Evidence-based Practice Center Systematic Review Protocol

Project Title: PCA3 Testing for the Diagnosis and Management of Prostate Cancer

I. Background and Objectives

Prostate cancer is the second most common cancer in males, accounting for 217,000 new cases per year in the United States and causing 32,000 deaths per year.[1] The disease is unpredictable. Most patients have indolent tumors and may live for years with no untoward effects, ultimately dying of other causes. But some patients have aggressive tumors that spread beyond the prostate, resulting in significant discomfort and death.

The paramount diagnostic challenge in dealing with prostate cancer is deciding which patients to biopsy and when. The most pressing challenge in managing clinically localized disease is distinguishing between men who have aggressive disease and need aggressive therapy and men who have indolent disease and can be safely managed by active surveillance.

Screening programs that use the prostate-specific antigen (PSA) test have been in place since the late 1980s and have sparked interest and controversy. These programs are based on the premise that PSA testing can lead to early detection of prostate cancer and that effective treatments can be initiated to improve clinical outcomes. However, testing total PSA levels to decide which patients should undergo a biopsy 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 results may undergo unwarranted biopsy that also yields negative results. These men may experience unnecessary anxiety, discomfort, and occasionally significant procedural complications such as infection or hemorrhage.

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

The prostate cancer antigen 3 gene (*PCA3*), formerly known as *DD3*, was first identified in 1999.[2] PCA3 is a non–protein-coding messenger RNA (mRNA) that is highly overexpressed in prostate cancer tissue when compared to normal prostate tissue or to benign prostatic hyperplasia. In 2003, the strong association between PCA3 mRNA levels and prostate cancer led to the development of a urinary assay to measure this analyte to aid in cancer detection.[3] Currently, no PCA3 mRNA test has been approved by the U.S. Food and Drug Administration (FDA). However, several reference laboratories offer PCA3 testing as laboratory-developed tests (tests developed by and used at a single laboratory testing site). PCA3 testing has two proposed clinical utilities: 1) to inform decisions about when to biopsy or rebiopsy patients versus when to wait; and 2) to determine in patients with cancer-positive biopsies whether the disease is indolent or aggressive so that an optimal treatment plan can be developed.[4]

Current decisionmaking about potential prostate disease status and whether to biopsy or rebiopsy is not standardized but depends on consideration of a variety of clinical (e.g., age, family history, race) or laboratory factors.[5] Recently, attention has been directed at creating algorithms or nomograms that combine multiple clinical and laboratory features into risk scores to help in clinical decisionmaking.[5-8] Nomograms are intended to exploit the incremental value of running multiple tests, each with independent contributions to estimating the risk of biopsy outcomes for patients. Although there is wide variation in the manner in which the algorithms

Source: www.effectivehealthcare.ahrq.gov
Published Online: October 14, 2011

and nomograms have been developed and validated, a recent systematic review has suggested that these tools tend to provide more accurate diagnostic predictions for cancer-positive biopsies than the use of PSA testing or other factors alone.[6]

Recent reports describe the use of PCA3 testing to identify patients with aggressive versus indolent prostate cancer. Results of studies have been mixed. If the PCA3 test is found to correlate with disease prognosis, it could be a valuable tool in identifying patients who are better treated with expectant management (e.g., aggressive followup of the tumor without radical treatment intervention) versus those better treated with curative therapy (e.g., surgery or radiation therapy).[9]

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

II. The Key Questions

The overarching question to be addressed in our systematic review is the following:

What is the evidence informing the use of PCA3 mRNA testing in three clinically distinct subpopulations: 1) as a replacement or add-on test to aid in decisionmaking about initial prostate biopsy in patients with positive PSA test and/or digital rectal examination (DRE) results; 2) as a replacement or add-on test to aid in decisionmaking about repeat prostate biopsy in patients with positive PSA test and/or DRE results; and 3) as a prognostic or monitoring test in patients with a positive biopsy to aid in decision-making about progression to active surveillance or treatment.

The Key Questions (KQs) present three proposed scenarios in which this testing may be used.

Question 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 (e.g., elevated total PSA values, decreased percent-free PSA levels, elevated PSA velocities, complexed PSA, or externally validated nomograms) with regard to:

- Diagnostic accuracy (clinical validity) for identifying prostate cancer (e.g., sensitivity, specificity, predictive values)?
- Intermediate outcomes, including improved decisionmaking in selecting patients for biopsy that leads to reduction in negative biopsies and increased identification of prostate tumors (particularly cancerous tumors with aggressive features)?
- Long-term health outcomes (clinical utility), including reduced mortality/ morbidity and improved quality of life, and the potential clinical and personal harms of PCA3 testing and/or related interventions (e.g., anxiety over false-positive results, anxiety about and discomfort of biopsy, complications of biopsy, misdiagnosis)?

Question 2

Source: www.effectivehealthcare.ahrq.gov
Published Online: October 14, 2011

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 screening tests (e.g., elevated total PSA values, decreased percent-free PSA levels, elevated PSA velocities, complexed PSA, or externally validated nomograms) with regard to:

- Diagnostic accuracy (clinical validity) for identifying prostate cancer (e.g., sensitivity, specificity, predictive values)?
- Intermediate outcomes, including improved decisionmaking in selecting patients for biopsy that leads to reduction in negative biopsies and increased identification of prostate tumors (particularly cancerous tumors with aggressive features)?
- Long-term health outcomes (clinical utility), including reduced mortality/ morbidity, improved quality of life, and the potential clinical and personal harms of PCA3 testing and/or related intervention (e.g., anxiety over false-positive results, anxiety about and discomfort of biopsy, complications of biopsy or misdiagnosis)?

Question 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 identifying aggressive prostate cancer (e.g., sensitivity, specificity, predictive values) as part of a diagnostic workup?
- Intermediate outcomes, such as on impact on decision-making about prognosis or triage for active surveillance and/or aggressive treatment?
- Long-term health outcomes (clinical utility) from decisions regarding the choice of active surveillance and/or aggressive treatment. These decisions may be based on a single PCA3 value after biopsy diagnosis of cancer or on PCA3 monitoring over time. The outcomes of interest include reduced mortality/morbidity, improved quality of life, and the potential clinical and personal harms of PCA3 testing (e.g., anxiety about treatment decision, misdiagnosis resulting in overtreatment or undertreatment, adverse events related to treatment)?

The proposed KQs were posted for public comment on the Effective Health Care Program Web site (www.effectiveheatlhcare.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.

Table 1: PICOTS Framework

Population(s)

Source: www.effectivehealthcare.ahrq.gov
Published Online: October 14, 2011

KQ 1: Adult male patients who are candidates for initial prostate biopsy based on elevated prostate-specific antigen (PSA) and/or abnormal digital rectal examination (DRE).
KQ 2: Adult male patients with one or more previous negative prostate biopsies who are candidates for repeat biopsy based on elevated PSA and/or abnormal DRE.
KQ 3: Adult male patients with a positive prostate biopsy.

Interventions

- Testing for the prostate cancer antigen 3 gene (*PCA3*) alone or in conjunction with other tests. Note: PCA3 testing has not been approved by the FDA but is offered as a laboratory-developed test. Testing complies with requirements of the Clinical Laboratory Improvement Act (Public Law 100-578).
- PCA3 comparators
  - Total PSA
  - Percent free PSA
  - PSA velocity
  - Complexed PSA
  - Externally validated nomograms
- Prostate biopsy

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.

