Evaluation of New Tests and Interventions for Prostate Cancer Management: A Systematic Review

Ghadeer Olleik, MSc\textsuperscript{a,b}; Wassim Kassouf, MD\textsuperscript{c}; Armen Aprikian, MD\textsuperscript{a,c}; Jason Hu, MSc\textsuperscript{a,b}; Marie Vanhuyse, MSc, MD\textsuperscript{e}; Fabio Cury, MD\textsuperscript{c,e}; Stuart Peacock, PhD\textsuperscript{f}; Elin Bonnevier, MSc\textsuperscript{g}; Ebba Palenius, MSc\textsuperscript{e}; and Alice Dragomir, MSc, PhD\textsuperscript{a,b}

Prostate cancer (PCa) is the second most frequently diagnosed male malignancy worldwide.\textsuperscript{1} In Canada, PCa accounted for an estimated 21% of all new cancer cases in 2016, with approximately 21,600 new cases diagnosed.\textsuperscript{2} In the United States, approximately 220,800 men are diagnosed with PCa and 27,000 men die of the disease per year.\textsuperscript{3}

Abstract

**Background:** Inaccurate risk classification and the burden of unnecessary biopsies are a challenge due to the limited ability of current risk assessment tools and modalities to diagnose prostate cancer (PCa) and distinguish indolent from aggressive disease. This systematic review assesses newly developed tests and interventions with high evidence of clinical utility that might be adopted in clinical practice during PCa management before initial and repeat biopsy, after positive biopsy, and after radical treatment. **Methods:** The Cochrane, Embase, MEDLINE, and Web of Science databases were searched for studies pertaining to the clinical utility of PCa diagnostic tests. Outcomes of interest were (1) a measure of the percentage of altered decision-making, (2) decrease in number of unnecessary biopsies, (3) decrease or increase in treatment intensity, and (4) risk reclassification after test results. **Results:** The search yielded 2,940 articles, of which 46 met the inclusion criteria. We found clinical utility evidence on the Prostate Health Index (PHI), 4Kscore test, MRI, OncotypeDX, Decipher test, Prolaris, ConfirmMDx, Progensa PCA3, NADIA ProsVue, and ProMark. No evidence was identified for Prostarix, ProstaVysion, Prostate Core Mitc Test, and Mi-Prostate Score. The interventions demonstrated their clinical utility in terms of change in treatment recommendations, decrease/increase in interventional treatment, decrease in biopsy, and risk reclassification. At diagnosis after a positive biopsy, ProMark, OncotyPE DX, Prolaris, and MRI guided the use of active surveillance. Use of NADIA ProsVue, Decipher, and Prolaris aided in the decision to add adjuvant therapy post-prostatectomy. PHI, 4Kscore, and MRI used prior initial and repeat biopsies, and ConfirmMDx and Progensa PCA3 used prior repeat biopsies to improve prediction of biopsy outcome, allowing a decrease in unnecessary biopsies. **Conclusions:** This systematic review suggests that implementation of these tests in clinical practice could effectuate personalized treatment of PCa. Further clinical and economic evaluation studies of long-term PCa outcomes are warranted to provide further guidance.

Prostate cancer (PCa) is the second most frequently diagnosed male malignancy worldwide.\textsuperscript{1} In Canada, PCa accounted for an estimated 21% of all new cancer cases in 2016, with approximately 21,600 new cases diagnosed.\textsuperscript{2} In the United States, approximately 220,800 men are diagnosed with PCa and 27,000 men die of the disease per year.\textsuperscript{3}

Early detection of PCa results in high cure rates, better outcomes, and lower costs.\textsuperscript{4,5} Currently, detection and clinical staging depend on serum prostate-specific antigen (PSA) levels, biopsy Gleason score, and T staging (through digital rectal examination [DRE] and imaging).\textsuperscript{6,7} However, these assessment tools do not accurately stratify patients with PCa, leading to overdiag-

\textsuperscript{a}Division of Urology, Department of Surgery, McGill University, \textsuperscript{a}Division of Urology, Department of Surgery, McGill University, \textsuperscript{b}Research Institute of the McGill University Health Centre, \textsuperscript{c}McGill University Health Centre, \textsuperscript{c}Division of Medical Oncology and \textsuperscript{d}Division of Radiation Oncology, Department of Oncology, McGill University, Montreal, Quebec, Canada; \textsuperscript{e}BC Cancer Agency, Vancouver, British Columbia, Canada; \textsuperscript{f}Lund University, Lund, Sweden.

Submitted March 1, 2018; accepted for publication June 27, 2018.

The authors have disclosed that they have no financial interests, arrangements, affiliations, or commercial interests with the manufacturers of any products discussed in this article or their competitors.

