for PCA3 in 11 of the 15 studies (73 percent). Among the remaining four studies, three did not report the assay method, and one used the "Gen-Probe" test. The median AUC differences for the three commonly used PCA3 assay methods were 0.126, 0.054 and 0.080, respectively. Two studies that found a negative AUC difference used the Progensa assay methodology.[65, 76]

Each reviewed study was assigned a QUADAS quality score of good, fair, or poor.[100] Among the 15 included studies[20, 51, 65, 68, 70, 71, 74-80, 82, 96] in the AUC differences computation (Table 10), all were rated poor. Only one of the studies was blinded in both directions (i.e., laboratory blinded to outcome, and clinicians blinded to laboratory results) and only two blinded in a single direction (one in each direction).

PCA3 and tPSA: Reported Medians and Standard Deviations

Seven studies (Table 11) reported some information concerning the distributions of PCA3 and tPSA elevations among men with screen positive test results (elevated tPSA with or without positive DRE) who subsequently had positive or negative needle biopsy test results. Results from these studies and related information are also shown in Table 11, sorted by effect size. The distributions of both markers are highly right skewed and have been shown to be reasonably Gaussian after a logarithmic transformation. For this reason, we chose to include for analysis only those studies in which the median or logarithmic mean could be determined along with the logarithmic standard deviation. In some instances the standard deviation was estimated using reported centiles (e.g., inter-quartile range). If a study only reported the range, the standard deviation was computed assuming the range represented 6 standard deviations.[101] For each study, the difference in marker levels in those with positive or negative biopsies was expressed as a z-score, using a study-specific pooled standard deviation.

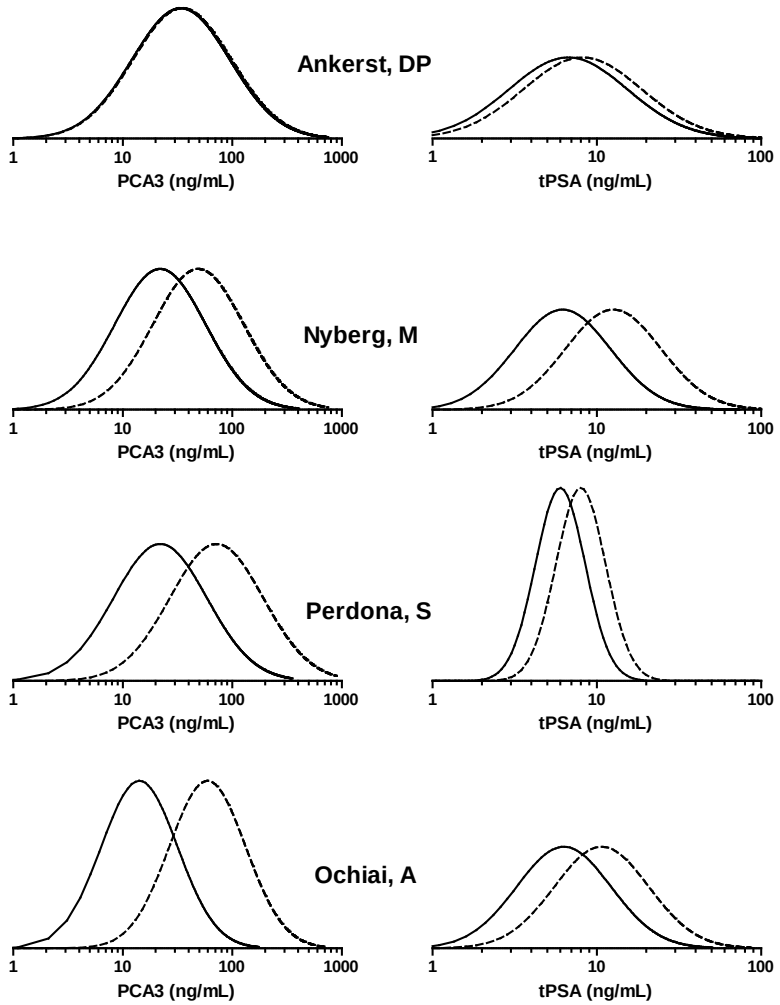
It was possible to obtain a median and pooled log standard deviation for both markers using data from four studies[20, 51, 76, 77] (Table 11). Three additional studies[66, 70, 74] had partial data and these have also been summarized in Table 11. For each of the four studies, the z-score for tPSA was subtracted from the PCA3 z-score to give a summary comparator (last column in Table 11). In one study,[51] PCA3 was not found to not have any separation, and that study also found the tPSA separation to be small (z=0.24). The overall estimate was a z-score of -0.24, indicating that tPSA has better separation than PCA3. This was also the finding of the second study[76], although that study found both markers to perform better. The second two studies found that PCA3 was consistently higher than tPSA measurements, with overall estimates of 0.38[20] and 0.98.[77]

Table 11. Comparison of PCA3 and tPSA differences in central estimates in men with positive and negative prostate biopsy results, after accounting for study-specific variability in measurements

			PCA3 Scores				tPSA Results (ng/mL)				
			Median for Bx		Pooled	Z_{PCA3}[1]	Median for Bx		Pooled	Z_{tPSA}[1]	
Author	**Year**	**N**	**Pos**	**Neg**	**Log SD**		**Pos**	**Neg**	**Log SD**		$Z_{PCA} - Z_{tPSA}$
Ankerst[51]	2008	443	34.3	34.2	0.4480	0.00	8.2	6.7	0.3604	0.24	-0.24
Nyberg[76]	2010	62	49.0	22.0	0.4150	0.84	12.6	6.2	0.2918	1.06	-0.22
Perdona[20]	2011	218	72.0	22.0	0.4264	1.21	8.0	6.0	0.1514	0.83	0.38
Ochiai[77]	2011	105	59.5	14.2	0.3489	1.78	10.7	6.3	0.2875	0.80	0.98
Hessels[74]	2010	336	50.0	18.0	-	-	8.3	5.9	0.4836	0.31	-
Aubin[66]	2010	1072	33.8	16.7	-	-	-	-	-	-	-
de la Taille[70]	2011	516	50.0	18.0	-	-	5.8	5.2	-	-	-

Pos = prostate biopsy positive, Neg = prostate biopsy negative
Z score = (log (Pos median) – log (Neg median)) / pooled log SD
Shaded rows indicate studies focusing on the "grey zone" of tPSA

Figure 9. Distributions of PCA3 and tPSA in men with positive and negative prostate biopsies from four studies that reported such information in the same cohort



The logarithmic means and standard deviations needed for these figures were obtained from Table 11. The solid line indicates the distribution for those with a negative biopsy, while the dashed line represents the distribution in those with a positive biopsy. For each study (indicated by the first author), these distributions were obtained in the same group of men. The figures are sorted by the relative performance of PCA3 over tPSA.

However, given the need to estimate the population distribution parameters for these markers (Appendix J), the analyses may not accurately reflect the findings of these studies. Had they actually reported the median and logarithmic standard deviations for their populations, the results may have been considered more reliable. Figure 9 shows the overlapping distributions from the four studies, derived from the data in Table 11. The analyses were performed across each set of paired distributions. Given only four studies were analyzed, it was not possible to stratify results by race, region or test methodology. The three studies that truncated results above 10 ng/mL are shown in grey (Table 11). Truncation is the likely cause for the distributions to have such a small standard deviation for tPSA in Table 11.[20]

PCA3 and tPSA: Performance at a PCA3 Cut-Off Score of 35

Seven studies[65, 66, 70, 71, 76, 77, 86] listed in Table 12 reported the sensitivity and specificity of PCA3 score at a cut-off of 35 among men with positive initial screening results (elevated tPSA with or without positive DRE) who subsequently had positive or negative needle biopsy test results. These studies and the corresponding sensitivity of tPSA at the same specificity found for the PCA3 cut-off level are shown in Table 12, sorted by effect size. The difference in the two sensitivities (with the specificity held constant) is a comparison of the ability to distinguish prostate cancer between the two markers. In some instances, the tPSA results were estimated from a published ROC curve. In two studies, the specificity was incorrectly reported as 1-specificity, or vice versa. The most appropriate analysis comparing two tests on the same population is to use a matched analysis of the 2x2 table. However, all of the seven studies reported only independent evaluations of each marker.

Among the seven studies (Table 12), the PCA3 score cut-off level of 35 were associated with false-positive rates (1-specificity) ranging between 20 percent and 45 percent with corresponding sensitivities (detection rates) range from 48 percent to 77 percent. For each study, the corresponding sensitivity for tPSA (at the same false positive rate) was subtracted from the PCA3 sensitivity. For one study,[76] the difference was negative, while the six remaining studies showed PCA3 having higher sensitivities ranging from 3 percent to 31 percent. The median increase in sensitivity was 19.8 percent. Since the 2x2 matched tables were not available, the statistical significance was estimated in the following way. For the two smallest studies,[65, 76] all possible 2x2 tables were computed and the p values assigned. The p value associated with the "middle" table was taken as the overall significance. This is likely to be appropriate, given the knowledge that there is some correlation between the markers. For the larger tables, even the most negative of assumptions would still result in a significant finding in the direction of PCA3. For two studies, neither of these methods produced a reliable estimate of significance.[66, 77]

Given only seven studies were analyzed, it was not possible to stratify results by race, region or test methodology. Of interest, however, is that the two studies finding PCA3 to be the least useful were both quite small with only 18[76] and 12[65] positive biopsy results, leading to wide confidence intervals on the sensitivity estimates.

Table 12. A comparison of PCA3 and tPSA in identifying a positive prostate biopsy among matched studies: difference in sensitivities at the fixed specific associated with the commonly used PCA3 score cut-off of 35

			Prostate Biopsy			PCA3 score			tPSA[1]	Difference
Author	Year	Number	Initial	Positive		1-Spec (%)	Sens (%) A		Sens (%) B	(A-B) (%)
Nyberg[76]	2010	62	45.0	29.0		45.4	66.7		90.0	-23.3 (NS)
Adam[65]	2011	27	18.1	42.9		35.1	77.1		74.0	3.1 (NS)
Aubin[66]	2010	1072	100	17.7		21.4	48.4		35.5	12.9 (?)
Ochiai[77]	2011	105	19.0	36.0		25.4	74.3		54.5	19.8 (?)
Deras[71]	2008	570	50.3	36.0		26.1	53.9		32.5	21.4 (S)
Wang[86]	2009	187	27.3	46.5		20.0	52.9		24.5	28.4 (S)
de la Taille[70]	2011	516	-	40.0		24.0	64.0		33.0	31.0 (S)

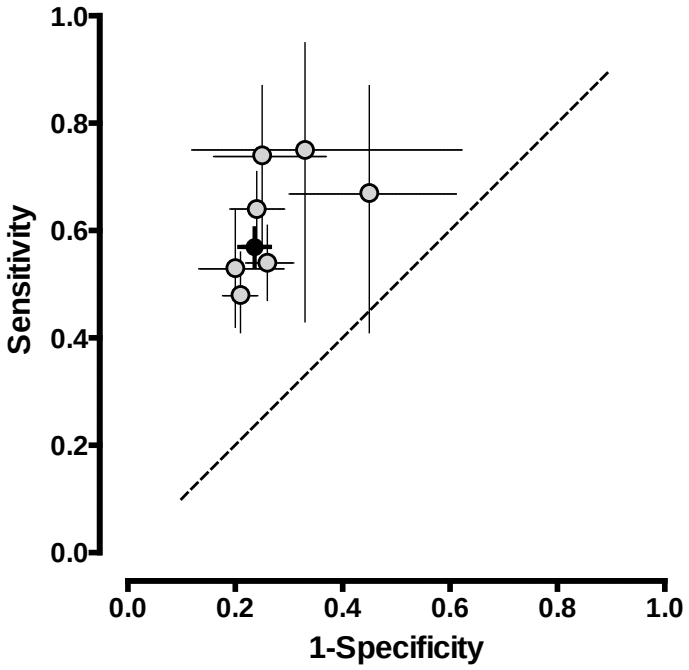
Spec=specificity (1-specificity=false positive rate), Sens = sensitivity (detection rate), NS = not significant, S = significant (see text), ?= significance level cannot be estimated.