KQ 3

- *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.

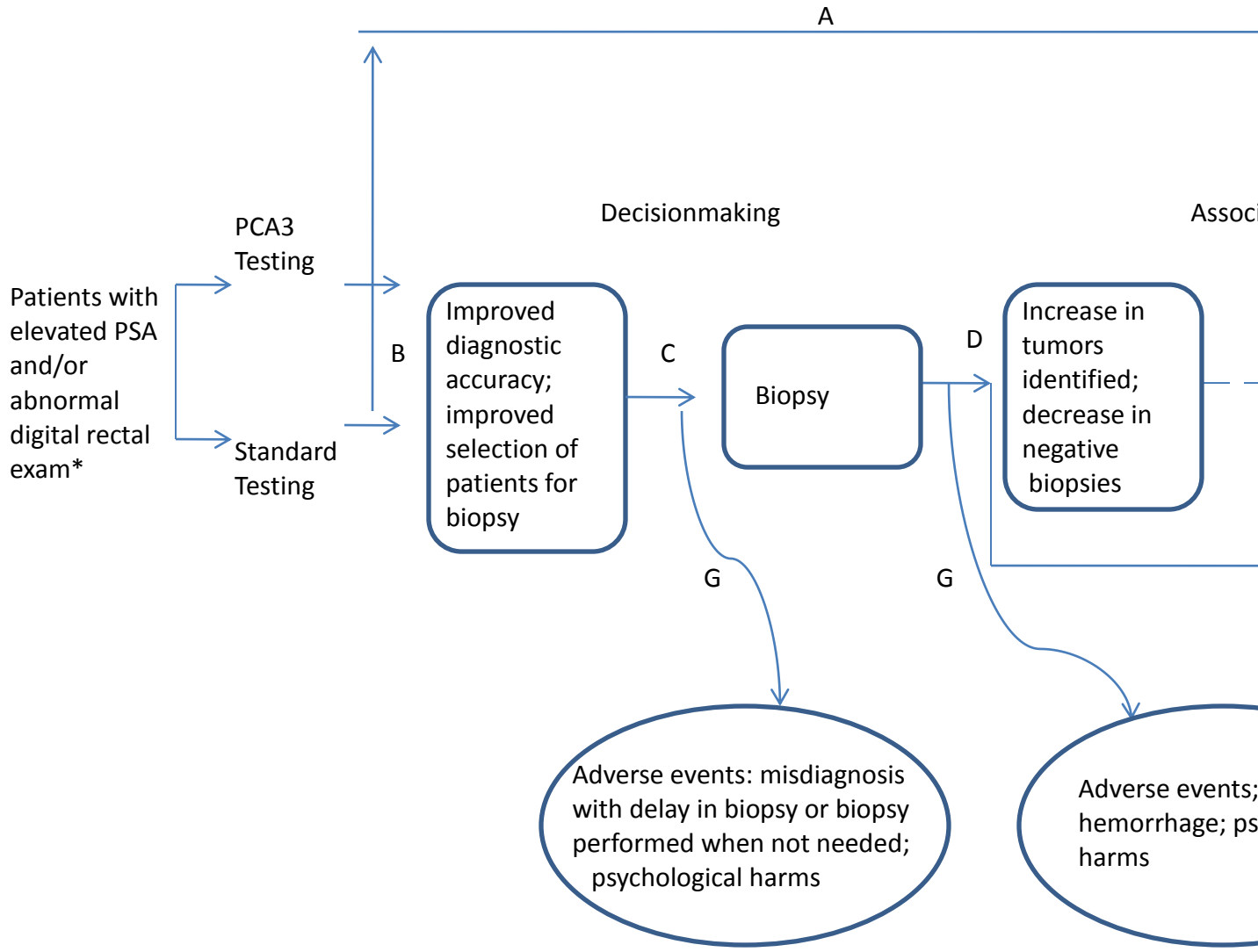
Setting

- All settings.

Source: www.effectivehealthcare.ahrq.gov
Published Online: October 14, 2011

III. Analytic Frameworks

PCA3 as a Diagnostic Indicator for Biopsy or Re-Biopsy in Patients with Elevated PSA and/or Abnormal Digital Rectal Examination (KQs 1 and 2)



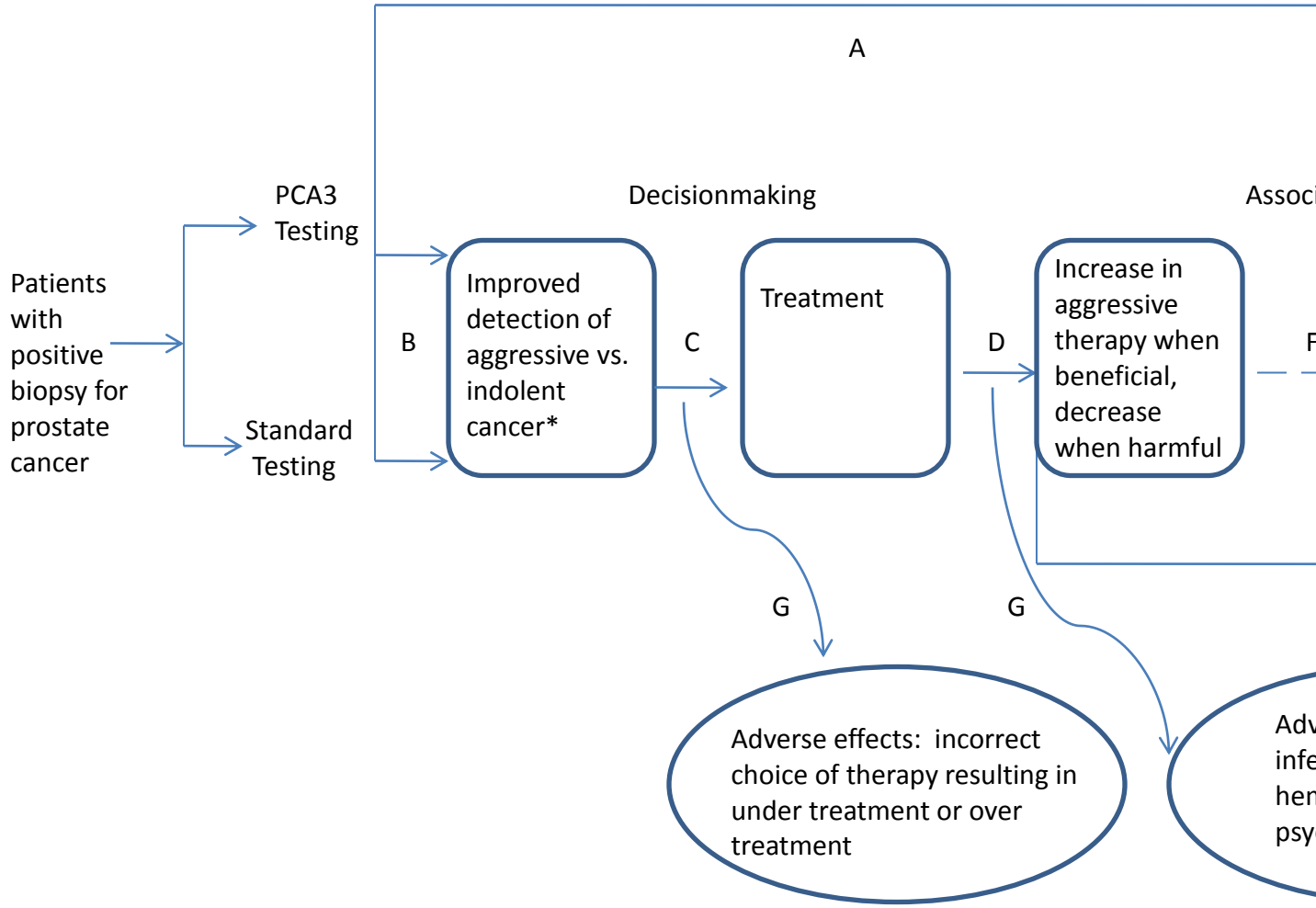
*Patients may be evaluated for initial biopsy or rebiopsy after one or more negatives.