This study was funded by Prostate Cancer Canada – Discovery Grant 2015.


**Correspondence:** Alice Dragomir, MSc, PhD, Research Institute of the McGill University Health Centre, RI-MUHC/CORE (2B.45), 5252 Boulevard de Maisonneuve O, Montréal, Quebec H4A 3S5, Canada. Email: alice.dragomir@mcgill.ca

© JNCCN—Journal of the National Comprehensive Cancer Network | Volume 16 Number 11 | November 2018
nosis and overtreatment. Although PSA screening was associated with declining PCA-specific mortality (PCSM), it increased PCA incidence, leading to overtreatment of clinically insignificant tumors. In addition to the lack of specificity of PSA testing, biopsy undersampling raised additional concerns. Upgrading and downgrading Gleason score post-prostatectomy is a clear reflection of biopsy sampling error, which causes overtreatment of some cases and undertreatment of others.

Although some patients need immediate treatment, up to 60% of patients diagnosed with PCa according to current practice can be managed safely with active surveillance (AS). Thus, significant efforts have been made to find new tests and interventions that can differentiate between indolent and aggressive cancer, optimize the use of biopsy, and support the treatment decision.

Many new tests have demonstrated clinical utility and benefit in PCa management. These tests are applied in 4 main decision points: screening, after negative biopsy, after positive biopsy, or after radical treatment. To spare a patient with indolent PCa from an unnecessary biopsy after an elevated PSA test, interventions such as the Prostate Health Index (PHI; Beckman Coulter, Inc.), 4Kscore test (OPKO Health, Inc.), MRI, and Prostarix (Metabolon, Inc.) could be used for screening. Furthermore, once diagnosed with PCa, patients might benefit from a group of interventions developed to distinguish aggressive cancers (that require treatment) from non-aggressive cancers (that could be observed safely); these interventions include Prolaris (Myriad Genetic Laboratories, Inc.), Decipher test (GenomeDX Biosciences Inc.), OncotypeDX (Genomic Health, Inc.), ProstaVision (Bostwick Laboratories), MRI, Mi-Prostate Score (MiPS; MLabs), and ProMark (Metamark Genetics, Inc.). Other interventions could help overcome false-positive screening results and sampling errors after a negative biopsy to identify candidates for a repeat biopsy, and these include Progensa PCA3 (Gen-Probe Incorporated), 4Kscore, PHI, MRI, Prostate Core Mitomic Test (PCMT; MDNA Life Sciences Inc.), and ConfirmMDx (MDx Health). Finally, tests have been developed for after radical treatment to assess whether additional treatment is necessary depending on pathologic findings, including Decipher, NADiA ProsVue (IRIS International, Inc.), ProMark, and Prolaris.

Clinicians would undoubtedly like to consider some of these interventions, due to their potential role in improving risk stratification and outcomes prognostication. Unfortunately, most of these interventions are not used in clinical practice, mainly due to lack of evidence of their clinical benefit, clinical utility, and cost-effectiveness. Many analytical validity and clinical validity studies have been published; however, not many clinical utility studies can be found, which reflect the interventions’ usefulness in clinical practice.

The objective of this systematic review was to assess the clinical utility of newly marketed tests for use in PCa management before initial or repeat biopsy (after negative biopsy), after positive biopsy, and post-prostatectomy.

**Methods**

**Literature Search**

The Cochrane, Embase, MEDLINE, and Web of Science databases were systematically searched by an experienced librarian at McGill University. All search strategies were peer reviewed by a second experienced librarian at the same institution. The search strategy included vocabulary and text built around the research question according to the PICO (Patient, Intervention, Comparator, Outcome) framework. The search was conducted on November 22, 2016, and updated on February 24, 2017, to identify studies on the clinical utility of new PCa tests. A research protocol was established and followed for each step of the systematic review. The appropriate strategy was used to perform the study using selected MeSH terms and keywords. The MEDLINE search strategy (see supplemental eAppendix 1, available with this article at JNCCN.org) was adapted for the Cochrane, Embase, and Web of Science databases.

All published studies written in English or French were considered. Our search was not restricted by year of publication in order to include all articles about the issue of concern. Reference lists of the included articles were screened for additional eligible articles. Search terms included “prostate cancer,” “prostatic neoplasms,” “4Kscore,” “Progensa,” “Prostate Core Mitomic Test,” “ConfirmMDx,” “Decipher,” “NADiA ProsVue,” “Prostarix,” “Oncotype,” “ProMark,” “MRI,” “Mi-Prostate
Score,” “Prolaris,” “Prostate Health Index,” and “ProstaVysion,” as well as acronyms or other terms for these words. Duplicates were identified and excluded using EndNote’s (Clarivate Analytics) Author/Title-Year duplicate checker, followed by a manual verification. Truncation and wild cards were used to avoid missing any article that might include tests of interest. We included all possible study types that could include clinical utility evidence.