Shaded rows indicate studies focusing on the 'grey zone' of tPSA

[1] Sensitivity for tPSA elevation at the same false positive rate (1-specificity) found for a PCA3 score at a cut-off of 35.

As a way of estimating whether the seven studies were reasonably consistent in their estimates of sensitivity/specificity, a simple random effects meta-analysis was performed for PCA3 (Figure 10). The weighted consensus for sensitivity and 1-specificity are 56.9 (95 percent CI: 53.3 to 60.4) and 23.6 percent (95 percent CI: 21.6 to 25.6 percent), respectively. Assuming no threshold effect (the PCA3 cut-off score was held constant), the diagnostic odds ratios associated with these studies range from 2.4 to 8.2, with a summary estimate of 4.17 (95 percent CI: 3.26 to 5.34). Heterogeneity is moderate (I^2=28 percent). This analysis is not biased by the tPSA upper cut-off of 10 ng/mL used in two of the studies[66, 70] as PCA3 is independent of tPSA measurements.

Figure 10. The performance of PCA3 score to identify subsequent positive prostate cancer biopsies from seven studies



The sensitivity (y-axis) versus the false positive rate (1-specificity, x-axis) is shown for the seven studies (open circles) along with the weighted consensus estimate (dark circle). Solid lines indicate the 95% confidence intervals of the estimates. The dashed line indicates where the sensitivity and 1 - specificity are equal.

PCA3 and tPSA: ROC Curves - Sensitivity / Specificity

Ten studies[20, 65, 66, 70, 71, 76, 77, 79, 80, 82] (Table 13) provided ROC curves for both PCA3 and for tPSA elevations among men with positive initial screening test results (elevated tPSA with or without positive DRE) who subsequently had positive or negative needle biopsy test results. These studies and the corresponding performance of tPSA testing are presented in Table 13. In each study, the sensitivities of each marker at preselected false positive (1-specificity) rates were estimated from published ROC curves. These values were recorded to the nearest percent (e.g., sensitivity of 55 percent).

Table 13 shows details of test performance from the 10 included studies, sorted from smallest to largest number of enrolled men. The table entries showing test performance are the PCA3 sensitivity, followed, in parentheses, by the incremental increase, or decrease, of tPSA sensitivity. For example, at a false-positive rate (1-specificity) of 20 percent, the first study found

a PCA3 sensitivity of 50 percent, with the corresponding difference in tPSA sensitivity being -19 percent, (i.e., 69 percent). Negative numbers indicate that tPSA is performing better; positive numbers indicated PCA3 is performing better. The last line in Table 13 is the median of these differences, ignoring matching. That is, the median PCA3 sensitivity is provided along with the median difference computed separately. Of interest is that both studies that found tPSA to be better[65, 76] (negative differences) along with the study reporting the largest advantage to PCA3[82] are the three smallest studies. All studies of 200 or more subjects tend to agree on the PCA3 sensitivity and the relatively advantage over the range of (1-specificities).

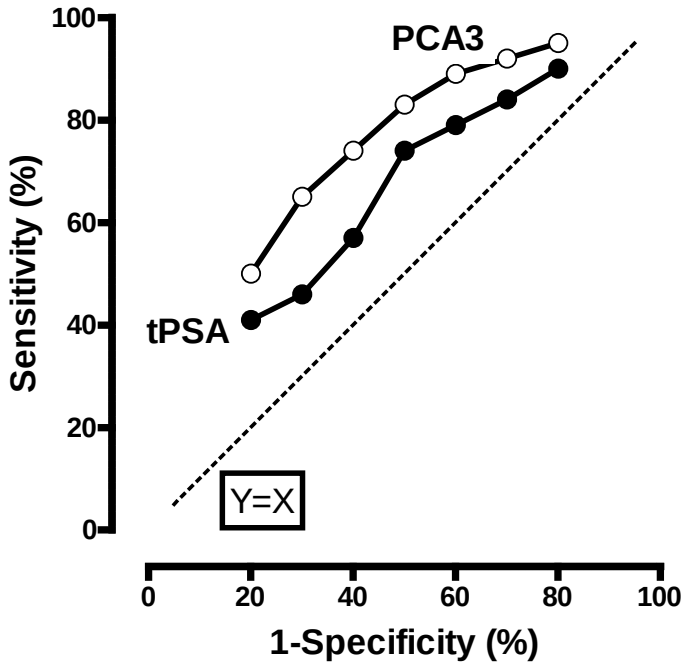
Table 13. Matched studies of PCA3 and tPSA testing performance to identify positive biopsy men

			PCA3 Sensitivity (Difference)[a] at PCA3 (1-Specificity) Rates of:						
Author	**Year**	**Number**	**20%**	**30%**	**40%**	**50%**	**60%**	**70%**	**80%**
Adam[65]	2011	27	50 (-19)	68 (- 7)	75 (0)	78 (1)	78 (- 2)	81 (1)	90 (0)
Shilling[82]	2010	32	57 (19)	71 (29)	90 (45)	95 (39)	95 (17)	100 (6)	100 (6)
Nyberg[76]	2010	62	50 (-11)	61 (-11)	73 (- 4)	89 (0)	94 (- 3)	94 (- 6)	94 (- 6)
Ochiai[77]	2011	105	74 (23)	87 (30)	92 (20)	97 (7)	98 (9)	100 (8)	100 (8)
Rigau[80]	2010	215	32 (- 2)	57 (15)	64 (14)	77 (11)	84 (11)	87 (9)	93 (10)
Perdona[20]	2011	218	70 (24)	79 (21)	82 (21)	88 (25)	97 (20)	100 (17)	100 (12)
Ploussard[79]	2010	301	40 (12)	56 (21)	71 (25)	80 (21)	83 (13)	89 (11)	93 (11)
de la Taille[70]	2011	516	57 (27)	71 (29)	77 (27)	85 (26)	89 (21)	93 (17)	95 (7)
Deras[71]	2008	570	45 (21)	55 (22)	63 (19)	74 (20)	82 (25)	89 (13)	95 (5)
Aubin[66]	2010	1072	49 (13)	60 (18)	68 (11)	74 (7)	79 (5)	85 (2)	94 (3)
Median (all data, N=10)			**50 (16)**	**65 (21)**	**74 (19)**	**83 (15)**	**87 (12)**	**91 (9)**	**95 (7)**
Median (only 'grey zone', N=4)			**53 (19)**	**65 (18)**	**74 (23)**	**83 (23)**	**86 (16)**	**91 (14)**	**95 (9)**
Median (excluding 'grey zone', N=6)			**50 (9)**	**65 (18)**	**74 (17)**	**83 (9)**	**89 (10)**	**91 (7)**	**95 (5)**

[a](PCA3 sensitivity – tPSA sensitivity) when (1-specificity) is held constant at values ranging from 20% to 80%.
Shaded rows indicate studies focusing on the 'grey zone' of tPSA. It also indicates the summary line where the results of those studies have been removed.

Figure 11 displays the summary ROC curves using the PCA3 median sensitivities and median differences provided in the last row of Table 13 This can be taken as a simple summary of performance for the two tests, under similar circumstances. The second to the last line in Table 13 shows the summary results only if the four studies[20, 66, 70, 79] having an upper limit set for tPSA were to be used. On average, the PCA3 sensitivity does not change, but the difference for tPSA sensitivity is smaller, but PCA3 sensitivities are consistently higher.

Figure 11. Summary ROC curves for PCA3 and tPSA based on studies applying the tests to the same population



Data are presented in Table 13. The open circles (solid line) indicates performance of PCA3 scores, while the filled circles (solid line) indicated tPSA performance. The dashed line indicated where the sensitivity equals 1-specificity, indicating a test with no predictive ability.

PCA3 and tPSA: Regression Analysis

Two studies[20, 70] reported sufficient results of regression analysis separately for PCA3 and for tPSA elevations in the same population of men to be included in these analyses. Each of the studies reported the odds ratio (OR) for each marker when that marker was assumed to be a continuous variable. That is, the antilog of the OR will be the regression coefficient per unit increase of the marker (e.g., increase of PCA3 score from 30 to 31). This makes comparison of PCA3 and tPSA difficult, as the range of results for the two markers differs (Table 14). To account for this, the coefficients will be used to compute the ratio of the ORs at the 25th and 75th centiles for each marker. This is a measure of the change in odds over the inter-quartile range. This ratio of ORs for PCA3 will then be divided by the corresponding ratio for tPSA. Values greater than 1 indicates that PCA3 provides more discrimination than tPSA. This normalization also allows for comparisons between studies, where the coefficient is dependent on the range of tPSA values studied. Both studies included in this analysis restricted tPSA to <10 ng/mL).[20, 70]

Only one of the included studies[20] provided the inter-quartile ranges for both markers. It was necessary to estimate those ranges for the second study.[70] For PCA3, this was done by extrapolating the log mean and SD from two centiles provided as part of the sensitivity/ specificity results. For tPSA, this was done by using the inter-quartile range from the first study[20] and adjusting for a minor difference in the mean values reported. Both of the studies restricted the upper limit of tPSA to <10 ng/mL.

Two additional studies provide some further insight. One[66] showed similar coefficients for PCA3 and tPSA, but it was not possible to compute the ratio of the ORs for tPSA as no data were provided to estimate the 25th and 75th centiles. However, given that the inter-quartile range of PCA3 scores are generally larger than the corresponding range of tPSA results, these coefficients are likely to have shown an overall finding of PCA3 being more discriminatory. Another study[68] provided only the continuous OR estimates. The PCA3 OR was the highest reported among the four studies in Table 14, and the corresponding OR for tPSA was slightly under 1.0. This would have to be associated with PCA3 being more discriminatory, but estimates of effect size cannot be provided because information about the distributions were not provided.

Five studies[20, 66, 70, 77, 68] provided some information on the independence of PCA3 and tPSA as markers for prostate biopsy status. Two specific measures were sought. The most useful would be bivariate correlations (parametric or non-parametric) between the two markers for those with positive, and for those with negative, prostate biopsies. Alternatively, logistic regression coefficients (or the corresponding ORs) with, and without, adjusting for tPSA were evaluated. In many of the studies reporting logistic regression models, additional factors such as history, prostate volume or other variables were also included. If both PCA3 and tPSA coefficients remain essentially constant after adjusting for the other marker (and possibly additional markers), this can be taken as evidence that the two markers together are more predictive that either alone (independent).

Three studies[20, 68, 70] reported information on correlation coefficients (Table 15). One reported the two correlation coefficients (non-parametric estimates), one reported a single merged correlation (parametric), and the third just reported that the correlations were ‘low’ for both groups. One potential problem with these estimates is that reliable correlation estimates for both PCA3 and tPSA would require a logarithm transformation prior to computing a parameter estimate such as the Pearson’s correlation coefficient. None reported that the data were

transformed. Overall, it appears that the two markers are not highly correlated in either of the groups of interest.

Five studies[20, 66, 68, 70, 77] provided information on coefficients for PCA3 and/or tPSA from univariate and multivariate logistic regression modeling (Table 15). Four of the five[20, 66, 70, 77] found the PCA3 coefficients unchanged after accounting for tPSA (and often other variables as well). The remaining study[68] found a reduction in the coefficient, but it was still the most significant predictor. In addition, this study did not include tPSA in the multivariate model, as it was not statistically significant in the univariate logistic regression (p=0.08). The results were less consistent for tPSA. Three studies[20, 66, 70] found tPSA essentially unchanged after accounting for PCA3. Interestingly, these three studies all restricted tPSA levels to under 10 ng/mL. This may reduce the correlation between the two markers, if PCA3 and tPSA are concordant when tPSA elevations are relatively high. A fourth,[77] did not report the coefficients but did report the p-value was reduced from being highly significant (p<0.001) to no significance (p=0.52). The fifth did not report results for tPSA after adjustment, as it only included variables found to be significant in univariate modeling.