Figure 1. Generic analytic framework for Key Questions 1 and 2. The patient population for Key Question (KQ) 1 is male patients who have an elevated PSA and/or an abnormal digital rectal examination (DRE). This figure depicts the comparative effectiveness of using PCA3 testing versus other screening tests (e.g., PSA) for intermediate outcomes and long-term health outcomes. 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 the diagnostic accuracy (clinical validity) of the PCA3 test and its designated comparators. Link C addresses the impact of test results on the decision to proceed to the initial prostate biopsy, which, in turn, impacts intermediate outcomes (Link D) and may affect health outcomes (Link E). Intermediate outcomes may have an association with health outcomes (Link F). Link G on the

Source: www.effectivehealthcare.ahrq.gov
Published Online: October 14, 2011

left addresses potential personal and clinical harms related to the effect testing has on the biopsy decision; Link G on the right focuses on adverse events/personal harms related to biopsy. The framework for KQ 2 is essentially the same, except the patient population is male patients who have elevated PSA and/or a positive DRE, as well as one or more previous prostate biopsies that were negative. Link C in this case addresses the impact of test results on the decision to perform a repeat prostate biopsy.

PCA3 to Distinguish Aggressive vs. Indolent Cancers of the Prostate (Key Question 3)



* Diagnostic accuracy

Figure 2. Analytic framework for KQ 3. The patient population for Key Question (KQ) 3 is male patients who have a biopsy positive for prostate cancer. This figure depicts the

Source: www.effectivehealthcare.ahrq.gov
Published Online: October 14, 2011

comparative effectiveness of using PCA3 testing versus other tests (e.g., Gleason score, tumor burden) for intermediate outcomes and long-term health outcomes. Direct evidence of the impact of testing on health outcomes (e.g., improved 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 distinguishing between aggressive and indolent tumor types. Link C addresses the impact of test results on decisionmaking related to prognosis or triage for active surveillance and/or aggressive treatment, which, in turn, impacts intermediate outcomes (Link D) and indirectly affects health outcomes (Link E). Intermediate outcomes may have an association with health outcomes (Link F). Link G addresses potential personal and clinical harms related to the effect of testing on treatment decisions and the clinical and personal harms related to treatment.

Source: www.effectivehealthcare.ahrq.gov
Published Online: October 14, 2011

IV. Methods

Methodological practices to be followed in this review will be derived from the AHRQ *Methods Guide for Effectiveness and Comparative Effectiveness Reviews*[10] (hereafter *Methods Guide*) and the AHRQ *Methods Guide for Medical Test Reviews* (hereafter AHRQ *Test Review Guide*)[11].

A. Criteria for Inclusion/Exclusion of Studies in the Review

Randomized Controlled Trials (RCTs) will be included, as well as observational and diagnostic accuracy studies that assess comparative effectiveness. Study designs will be classified according to the Community Guide model[12] and the internal validity hierarchy developed by the U.S. Preventive Services Task Force (USPSTF)[13]. Table 2 presents a preliminary set of criteria developed for inclusion and exclusion of studies based on our understanding of the literature.

B. Searching for the Evidence: Literature Search Strategies for Identification of Relevant Studies To Answer Key Questions

The research librarian in collaboration with the review team will develop and implement search strategies designed to identify evidence relevant to each KQ. The bibliographies of both primary studies and systematic reviews will be hand searched to assure complete identification of relevant articles. The time frame for the search will be limited to literature published after January 1, 1990 (PSA testing for early detection of prostate cancer was approved by the FDA in 1993) through the final search date in 2011. Literature searches will be restricted to the English language, with the exception of published articles in other languages for which English translations are made available. Two studies have demonstrated that excluding non–English-language studies has little impact on effect size estimates or conclusions relative to the resources required for translation.[14, 15] In this case, the results of most studies of PCA3 conducted in other countries are being published in English-language journals, so a small impact on results is anticipated.

Comprehensive searches for primary studies will be conducted in the following databases:

- MEDLINE®
- EMBASE®
- Cochrane Central Register of Controlled Trials (CENTRAL)

The databases will be searched for RCTs, nonrandomized controlled studies, nonrandomized comparative studies, systematic reviews, and prospective or retrospective cohort and case-control studies. The search strategy will use the National Library of Medicine's Medical Subject Headings (MeSH®) keyword system. As an example, the MEDLINE search will use combinations of the terms:

> "Prostatic Neoplasms"[Mesh] (OR text words "prostate cancer", "prostatic cancer", "prostate neoplasms", "prostatic neoplasms") AND ("prostate cancer antigen 3, human"[SUBSTANCE NAME] OR pca3[TIAB] OR "DD3 antigen, human"[SUBSTANCE NAME] OR ….additional synonyms)

Source: www.effectivehealthcare.ahrq.gov
Published Online: October 14, 2011

Table 2: Inclusion/exclusion criteria

Category	Inclusion/Exclusion Criteria
Population	KQ1: Adult male patients with elevated PSA and/or abnormal DRE ; no previous biopsies KQ2: Adult male patients with elevated PSA and/or abnormal DRE; one or more previous biopsies KQ3: Adult male patients with prostate cancer-positive biopsies
Interventions	KQs 1–3: PCA3 mRNA expression testing alone or in combination with other standard validated tests for prostate cancer KQs 1 & 2: Prostate biopsy Exclusion: Studies with biopsy protocols that do not require at least six cores per biopsy
Comparators	KQs 1 & 2: No PCA3 testing; use of standard validated tests for prostate cancer (total PSA, percent free PSA, PSA velocity, complexed PSA, externally validated nomograms) KQ3: No PCA3 testing; tumor volume, Gleason score, and clinical staging Exclusion: Studies using nomograms that have not been externally validated
Outcomes	KQs 1 & 2: Intermediate outcomes – diagnostic accuracy (e.g., sensitivity, specificity), biopsy decisionmaking, and number of unnecessary biopsies Long-term outcomes – mortality, morbidity, function, and quality of life Harms – complications of biopsy (infection or hemorrhage), anxiety about and discomfort of biopsy, and anxiety over false-positive results or misdiagnosis KQ 3: Intermediate outcomes – diagnostic accuracy (e.g., sensitivity, specificity), treatment decisionmaking, and monitoring/triage for active surveillance or aggressive treatment. Long-term outcomes – morbidity, mortality, function, and quality of life Harms – treatment-related harms and anxiety over false-positive results or misdiagnosis
Time Period	KQs 1–3: January 1,1990, to search date in 2011
Setting	KQs 1–3: All practice settings
Publication language	KQs 1–3: English or an English translation when available
Study designs	KQs 1–3: Systematic reviews, randomized or nonrandomized controlled trials, prospective or retrospective cohort studies, diagnostic accuracy studies, or case-control studies. The strength of association between the test results and the outcomes of interest will be assessed as part of both the study selection and analyses.
Followup duration	All durations of followup