**Study Selection and Data Extraction**

Our systematic review included articles that have clinical utility evidence. Clinical utility studies assess the ability of the test to affect patient outcomes and treatment decisions. The best way to demonstrate the clinical utility of a test is by showing its ability to decrease PCSM or metastasis. Other important outcomes in contemporary PCa management are overtreatment and overdiagnosis, and showing how testing affects these would be essential. However, because PCa is a long-term disease, present studies might not be able to demonstrate these outcomes within their relatively short follow-up periods. Thus, clinical utility evidence concerning PCa will logically focus on short-term outcomes, such as change in treatment decision, patient stratification, or decrease in interventional treatment. These outcomes will clarify the ability of each test to alter treatment decisions at each disease phase.

After duplicate removal, 2 reviewers independently screened all titles and abstracts to exclude irrelevant studies as part of step 1. The screening procedure continued as part of step 2; however, full texts were assessed for relevancy using predetermined eligibility criteria (Table 1). Inclusion criteria for evidence of clinical utility evidence included tests that demonstrated a measure of the percent of altered clinical decision-making after addition of the tests (how many patients had a change in treatment), a quantification of the decrease/increase in interventional treatment after performing these tests, an evaluation of the number of unnecessary prostate biopsies and number of missed PCa diagnoses (before initial biopsy and after a negative biopsy), or a measure of patients’ reclassification into risk groups.

Risk group reclassification was an important outcome in our systematic review, because risk stratification is a crucial aspect in clinical decision and PCa management. It is important to note that many contemporary patient stratification systems exist, which help determine appropriate treatment. Although the systems are similar, some differences exist. Introduced in 1998, the D’Amico stratification system divides patients with PCa into 3 groups. The low-risk group includes patients with T1–T2a staging, PSA level ≤10 ng/mL, and Gleason score ≤6; patients with T2b staging, PSA level of 10 to 20 ng/mL, or Gleason score 7 are identified as the intermediate-risk group; and the high-risk group is defined as patients with stage ≥T2c disease, PSA level ≥20 ng/mL, or Gleason score 8 to 10. Organizations, such as NCCN, the National Institute for Health and Clinical Excellence, American Urological Association (AUA), and the European Association of Urology (EAU), have developed their own classification systems. The AUA and EAU agree with the D’Amico classification, whereas NCCN added a fourth group to its classification: the very-low-risk group (all of the following: T1c, PSA <10 ng/mL, Gleason score ≤6, positive biopsy cores <3, each core ≤50% cancer involvement, and PSA density <0.15 ng/mL/g). The NCCN classification is the most commonly used worldwide. The NCCN and D’Amico systems were the main ones used in the studies included in this review.

Long-term outcomes, such as effect of test use on morbidity and mortality, were included if available; any study that did not tackle one of these issues was excluded. The reviewers assessing the studies were initially blinded to each other’s results. At step 3, data were extracted from the eligible full-text articles using a prepared data extraction sheet (supplemental eAppendix 2). Each reviewer completed each step independently, followed by a discussion at the end of each step to solve any disagreement.

<table>
<thead>
<tr>
<th>Table 1. Inclusion and Exclusion Criteria for Study Eligibility</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Inclusion Criteria</strong></td>
</tr>
<tr>
<td>Any article related to prostate cancer treatment, screening, or diagnosis</td>
</tr>
<tr>
<td>Any article related to test of interest</td>
</tr>
<tr>
<td>Any article with clinical utility evidence</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>
Quality Assessment
Eligible studies were assessed for quality using a modified version of the scale developed by Rector et al. (supplemental eAppendix 3). The modified checklist is composed of 17 questions evaluating study design, methodology, intervention, bias risk, and outcomes. An ordinal scale was used to give values for each question: 0 if “not clear,” or “not a relevant item;” 1 for “good quality;” and 2 for “excellent quality.” Scoring was performed by 2 reviewers independently, and a discussion occurred at the end. Based on overall score, the studies were categorized as excellent quality if scoring was >75%; good quality if scoring was between 50% and 75%; and poor quality if scoring was <50%.