Table 14. Comparison of modeled univariate continuous odds ratios (OR) for PCA3 and tPSA in matched studies

			PCA3 Odds Ratio (OR)				tPSA Odds Ratio (OR)				Ratio
Author	**Year**	**Number**	**report**	**@25th**	**@75th**	**Ratio[A]**	**report**	**@25th**	**@75th**	**Ratio[B]**	**(A/B)**
Perdona[20]	2011	218	1.030	16	70	4.93	1.258	5.0	9.0	2.50	1.97
de la Taille[70]	2011	516	1.020	13	53	2.16	1.120	4.0*	8.0*	1.57	1.38
Aubin[66]	2010	1072	1.019	9	35	1.63	1.106				
Cao[68]	2011		1.075				0.964				

* Values estimated from those reported in Perdona (16) after subtracting 1 (the difference in mean tPSA values in those with positive versus negative biopsies). Both studies only included patients between 2.5 and 10 ng/mL.

Table 15. Measures of independence of PCA3 and tPSA in identifying men having a positive biopsy finding in matched studies

			Corr (PCA3/tPSA)		PCA3 OR	tPSA OR	Other relevant findings
Author	**Year**	**Number**	**Bx Pos**	**Bx Neg**	**(raw, adj)**	**(raw, adj)**	
Perdona[20]	2011	218	'low'	'low'	1.030/1.030	1.258/1.239	Accuracy improves 3%
Aubin[66]	2010	1072	-	-	1.019/1.015	1.106/1.087	AUC improves; 69 - 75%
de la Taille[70]	2011	516	0.042		1.020/1.010	1.120/1.150	Accuracy improves 5.5%
Ochiai[77]	2011	105	-	-	P<0.001/<0.001	P<0.001/0.52	-
Cao[68]	2011	131	-0.079	-0.372	1.075/1.055		-

Bx = prostate biopsy, Corr = Correlation coefficient, OR = odds ratio, raw = as observed, adj = after adjustment for other markers
Shaded rows indicate studies focusing on the 'grey zone' of tPSA.

PCA3 and tPSA Elevations: Diagnostic Accuracy

PCA3 and tPSA GRADE Strength of Evidence: LOW

- Risk of Bias: Moderate.
  Two biases were identified that could potentially impact this analysis: partial verification bias, and a sampling bias (some studies restrict the range of tPSA measurements to be enrolled). Partial verification bias was clearly present for tPSA elevation, but our analyses and a review of the literature indicated that in this setting, the ROC curve was unlikely to be biased (Appendix J). Thus, the focus was towards the ROC and related measurements. Monte Carlo modeling was used to account for the verification bias related to the specific cut-off level at which a certain performance was obtained (Appendix J). The second bias was accounted for by stratifying the analyses, when possible. Although there was a relatively high potential for bias to affect select measurements and their interpretation, the measures taken as part of the analyses result in a low risk of those biases influencing the final interpretation. Publication bias was informally evaluated and not considered to be an important source of potential bias. However, given the poor quality of all included studies, there is potential for unidentified biases to have occurred.
- Consistency: CONSISTENT
  Due to the matched nature of our analyses, it was not possible to formally test for heterogeneity, as original data were not available. No study actually reported a matched analysis.
- Directness: INDIRECT
  Diagnostic accuracy, an intermediate outcome, can be considered a direct outcome. However, some biopsies may more indicative of serious existing (or future) disease than others. None of the included studies provided evidence that the positive biopsies identified by PCA3 were at least as serious as those identified by tPSA.
- Precision: PRECISE
  A formal analysis of precision (e.g., confidence intervals) was not able to be computed due to the matched nature of our analyses and the lack of original data. In one analysis that included 15 studies (AUC difference), it was possible to see the reduction in performance for tPSA in that subset of ‘grey zone’ studies, where the AUC difference expanded to a median of 18 percent, compared to the 7.9 percent found in the studies with no sampling bias. Overall, analysis showed that PCA3 measurements had higher sensitivity at any specificity compared to tPSA, and higher specificity at any sensitivity.
- Dose-Response Relationship: PRESENT
  In the largest analysis of studies (after excluding those in the ‘grey zone’) 10 of 12 AUC differences indicated PCA3 to be better (median +0.079). That analysis was also able to see the reduction in performance for tPSA in that subset of ‘grey zone’ studies, where the difference expanded to a median of 0.18 percent.
- Strength of Association: WEAK
  Although there is evidence that PCA3 will be slightly better at identifying high risk individuals with a prostate cancer, both PCA3 and tPSA are relatively weak predictors with low sensitivity and low specificity.

Summary of the Remaining Key Questions 1 and 2 "Combined" Analyses

In this combined set of analyses PCA3 was to be compared with a total of six comparators (tPSA being one) against three intermediate outcome (diagnostic accuracy being one) and one long term outcome (long term morbidity/mortality). This results in a total of 24 comparisons. Table 16 showed that for diagnostic accuracy, only tPSA and three other comparators have any available data. No reports were identified that provided data on any of the six markers and the other two intermediate outcomes of interest. Table 16 shows a summary of the numbers of studies available for each of the 18 analyses, along with the strength of evidence findings. More detail regarding the limited findings for diagnostic accuracy (fPSA, PSA density and validated nomograms) can be found in Appendix K.

Table 16. Key Questions 1 and 2 "Combined" Analyses: Summary Studies Available and Strength of Evidence for the Six PCA3 Comparators versus the four Outcomes of Interest

	Outcomes of Interest							
PCA3 Comparators	**Diagnostic Accuracy**		**Patient Decisionmaking**		**Biopsy-related Harms**		**Long term Health**	
	Number of Studies	Strength of Evidence	Number of Studies	Strength of Evidence	Number of Studies	Strength of Evidence	Number of Studies	Strength of Evidence
tPSA elevations	17	Low	0	Insufficient	0	Insufficient	0	Insufficient
fPSA	4	Insufficient	0	Insufficient	0	Insufficient	0	Insufficient
tPSA velocity	0	Insufficient	0	Insufficient	0	Insufficient	0	Insufficient
PSA density	2	Insufficient	0	Insufficient	0	Insufficient	0	Insufficient
cPSA	0	Insufficient	0	Insufficient	0	Insufficient	0	Insufficient
Validated Nomogram	3	Insufficient	0	Insufficient	0	Insufficient	0	Insufficient

Strength of Evidence (Insufficient, Low, Moderate, High)

Key Question 3 – Testing for PCA3 and Comparators to Identify Patients with Low Risk/Indolent Cancer Who May Be Candidates for Active Surveillance

KQ3 presented a complex clinical scenario. 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 (Link B in the analytic framework, Figure 2). Variability was observed in the terminology and definitions of high risk/aggressive and low risk/indolent disease. Low risk tumors are variably referred to as "low risk,"[68] "indolent,"[70, 89] "insignificant,"[74, 94, 95] and "low volume/low grade."[92, 94] High risk tumors were referred to as "intermediate or high risk,"[68] "significant,"[70, 74, 92] "unfavorable,"[89] and "aggressive."[94] These endpoints are not clinical outcomes or legitimate surrogates for clinical outcomes, but rather risk categories, defined by the prognostic markers, that have been proposed as a clinical guide to decisionmaking about whether to proceed with treatment or active surveillance. What is primarily lacking in this literature is the formal evidentiary link between variable definitions of risk categories and subsequent management decisions and specific long-term clinical outcomes. Estimation of diagnostic accuracy for tests used to make clinical decisions requires the use of a specified clinical endpoint or validated surrogate. Therefore, estimation of diagnostic accuracy for this key question was not possible.

Description of Included Studies

The inclusion criteria for KQ3 were also set to select only matched studies. These are defined as studies that provide estimates of diagnostic test performance or other outcomes for PCA3 and at least one other comparator using the same sample set. Studies of PCA3 alone, or of other comparators without PCA3, were excluded. Eleven studies were identified that addressed KQ3 and reported on PCA3 and other preoperative and pretreatment markers for characterizing tumors based on biopsy or prostatectomy results.[67, 68, 74, 89-95]Three studies included data on biopsy results,[68, 70, 89] seven on prostatectomy results[67, 90-95] and one on both.[74] Two studies were conducted on subjects in an active surveillance program.[89, 90] Table 1 through Table 4 include descriptive information about these studies. Table 17 and Table 18 provide information on biopsy and prostatectomy marker results. Table 19 provides more detailed information about the wide variety of markers investigated in these studies for association with low and high risk disease.

Only two studies had a longitudinal component and describe a clinical outcome.[89, 90] None of the other biomarkers or pathological markers used met criteria for validated intermediate or surrogate outcomes.

- Lymph node involvement in a prostate cancer patient is an indicator of poor clinical outcome. One study[90] attempted to identify "micrometastases," based on identifying tumor cells within the lymph nodes that are producing the prostate cancer markers tPSA and PCA3. The method used to quantify these markers in lymph node extracts was real time reverse transcriptase PCR (RT-PCR) for both PCA3 and PSA mRNA. The study followed 120 patients with localized prostate cancer for four to six years and used biochemical recurrence (any serum tPSA greater than 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 to

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) RT-PCR testing. The RT-PCR PSA test had a sensitivity for biochemical recurrence of 73 percent and a false positive rate of 22 percent ($p<0.001$). The corresponding PCA3 test had a lower detection (42 percent) and a comparable false positive rate (23 percent), but the effect was not significant ($p=0.095$). While 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. The authors provided no information on validation of quantitative RT-PCR for these biomarkers in this sample type or on confirmation of the results using a published method (immunohistochemistry).

- Based on no more than 2-year followup of patients in an active surveillance program, Tosoian et al. reported PCA3 and tPSA results (mean, standard deviation, median) for the 38 of 294 patients progressing to treatment based on yearly biopsy results.[89, 90] Epstein criteria were used for initial enrollment in the surveillance program. Progression to treatment was recommended for "unfavorable" findings", defined as any Gleason pattern 4 or 5, greater than 2 positive biopsy cores, or more than 50 percent involvement of any core with cancer (modified Epstein criteria). No difference in PCA3 and tPSA levels was observed between the 13 percent 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, and 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).

All studies were judged to be poor quality, either because of missing followup to clinical endpoints, unclear data presentation and/or blinding. One study was funded by GenProbe[91] and the other[93] did not report on source of funding or conflicts of interest (Table 1). No studies were identified that reported on other intermediate outcomes (e.g., diagnostic accuracy, decisionmaking, harms) or long-term clinical outcomes (e.g., mortality/survival, morbidity, quality of life). The detailed results of assessment of quality of individual studies addressing KQ3 are presented in Table F-2 in Appendix F.

PCA3 and Comparators - Intermediate Outcome: Diagnostic Accuracy

No studies were identified that reported on matched data for PCA3 and comparator results, and also reported specific clinical outcomes of patients with tumors characterized as 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.

Strength of Evidence: Insufficient

PCA3 and Comparators - Intermediate Outcome: Impact on Decisionmaking

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 impacted decisions made with regard to selection of active surveillance versus aggressive treatment.

Strength of Evidence: Insufficient

PCA3 and Comparators - Intermediate Outcome: Treatment-Related Harms

Studies have been conducted that document treatment-related clinical harms such as incontinence, impotence, and prostatitis. Based on 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.

Strength of Evidence: Insufficient

PCA3 and Comparators – Intermediate and Long-Term Health Outcomes

No studies were identified that reported PCA3 and comparator results and the association of low and high risk categorization with long-term outcomes such as mortality/survival and morbidity (e.g., function, quality of life) of the selected course of treatment. However, two poor quality studies reported on relatively short-term health outcomes, biochemical recurrence and progression from surveillance to treatment in an active surveillance program.

Strength of Evidence: Insufficient

Strength of evidence was not evaluated. Only two studies were identified, and they did not study the same setting, have the same sample type or have comparable outcome measures.

.