Source: www.effectivehealthcare.ahrq.gov
Published Online: October 14, 2011

Sample size	No studies of PCA3 will be excluded based on sample size but this element will be considered as part of quality assessment. Studies of designated comparators may be limited to meta-analyses and to studies with 100 patients or more — based on an estimate of 10 covariables (e.g., age, PSA levels, family history) in the study with a minimum of 10 subjects per covariable.

along with key terms in titles and abstracts (e.g., biops*, aggressive, significan*, clinical*, prognos*). Proposed search strategies are shown in Appendix A. Limits include "Humans" and "Publication Date 1990/01/01 to 2011/08/29."

For the purposes of our review, the grey literature comprises information that is not controlled by commercial publishing. This literature includes: abstracts presented at major oncology and urology meetings; regulatory documents (e.g., FDA reviews); proprietary data via submitted scientific information packets; and manufacturer Web sites and clinical trial registries (ClinicalTrials.gov). Electronically retrievable abstracts from relevant conferences (e.g., American Urological Association, American Society of Clinical Oncology, National Comprehensive Cancer Network Annual Congress, Prostate Cancer Research Institute) held within the last 2 years will also be searched. Scientific Information Packets from Gen-Probe will be sought through the Scientific Resource Center. Other information reviewed will include grey literature from ClinicalTrials.gov, and relevant information on tests from the FDA. This information will be used to determine if studies of intermediate or long-term outcomes are in the pipeline.

Searches for systematic reviews will be conducted in MEDLINE®, the Cochrane Database of Systematic Reviews, and the Web sites of the National Guideline Clearinghouse (www.guidelines.gov), and the Health Technology Assessment Programme (www.hta.ac.uk) and the National Institute for Clinical Excellence (www.nice.org.uk), both of which are in the United Kingdom**.**

Unpublished data will be included as evidence only if published in a peer-reviewed journal during the course of our review. We will conduct an updated search of the published literature upon submission of the draft report to determine if new information has been published since completion of the previous search. Newly identified studies will be added to the draft report as needed. The Technical Expert Panel (TEP) and individuals providing peer review will be asked to inform the project team about any studies relevant to the KQs that they did not find on the list of selected studies. Results from the updated search will be stored in an EndNote9® library.

Data Selection

After we upload the EndNote library that contains the results of our literature searches, we will use DistillerSR software to determine study eligibility. Using three preset questions that reflect our inclusion criteria, the reviewers will first screen abstracts, marking each of the three questions 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). DistillerSR provides a report on discrepancies between reviewers that will be resolved by discussion and consensus opinion; a third reviewer may be consulted. Using a second set of six selection questions, full articles will be reviewed using the same approach to determine eligibility for data abstraction. As in the abstract selection step, a "No" response to any of the article selection questions in DistillerSR makes the article ineligible, but all the reasons for excluding each article will be captured by DistillerSR and will be available by report.

Source: www.effectivehealthcare.ahrq.gov
Published Online: October 14, 2011

Only studies reporting results of pretreatment markers will be reviewed. Articles that meet the selection criteria (e.g., populations, interventions, outcomes of interest) for abstraction will provide information on:

- *Intermediate outcomes of interest,* including diagnostic accuracy (i.e., comparing the ability of two tests to predict biopsy results) and impact on decisionmaking (e.g., reduction in negative biopsies and increased detection of prostate tumors).

    - Matched studies

    The highest quality studies for this purpose will be "matched" studies of PCA3 and one or more designated comparator (e.g., primary matched studies or systematic reviews/meta-analyses of primary matched studies). In general, these studies will include a cohort of men with elevated PSA and/or a positive DRE for whom several test results are available (e.g., PCA3, PSA and fPSA), as well as prostate biopsy or prostatectomy results. Strength of evidence for each PCA3/comparator pair will be assessed as described in Section F of this protocol. No primary matched studies will be excluded based on sample size, but this will be considered part of quality assessment. If identified, systematic reviews/meta-analyses of two or more comparators other than PCA3 will also be selected.

    - Unmatched studies

    If the strength of evidence is "low" or "insufficient" (Table 3), the option of abstracting and analyzing lower quality studies of PCA3 (e.g., unmatched diagnostic accuracy studies or systematic reviews/meta-analyses of such primary studies) will be considered, and a decision will be made with the Evidence-based Practice Center (EPC) team. In general, these studies will include only a single test (e.g., PCA3 or fPSA), and the comparison will be made between, rather than within, studies.

- *Long-term health outcomes of interest,* including reduced morbidity/mortality, improved quality of life, and/or fewer clinical and personal harms from the interventions.

    - Matched studies

    Matched studies will be those that report on PCA3 testing and at least one other designated comparator and on health outcomes beyond biopsy results. Such studies would include matched comparator studies and RCTs and systematic reviews/meta-analyses of these studies.

    - Unmatched studies

    Unmatched studies will be nonrandomized trials or other comparative studies that report on PCA3 testing and at least one other comparator and on health outcomes beyond biopsy results. In general, comparisons will be made between, rather than within, studies.

C. Data Abstraction and Data Management

Data from all included studies will be abstracted using DistillerSR software into standard evidence tables by one reviewer, and checked for accuracy and completeness by a second reviewer. Data abstraction tables will be pilot tested for completeness on a group of select

Source: www.effectivehealthcare.ahrq.gov
Published Online: October 14, 2011

studies and will be revised as necessary before full data abstraction begins. Project staff will meet regularly to discuss the results at each phase, review studies that are difficult to classify, and address any questions that the team may have. Authors of included studies will be contacted if clarification of methods or results is needed.

The data elements to be abstracted were established in consultation with the TEP and include:

Study Description:

- Author, Year
- Country
- Institution type
- Institutions
- Enrollment period
- Source of Funding
- Authors' disclosures of industry relationship(s)

Study Design, Setting, Enrollment and Followup Numbers, Interventions

- Study design
- Setting
- Number of patients studied at enrollment, after PCA3 testing (or comparator testing), at prostate biopsy, and at followup
- Average followup in months
- Prostate biopsy findings
  - Number of cores per biopsy
  - Number of positive cores per biopsy
  - Percentage of cancer per biopsy core
  - Gleason grades, scores
  - PSA density
  - Pathology stage
  - Prostate volume
  - Percentage of "insignificant findings" based on pathology
  - Results of magnetic resonance imaging
- Prostatectomy findings
  - Pretreatment clinical stage
- Interpretation blinded to study categories

Long-term Outcomes

- Mortality
  - Overall and prostate cancer-specific
  - 10-year survival
- Morbidity
  - Local progression
  - Distant metastases
  - Pain
  - Biochemical failure
- Treatment-related morbidity
  - Urinary incontinence