Types of Interventions
The interventions were divided into 4 groups (supplemental eTable 1). Group 1, developed for screening, was used to reduce the number of unnecessary (negative) biopsies in patients with elevated PSA levels, and include MRI, PHI, 4Kscore, and Prostarix. These interventions are usually used in men suspected of having PCa in whom screening is indicated (eg, men aged >50 or those aged >45 years with PCa family history; African Americans; men without previous PCa treatment or biopsy with a PSA level of 2–10 ng/mL; or suspicious DRE results). Group 2 interventions, such as OncotypeDX, Decipher, Prolaris, ProstaVysion, ProMark, and MiPS, are used in patients with previous positive biopsy results (to distinguish aggressive cancers that require treatment) from nonaggressive cancers, that do not. The third group includes ConfirmMDx, Progensa PCA3, 4Kscore, PHI, MRI, and PCMT, which are used in patients with negative biopsy results to identify those for whom a repeat biopsy is needed. Usually Group 3 tests are used in patients suspected of having PCa in whom screening is indicated (eg, men aged >50 or those aged >45 years with PCa family history; African Americans; men without previous PCa treatment or biopsy with a PSA level of 2–10 ng/mL; or suspicious DRE results). Finally, Group 4 tests, such as Decipher, ProMark, Prolaris, and NADiA ProsVue, are used after radical treatment to assess whether additional treatment is necessary depending on pathologic findings, and are used in patients at risk for recurrence or PCSM (eg, men with adverse postsurgery pathology, pathologic T3 disease, increasing PSA levels, or positive surgical margins) in whom treatment amelioration/addition is suggested.

Grouping of the interventions was based on published literature and material from manufacturers’ websites. Figure 1 correlates each intervention with the different PCa stages throughout screening, diagnosis, and treatment.

Data Synthesis
Figure 2 illustrates the different stages of the systematic review and outlines the number of screened articles, abstracts, full texts, and exclusions.

Outcomes varied depending on the intervention and when it was used (eg, screening, diagnosis, treatment). The change in decision to perform a biopsy was assessed as an important outcome for tests categorized as Group 1. Similarly, this was studied for Group 3 to determine the change in decision regarding who to rebiopsy. Additionally, the percentages of treatment change, decrease/increase in intervention, and patient reclassification were assessed to quantify the importance of the tests in Groups 2 and 4.

Results
After duplicate removal, a total of 2,940 citations were identified and screened for relevance, 170 were selected for full-text assessment (Figure 2), yielding 41 articles; after review of the reference lists, 5 additional articles were identified, resulting in a total of 46 articles considered for this systematic review.

Study Characteristics
Included articles were either clinical utility articles or articles that included some clinical utility evidence. We further categorized the articles retrieved based on whether the intervention was used in screening, after negative biopsy, after positive biopsy, or after radical treatment, which is clarified in supplemental eTable 1. The included articles were published between 2001 and 2016. Sample sizes ranged between 11 and 2,914 patients. There were 5 articles on MRI screening, 7 on PHI, 10 on 4Kscore, 2 on OncotypeDX, 7 on Decipher, 6 on Prolaris, 1 on ConfirmMDx, 6 on Progensa PCA3, 1 on NADiA ProsVue, and 1 on ProMark. No clinical utility evidence was found for Prostarix, ProstaVysion, and MiPS (Figure 3).
Outcomes
Clinical utility outcomes studied were different based on which intervention was used and how it affected those outcomes. A total of 24 articles studied the reduction in unnecessary biopsy, considering either initial or repeat biopsy [18,19,22,32,34,53,54,58–65,67–69,72–75,80,82]; 1 discussed avoiding overtreatment [17]; 1 evaluated likelihood of risk reclassification [35]; 19 considered change in treatment recommendations, and 10 of these also considered risk restratification [22,25,36,41,47,49–52,55–57,66,70,71,76–78,81]; and 1 considered reduction of under- and overstaging [79]. The findings reported in this section are presented in Table 2 and supplemental eTable 2.

4Kscore PCa Test: 4Kscore is a blood test that measures a panel of kallikrein markers: total PSA, free PSA, intact PSA, and human kallikrein 2. Many published studies have demonstrated 4Kscore’s ability to detect clinically insignificant cancer and predict metastatic disease compared with PSA alone. [83] Our literature search identified 10 publications reporting clinical utility evidence on the 4Kscore; 9 publications [17,18,34,58–64] demonstrated its ability to reduce unnecessary biopsies by predicting biopsy histopathology and occurrence of metastatic and aggressive disease. The first publication on 4Kscore [63] consisted of 740 unscreened men who underwent biopsy for an elevated PSA level, and showed that application of this test led to a 60% reduction in unnecessary biopsies at a threshold of >20%. Similar results were seen in subsequent studies: reductions in biopsies were seen in 64% [34], 49% [58], 82% [60], 25% [61], 51% [61], and 41% [64] of patients. In addition, Braun et al [59] reported a 25% reduction, but at a threshold of ≥8%. Hence, reduction of unnecessary biopsies ranged between 25% and 82%.