Table 17. Characteristics of biopsy studies

Author, Year	Number of Biopsies	% Repeat Biopsies	Cores per Biopsy by Protocol*	% Positive Biopsies	Biopsy Gleason Scores Cutoff, %	Pathological Stage, %
Adam, 2011[65]	105	18.1	13	42.9	GS ≤ 6, 64; GS = 7, 27; GS ≥ 8, 9	NR
Ankerst, 2008[51]	443	100	NR	27.8	GS ≤ 6, 55.3; GS = 7, 33.3; GS = 8, 5.7; GS = 9, 1.6	NR
Aubin, 2010[66]	1072	100	10	17.7	GS = 6, 69.5; GS = 7, 28.9	NR
Cao, 2011[68]	131	NR	10	66	GS ≤ 6, 50; GS = 7, 38; GS ≥ 8, 12	≤T2a, 25; T2b, 52; T3, 23
de la Taille, 2011[70]	516	0	8	40	GS < 6. 1; GS = 6, 52; GS = 7, 42; GS >7, 5	T1c, 86; T2, 13; T3, 1
Deras, 2008[71]	570	50.3	10	36	GS ≤ 6, 54.9; GS = 7, 29.1; GS = 8-10, 12	NR
Fradet, 2004[72]	443	20.5	NR	34	GS = 6 or 7, 72	NR
Hessels, 2010[74]	336	NR	8	40%	GS < 8.2; GS =5,54; GS = 7, 26.7%; GS 8 - 5.2; GS>8 5.9%	T2,59; T3, 42
Mearini, 2009[75]	96	NR	NR	72.9	GS < 7, 64.3; GS = 7, 27.1; GS > 7: 6 (8.6)	NR
Nyberg, 2010[76]	62	45	NR	29	GS ≤ 6, 22; GS = 7, 56; GS ≥ 8, 22	NR
Ochiai, 2011[77]	105	19	8	36	GS = 6, 37; GS = 7, 37; GS = 8, 26	NR
Ouyang, 2009[78]	92	NR	NR	46.7	GS = 5 or 6: 83.7; GS = 7 or 8, 16.3	NR
Perdona, 2011[20]	218	38.5	12	33.5	GS ≥ 7, 37	NR
Ploussard, 2010[79]	301	100	10	23.6	GS ≤ 6, 65.5; GS = 7, 30.2; GS ≥ 8, 4.3	T2, 17.4
Rigau, 2010[80]	215	NR	10-18	34	GS <7: 24.7; GS= 7: 57.5; GS > 7: 17.8	NR
Roobol, 2010[96]	721	29.4	6	16.9	NR	NR
Schilling, 2010[82]	32	13.6	10	56	NR	NR
Tinzl, 2004[83]	158	26.5	NR	39	GS < 7, 60; GS=7, 22; GS > 7, 18	T2, 74; ≥T3, 26
Tosoian, 2010[89]	294	100	12	12.9	GS < 7, 57.9; GS ≥ 7, 42	NR
Wang, 2009[86]	187	27.3	12	46.5	GS=6, 36.8; GS= 7, 37.9; GS ≥ 8, 25.3	T1c, 69; T2a, 22; ≥T2b, 9

GS, Gleason Score; NR, Not Reported; SD = standard deviation

Table 18. Characteristics of prostatectomy only studies

Author, Year	N RP[1]	Biopsy Gleason Score (GS) Cutoffs, %	RP Gleason Score Cutoffs, %	Pathologic Stage, (%)	Low Risk Prostate Cancer Category, %	Other Post-RP Pathological Parameters[6] Reported
Auprich, 2011[67]	305	GS <7, 50.8 GS ≥7, 49.2	GS <7, 27.9 GS ≥7, 72.1	NR	Insignificant[2] 10	ECE, SVI, TV ≥ 0.5
Hessels, 2010[74]	70	GS <7, 61 GS ≥7, 39	GS <7, 43 GS ≥7, 57	pT2,59 p T3, 42	Low volume/low grade	ECE
Kusuda, 2011[90]	120	NR	GS ≤ 6, 56.7 GS = 7, 33.3 GS ≥ 8, 10.0	pT2, 55.8 pT3, 42.5 pT4, 1.7	NR	NR
Liss, 2011[91]	98	GS ≤ 6, 53.1 GS > 6, 47.9	GS ≤ 6, 29.9 GS > 6, 71.1	pT2, 77.3 pT3a, 16.5 ≥pT3b, 6.2	NR	ECE, PNI
Nakanishi, 2008[92]	96	GS = 6, 40.6 GS = 7, 53.1 GS = 8, 6.3	GS = 6, 15.6 GS = 7, 77.1 GS = 8, 3.1 GS = 9, 4.2	pT2, 82.3 pT3a, 10.4 pT3b, 7.3	Low volume / low grade 11[3]	TV ≥ 0.5
Ploussard, 2011[93]	106	GS =6, 100	GS = 6, 58.4 GS = 7, 41.6	pT0, 0.9 pT2, 75.5 pT3, 23.6	Criteria for active surveillance[4] 18.9	TV ≥ 0.5
Vlaeminck-Guillem, 2011[94]	102	GS = 6, 47 GS = 7, 45 GS = 8/9, 8	NR	NR	Low volume / low grade[5] 8.2	TV ≥ 0.5
Whitman, 2008[95]	72	GS = 6, 71.2 GS = 7, 20.8 GS = 8/9, 9.7	GS = 6, 58.3 GS = 7, 32.0 GS = 8/9, 9.7	pT2, 70.9 pT3a, 20.8 pT3b, 8.3		ECE, TV ≥ 0.5

1 RP, radical prostatectomy;
2 Defined by Epstein criteria: organ confined, tumor volume less than 0.5 cc, absence of Gleason grade 4 or 5.
3 Defined as organ confined, dominant tumor volume less than 0.5 cc, and absence of Gleason grade 4 or 5.
4 Defined by Epstein criteria: organ defined, tumor volume less than 0.5 cc, absence of Gleason grade 4 or 5.
5 Organ confined cancer (pT2 or less) with a total tumor volume of less than 0.5 ml, and absence of Gleason grade 4 or 5 disease.
6 ECE, ExtraPCasular Extension; PNI, Perineural Invasion; SVI, Seminal Vesicle Invasion; TV, Tumor Volume; NR, Not Reported

Table 19. Comparators investigated along with PCA3 levels in studies addressing KQ3

	Auprich 2011 [67]	Cao 2010 [68]	de la Taille 2011 [70]	Hessels 2010 [74]	Kusuda 2011 [90]	Liss 2011 [91]	Nakanishi 2008 [92]	Ploussard 2011[79]	Tosoian 2010[89]	Vlaeminck-Guillem 2011[94]	Whitman 2008[95]
No. of biopsy patients	--	131	515	70	--	--	--	--	294	--	--
Number of prostatectomy patients	305	--	--	336	120	98	96	106	--	102	72
Biopsy											
Gleason score	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Biopsy clinical T stage	No	Yes	Yes	Yes	No	No	No	Yes	No	Yes	Yes
Percent tumor positive cores	No	No	Yes	No	No	No	Yes	Yes	No	No	No
Percent tumor per core	No	No	No	No	No	No	No	Yes	No	No	No
Prostate volume	No	Yes	No	Yes	No	No	No	Yes	No	Yes	No
Prostatectomy											
Gleason score	Yes	No	No	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes
Pathological stage	No	No	No	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes
Tumor volume	Yes	No	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Sets of criteria for identification of low risk tumors	Yes	No	Yes	Yes RP	No	No	Yes	Yes	Yes	Yes	No
Other markers											
Multifocality	No	NA	NA	No	No	No	No	No	No	Yes	No
Extracapsular extension	Yes	NA	NA	No	No	Yes	No	No	No	No	Yes
Seminal vesicle invasion	Yes	NA	NA	No	No	No	No	No	No	No	No
Perineural invasion	No	NA	NA	No	No	Yes	No	No	No	No	No
Surgical margin status	No	NA	NA	No	Yes	No	No	No	No	Yes	No
Lymph node mRNA expression	NA	NA	NA	NA	Yes	NA	NA	NA	NA	NA	NA

NA = Not applicable; RP = radical prostatectomy:

Discussion

Key Findings and Strength of Evidence

This 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 with high risk (i.e., aggressive) and low risk (i.e., indolent) prostate cancers. The first key question (KQ1) focused on men having an initial biopsy, while KQ2 focused on men with at least one prior negative biopsy. KQ3 focused on men with a positive prostate biopsy, in order 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. 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, decisionmaking) and long term (morbidity/mortality related to prostate cancer). All included studies were considered to be of poor quality. This discussion considered what was learned about PCA3 testing in the context of other potential markers (comparators) for these intended uses.

Key Question 1: Testing PCA3 and Comparators to Identify Prostate Cancer in Men Having an Initial Biopsy

Comparison of PCA3 to Six Comparators

Only two studies provided results from both PCA3 testing and a comparator marker in populations of men at increased risk for prostate cancer who were having an initial biopsy. Strength of evidence[64] was considered insufficient to derive any conclusions about relative performance or about the combination of PCA3 and one or more of the comparators. This includes all intermediate and long-term outcomes of interest.

Key Question 2: Testing PCA3 and Comparators to Identify Prostate Cancer in Men Having a Repeat Biopsy

Comparison of PCA3 to Six Comparators

Only three studies 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 includes all intermediate and long-term outcomes of interest.

Potential to Combine Studies Addressing Key Questions 1 and 2

Given the limited data available for KQ1 and KQ2, we examined the pool of studies that provided comparison data, but were not limited to only men having an initial biopsy or men with no previous positive biopsy having a repeat biopsy. An additional 12 studies provided usable

comparative data for PCA3 and tPSA, but the enrolled population consisted of a mixture of men in both categories. The proportion of enrolled men having an initial biopsy (100 percent would indicate a study suitable for KQ1, while 0 percent would indicate a study suitable for KQ2) was then plotted against two separate measures of relative effect size (difference in AUC and sensitivity at selected specificities). In both analyses, the correlation was essentially zero, indicating that neither of the two relative effect size measures was dependent on the mixture of initial versus repeat biopsies. This provided sufficient evidence to perform a "combined" KQ1/KQ2 analysis.

Comparison of PCA3 to Extent of tPSA Elevations: Diagnostic Accuracy

Based on results from 15 studies, PCA3 was more discriminatory for detecting prostate cancer than extent of tPSA elevations among men identified as being at risk. At any set clinical sensitivity, the clinical specificity of PCA3 testing is higher than that of tPSA. Conversely, at any set clinical specificity, the clinical sensitivity of PCA3 was higher than that of tPSA. These two biomarkers appeared to be independent in 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.

Comparison of PCA3 to Five Other Comparators: Diagnostic Accuracy

Limited data were available for fPSA (four studies), PSA density (two studies) and externally validated nomograms (three studies). No studies were found suitable for PSA velocity or complexed PSA. The strength of evidence for all five comparisons was considered insufficient.

Comparison of PCA3 to Six Comparators: Other Intermediate Outcomes

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 decisionmaking 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.

Comparison of PCA3 to Six Comparators: Long-Term Outcomes

The data were missing or inadequate for comparison of PCA3 testing to the other selected biomarkers with regard 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.

Key Question 3: Testing PCA3 and Comparators Among Men with a Positive Prostate Biopsy to Identify Low Risk/Indolent Patients Who May Be Candidates for Active Surveillance

The third Key Question focuses on intermediate and long term outcomes and the reference standard must also be a longer term clinical endpoint, in order to investigate outcomes in the context of categorization of risk. These endpoints might include measures of progression, metastasis, and prostate cancer related morbidity (e.g., function, quality of life) or mortality. This is unlike the situation in KQ1 and KQ2, where positive or negative prostate biopsy results were

the endpoint for diagnostic accuracy. A validated surrogate could be used in this situation, and we considered biochemical recurrence (specific rise in tPSA levels) a validated surrogate. It is commonly used as an indicator of risk for disease recurrence, but is not a highly sensitive and specific marker of prostate cancer. Progression from active surveillance to treatment appears to be a commonly used intermediate marker of overall disease progression.

Seven prospective cohorts studies of men in active surveillance are currently active.[52] The one partially informative study described in Results was derived from one of these seven ongoing studies.[89] More time will be needed for assessment of progression free survival, mortality and other long-term outcomes.[52] Given the relatively recent advent of PCA3 testing, it is not surprising that no studies were identified that provided intermediate or long term outcomes based on PCA3 and one or more comparators.