Source: www.effectivehealthcare.ahrq.gov
Published Online: October 14, 2011

  - Impotence
  - Rectal incontinence
  - Proctitis
- Quality-of-life measures

Participant Characteristics

- Age
- Race
- Benign prostatic hyperplasia
- High-grade prostatic intraepithelial neoplasia
- Atypical small acinar proliferation
- Criteria for study inclusion
  - Elevated total PSA
  - Positive DRE
  - ≥ 1 negative biopsies
  - Family history
  - African American ancestry
  - Age
  - Scheduled for biopsy – type not specified
  - Positive biopsy
  - Other

PCA3 Specimen Collection

- Method of collection – attentive prostate message
- Specimen
  - Urine – sedimented
  - Urine – unsedimented
  - Prostatic ejaculate
- Handling information
  - Time to transport
  - Holding temperature
  - Type of transport media
  - Storage temperature

PCA3 Assay

- Assay used
  - First generation
  - Second generation
  - Third generation
- Housekeeping gene
- Result unit
- Testing blinded to outcomes

PCA3 Test Results

- Cut-off point(s)
- Summary statistics
  - Stratified by negative or positive biopsy result

Source: www.effectivehealthcare.ahrq.gov
Published Online: October 14, 2011

    - Stratified by PCA3 cut-off
- Performance parameters (e.g., sensitivity/specificity, PPV/NPV, odds ratio [OR])
    - stratified by negative or positive biopsy result
    - stratified by PCA3 cut-off
- Area under the curve (AUC)

Comparator Results

- Total PSA
    - Test/vendor
    - PSA cut-off points
    - Summary statistics
        - Stratified by negative or positive biopsy result
        - Stratified by PSA cut-off
        - *p* value versus PCA3
    - Performance parameters (e.g., sensitivity/specificity, PPV/NPV, OR)
        - stratified by negative or positive biopsy result
        - stratified by PSA cut-off
    - AUC
- PSA velocity
    - Action point
    - Calculation described over designated period of time
    - Summary statistics
        - Stratified by negative or positive biopsy result
        - *p* value versus PCA3
    - Performance parameters (e.g., sensitivity/specificity, PPV/NPV, OR)
        - Stratified by negative or positive biopsy result
        - AUC
- Free PSA
    - Calculation of result
    - Increased risk <10 percent or <25 percent
    - Summary statistics
        - Stratified by negative or positive biopsy result
        - Stratified by percent fPSA cut-off
        - *p* value versus PCA3
        - Performance parameters (e.g., sensitivity/specificity, PPV/NPV, OR)
        - Stratified by negative or positive biopsy result
        - Stratified by percent fPSA cut-off
    - AUC
- Complexed PSA
    - Summary statistics
        - Stratified by negative or positive biopsy result
        - *p* value versus PCA3
    - Performance parameters (e.g., sensitivity/specificity, PPV/NPV, OR)
        - Stratified by negative or positive biopsy result
    - AUC
- Externally validated nomograms

Source: www.effectivehealthcare.ahrq.gov
Published Online: October 14, 2011

- Variables in nomogram
- Parameters of nomogram
- Calculation of score
- Summary statistics
  - Stratified by negative or positive biopsy result
  - *p* value versus PCA3
- Performance parameters (e.g., sensitivity/specificity, PPV/NPV, OR)
  - Stratified by negative or positive biopsy result
- AUC

Diagnostic Accuracy Data

- Reference standard
  - Prostate biopsy
  - Prostatectomy
- Test
  - N, TP, FN, FP, TN
  - Sensitivity (95% confidence interval [CI])
  - Specificity
  - PPV
  - NPV
  - OR
  - AUC

Quality Ratings

- Randomized prospective interventional study
- Nonrandomized comparative intervention study
- Diagnostic accuracy (QUADAS)
- Systematic review/meta-analysis (modified AMSTAR)

Evidence Tables

Templates for evidence tables will be created in Microsoft Access® and Excel®. One reviewer will perform primary data abstraction of all data elements into the evidence tables, and a second reviewer will review articles and evidence tables for accuracy. Disagreements will be resolved by discussion and if necessary by consultation with a third reviewer. When small differences occur in quantitative estimates of data from published figures, the values will be obtained by averaging the two reviewers' estimates.

D. Assessment of Methodological Quality of Individual Studies and Reviews

To assess the methodological quality of included studies, we will use study design specific quality criteria proposed by the USPSTF.[13,16] In all cases, quality assessment of included studies will be performed independently by two senior staff, and disagreements will be resolved through discussion or third-party adjudication as needed. Quality assessments will be summarized for each study and recorded in tables.

Quality will be assessed on the basis of the following criteria for randomized prospective interventional (intention to test and treat) studies[16]:

Source: www.effectivehealthcare.ahrq.gov
Published Online: October 14, 2011

- Initial assembly of comparable groups: adequate randomization, including concealment and whether potential confounders (e.g., other concomitant care) were distributed equally among groups
- Maintenance of comparable groups (includes attrition, crossovers, adherence, and contamination)
- Important differential loss to followup or overall high loss to followup
- Measurements: equal, reliable, and valid (includes masking of outcome assessment)
- Clear definition of interventions
- All important outcomes considered and defined
- Analysis: adjustment for potential confounders and intention-to-treat analysis

Quality will be assessed on the basis of the following criteria for nonrandomized comparative intervention studies[13,17]:

- Were the sample definition and selection prospective or retrospective?
- Were inclusion/exclusion criteria clearly described?
- Were participants selected to be representative?
- Was there an attempt to balance groups by design?
- Were baseline prognostic characteristics clearly described and groups shown to be comparable?
- Were interventions clearly specified?
- Were participants in treatment groups recruited within the same time period?
- Was there an attempt by investigators to allocate participants to treatment groups in an attempt to minimize bias?
- Were concurrent/concomitant treatments clearly specified and given equally to treatment groups?
- Were outcome measures clearly valid, reliable, and equally applied to treatment groups?
- Were outcome assessors blinded?
- Was the length of followup adequate?
- Was subject attrition below an overall high level (<20%)?
- Was the difference in attrition between treatment groups below a high level (<15%)?
- Did the analysis of outcome data incorporate a method for handling confounders such as statistical adjustment?

The rating of intervention studies encompasses three quality categories:

Good studies meet all criteria:

- Comparable groups are assembled initially and maintained throughout the study (followup at least 80%)
- Reliable and valid measurement instruments are used and applied equally to the groups
- Interventions are spelled out clearly
- All important outcomes are considered
- Appropriate attention is given to confounders in analyzing data
- Intention-to-treat analysis is used for RCTs

Fair studies have any or all of the following problems, but without the fatal flaws noted in the "poor" category below:

Source: www.effectivehealthcare.ahrq.gov
Published Online: October 14, 2011

- Comparable groups are assembled initially, but some questions remain about whether some (although not major) differences occurred with followup
- Measurement instruments are acceptable (although not the best) and are generally applied equally
- Some but not all important outcomes are considered
- Some but not all potential confounders are accounted for
- Intention-to-treat analysis has been done for RCTs

Poor studies have any of the following fatal flaws:

- Groups assembled initially are not close to being comparable or maintained throughout the study
- Unreliable or invalid measurement instruments are used or not applied at all equally among groups (including not masking outcome assessment)
- Key confounders are given little or no attention
- Intention-to-treat analysis is lacking for RCTs

Diagnostic Accuracy Studies

The quality of included diagnostic accuracy studies will be assessed by the Quality Assessment of Diagnostic Accuracy Studies (QUADAS) tool, which underwent a rigorous development process by Whiting and colleagues[18] and includes the following items:

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?