Another study [17] which included 392 patients who underwent radical prostatectomy (RP), showed that use of 4Kscore led to a 14% reduction in unnecessary surgeries, thus avoiding overtreatment. 4Kscore was supported by a high number of excellent and good quality articles, with 4 studies of excellent quality [60,62–64] showing a reduction in unnecessary biopsies between 41% and 57%.
Prostate Health Index: Seven studies have investigated the utility of the PHI (supplemental eTable 2). Lazzeri et al\textsuperscript{19} prospectively evaluated 646 patients who underwent initial biopsy and showed that the PHI cutoff of 27.6 would have avoided 15.5% biopsies and missed 9.8%. The same group\textsuperscript{55} also reported a 16.5% reduction in biopsies when using a cutoff of 25.5, while missing 8.5%. Furthermore, Filella et al\textsuperscript{65} and Ng et al\textsuperscript{69} reported a 19% (cutoff 31.94) and 45.2% (cutoff 27.6) reduction in biopsies, respectively, while missing 9.8% of cases. Two articles of excellent quality\textsuperscript{19,69} and 5 of good quality\textsuperscript{55-58} assessed the clinical utility of PHI in decreasing biopsy (Table 2); however, results were stated at different cutoffs.

MRI: Studies on MRI showed high specificity and sensitivity in predicting postoperative pathology\textsuperscript{34} in addition to high detection rates\textsuperscript{85,86}. Five articles on MRI use in screening were included in this review (supplemental eTable 2)\textsuperscript{22,79,82}; 3 of these studied its effect on number of biopsies performed\textsuperscript{22,80,82}, reporting a decrease in biopsies ranging between 51% and 70%. Confirming previous publications, MRI demonstrated a role in reclassification and staging\textsuperscript{51,87} allowing physicians to directly monitor PCa and identify high-grade disease that requires treatment.\textsuperscript{88} The published literature supports the importance of using MRI as a screening method, with 3 articles of excellent quality\textsuperscript{79-81} and 1 of good quality\textsuperscript{22} (Table 2).

**Progensa PCA3 Assay:** Six identified studies investigated clinical utility evidence on Progensa, a test that calculates the ratio between PCa antigen gene (PCA3) mRNA and PSA mRNA found in urine samples post-DRE \textsuperscript{53,54,72-75} All of these publications correlated PCA3 score to reduced repeat biopsy. Malavaud et al\textsuperscript{53} estimated a 37% reduction in repeat biopsy if the PCA3 test was used. Similarly, 49.51% and 63% reductions were reported at a cutoff of 25 by Gittelman et al\textsuperscript{54} and Tombal et al\textsuperscript{75}, respectively. Crawford et al\textsuperscript{53} confirmed previous published results by reporting a reduction of 77.1% at a cutoff of 35; however, the PCA3 test missed 21.6% of PCa diagnoses. Similarly, de la Taille et al\textsuperscript{72} and Haese et al\textsuperscript{71} reported approximately 60% and 40% reductions at a cutoff of 35 and 20, while missing between 9% and 21% of diagnoses, respectively. Similar to PHI, studies on PCA3 used a variety of cutoffs. Two articles were of excellent quality\textsuperscript{54,74} and 4 were of good quality (Table 2).\textsuperscript{53,72,73,85}

**ConfirmMDx:** We found one clinical utility study\textsuperscript{32} on ConfirmMDx, a biopsy-based test that measures methylation levels of 3 genes.\textsuperscript{89} Wojno et al\textsuperscript{2} reported a reduced rate of repeated biopsies in patients at risk for malignancy and with a previous negative biopsy. Only 6 of the 138 men (4.3%) with a ConfirmMDx negative result performed a repeat biopsy. A 10-fold decrease in repeat biopsies was observed. Only one published good-quality article assessed the usefulness of ConfirmMDx.

**Prolaris:** Prolaris is a genomic test that measures cell cycle progression signature consisting of 46 genes to predict disease mortality and progression. This tissue-based test could be used after a positive biopsy and post-RP\textsuperscript{90}; 5 articles showed evidence of its clinical utility after a positive biopsy\textsuperscript{97-10,70} and 1 showed evidence post-RP\textsuperscript{25} (Figure 3 and supplemental eTable 2).

Two observational prospective studies\textsuperscript{47,49} were conducted to evaluate the change in treatment recommendations pre- and post-Prolaris. Crawford et al\textsuperscript{47} showed that Prolaris altered 64.9% of the treatment recommendations, 37.2% had a reduction of interventional treatment, and 23.4% had an increase.
Notably, RP and radiation decreased by 49.5% and 29.6%, respectively. However, Shore et al\textsuperscript{49} reported a 47.8% change in treatment recommendations, noting that 72.1% of the change was a decrease in interventional treatment and 26.9% was an increase.