Findings in Relationship to What is Already Known

Systematic Review of PCA3

One prior structured systematic literature review of PCA3 testing and prostate cancer was identified.[102] That review covered the time period 2000 to 2009 and included only studies of diagnostic accuracy, with prostate biopsy as the gold standard. No comparisons were made to other biomarkers. Included were studies reporting on cohorts of adult men undergoing prostate cancer screening; mean tPSA levels were provided for each study, but elevated tPSA was not an inclusion criterion. Fourteen studies were included in the Ruiz-Aragon review[34, 71-73, 78, 83, 87, 92, 103-107] The four studies from our review that were also reviewed by Ruiz-Aragon were excluded from our review due to high likelihood of data duplication with other included studies.[34, 73, 87, 106]

Summary estimates of clinical sensitivity and specificity from the previously published meta-analysis (random effects model) were 63 percent (95 percent CI: 60-66 percent) and 75 percent (95 percent CI: 73 to 76 percent), respectively. Heterogeneity was high and significant ($p < 0.001$). The SROC AUC was 0.783. The authors acknowledged that no consensus existed on the most appropriate PCA3 cutoff for clinical decisionmaking. Based on QUADAS criteria,_ENREF_101 the authors reported the studies to be of moderate to high quality, in spite of the acknowledged lack of blinding. However, the review did not compare PCA3 performance with any other biomarkers, used reported sensitivity/specificity results from studies with varying PCA3 cutoffs (range, scores of 19 to 66), included studies using probability cutoffs, and included a study of prostatic fluid samples. These inclusion criteria may account for a portion of the observed heterogeneity.

Applicability

Key Questions 1 and 2: Biopsy Decisionmaking

To determine the effectiveness of PCA3 and comparator tests in predicting risk of prostate cancer at initial or repeat biopsy, or in risk categorization to inform decisions about treatment, it is useful to assess the applicability of the findings in this review.

Population and Settings of Care

The populations studied in these papers were largely drawn from academic medical centers where patients with elevated tPSA results and/or other risk factors (e.g., positive DRE, family

history, African American) are seeking referral or specialty care. Spectrum effects should be considered when evaluating data on diagnostic tests generated from referral centers. For example, these studies may have represented a group of men at higher risk of prostate cancer than the total cohort of men with elevated tPSA and/or other risk factors. The positive biopsy rate in such referral populations will depend on multiple factors, including the tPSA cutoff, the number of men with elevated tPSA who opt out of biopsy (e.g., men with lower tPSA levels and lower risk), and/or the proportions of men with other important risk factors. In the 17 included studies, biopsy positive rates ranged from 16.9 to 72.9 percent, with a median of 36 percent. It is unlikely that the clinical sensitivity and specificity estimates derived from these studies will be affected. However, positive and negative predictive values do depend on disease prevalence, and the reported predictive values will vary.

A second spectrum effect of more concern was potential difference in the severity of disease between a referral group and the total high risk cohort. If the more severe cases were more likely to be referred and/or accept biopsy, or if PCA3 was more likely to identify severe disease, clinical sensitivity could be overestimated. No studies addressed these potential biases.

Total PSA is currently the standard initial screening test for the identification of men at increased of prostate cancer. However, studies of proficiency testing data have shown that tPSA test kits provide variable results that could result in significant differences in proportion of identified cancers at recommended tPSA cutoffs. Therefore, an inherent limitation or source of heterogeneity for the PCA3 test may be the use of PSA for first order screening.

Our finding that the relative performance of PCA3 versus tPSA elevations is not dependent on the biopsy history is a new finding. The current FDA approval for PCA3 restricts it use to decision-making regarding a repeat biopsy in men with one or more negative biopsies. This review suggests that it might be just as useful in men making decisions about an initial biopsy. This would greatly broaden the settings in which PCA3 might be offered. Whether or not this occurs will involve other dynamics, such as costs and availability, as well as the validation of models incorporating PCA3 testing with other markers in the setting of the initially positive patient.

Interventions

No publications were identified that addressed the impact of adding results of PCA3 testing to the process of making decisions about proceeding to initial or repeat biopsies. Therefore, uncertainty remained about how the test would, or should, impact practice (e.g., all screen positive men, men in the “grey zone,” only for repeat biopsy decisions). Other issues that need clarification include: how the testing would be integrated into protocols for management of men with elevated tPSA or other risk factors, and whether physicians and patients are receptive to using PCA3 results in the biopsy decision.

Comparisons

For this intended use, analysis of 15 matched studies showed that PCA3 had slightly higher performance compared to the extent of tPSA elevations. However, this result is based on a summary analysis of matched within-study differences. Some studies reported the sensitivity and specificity for PCA3 and tPSA, but no studies actually reported a full matched analysis. A full matched analysis might provide 2x2 tables (or raw data for analysis) that allowed for direct comparison of PCA3 and a comparator in biopsy positive and biopsy negative samples (i.e.,

identifying those that both tests called positive, those both called negative, and those on which results were inconsistent).

Studies of PCA3 and comparators other than tPSA (e.g., free PSA, PSA density, PSA velocity or doubling time, complexed PSA, and externally validated nomograms) were inadequate or completely lacking. Even if PCA3 were a better secondary screening test than tPSA elevations (i.e., detected the same number of cancers, but with fewer biopsies in men without cancer), no data has shown that the identified men have the same type of prostate cancer (e.g., aggressive, indolent). Matched analyses of existing data could help answer this question.

Although the focus is generally on comparing PCA3 with other prostate cancer markers, neither PCA3 nor tPSA have high performance. To improve overall performance, both biomarkers and demographic information have been combined to predict the risk of prostate cancer. The patient-specific risk derived could be used in personalized decisionmaking regarding the benefits and harms of prostate biopsy.

Outcomes

Applicability was limited by the lack of information on the impact that reporting PCA3 results in a clinical setting can have on intermediate and long-term outcomes. Of primary interest is whether patient decision-making regarding prostate biopsy would be impacted by reporting PCA3 results (improved diagnostic accuracy), and whether those decisions result in higher positive predictive values (fewer biopsies performed in men without prostate cancer). Future studies might consider routinely combining the tPSA elevations with PCA3 results according to one of several validated nomograms contained in this review. The population could be randomized into those with PCA3 included versus those receiving routine care. Both groups would have urine samples collected; one group would be immediately tested, while the other is tested at a later time. Cut-offs levels could be selected such that the sensitivities are similar in both groups, but fewer biopsies would be expected with the use of PCA3. This would be expected to result in a higher positive predictive value. Once the study is complete, patients could receive additional results as if they were had been randomized to the other group. This would allow for a fully matched study of which cancers were detected by both, as well as comparing the cancer detected by one or the other.

The remaining intermediate and longer term outcomes are more challenging due to the difficulty of collecting this information and the lack of validated surrogate measurements.

Timeframes

After the introduction of routine PSA screening for prostate cancer, it was thought that the harms associated with unnecessary biopsies among false positive PSA tests were outweighed by the benefits of identifying prostate cancer in its early stage. Since that time, more information on the benign nature of many of these cancers, the harms associated with treatments, including biopsy itself, and more attention is being focused on the avoidance of unnecessary biopsy. This area is the topic of KQ1 and KQ2. Avoiding harms of treatments associated with less aggressive cancers is the topic of KQ3.

Key Question 3: Management of Biopsy Positive Patients

To determine the effectiveness of PCA3 and comparator tests in identifying patients who are candidates for active surveillance versus aggressive therapy (e.g., radical prostatectomy, brachytherapy, external radiation therapy), it is useful to assess the applicability of the findings

in this review in a systematic manner, reflecting both on the design and execution of the included studies. However, no studies reported clinical outcomes of patients classified as low risk and enrolled in active surveillance versus those of patients for whom treatment was recommended or elected. Two poor quality longitudinal studies reported on PCA3 and tPSA, but data could not be compared or combined. Detection of micrometastases in lymph node tissue as a predictor of biochemical recurrence was outside the inclusion criteria because no PSA serum testing or PCA3 urine testing was performed along with lymph node extracts. The study of 2-year followup on progression of patients from active surveillance to treatment had design flaws (e.g., partially matched, incomplete data).

Implications for Clinical and Policy Decisionmaking

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 impacted 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, in order to achieve the 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 decisionmaking.

Limitations of the Comparative Effectiveness Review Process

One limitation of the review process was our decision to craft two separate Key Questions on biopsy decisionmaking – one for initial biopsies and one for repeat biopsies. We were able to identify only a small number of studies specifically targeting use for these specific populations. A majority of studies included populations with mixed biopsy histories (some patients were candidates for initial and others for repeat biopsy). This was addressed in our analysis, but reflected a characteristic of studies that was not anticipated when we scoped the Key Questions.

A second limitation was the need to develop an expanded QUADAS framework to address the quality of studies for the three Key Questions. QUADAS asks if the reference standard results are interpreted without knowledge of the index test results (in this case PCA3). However, it was also important to know if this was also true for the tPSA comparator test, and if partial verification bias was identified. For Key Question 3 we observed that the endpoint of interest (patient clinical outcomes) seldom had appropriate clinical information. Again we added an additional QUADAS item to ensure we accounted for this information.

Finally, a procedural limitation was the implementation of the DistillerSR application (and the associated learning curve) concurrently with the beginning of this review. This was balanced by strengths of DistillerSR, such as improved efficiency of abstract review.

Limitations of the Evidence Base

Key Questions 1 and 2

There were several severe limitations of the evidence base identified during the review. First, our inclusion criteria required that PCA3 and the comparator both be reported from the same population. These studies were identified as being 'matched'. The aim was to reduce the well known variability in diagnostic test performance that both the underlying population and study entry characteristics can cause. By requiring a matched comparison, the analyses might result in less between-study heterogeneity and we believe that this was the case. However, requiring this matching reduced the number of included studies. It is possible that some conclusions may have been possible for some of the comparators if unmatched data had been included. On the other hand, the likely variability in results may have rendered those data unhelpful.

Another challenging aspect of this review was integrating and interpreting the information for diagnostic accuracy. We found data to support the conclusion that PCA3 had slightly higher performance compared to the extent of tPSA elevation. Based on limited data, the two markers seemed to be relatively independent. Better performance would be expected by combining the two, and this was supported by one report. We could have extended our modeling to include both markers, but chose not to because that was beyond the scope of the report. Such modeling was identified as a gap in knowledge.

There were limited data available when focusing only on initial, or on repeat, prostate biopsies. Most studies reported a mixture of initial and repeat biopsies. Two analyses (difference between the matched AUCs, comparison of ROC curves) allow for a regression of the effect size on proportion with repeat biopsies in nine studies. The slope was not significant, and based on this result we concluded that any difference between PCA3 and tPSA elevations for men being offered an initial versus repeat biopsy would likely be small.

The issue of potential verification bias for analyses related to tPSA was raised early in discussion of the analytical framework and key questions, and discussed with members of the TEP. Although this potential bias was not acknowledged by any of the included studies, we chose to proceed with review of the data sets. A modeling exercise was undertaken to investigate the magnitude of the effect, and whether it affected the overall estimate of performance, performance estimate at a given cut-off, or both. We concluded that the effect of verification bias, while present, was likely to influence the overall estimate of performance (e.g., sensitivity at a given specificity), but had strong influence on performance at a given cut-off (e.g., sensitivity / specificity at >6 ng/mL of tPSA). This latter issue was potentially solvable by additional modeling (Appendix J).

Key Question 3

The review of the literature revealed a lack clinical followup after patients were placed into risk categories (low risk/indolent, high risk/aggressive) that were defined by the results of PCA3 and other biomarker and pathological tests. In nine of 11 studies, a reference or gold standard clinical endpoint (or validated surrogate) was lacking; in one poor-quality short-term study, PCA3 levels were not associated with disease progression.