Quality assessment of included studies will be performed independently by two senior staff, with disagreements resolved through discussion or third-party adjudication as needed. Quality assessment will be performed by analyzing the individual items described above and their

Source: www.effectivehealthcare.ahrq.gov
Published Online: October 14, 2011

potential for contributing to study bias, and the results will then be summarized for each study and recorded in tables.

Quality ratings are summarized into general quality classes (from Table 5-4, Paper 5, AHRQ *Test Review Guide*)[11]:

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 that may invalidate the results.

Systematic Reviews and Meta-Analyses

The quality of included systematic reviews/meta-analyses will be assessed with the AMSTAR tool[19], with minor modifications to accommodate our comparative effectiveness review (CER) and its topic. Specifically, we integrated some AMSTAR guidance elements into the KQs, used the AMSTAR changes suggested for CERs in Chapter 11 of the *Methods Guide*[10], and added questions on meta-analysis (based on validated concepts from the PRISMA checklist[19] for reporting systematic reviews/meta-analyses).

1. Did the authors provide a rationale and ''a priori'' design for the review, and were the research questions and inclusion criteria established before the conduct of the review?
2. Was there dual review for study selection and data extraction, and was there a consensus procedure for resolving disagreements?
3. Was there a comprehensive literature search that adequately addressed the KQs, included at least two electronic databases, described all information sources with coverage dates, and stated MeSH terms/key words along with an electronic search strategy for at least one database?
4. Were the electronic searches supplemented by reviewing the references in the studies and reviews found, by consulting experts in the particular field of study, or other approaches if relevant?
5. Was the status of publication (i.e., grey literature) used as an inclusion criterion? If a grey literature search was conducted, was the search strategy provided, and how was quality assessed?
6. Was a list of studies (included and excluded) provided? Were inclusion/exclusion criteria and their rationale clearly stated—along with the numbers of studies screened, assessed for eligibility, and included in the review—with reasons for exclusions at each stage (ideally with a flow diagram)?
7. Were any discrepancies between data from primary papers and the published systematic review/meta-analysis identified?
8. Were the characteristics of the included primary studies provided, including those of the study participants (e.g., age, race, sex, relevant socioeconomic data, disease status, duration, severity, or other diseases), the interventions, and the outcomes?
9. Was the scientific quality (e.g., sample size, study design, blinding, biases, and confounders) of the included individual studies assessed and documented? For study designs other than RCTs, were design limitations discussed?
10. Was the quality of included primary studies high?
11. Was the scientific quality of the included studies used appropriately in formulating conclusions? [Note: Irrelevant if only the data and analysis from the systematic review are used in this CER and the conclusions are not.]
12. Were the methods used to combine the findings of studies appropriate? For the pooled results, was a test done to assess homogeneity (i.e., Chi-squared test, I^2)? If heterogeneity

Source: www.effectivehealthcare.ahrq.gov
Published Online: October 14, 2011

was observed, was a random effects model used and/or the clinical appropriateness of combining studies taken into consideration?

13. For each study, and for all outcomes considered, were simple summary data estimates and confidence intervals provided? Did each meta-analysis include the model used; results with confidence intervals; homogeneity measures; description of methods used for any additional analyses (e.g., sensitivity or subgroup analyses); and assessment of bias across studies?
14. For each study, and for all outcomes considered, were simple summary data estimates and confidence intervals provided? Did each meta-analysis include the model used; results with confidence intervals; homogeneity measures; description of methods used for any additional analyses (e.g., sensitivity or subgroup analyses); and assessment of bias across studies?
15. Was potential conflict of interest addressed by acknowledging sources of support in the systematic review and the included studies?

Quality assessment of included studies will be performed independently by two senior staff, with disagreements resolved through discussion or third-party adjudication as needed. Quality assessments will be summarized for each study and recorded in tables.

After taking into account the number and seriousness of any methodological shortcomings, categorical quality classes can be defined (according to Chapter 11 in the *Methods Guide*[10]):

Good. The study has few or no methodological shortcomings and a low risk of bias.
Fair. The study has some methodological flaws, but the EPC staff determined that the flaws will not seriously bias or invalidate the results.
Poor. The study has one or more flaws that, in the judgment of the EPC staff, are highly likely to bias or invalidate the results.

E. Data synthesis

The strongest study design for determining pair-wise relative performance of PCA3 testing and each comparator (e.g., total PSA, % free PSA, PSA velocity, complexed PSA, or externally validated nomograms) in predicting initial or repeat biopsy results or tumor prognosis will be derived from the direct comparisons available in matched studies.

Matched Analysis

For each individual matched study, the data can most likely be summarized in one or more 2×2 tables. These data can be analyzed with McNemar's test to determine the statistical significance of PCA3 versus the comparator. The effect sizes for these comparisons can be provided as ORs, estimated by the ratio of the values in the off diagonal (b/c) with an associated confidence interval. Using standard statistical methods, the data might be pooled (random effects meta-analysis), with the ORs between similar studies combined for an overall estimate and associated 95% CI (Comprehensive Meta-analysis Version 2, Biostat Inc., Englewood, NJ). This software supports meta-analysis of both matched and unmatched data. Other methods could also be used such as comparisons of areas under the curve (AUC), areas under a limited portion of the curve, and differences or ratios of correct proportions, depending on ease of interpretation. If the number of matched studies per pair (e.g., PCA3 and %fPSA) is limited, the relatively small cumulative numbers may make the exploration of heterogeneity difficult. Within these matched analyses, it might be possible to consider a chain of evidence that evaluates the totality of the evidence using informal methods.

Source: www.effectivehealthcare.ahrq.gov
Published Online: October 14, 2011

Potential Unmatched Analyses

If the matched comparisons described above are not identified (or the cumulative strength of evidence is low or insufficient), unmatched comparisons could be considered. For example, the results of a meta-analysis of studies that investigated the diagnostic accuracy of PCA3 testing could be compared to a meta-analysis of studies that investigated the diagnostic accuracy of free PSA testing. The comparison is complicated, however, by known and unknown variables such as differences in study populations, cut-off levels, assays used, and outcomes included. These differences could impact the pooling of studies and/or the comparison of summary results.

If sufficient studies that involve PCA3 testing are identified with reasonably consistent study designs, patient populations, comparators, and outcomes of interest, a meta-analysis(es) will be performed that is consistent with the approach outlined in Chapter 9 of the *Methods Guide*[10] and Paper 8 in the AHRQ *Test Review Guide.*[11] Such analyses will include: summary performance estimates with 95% CIs, use of a random effects model (Comprehensive Meta-analysis Version 2, Biostat Inc., Englewood NJ), generation of forest plots as recommended[11], and evaluation of between-study heterogeneity (Q test, I^2). All *p*-values are two sided at the 0.05 level. Should heterogeneity be found and if sufficient studies are available, subgroup analyses will be performed to investigate the source. Potential covariates could include: study size, PCA3 assay method or type, test cut-off levels, characteristics of comparator tests, geographical study sites, study subject inclusion criteria, biopsy protocol (e.g., number of cores taken per biopsy), pathology findings, blinding, and study quality. Tests will be performed to determine whether a threshold effect is present, and if so, an analysis of summary receiver operating characteristics (SROCs) can be used to synthesize information from multiple studies on diagnostic accuracy, including summary AUC and *p*-value (SAS 9.2 software [SAS Institute, Inc., Cary, NC] can also generate hierarchical SROCs if needed).[11] It is likely that a range of cut-off levels will be evaluated to allow flexibility in use of the test in varying clinical scenarios.