Another study by Shore et al\textsuperscript{48} evaluated “possible” change of treatment by sending physicians biopsy findings with and without the Prolaris test results for patients who were already treated, and found that test results led to definite or possible change in 32% of patients.

The remaining 3 articles assessed reclassification by Prolaris\textsuperscript{42,51,76}—2 were after a positive biopsy\textsuperscript{50,70} and 1 was after RP\textsuperscript{25}—although only 2 of these studies scored as good quality.\textsuperscript{25,70} Cuzick et al\textsuperscript{70} reported reclassification in 14% of patients with a low Cancer of the Prostate Risk Assessment (CAPRA) score and 44% of those with an intermediate CAPRA score to higher- and lower-risk groups, respectively. Cooperberg et al\textsuperscript{25} reported reclassification of 56% of patients with low-risk CAPRA scores based on Prolaris test results. Overall, 5 good-quality publications highlight the usefulness of Prolaris in clinical practice for reclassifying patients and changing treatment decisions (Table 2).

**ProMark:** ProMark is a biopsy-based prognostic test that detects 8 protein biopsy markers to predict disease aggressiveness and outcome in patients with PCa.\textsuperscript{91} The test was developed in a study of 381 patient biopsies matched with prostatectomy tissues, then validated in another part of the same study consisting of 256 men to distinguish between favorable and nonfavorable pathology at RP.\textsuperscript{71} The primary goal of this study was to demonstrate a model for distinguishing candidates for AS from those requiring prostatectomy, in addition to identifying favorable versus nonfavorable pathology. Results showed that frequency of favorable pathology decreases with increasing ProMark scores. This study also had clinical utility evidence on ProMark, showing that the net reclassification improvement was 0.34 (P<.00001; 95% CI, 0.20–0.48) and 0.24 (P<.0001; 95% CI, 0.12–0.35) for the NCCN and D’Amico risk categories, respectively. Only 1 study of excellent quality was found to support the ProMark test (Table 2).

**OncotypeDX:** OncotypeDX is a genomic test that can be used after a positive biopsy on prostate tumor samples as small as 1 mm.\textsuperscript{23} Two articles studying the effect of OncotypeDX on treatment patterns provided evidence of its clinical utility. Albala et al\textsuperscript{51} reported a 21% reduction in interventional treatment, mainly a 13% decrease in radiation and 10% in RP.

---

**Table 2. Quality of Publications**

<table>
<thead>
<tr>
<th>Intervention</th>
<th>Excellent Quality (&gt;75%)</th>
<th>Good Quality (50%-75%)</th>
<th>Poor Quality (&lt;50%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>4Kscore</td>
<td>4(92-95)</td>
<td>47(19,43,33-121)</td>
<td></td>
</tr>
<tr>
<td>PHI</td>
<td>27(69)</td>
<td>5(25-46)</td>
<td></td>
</tr>
<tr>
<td>MRI</td>
<td>3(70-81)</td>
<td>1(0)</td>
<td></td>
</tr>
<tr>
<td>Progensa PCA3</td>
<td>2(94-74)</td>
<td>4(25,23,71,35)</td>
<td></td>
</tr>
<tr>
<td>ConfirmMDx</td>
<td>1(0)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prolaris</td>
<td>5(25-47,49-79)</td>
<td>1(0)</td>
<td></td>
</tr>
<tr>
<td>OncotypeDX</td>
<td>1(0)</td>
<td>1(0)</td>
<td></td>
</tr>
<tr>
<td>Decipher</td>
<td>4(55,96-78)</td>
<td>2(23,56)</td>
<td></td>
</tr>
<tr>
<td>NADiA ProsVue</td>
<td>1(0)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ProMark</td>
<td>1(0)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Abbreviation: PHI, Prostate Health Index.
In addition, they reclassified 4.3% of very low-risk patients and 35.7% of low-risk patients as intermediate-risk based on OncotypeDX findings. Similarly, a 24% reduction in interventional treatment was observed in Dall’Era et al.52 Only one publication of good quality was found assessing the clinical utility of OncotypeDX based on its effect in decreasing interventional treatment.

**Decipher:** Decipher, or genomic classifier (GC), is a genom test that uses the expression of 22 RNA markers (coding and noncoding) to predict metastasis and PCSM. Moreover, it allows risk stratification of patients post-RP and guides the treatment decision regarding adjuvant therapy.77,92 Seven studies were found to have clinical utility evidence (Figure 3, supplemental eTable 2) on Decipher after prostatectomy,56,55-57,76-78 and none were found on its use after a positive biopsy.