Research Gaps

Overview

The PCA3 test is the key marker of this comparative effectiveness review. Its performance has been evaluated against/with six comparators (tPSA, fPSA, PSA density, PSA velocity and doubling time, complexed PSA and externally validated nomograms) in three clinical scenarios (diagnostic accuracy and other intermediate outcomes, and long term outcomes). With the exception of analyses that include PCA3 and tPSA for the intermediate outcome of diagnostic accuracy, evidence was insufficient to answer the key questions. These then, remain as gaps in evidence. Even for the PCA3/tPSA comparison and diagnostic accuracy, the strength of evidence was low. Thus, virtually all comparisons for all outcomes in this review could be considered gaps in knowledge. As these comparisons have been extensively reviewed in the Results and Discussion, they will not be repeated here.

Instead, the following sections deal with gaps in knowledge and their associated research question/future study design for cross-cutting issues, statistical/methodological issue related to multiple comparisons, or clinical issues that are relevant to multiple comparisons. The first section focuses on gaps for KQ1 and KQ2, while the second section focuses on KQ3. Within each section, the specific gaps are gathered into two general areas: gaps relating to clinical issues, and those relating to methodological and statistical issues. For each identified gap, a potential study is designed and discussed.

Gaps in Knowledge for Key Questions 1 and 2

Clinical Gaps in Knowledge

1. Does the addition of PCA3, either alone, or in combination with other markers, change prostate cancer biopsy decisionmaking for the patient or physician? Several studies (and our 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.

   Future Studies: Researchers might consider routinely combining the tPSA elevations with PCA3 results according to one of several validated nomograms contained in this review. The population could be randomized into those with PCA3 included as part of care (intervention) versus those receiving current care (control). Both groups would have urine samples collected; with the intervention group being immediately tested, while the control group members are tested later. Cut-offs levels in the two groups could be selected separately, such that the sensitivities (proportion of cancers detected) are expected to be similar in both groups. However, fewer biopsies would be expected among the intervention group. This would be expected to result in a higher positive predictive value. Once the study is complete (perhaps one year later), patients could receive additional test results, as if they had been randomized to the other group. This would allow for a fully matched analysis of which cancers were detected by both, as well as comparing the cancer detected by protocol versus the other. These data could provide the type of data that would allow PCA3 to become a routine test, or be rejected as a potential contributor to the testing process.

2. What improvement in diagnostic accuracy is needed for any new test (e.g., PCA3) to provide sufficient value to impact biopsy decisionmaking? Were there clear guidance on how much improvement in diagnostic accuracy would be required to impact clinical protocols, the methods required to assess and accept/reject prospective markers would be streamlined. The relative importance of other factors to be considered (e.g., convenience, cost) would also be useful.
   Future Studies: Models can be created for various types of markers (e.g., continuous, categorical) with varying performance characteristics (e.g., low sensitivity but high specificity, high sensitivity but low specificity) that utilize, when possible, existing prostate cancer markers. Health care providers could evaluate the relative improvement in test performance and also provide guidelines regarding other factors that assay developers need to bear in mind.
3. How does PCA3 compare to the two commonly used add on tests of fPSA and tPSA velocity/doubling time? These comparisons have been singled out, because both comparators have been recommended for clinical implementation (NCCN guidelines) but their use has generated some controversy rather than bringing consensus. Special attention should be paid toward looking at the relative performance of PCA3 against these two comparators at outcomes of decisionmaking as a way to avoid even further fracturing of protocols based on limited evidence.
   Future Studies: Since our restriction to "matched" studies identified no suitable data, it may be necessary to move to a broader set of inclusion criteria that does not require PCA3 and fPSA/tPSA velocity to be measured in the same dataset. However, prior to this undertaking, it would be prudent to validate that such a methodology is likely to provide reliable information (see Methodological and Statistical Gaps in Knowledge, Number 4).
4. 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 (as well as for some of the comparators) is unknown, but may be important.
   Future Studies: It may be possible to create a collaboration between groups studying PCA3 such that pooled raw data might be created. If funding is available, it may also be possible to collect some of the relevant data that might be missing. From this combined dataset, it might be possible to answer some of these important questions.
5. 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.
   Future Studies: Few, if any studies included in this review identified a well-defined cohort of consecutive men identified as being at risk for prostate cancer and then followed their subsequent screening/diagnostic decision-making. This may require external funding, as a relatively high proportion of men identified as being at high risk will not chose to have a biopsy, according to current guidelines. However, with the collection of data that could reduce partial verification bias, and longer followup, it may be possible to account for, or more clearly delineate the effect of this bias on tPSA and on other related biomarkers.

Methodological and Statistical Gaps in Knowledge

1. What modeling approach/algorithm would allow for the easiest inclusion of new markers while reducing the need for independent verification? Most reported multivariate modeling of prostate cancer risk relies on logistic regression. These models are difficult to compare across studies and do not allow for simple inclusion of new variables without re-computing all coefficients. Other models, such as multivariate overlapping Gaussian distributions, may fit the markers of interest and might allow for easier comparisons as well as the ability to easily add (or subtract) markers as knowledge increases. This could also allow for validation of partial models, if some markers have not been measured.
   Future Studies: A review of select literature might provide a model methodology to use in designing an approach to multivariate modeling. One that has been successful in the area of prenatal screening for select congenital abnormalities (e.g., Down syndrome) is based on multi-dimensional overlapping Gaussian distributions. Such testing began in the early 1980s with both a demographic (maternal age) and a second trimester biochemical marker (alpha-fetoprotein). Currently, the most advanced protocol relies on these and three more second trimester and two first trimester biochemical markers, and a first trimester ultrasound markers. All are successfully combined to produce one risk that has been extensively validated as part of routine screening. Other models to explore might include those used in testing for BRCA1/2 mutations and breast cancer.
2. What factors influence whether partial verification bias impacts the tPSA and/or the matched tPSA/PCA3 ROC curves? Factors that could be explored include the range of cancer rates, the range of verification rates, and the use of continuous versus categorical verification corrections. There have been only a handful of reports on tPSA use that address partial verification bias. A better understanding of this issue is needed if PCA3 is to be properly evaluated in the context of the widespread use of tPSA as triage test for treatment decisions.
   Future Studies: Both modeling and examining select datasets might help provide an answer to this question in the specific setting of interest. Our modeling suggests that verification bias introduced by decision-making relying on tPSA elevations has a strong impact on the sensitivity/specificity at a given tPSA cut-off. However, we found that it does not distort the ROC curve to any great extent. If this can be verified, it might result in the reexamination of previous studies where verification bias was thought to have made the reliable interpretation of results impossible.
3. What absolute cut-offs or continuous values can be assigned to the PCA3 assay across the ROC curve? While the analyses in this review provide an approximate ROC curve allowing interpolation of sensitivity and false positive rates across the range of values, the absolute PCA3 and tPSA cut-offs may not be appropriate in every setting.
   Future Studies: The modeling studies suggested for the previous gap could also help resolve this issue. Is it possible to account for verification bias to such an extent that reliable performance estimates can be computed?
4. Does our review's literature restriction to matched studies provide more consistent and reliable comparisons than had the review used independent summaries of each marker's performance? Given the increasing emphasis on comparative effectiveness analyses, a formal comparison of these two methods might provide useful guidance to future reviews.

Future Studies: A structured evidence review between the performance of PCA3 and tPSA could be performed without restricting data to matched studies. That is, a summary review of PCA3 performance (via an ROC curve) could be compared to a similar independent summary of the performance of tPSA elevations to identify prostate cancer. The two independent ROC curves could then be compared and the difference compared to our matched differences. As a check on precision, one could examine the tPSA analysis and determine whether the "grey zone" studies differ from the studies having no restriction on the tPSA elevation. Our matched analyses clearly show this difference and are likely an indicator of reliability.

5. Does the reporting of matched analyses improve the usefulness of the dataset? Although our inclusion criteria required PCA3 and a comparator to be measured in the same population, it did not require a formal matched analysis to be reported. Thus, the reports did not allow for a comparison of how many men with cancer were identified by both markers, neither of the markers, or only one or the other marker. Requesting such analyses be performed using existing datasets would help answer this question.
   Future Studies: It may be possible to collaborate with authors of selected studies to obtain original data, in which a true matched analysis can be performed. The study (or studies) should be large, and have information available about the 'aggressiveness' of the cancers identified. Then, two 2x2 tables could be created (one for positive biopsies, one for negative) comparing the results of the two tests at chosen cut-offs. A series of these tables could be created at fixed sensitivities or fixed false positive rates. One could imagine that the data examined this way will provide insights into the types of cancers identified by both, one, or neither of the tests.
6. How can researchers studying PCA3 and other comparators be encouraged to provide proper reporting of statistical details? Proper reporting of statistical information on studies of PCA3 and the comparators was often absent in articles evaluated for this review. These include: confidence intervals, standard errors, prediction limits and other measure of dispersion and precision for all effect measures as well as good summary parameters for their data (e.g., selected centiles, medians, geometric means and trimmed logarithmic standard deviations). All studies identified were of poor quality when rated by QUADAS.
   Future Studies: Statisticians, epidemiologists and others with experience in analyzing screening and diagnostic tests are in the process of setting guidelines for evaluation. Those guidelines should be reviewed with an eye towards ensuring that 1) appropriate statistics are used (e.g., needed transformations performed), 2) methods to allow for raw data to be available for additional analyses, and 3) confidence intervals / standard errors be provided more consistently, when possible.
7. How can systematic differences in marker levels due to reagents/manufactures be minimized or accounted for by analysis? Systematic differences between reagents/manufacturers exist for at least some of the markers that can influence the tests performance at fixed mass unit cut-offs.
   Future Studies: Review of external proficiency testing results may provide guidance as to what steps may be needed. Attempting to harmonize results is one possibility, but this can take time and may not ever succeed. Mathematical methods that use a normalizing function (z-score, multiples of the median) could also be explored and tested against

existing datasets to see whether more consistent patient-specific information can be provided from the test results.

Gaps in Knowledge for KQ3

Clinical Gaps in Knowledge

1. 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 to 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 between studies.
   Future studies: A consensus conference or statement from a relevant professional organization might help harmonize the definitions of intermediate outcomes such that research studies can be combined. However, the only sure way to determine the appropriateness of intermediate outcomes is to engage in long-term followup of a select cohort of men with sufficient numbers that even relatively uncommon events (e.g., death due to prostate cancer) can be quantified.
2. How can PCA3 alone or when integrated with one or more comparators be used to improve decisionmaking about whether to choose active surveillance or aggressive treatment for biopsy positive men? No studies have yet examined the impact of PCA3 on decisionmaking when compared to existing criteria such as the Epstein Criteria. There have been no outcomes studies performed to determine how well PCA3 scores predict the behavior of a particular tumor over time.
   Future Studies: This is similar to question 1 (Clinical Gap in Knowledge for KQ1/KQ3), but studies would focus on other intermediate outcomes. Based on this review, and the recent FDA approval of PCA3, it should be possible to establish protocols that use PCA3 and begin developing studies that examine and validate intermediate health outcomes.

Methodological and Statistical Gaps in Knowledge

1. Can intermediate outcomes, such as cancer classifications of aggressive or indolent tumors be properly validated? Given that current clinical practice guidelines employ unvalidated, or partially validated intermediate outcomes, it is difficult to design studies that would provide proper validation. Exploration of what study designs or re-analyses of existing dataset might provide stronger validation of select intermediate measures could be undertaken.
   Future Studies: To validate intermediate outcomes, they must be standardized (addressed earlier) and then associated with longer term health outcomes. This might require longitudinal studies with followup of five to ten years. Some of these studies are underway and preliminary results have been reported, but longer timeframes need to be utilized.
2. What is the impact of use of PCA3 on long-term health outcomes when used to help select patients for active surveillance versus aggressive treatment?
   Future Studies: Longitudinal studies that track patients under both active surveillance and aggressive treatment would be helpful in determining this impact. Many of the ongoing

studies are focused on only the active surveillance group, and these should be expanded to at least include a subset of those undergoing treatment for comparison.

References

1. Jemal A, Siegel R, Xu J, et al. Cancer statistics, 2010. CA Cancer J Clin. 2010 Sep-Oct;60(5):277-300. PMID: 20610543.