Assuming that two unmatched analyses can be performed for PCA3 testing and a comparator, further analyses will attempt to compare performance as reported in the two studies. This comparison might be made in several ways, depending on the characteristics of the tests and the available data. First, AUCs could be compared, although this approach has the disadvantage of containing information about the performance of the test at thresholds that would not be clinically relevant. Alternatively, selected specificities could be chosen (e.g., 95%, 90%, and 85%), and the corresponding sensitivities (with CIs) be determined from the respective SROC. In many of these pair-wise unmatched comparisons it may not be possible to provide a reliable comparison of test performance. In such instances, a qualitative conclusion might be possible. For all comparisons, gaps in knowledge will be identified, and study designs relevant to addressing these gaps will be presented.

Although we do not anticipate that we will identify RCTs or nonrandomized clinical trials, if we identify such studies we will use the statistical techniques recommended in the *Methods Guide*.[10]

F. Grading the Evidence for Each Key Question

We will grade the strength of evidence for primary outcomes by using the standard process of the Evidence-based Practice Centers as outlined in the *Methods Guide*.[10,20] The grade will be based on four major domains: risk of bias, consistency, directness, and precision of the evidence. We will classify the bodies of evidence pertaining to each primary outcome into four basic grades: high, moderate, low, and insufficient (Table 3).[20] Additional domains such as

Source: www.effectivehealthcare.ahrq.gov
Published Online: October 14, 2011

dose-response association, plausible confounding, strength of association, and publication bias will be assessed and reported if applicable.

G. Assessing Applicability

In addition to quality assessment, we will also assess the applicability of studies. Judgments of applicability for each outcome (including harms) will be performed separately from assessments of the other domains of strength of evidence.[10,21] We will identify and abstract factors in individual studies that might affect applicability, particularly factors related to the

Table 3. Strength of evidence grades and definitions[20]

Grade	Definition
High	High confidence that 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.
Low	Low confidence that the evidence reflects the true effect. Further research is likely to change the confidence in the estimate of effect and is likely to change the estimate.
Insufficient	Evidence either is unavailable or does not permit a conclusion.

populations studied (e.g., selection criteria, geography, race/ethnicity), the tests used, and/or the specific details of the gold standard used (e.g., number of cores per biopsy).

It is anticipated that several factors may produce a major impact on applicability. Most studies have taken place in either the United States or Europe. Prostate cancer screening practices in the general population at risk (men over the age of 50) may not be the same in these two geographic locations and may differ from country to country in Europe. Populations with higher frequency of screening (e.g., larger number of men; shorter intervals between PSA testing) are likely to produce patient study populations with cancer that is harder to detect (missed in prior screening) and may yield a reduced prevalence of disease.

Treatment interventions differ in frequency and type in different countries and at different practice centers. Treatments have been evolving over the past 10 years, with increased use of robotic surgical approaches and improved visualization of the prostate during radiation treatment. In addition, the identification of patients with indolent cancer and decisions about the best treatment approach (i.e., use of active surveillance vs. watchful waiting) has been the subject of both ongoing studies and practice evolution. Given this practice heterogeneity, it may be difficult to interpret test results from publications (especially older ones) and apply them to current best practices.

The number of core biopsies used in each study may impact the number of cancerous tumors identified. More core biopsies will increase the yield of cancerous tumors and will potentially increase test sensitivity but decrease test specificity. While the current standard appears to be 12 cores per biopsy per patient, some reports are based on the use of a smaller number of cores (6, 8, or 10) per biopsy.

Based on these characteristics and any others identified during the literature review, we will note any potential limitations to applicability based on the interpretation of individual studies and evaluation of the applicability of the total body of evidence.

References

Source: www.effectivehealthcare.ahrq.gov
Published Online: October 14, 2011

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. Erratum in: CA Cancer J Clin 2011 Mar-Apr;61(2):133-134.
2. 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):5975-5979. PMID 10606244.
3. Hessels D, Klein Gunnewiek JM, van Oort I, et al. DD3(PCA3)-based molecular urine analysis for the diagnosis of prostate cancer. Eur Urol 2003 Jul;44(1):8-15; discussion 15-16. PMID 12814669.
4. Vlaeminck-Guillem V, Ruffion A, Andre J. [Value of urinary PCA3 test for prostate cancer diagnosis French]. Prog Urol 2008 May;18(5):259-265. PMID 18538269.
5. Amling CL, Catalona WJ, Klein EA. Deciding whom to biopsy. Urol Oncol 2010 Sep-Oct;28(5):542-545. PMID 20816613.
6. 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. Eur Urol 2008 Aug;54(2):274-290. PMID 18511177.
7. 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. Eur Urol 2009 Oct;56(4):659-667. PMID 19304372.
8. 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):1303-1308; discussion 1308. PMID 18707724.
9. Ficarra V, Novara G, Zattoni F. The role of the prostate cancer antigen 3 (PCA3) test for the diagnosis of prostate cancer in the era of opportunistic prostate-specific antigen screening. Eur Urol 2010 Oct;58(4):482-484; discussion 484-485. PMID 20685032. Comment on: Eur Urol 2010 Oct;58(4):475-481.
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. Methods Guide for Medical Test Reviews. Rockville, MD: Agency for Healthcare Research and Quality; Draft Posted November 2010. Available at: http://www.effectivehealthcare.ahrq.gov/tasks/sites/ehc/assets/File/methods_guide_for_medical_tests.pdf.
12. Briss PA, Zaza S, Pappaioanou M, et al. Developing an evidence-based Guide to Community Preventive Services--methods. The Task Force on Community Preventive Services. Am J Prev Med 2000 Jan;18(1 Suppl):35-43. PMID 10806978.
13. U.S. Preventive Services Task Force Procedure Manual. Rockville, MD: Agency for Healthcare Research and Quality, July 2008. AHRQ Publication No. 08-051180-EF.
14. Juni P, Holenstein F, Sterne J, et al. Direction and impact of language bias in meta-analyses of controlled trials: empirical study. Int J Epidemiol 2002 Feb;31(1):115-123. PMID 11914306.
15. 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):964-972. PMID 11004423.
16. Harris RP, Helfand M, Woolf SH, et al; Methods Work Group, Third Preventive Services Task Force. Current methods of the US Preventive Services Task Force: a review of the process. Am J Prev Med 2001 Apr;20(3 Suppl):21-35. PMID 11306229.
17. Deeks JJ, Dinnes J, D'Amico R, et al. Evaluating non-randomised intervention studies. Health Technol Assess 2003 Sep;7(27):1-173. PMID 14499048.