Most of these articles studied the effect of Decipher on decision-making. The first article, which consisted of 24 pathologically high-risk patients,46 studied the effect on salvage and adjuvant treatment recommendations. The urologic oncologists provided their treatment recommendations for each patient pre-and post-Decipher testing results. Treatment recommendations changed in 43% of the cases in the adjuvant group, of which 27% were a reduction in interventional treatment and 37% were an increase. In the salvage group, there was a 53% change in treatment recommendations, with a 16% reduction in interventional treatment and a 61% increase. In a similar context and with a larger number of urologists, another study reported that treatment was deintensified to observation for 40% of patients who were recommended for adjuvant radiation therapy and was intensified for 13% of those recommended for observation.55 In addition, Decipher reclassified 51% of the patients as low risk. Similarly, Michalopoulos et al56 reported that Decipher caused a change in treatment recommendations in 30.8% of patients, of whom 42.5% had a reduction in treatment intensity and 17.6% had an increase. These findings agree with those of Nguyen et al,57 which showed that Decipher results modified 35% and 45% of the treatment recommendations by oncologists and urologists, respectively.

Risk reclassification was also an important outcome in the remaining 3 articles. In the study by Cooperberg et al,76 Decipher reclassified 49 of 185 men as low to intermediate risk who were high risk according to CAPRA Postsurgical (CAPRA-S) score ≥6. Among those men, 3 PCSM events were observed, whereas 17 PCSM events were observed in those who were classified as high-risk by GC. In a study of 2,342 patients, Den et al77 showed that GC reclassified 52%, 76%, and 40% of patients into CAPRA-S low-, intermediate-, and high-risk groups, respectively. Likewise, Ross et al78 showed that Decipher reclassified 71%, 52%, and 19% of patients into CAPRA-S low-, intermediate-, and high-risk groups, respectively. In addition, GC correlated with increased cumulative incidence of biochemical recurrence, metastasis, and PCSM after RP (P<.01). Metastases were detected in 47% of those classified high risk by GC versus 12% in those with a low GC score. The strength of clinical utility evidence was high for Decipher: 4 studies were excellent55,76-78 and 2 were of good quality56 (Table 2); they demonstrated its ability in reclassifying patients and causing a change in treatment.

**NADiA ProsVue:** NADiA ProsVue is a blood-based test that determines the rate of total PSA change over time by measuring extremely low concentrations of PSA from 3 blood samples taken after RP.93 This test helps identify patients with low risk of recurrence post-RP.90,91 Only one study provided some clinical utility evidence of this test. In a prospective, multicenter clinical trial reporting on 225 men treated with RP, Moul et al90 showed that a score of ≤2 pg/mL reduced the secondary treatment recommendation in 63.4% of patients who were initially referred for secondary treatment. After NADiA ProsVue results, only 11.7% of the men were referred for secondary treatment. Only one article of good quality assessed the clinical utility of NADiA ProsVue.41

**ProstaVysion, MiPS, Prostarix, and PCMT:** No clinical utility evidence was found on these tests.

**Quality Assessment and Risk of Bias**

Supplemental eTable 2 presents the quality assessment scores; 16 articles were of excellent quality, scoring >754,19,54,55,60,62-64,69,71,74,76-81 whereas most of the others were of good quality (n=26), scoring between 50 and 75.17,18,22,25,32,34-36,41,47-49,51,53,56,58,59,61,65-68,70,72,73,75 Additionally, Table 2 presents the quality assessment score in terms of number of studies in each category (high-, good-, and low-quality). As part of the sys-
tematic review, potential biases were assessed for each of the publications, and were incorporated in the calculated quality scores (Table 3).

**Discussion**

Although tools used in current practice lack the precision to guide treatment decisions, Gleason score, T staging, PSA level, and DRE results continue to be important in the risk stratification, diagnosis, and management of patients with PCa. Finding and developing new prognostic tests and interventions will not be sufficient to directly improve treatment decisions. After these interventions are validated, they should be integrated into clinical practice to provide insight into their benefits and applicability; this depends on the ability to access the interventions, which mainly depends on the ability to understand an intervention’s results and scores, and to relate all this to patient outcomes and costs. After effective access to an intervention and adoption in real life, clinical utility can be evaluated, which demonstrates the usefulness of the test and the value that the intervention adds to clinical management.94

We performed this systematic review to assess clinical utility evidence regarding available interventions, with the hope of promoting their use in achieving better personalized treatment in PCa. Choosing the appropriate intervention throughout the disease course, from screening to treatment, is important for reducing the uncertainty related to diagnosis and treatment. Hence, we divided the interventions into groups based on the stage of treatment in which each could be used, and identified the tests with the highest evidence of clinical utility.