2. Sutcliffe P, Hummel S, Simpson E, Young T, Rees A, Wilkinson A, Hamdy F, Clarke N, Staffurth J. Use of classical and novel biomarkers as prognostic risk factors for localised prostate cancer: a systematic review. Health Technol Assess. 2009 13(1-260. PMID:

3. Freedland SJ. Screening, risk assessment, and the approach to therapy in patients with prostate cancer. Cancer. 2011 Mar 15;117(6):1123-35. PMID: 20960523.

4. Yin M, Bastacky S, Chandran U, et al. Prevalence of incidental prostate cancer in the general population: a study of healthy organ donors. J Urol. 2008 Mar;179(3):892-5; discussion 95. PMID: 18207193.

5. SEER Cancer Statitstics Review, 1975-2005, National Cancer Institute, Bethesda, MD, based on November 2007 SEER data submission, posted to the SEER web site, 2008. http://seer.cancer.gov/csr/1975_2005/.

6. Draisma G, Boer R, Otto SJ, et al. Lead times and overdetection due to prostate-specific antigen screening: estimates from the European Randomized Study of Screening for Prostate Cancer. J Natl Cancer Inst. 2003 95(868-78. PMID:

7. Schroder F, Kattan MW. The comparability of models for predicting the risk of a positive prostate biopsy with prostate-specific antigen alone: a systematic review. European urology. 2008 Aug;54(2):274-90. PMID: 18511177.

8. Wolf AMD, Wender RC, Etzioni RB, et al. American Cancer Society Guideline for the Early Detection of Prostate Cancer: Update 2010. CA Cancer J Clin. 2011 60(70-98. PMID:

9. National Comprehensive Cancer Network Guidelines - Prostate Cancer Early Detection, Version I.2011. www.NCCN.org. 2011 PMID:

10. Prostate-Specific Antigen Best Practice Statement: 2009 Update. American Urological Association. http://www.auanet.org/content/media/psa09.pdf. February 10, 2011.

11. Vlaeminck-Guillem V, Ruffion A, Andre J, et al. Urinary prostate cancer 3 test: toward the age of reason? Urology. 2010 Feb;75(2):447-53. PMID: 19586654.

12. Lin DW. Beyond PSA: utility of novel tumor markers in the setting of elevated PSA. Urol Oncol. 2009 May-Jun;27(3):315-21. PMID: 19414121.

13. Lin K, Lipsitz R, Miller T, et al. Benefits and harms of prostate-specific antigen screening for prostate cancer: an evidence update for the U.S. Preventive Services Task Force. Ann Intern Med. 2008 Aug 5;149(3):192-9. PMID: 18678846.

14. Rosenthal SA, Sandler HM. Treatment strategies for high-risk locally advanced prostate cancer. Nat Rev Urol. 2010 Jan;7(1):31-8. PMID: 20062072.

15. Thompson IM, Ankerst DP, Chi C, et al. Assessing prostate cancer risk: results from the Prostate Cancer Prevention Trial. J Natl Cancer Inst. 2006 Apr 19;98(8):529-34. PMID: 16622122.

16. Amling CL, Catalona WJ, Klein EA. Deciding whom to biopsy. Urol Oncol. 2010 Sep-Oct;28(5):542-5. PMID: 20816613.

17. Simmons MN, Berglund RK, Jones JS. A practical guide to prostate cancer diagnosis and management. Cleveland Clinic Journal of Medicine. 2011 78(321-31. PMID:

18. D' Amico AV, Whittington R, Malkowicz SB, et al. Biochemical outcome after radical prostatectomy, external beam radiation therapy, or interstitial radiation therapy for clincally localized prostate cancer. JAMA. 1998 280(969-74. PMID:

19. Epstein JI, Walsh PC, Carmichael M, et al. Pathologic and clinical findings to predict tumor extert of nonpalpable (stage T1c) prostate cancer. Jama. 1994 271(368-74. PMID:

20. Perdona 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. European urology. 2011 Jan;59(1):81-7. PMID: 20947244.

21. Cheng L, Montironi R, Bostwick DG, et al. Staging of prostate cancer. Histopathology. 2012 60(87-117. PMID:

22. Chun FK, de la Taille A, van Poppel H, et al. Prostate cancer gene 3 (PCA3): development and internal validation of a novel biopsy nomogram. European urology. 2009 Oct;56(4):659-67. PMID: 19304372.

23. Eastham JA, Scardino PT, Kattan MW. Predicting an optimal outcome after radical prostatectomy: the trifecta nomogram. J Urol. 2008 Jun;179(6):2207-10; discussion 10-1. PMID: 18423693.

24. Huang Y, Isharwal S, Haese A, et al. Prediction of patient-specific risk and percentile cohort risk of pathological stage outcome using continuous prostate-specific antigen measurement, clinical stage and biopsy Gleason score. BJU Int. 2011 May;107(10):1562-9. PMID: 20875091.

25. Hinnen KA, Roeloffzen EM, Battermann JJ, et al. Survival after prostate brachytherapy in patients aged 60 years and younger. BJU Int. 2011 107(1906-11. PMID:

26. Boorjian SA, Karnes RJ, Viterbo R, et al. Long-term survival after radical prostatectomy versus external-beam radiotherapy for patients with high-risk prostate cancer. Cancer. 2011 Jan 10; PMID: 21225881.

27. Bussemakers MJ, van Bokhoven A, Verhaegh GW, et al. DD3: a new prostate-specific gene, highly overexpressed in prostate cancer. Cancer Res. 1999 Dec 1;59(23):5975-9. PMID: 10606244.

28. Durand X, Moutereau S, Xylinas E, et al. Progensa PCA3 test for prostate cancer. Expert Rev Mol Diagn. 2011 Mar;11(2):137-44. PMID: 21405964.

29. Makarov DV, Loeb S, Getzenberg RH, et al. Biomarkers for prostate cancer. Annual review of medicine. 2009 60(139-51. PMID: 18947298.

30. de Kok JB, Verhaegh GW, Roelofs RW, et al. DD3(PCA3), a very sensitive and specific marker to detect prostate tumors. Cancer Res. 2002 May 1;62(9):2695-8. PMID: 11980670.

31. 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 Sep;53(9):964-72. PMID: 11004423.

32. Day JR, Jost M, Reynolds MA, et al. PCA3: from basic molecular science to the clinical lab. Cancer Lett. 2011 Feb 1;301(1):1-6. PMID: 21093148.

33. de la Taille A. Progensa PCA3 test for prostate cancer detection. Expert Rev Mol Diagn. 2007 Sep;7(5):491-7. PMID: 17892357.

34. Groskopf J, Aubin SM, Deras IL, et al. APTIMA PCA3 molecular urine test: development of a method to aid in the diagnosis of prostate cancer. Clin Chem. 2006 Jun;52(6):1089-95. PMID: 16627561.

35. Sokoll LJ, Ellis W, 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 Mar;389(1-2):1-6. PMID: 18061575.

36. Boorjian SA, Karnes RJ, Viterbo R, et al. Long-term survival after radical prostatectomy versus external-beam radiotherapy for patients with high-risk prostate cancer. Cancer. 2011 Jul 1;117(13):2883-91. PMID: 21692049.

37. Thompson IM, Ankerst DP, Chi C, et al. Operating characteristics of prostate-specific antigen in men with an initial PSA level of 3.0 ng/ml or lower. JAMA. 2005 Jul 6;294(1):66-70. PMID: 15998892.

38. Sokoll LJ, Witte DL, Klee GG, et al. Redesigned Proficiency Testing Materials Improve Survey Outcomes for Prostate-Specific Antigen. Arch Pathol Lab Med. 2000 124(1608-13. PMID:

39. Schreiber WE, Endres DB, McDowell GA, et al. Comparison of fresh frozen serum to proficiency testing material in College of American Pathologists surveys: alpha-fetoprotein, carcinoembryonic antigen, human chorionic gonadotropin, and prostate-specific antigen. Arch Pathol Lab Med. 2005 129(331-337): PMID:

40. Garg UC, Howanitz JH, Nakamura RM, et al. Production, analysis, and characterization of reference materials for prostate-specific antigen. Arch Pathol Lab Med. 1995 119(1104-08. PMID:

41. Klee GG, Schryver PG, Kisabith RM. Analytic bias specifications based on the analysis of effects on performance of medical guidelines. Scand J Clin Lab Invest. 1999 59(509-12. PMID:

42. Westgard JO, Westgard SA. An Assessment of Metrics for Analytic Quality Using σ Performance Data From Proficiency Testing Surveys and the CLIA Criteria for Acceptable Performance. Am J Clin Pathol. 2006 125(343-54. PMID:

43. Tosoian J, Loeb S. PSA and beyond: the past, present, and future of investigative biomarkers for prostate cancer. TheScientificWorldJournal. 2010 10(1919-31. PMID: 20890581.

44. Roddam AW, Duffy MJ, Hamdy FC, et al. Use of prostate-specific antigen (PSA) isoforms for the detection of prostate cancer in men with a PSA level of 2-10 ng/ml: systematic review and meta-analysis. European urology. 2005 Sep;48(3):386-99; discussion 98-9. PMID: 15982797.

45. Lee R, Localio AR, Armstrong K, et al. A meta-analysis of the performance characteristics of the free prostate-specific antigen test. Urology. 2006 Apr;67(4):762-8. PMID: 16600352.

46. Benson MC, Whang IS, Pantuck A, et al. Prostate specific antigen densivity; a means of distinguishing benigh prostatic hypertrophy and prostate cancer. J Urol. 1992 147(815-16. PMID:

47. Giannarini G, Scott CA, Moro U, et al. Are PSA density and PSA density of the transition zone more accurate than PSA in predicting the pathological stage of clinically localized prostate cancer? Urol Oncol. 2008 26(353-60 PMID:

48. Kundu SD, Roehl KA, Yu X, et al. Prostate specific antigen density correlates with features of prostate cancer aggressiveness. J Urol. 2007 177(505-09. PMID:

49. Thanigasalam R, Mancuso P, Tsao K, et al. Prostate-specific antigen velocity (PSAV): a practical role for PSA? ANZ journal of surgery. 2009 Oct;79(10):703-6. PMID: 19878164.

50. Vickers AJ, Savage C, O'Brien MF, et al. Systematic review of pretreatment prostate-specific antigen velocity and doubling time as predictors for prostate cancer. J Clin Oncol. 2009 Jan 20;27(3):398-403. PMID: 19064972.

51. Ankerst DP, Groskopf J, Day JR, et al. Predicting prostate cancer risk through incorporation of prostate cancer gene 3. J Urol. 2008 Oct;180(4):1303-8; discussion 08. PMID: 18707724.

52. Cooperberg MR, Carroll PR, Klotz L. Active surveillance for prostate cancer: progress and promise. J Clin Oncol. 2011 29(3669-76. PMID:

53. Methods Guide for Effectiveness and Comparative Effectiveness Reviews. Rockville, MD: Agency for Healthcare Research and Quality; August 2011. AHRQ Publication No. 10(11)-EHC063-EF. Chapters available at: www.effectivehealthcare.ahrq.gov.

54. Methods Guide for Medical Test Reviews. Rockville, MD: Agency for Healthcare Research and Quality; Draft Posted November 2010. http://www.effectivehealthcare.ahrq.gov/tasks/sites/ehc/assets/File/methods_guide_for_medical_tests.pdf. February 8, 2012.

55. Jüni P, Holenstein F, Sterne J, et al. Direction and impact of language bias in meta-analyses of controlled trials: empirical study. International journal of epidemiology. 2002 February 1, 2002;31(1):115-23. PMID:

56. Shea BJ, Hamel C, Wells GA, et al. AMSTAR is a reliable and valid measurement tool to assess the methodological quality of systematic reviews. J Clin Epidemiol. 2009 Oct;62(10):1013-20. PMID: 19230606.

57. U.S. Preventive Services Task Force Procedure Manual. Rockville, MD: Agency for Healthcare Research and Quality, July 2008. AHRQ Publication No. 08-051180-EF.