Source: www.effectivehealthcare.ahrq.gov
Published Online: October 14, 2011

18. 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 Mar 6;6:9. PMID 16519814.
19. 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):1013-1020. PMID 19230606.
20. 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):513-523. PMID: 19595577.
21. Atkins D, Chang S, Gartlehner G, et al. Assessing applicability when comparing medical interventions. Agency for Healthcare Research and Quality and the Effective Health Care Program. J Clin Epidemiol 2011Apr 2 [Epub ahead of print]. PMID: 21463926.

VI. Definition of Terms

Area under the curve (AUC) — area under the receiver operating characteristic (ROC) curve; as applied to ROC curves, the area subtended by the ROC curve and bounded by the x-axis (false-positive fraction) and the y-axis (true-positive fraction) (I/LA21).

Complexed PSA — the faction of prostate-specific antigen (PSA) bound to serum protease inhibitors (especially alpha-1 antichymotrypsin); some reports have suggested that this measurement helps distinguish benign from malignant prostate disease.

Free PSA — PSA in the circulation that is unbound to its usual carrier molecules, the protease inhibitors. Free PSA is used to distinguish prostate cancer from benign prostatic hyperplasia or other benign conditions of the prostate.

Negative likelihood ratio — the ability of the diagnostic test to accurately "rule out" the presence of prostate cancer.

Odds ratio — the ratio indicating the odds of a particular event occurring in one population to the odds of it occurring in another population.

Positive likelihood ratio — the ratio of the diagnostic test to accurately predict the presence of prostate cancer

Predictive value of a positive result — the probability of a man actually having prostate cancer after testing positive for cancer. Positive predictive value = (true positives)/(true positives + false positives).

Predictive value of a negative result — the probability of actually not having prostate cancer after testing negative for prostate cancer. Negative predictive value = (true negatives)/(true negatives + false negatives)

Prostate cancer antigen 3 (PCA3) — a gene associated with prostate cancer; no associated protein product has been identified and function of the gene is unknown; RNA levels can be measured in urine to detect prostate cancer.

Source: www.effectivehealthcare.ahrq.gov
Published Online: October 14, 2011

Prostate-specific antigen (PSA) — a glycoprotein secreted by prostate gland epithelial cells; increased levels of which are found in the blood of patients with cancer of the prostate; levels may be elevated in association with other pathologies affecting the prostate gland, including benign prostatic hyperplasia and prostatitis.

Prostate-specific antigen velocity (PSA velocity) — a measure of the rapidity of change in a man's PSA level. It is associated with the presence and activity of prostate cancer.

Receiver operating characteristics curve (ROC curve) — a graph of sensitivity against 1 – specificity.

Sensitivity — the proportion of men with prostate cancer who test positive for cancer. Sensitivity = (true positives)/(true positives + false negatives).

Specificity — the proportion of men without prostate cancer who test negative for cancer. Specificity = (true negatives)/(false positives + true negatives).

VII. Summary of Protocol Amendments

In the event of protocol amendments, the date of each amendment will be accompanied by a description of the change and the rationale.

VIII. Review of Key Questions

For all EPC reviews, key questions were reviewed and refined as needed by the EPC with input from Key Informants and the Technical Expert Panel (TEP) to assure that the questions are specific and explicit about what information is 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.

IX. Key Informants

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 healthcare decisions. The EPC solicits input from Key Informants when developing questions for systematic review or when identifying high priority research gaps and needed new research. Key Informants 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 must 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, individuals are invited to serve as Key Informants and those who present with potential conflicts may be retained. The TOO and the EPC work to balance, manage, or mitigate any potential conflicts of interest identified.

X. Technical Experts

Source: www.effectivehealthcare.ahrq.gov
Published Online: October 14, 2011

Technical Experts comprise a multi-disciplinary group of clinical, content, and methodological experts who provide input in defining populations, interventions, comparisons, or outcomes as well as identifying particular studies or databases to search. They are selected to provide broad expertise and perspectives specific to the topic under development. Divergent and conflicted opinions are common and perceived as health scientific discourse that results in a thoughtful, relevant systematic review. Therefore study questions, design and/or methodological approaches do not necessarily represent the views of individual technical and content experts. Technical Experts provide information to the EPC to identify literature search strategies and recommend approaches to specific issues as requested by the EPC. Technical Experts do not do analysis of any kind nor contribute to the writing of the report and have not reviewed the report, except as given the opportunity to do so through the public review mechanism.

Technical Experts must disclose any financial conflicts of interest greater than $10,000 and any other relevant business or professional conflicts of interest. Because of their unique clinical or content expertise, individuals are invited to serve as Technical Experts and those who present with potential conflicts may be retained. The TOO and the EPC work to balance, manage, or mitigate any potential conflicts of interest identified.

XI. Peer Reviewers

Peer reviewers are invited to provide written comments on the draft report based on their clinical, content, or methodological expertise. Peer review comments on the preliminary draft of the report are considered by the EPC in preparation of the final draft of the report. Peer reviewers do not participate in writing or editing of the final report or other products. The synthesis of the scientific literature presented in the final report does not necessarily represent the views of individual reviewers. The dispositions of the peer review comments are documented and will, for CERs and Technical briefs, be published three months after the publication of the Evidence report.

Potential Reviewers must disclose any financial conflicts of interest greater than $10,000 and any other relevant business or professional conflicts of interest. Invited Peer Reviewers may not have any financial conflict of interest greater than $10,000. Peer reviewers who disclose potential business or professional conflicts of interest may submit comments on draft reports through the public comment mechanism.

Appendix A:

02/09/11

((prostate cancer antigen 3, human [SUBSTANCE NAME] OR pca3 [TIAB] OR (prostate cancer antigen 3 [TIAB]) OR (prostatic cancer antigen 3 [TIAB]) OR DD3 antigen, human [SUBSTANCE NAME] OR (differential display code 3 [TIAB]) OR dd3 [TIAB]) OR (prostatic neoplasms [MH] AND ((clinical* [TIAB] AND (significan* [TIAB] OR importan* [TIAB])) OR aggressive [TIAB] OR biops* [TIAB]) AND (nomogram [TIAB] OR (neural [TIAB] AND network [TIAB])))) OR ((((((clinical* [TIAB] AND (significan* [TIAB] OR importan* [TIAB])) OR aggressive [TIAB] OR biops* [TIAB]) AND prostate-specific antigen [MH]) AND prostatic neoplasms [MH]) AND (predict* [TIAB] OR prognos* [TIAB])))

#2505

Source: www.effectivehealthcare.ahrq.gov
Published Online: October 14, 2011

Revised Searches (PubMed only)

"prostate cancer antigen 3, human" [Supplementary Concept] OR
("differential display code 3 antigen" OR DD3) Field: Title/Abstract OR
(PCA3 OR "prostate cancer antigen 3") Field: Title/Abstract

- This was the test-specific set= 208 in PubMed

Additionally -
"Prostatic Neoplasms"[Mesh] OR "prostatic neoplasms" OR "prostate neoplasms" OR "prostatic cancer" OR "prostate cancer"
AND
(nomogram OR (neural AND network) OR antigen OR antigens) Field: Title/Abstract
AND
((clinical* AND (significan* OR importan* OR aggressive OR biops*)) OR predict* OR prognos* OR (select* OR decid* OR decision* OR choos* OR choice*)) Field: Title/Abstract
NOT the test-specific set
AND Limits: Humans, Publication Date from 19

Source: www.effectivehealthcare.ahrq.gov
Published Online: October 14, 2011