Our results showed that some interventions may have as many as 10 publications reporting clinical utility evidence, such as 4Kscore, whereas others had none. The quality of the articles also differed between interventions, with 4Kscore, PHI, MRI, Progensa PCA3, Prolaris, and Decipher having the highest levels of clinical utility evidence.

PHI and 4Kscore could be used for screening and after a negative biopsy. These tests demonstrated the ability to prevent between 15% and 64% of unnecessary biopsies, respectively, at varying thresholds and cutoffs, although missing some cancers.17,19,34,69 These findings agree with those of many publications that correlate PHI to Gleason score95,96 and the ability to avoid unnecessary biopsies. MRI is an intervention that can accurately identify significant cancer, even tumors missed in the anterior region.84 Retrieved articles showed clinical utility evidence on MRI at different disease stages.79,90 Progensa PCA3 is able to identify men with a higher risk of cancer. This test demonstrated a reduction in repeat biopsy up to 77% at a cutoff of 35.53 Despite these findings, we had difficulty choosing the most appropriate cutoff for predicting PCa aggressiveness; our literature search yielded articles reporting on the ability of PCA3 to reduce biopsies, but at different cutoffs. This agrees with the findings of Roobol et al,97 which concluded that PCA3 cannot replace PSA, but emphasizing that it could be used in conjunction with other assessment tools.

Decipher, for instance, has the potential ability to identify patients with a higher risk of metastasis and death post-RP, thus resolving uncertainties regarding who will benefit from adjuvant therapy. Clinical utility evidence showed that 31% to 53% of post-RP treatment recommendations were changed.

<table>
<thead>
<tr>
<th>Table 3. Risk of Bias for Included Publications</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intervention</td>
</tr>
<tr>
<td>4Kscore</td>
</tr>
<tr>
<td>PHI</td>
</tr>
<tr>
<td>MRI</td>
</tr>
<tr>
<td>Progensa PCA3</td>
</tr>
<tr>
<td>ConfirmMDx</td>
</tr>
<tr>
<td>Prolaris</td>
</tr>
<tr>
<td>OncotypeDX</td>
</tr>
<tr>
<td>Decipher</td>
</tr>
<tr>
<td>NADiA ProsVue</td>
</tr>
</tbody>
</table>

Abbreviation: PHI, Prostate Health Index.
New Tests for Prostate Cancer Management

with 16% to 43% changing from any to no treatment. In addition, it reclassified up to 60% of high-risk patients to low-risk. Similarly, Prostates was associated with postoperative adverse outcome prediction and with a change in treatment rate between 48% and 65%, while reclassifying up to 56% of patients with low-risk CAPRA scores.

Our study had some limitations. First, some studies selected in the final step of the systematic review were not blinded, and others had potential biases. However, all of these issues were taken into consideration in the quality appraisal score, and this was reflected in the final score. Second, the studies did not demonstrate the tests’ effect on PCa long-term outcomes (eg, recurrence-free survival, morbidity and mortality after PCa treatment based on these interventions, cancer survival, quality of life). Although our study highlights the tests with excellent and good evidence of clinical utility, further investigation is needed to determine the effect of these tests on long-term clinical outcomes. However, such studies will take years to have sufficient power to demonstrate differences, because localized PCas is typically a disease characterized by slow progression. Finally, we did not assess grey literature due to the challenges with these types of sources. However, we do not believe this had a significant effect on our study, because we examined interventions that were recently developed. Therefore, if we had included grey literature, it is likely that mostly only conference abstracts would have been identified, and conference abstracts generally lack important study information necessary for a rigorous appraisal of its methodological quality.

Our study also has important strengths. Many articles in the systematic review were clinical utility studies designed primarily to evaluate the new interventions. After organizing a protocol, 2 interpreters were able to assess the articles at each step, thus assuring our results. Most articles included in our review were of good and excellent quality, thus yielding important evidence on the new interventions that could be used in PCa management.

Conclusions

This study provides an overview of the clinical utility evidence for several interventions that could significantly impact personalized treatment decisions and improve clinical outcomes and quality of life for men with PCas if adopted in clinical practice. However, their cost-effectiveness should be demonstrated before public access is enabled. This review suggests that use of these tests in clinical practice could help achieve personalized treatment of PCAs by adding new meaningful information for better risk assessment and disease prognostication. Additional clinical and economic evaluation studies of long-term PCa outcomes are warranted to provide further guidance.

References


102. New Urine Test for Prostate Cancer Available; Unlike PSA Test, is Ultra-Specific for Prostate Cancer [press release]. Santa Monica, CA; Prostate Cancer Foundation; September 25, 2013.


See JNCCN.org for supplemental online content.