58. Higgins JPT, Deeks JJ, Altman DG on behalf of the Cochrane Statistical Methods Group. Special topics in statistics. In: Cochrane Handbook for Systematic Reviews of Interventions. J. P. T. Higgins and S. Green (Eds). West Sussex, England:Wiley-Blackwell, pp. 481-529, 2010.

59. Whiting PF, Weswood ME, Rutjes AW, et al. Evaluation of QUADAS, a tool for the quality assessment of diagnostic accuracy studies. BMC Med Res Methodol. 2006 6:9. (PMID: 16519814.

60. Dodds PR, Boucher JD, Shield DE, et al. Are complications of transrectal ultrasound-guided biopsies of the prostate gland increasing? Connecticut medicine. 2011 Sep;75(8):453-7. PMID: 21980673.

61. Loeb S, van den Heuvel S, Zhu X, et al. Infectious Complications and Hospital Admissions After Prostate Biopsy in a European Randomized Trial. European urology. 2012 Jan 5; PMID: 22244150.

62. Begg CB, Greenes RA. Assessment of diagnostic tests when disease verification is subject to selection bias. . Biometrics. 1983 39(207-13. PMID:

63. Punglia RS, D'Amico AV, Catalona WJ, et al. Effect of verification bias on screening for prostate cancer by measurement of prostate-specific antigen. N Engl J Med. 2003 Jul 24;349(4):335-42. PMID: 12878740.

64. 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 May;63(5):513-23. PMID: 19595577.

65. Adam A, Engelbrecht MJ, Bornman MS, et al. The role of the PCA3 assay in predicting prostate biopsy outcome in a South African setting. BJU Intl. 2011 Apr 20; PMID: 21507188.

66. 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 Nov;184(5):1947-52. PMID: 20850153.

67. Auprich M, Chun FK, Ward JF, et al. Critical assessment of preoperative urinary prostate cancer antigen 3 on the accuracy of prostate cancer staging. Euro Urol. 2011 Jan;59(1):96-105. PMID: 20980098.

68. Cao DL, Ye DW, Zhang HL, et al. 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 May 15;71(7):700-10. PMID: 20957673.

69. Chun FK, Epstein JI, Ficarra V, et al. Optimizing performance and interpretation of prostate biopsy: a critical analysis of the literature. European urology. 2010 Dec;58(6):851-64. PMID: 20884114.

70. de la Taille A, Irani J, Graefen M, et al. Clinical evaluation of the PCA3 assay in guiding initial biopsy decisions. J Urol. 2011 Jun;185(6):2119-25. PMID: 21496856.

71. Deras IL, Aubin SM, Blase A, et al. PCA3: a molecular urine assay for predicting prostate biopsy outcome. J Urol. 2008 Apr;179(4):1587-92. PMID: 18295257.

72. Fradet Y, Saad F, Aprikian A, et al. uPM3, a new molecular urine test for the detection of prostate cancer. Urology. 2004 Aug;64(2):311-5; discussion 15-6. PMID: 15302485.

73. Haese A, de la Taille A, van Poppel H, et al. Clinical utility of the PCA3 urine assay in European men scheduled for repeat biopsy. European urology. 2008 Nov;54(5):1081-8. PMID: 18602209.

74. Hessels D, van Gils MP, van Hooij O, et al. Predictive value of PCA3 in urinary sediments in determining clinico-pathological characteristics of prostate cancer. Prostate. 2010 Jan 1;70(1):10-6. PMID: 19708043.

75. 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 Jun;14(4):235-43. PMID: 19489685.

76. Nyberg M, Ulmert D, Lindgren A, et al. PCA3 as a diagnostic marker for prostate cancer: a validation study on a Swedish patient population. Scand J Urol Nephrol. 2010 Dec;44(6):378-83. PMID: 20961267.

77. 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 Mar;18(3):200-5. PMID: 21332814.

78. Ouyang B, Bracken B, Burke B, et al. 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 Jun;181(6):2508-13; discussion 13-4. PMID: 19371911.

79. 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 Oct;106(8):1143-7. PMID: 20230386.

80. Rigau M, Morote J, Mir MC, et al. PSGR and PCA3 as biomarkers for the detection of prostate cancer in urine. Prostate. 2010 Dec 1;70(16):1760-7. PMID: 20672322.

81. Roobol MJ, Schroder 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. European urology. 2010 Oct;58(4):475-81. PMID: 20637539.

82. Schilling D, Hennenlotter J, Munz M, et al. Interpretation of the prostate cancer gene 3 in reference to the individual clinical background: implications for daily practice. Urol Int. 2010 85(2):159-65. PMID: 20424427.

83. Tinzl M, Marberger M, Horvath S, et al. DD3PCA3 RNA analysis in urine--a new perspective for detecting prostate cancer. European urology. 2004 Aug;46(2):182-6; discussion 87. PMID: 15245811.

84. van Gils MP, Cornel EB, Hessels D, et al. Molecular PCA3 diagnostics on prostatic fluid. Prostate. 2007 Jun 1;67(8):881-7. PMID: 17440939.

85. van Gils MP, Hessels D, Hulsbergen-van de Kaa CA, et al. Detailed analysis of histopathological parameters in radical prostatectomy specimens and PCA3 urine test results. Prostate. 2008 Aug 1;68(11):1215-22. PMID: 18500693.

86. Wang R, Chinnaiyan AM, Dunn RL, et al. Rational approach to implementation of prostate cancer antigen 3 into clinical care. Cancer. 2009 Sep 1;115(17):3879-86. PMID: 19517474.

87. Marks LS, Fradet Y, Deras IL, et al. PCA3 molecular urine assay for prostate cancer in men undergoing repeat biopsy. Urology. 2007 Mar;69(3):532-5. PMID: 17382159.

88. Schilling D, de Reijke T, Tombal B, et al. The Prostate Cancer gene 3 assay: indications for use in clinical practice. BJU Int. 2010 Feb;105(4):452-5. PMID: 19930176.

89. 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 Feb;183(2):534-8. PMID: 20006883.

90. Kusuda Y, Miyake H, Kurahashi T, et al. 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 May 18; PMID: 21600799.

91. Liss MA, Santos R, Osann K, et al. PCA3 molecular urine assay for prostate cancer: association with pathologic features and impact of collection protocols. World J Urol. 2011 Oct;29(5):683-8. PMID: 21152924.

92. 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 May;179(5):1804-9; discussion 09-10. PMID: 18353398.

93. 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. European urology. 2011 Mar;59(3):422-9. PMID: 21156337.

94. Vlaeminck-Guillem V, Devonec M, Colombel M, et al. Urinary PCA3 score predicts prostate cancer multifocality. J Urol. 2011 Apr;185(4):1234-9. PMID: 21334023.

95. Whitman EJ, Groskopf J, Ali A, et al. PCA3 score before radical prostatectomy predicts extracapsular extension and tumor volume. J Urol. 2008 Nov;180(5):1975-8; discussion 78-9. PMID: 18801539.

96. Roobol MJ, Schroder FH, Van Leeuwen P, et al. Performance of the prostate antigen 3 (PCA3) gene and prostate-specific antigen in prescreened men: exploring the value of PCA3 for a first-line diagnostic test. . Eur Urol. 2010 58(475-81. PMID:

97. Roobol MJ, Schroder 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). European urology. 2010 Dec;58(6):893-9. PMID: 20933321.

98. Roddam AW, Hamdy FC, Allen NE, et al. The impact of reducing the prostate-specific antigen threshold and including isoform reflex tests on the performance characteristics of a prostate-cancer detection programme. BJU Int. 2007 Sep;100(3):514-7. PMID: 17542987.

99. Otto SJ, Moss SM, Maattanen L, et al. PSA levels and cancer detection rate by centre in the European Randomized Study of Screening for Prostate Cancer. Eur J Cancer. 2010 Nov;46(17):3053-60. PMID: 21047586.

100. 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. Journal of clinical epidemiology. 2010 May;63(5):513-23. PMID: 19595577.

101. Hozo SP, Djulbegovic B, Hozo I. Estimating the mean and variance from the median, range, and the size of a sample. BMC Med Res Methodol. 2005 5(17. PMID: 15840177.

102. Ruiz-Aragon J, Marquez-Pelaez S. [Assessment of the PCA3 test for prostate cancer diagnosis: a systematic review and meta-analysis]. Actas urologicas espanolas. 2010 Apr;34(4):346-55. PMID: 20470697.

103. Hessels D, Klein Gunnewiek JM, van Oort I, et al. DD3(PCA3)-based molecular urine analysis for the diagnosis of prostate cancer. European urology. 2003 Jul;44(1):8-15; discussion 15-6. PMID: 12814669.

104. Laxman B, Morris DS, Yu J, et al. A first-generation multiplex biomarker analysis of urine for the early detection of prostate cancer. Cancer Res. 2008 Feb 1;68(3):645-9. PMID: 18245462.

105. Shappell SB, Fulmer J, Arguello D, et al. PCA3 urine mRNA testing for prostate carcinoma: patterns of use by community urologists and assay performance in reference laboratory setting. Urology. 2009 Feb;73(2):363-8. PMID: 18995890.

106. van Gils MP, Hessels D, van Hooij O, et al. The time-resolved fluorescence-based PCA3 test on urinary sediments after digital rectal examination; a Dutch multicenter validation of the diagnostic performance. Clin Cancer Res. 2007 Feb 1;13(3):939-43. PMID: 17289888.

107. Wang Y, Sun G, Pan JG, et al. Performance of tPSA and f/tPSA for prostate cancer in Chinese. A systematic review and meta-analysis. Prostate Cancer Prostatic Dis. 2006 9(4):374-8. PMID: 16926855.

Abbreviations and Acronyms

ACMG	American College of Medical Genetics
ACS	American Cancer Society
AHRQ	Agency for Healthcare Research and Quality
AMP	Association for Molecular Pathology
AS	active surveillance
AUC	area under the receiver operating characteristics (ROC) curve
AUA	American Urological Association
ASAP	atypical small acinar proliferation
ASCO	American Society of Clinical Oncology
BPH	benign prostatic hyperplasia
PCA	College of American Pathologists
cc	cubic centimeter (measure of volume)
CER	comparative effectiveness review
CI	confidence interval
CU	clinical utility
CV	clinical validity OR coefficient of variation
DA	diagnostic accuracy
DD3	differential display code 3
DOR	diagnostic odds ratio
DRE	digital rectal examination
ECE	extracapsular extension
EPICOT	evidence, population, intervention, comparison, outcome, timestamp
ERSPC	European Randomized Study of Screening for Prostate Cancer
EVN	externally validated nomogram
FDA	Food and Drug Administration
FPR	false positive rate
GRADE	grading of recommendations assessment, development and evaluation
GS	Gleason score
IQR	Inter-quartile range
LR	likelihood ratio
mg	milligram
ml	milliliter
mRNA	messenger RNA (ribonucleic acid)
ng	nanogram
NICE	UK National Institute for Health and Clinical Excellence
NIH	National Institutes of Health
NCCN	National Cancer Consortium Network
NCI	National Cancer Institute
NPV	negative predictive value
NR	not reported
OAPR	odds of being affected given a positive result
OANR	odds of being affected given a negative result
OR	odds ratio
PCa	prostate cancer, prostatic cancer
PCA3	prostate cancer antigen 3
PCPT	Prostate Cancer Prevention Trial
PICOTS	patient, intervention, comparator, outcome, timing, and setting
PCRI	Prostate Cancer Research Institute
PIN	prostatic intraepithelial neoplasia
PPV	positive predictive value
PSA	prostate-specific antigen
cPSA	complexed prostate-specific antigen
fPSA	free prostate-specific antigen
tPSA	total prostate-specific antigen
PSAD	prostate-specific antigen density
PSADT	prostate-specific antigen doubling time
PSAV	prostate-specific antigen velocity

RCT	randomized controlled trial
RP	radical prostatectomy
RT-PCR	real-time – polymerase chain reaction
SD	standard deviation
SER	systematic evidence review
TEP	technical expert panel
TMA	transcription-mediated amplification
TURP	transurethral resection of the prostate
TV	tumor volume
US	United States
USPSTF	US Preventive Services Task Force