IMPORTANT REMINDER

Medical Policies are developed to provide guidance for members and providers regarding coverage in accordance with contract terms. Benefit determinations are based in all cases on the applicable contract language. To the extent there may be any conflict between the Medical Policy and contract language, the contract language takes precedence.

PLEASE NOTE: Contracts exclude from coverage, among other things, services or procedures that are considered investigational or cosmetic. Providers may bill members for services or procedures that are considered investigational or cosmetic. Providers are encouraged to inform members before rendering such services that the members are likely to be financially responsible for the cost of these services.

DESCRIPTION

Gene expression profile analysis has been proposed as a means to risk-stratify patients with low-risk prostate cancer, diagnosed by needle biopsy, to guide treatment decisions.

MEDICAL POLICY CRITERIA

Gene expression analysis to guide management of prostate cancer is considered **investigational** in all situations, including but not limited to the following tests:

1. Oncotype DX Prostate Cancer Assay
2. Prolaris
3. Decipher

NOTE: A summary of the supporting rationale for the policy criteria is at the end of the policy.

CROSS REFERENCES
BACKGROUND

Prostate cancer is the second most common noncutaneous cancer diagnosed among men in the United States. According to the National Cancer Institute, nearly 180,000 new cases are expected to be diagnosed in the United States in 2016 and are associated with approximately 26,000 deaths.\[^{1}\]\[^{1}\] Autopsy studies in the era prior to the availability of prostate-specific antigen screening have identified incidental cancerous foci in 30% of men 50 years of age, with incidence reaching 75% at age 80 years.\[^{2}\]\[^{2}\] However, between 1975 and 1991 prostate cancer mortality rose and subsequently dropped 39% by 2007. The rise in mortality is unexplained, though it has been suggested as due to how cause of death was assigned.\[^{3}\]\[^{3}\] Regarding the subsequent decline, a number of potential explanations have been suggested as underlying reasons: improvements in treatment and screening, changes in assigning causes of death, and risk of cardiovascular death among men with prostate cancer treated with hormonal therapy.\[^{3}\]\[^{3}\]

Regulatory Status

Clinical laboratories may develop and validate tests in-house and market them as a laboratory service; laboratory-developed tests (LDTs) must meet the general regulatory standards of the Clinical Laboratory Improvement Amendments (CLIA). Prolaris®, Oncotype DX® Prostate, and Decipher® gene expression profiling test are available under the auspices of CLIA. Laboratories that offer LDTs must be licensed by CLIA for high-complexity testing. To date, the U.S. Food and Drug Administration (FDA) has chosen not to require any regulatory review of this test.

In November 2015, the FDA’s Office of Public Health Strategy and Analysis published a document on public health evidence for FDA oversight of LDTs.\[^{4}\]\[^{4}\] FDA argued that many tests need more FDA oversight than the regulatory requirements of CLIA. CLIA standards relate to laboratory operations, but do not address inaccuracies or unreliability of specific tests. Prolaris is among the 20 case studies in the document cited as needing FDA oversight. The document asserted that patients are potentially receiving inappropriate prostate cancer care because there is no evidence that results from the test meaningfully improve clinical outcomes.

EVIDENCE SUMMARY

Validation of the clinical use of any genetic test focuses on 3 main principles:

1. The analytic validity of the test, which refers to the technical accuracy of the test in detecting a mutation that is present or in excluding a mutation that is absent;
2. The clinical validity of the test, which refers to the diagnostic performance of the test (sensitivity, specificity, positive and negative predictive values) in detecting clinical disease; and
3. The clinical utility of the test, i.e., how the results of the diagnostic test will be used to change management of the patient and whether these changes in management lead to clinically important improvements in health outcomes.

The focus of this review is on evidence related to the ability of test results to:

- Guide decisions in the clinical setting related to either treatment, management, or prevention, and
- Improve health outcomes as a result of those decisions.

This policy was initially based on a 2013 BlueCross BlueShield Association Technology Evaluation Center (TEC) Assessment which was updated in January 2015 with a literature review through September 30, 2014.[5,6] Full-length publications were sought that described the analytic validity (technical performance), clinical validity (prognostic accuracy), and clinical utility (accurately identifying men experiencing improved health outcomes by avoiding treatment or undergoing more appropriate therapies) of Prolaris, Oncotype DX Prostate, and Decipher gene expression profiling.

**ONCOTYPE DX® PROSTATE**

**Analytic Validity**

In the only study of validity for the Oncotype DX® Prostate test, Knezevic et al., authors from Genomic Health, Inc., the developer of the test, reported analytical accuracy and reproducibility of Oncotype Dx® Prostate.[7] Estimates of analytic precision and reproducibility were derived from analysis of RNA prepared from 10 microdissected prostate tumor samples obtained by needle biopsy. Individual Gleason scores were assigned using the 2005 International Society of Urological Pathology Consensus guidelines.[8]

The results showed that the assay could accurately measure expression of the 12 cancer-related and 5 reference genes over a range of absolute RNA inputs (0.005-320 ng); the limit of detection in a sample was 0.5 ng/uL. The analytic accuracy showed average variation of less than 9.7% across all samples at RNA inputs typical of needle biopsy specimens. The amplification efficiency for the 17 genes in the test ranged from 88% to 100%, with a median of 93% (SD=6%) for all 17 genes in the assay. Analytic precision was assessed by examining variability between replicate results obtained using the same mRNA input. Reproducibility was measured by calculating both within and between mRNA input variation. A low input level of 5 ng mRNA was used to reflect the lowest 2.5 percentile of a tumor sample of 0.023 cm3. When converted to GPS units (unit measure for reporting test results), the standard deviation for analytic precision was 1.86 GPS units (95% confidence interval [CI], 1.60 to 2.20) on the 100-unit scale. The standard deviation for reproducibility was 2.11 GPS units (95% CI, 1.83 to 2.50) on the 100-point scale. This study provides sufficient evidence to establish the analytic validity of Oncotype Dx® Prostate.

**Clinical Validity**
In 2016, Brand et al combined the Klein et al (2014) and Cullen et al (2015) studies using a patient-specific meta-analysis.[9] The GPS was compared to the CAPRA score, NCCN risk group, and AUA/EAU risk group. The authors tested whether the GPS added predictive value for the likelihood of favorable pathology above the clinical risk assessment tools. The model including the GPS and CAPRA score provided the best risk discrimination; the AUC improved from 0.68 to 0.73 by adding the GPS to CAPRA score. The AUC improved from 0.64 to 0.70 by adding the GPS to the NCCN risk group. The improvements were reported to be significant but the confidence intervals for AUC were not provided.

Whalen et al (2016) prospectively evaluated the correlation of GPS with final pathology at RP in a clinical practice setting. Eligible men were 50 years of age and older with more than 10 years of life expectancy, PSA levels of 20 ng/mL or less, stage cT1c-cT2c newly diagnosed, untreated prostate cancer, and who met NCCN classifications as very low risk, low risk, or low-intermediate risk.[10] Men were enrolled from May 2013 to August 2014 at an academic medical center. After initial review at the institution, Genomic Health further reviewed biopsy samples to assign Gleason score and tumor length. Samples with Gleason grade discrepancy between initial and central review were excluded from analyses. Clinicians were blinded to GPS when counseling patients on management with active surveillance versus definitive treatment. Genomic Health reclassified patients' cancers as “less favorable,” “consistent with,” or “more favorable” than what would have been predicted by their NCCN risk group. The primary outcome was adverse pathology at RP defined as any pT3 stage and primary Gleason grade of 4 or any pattern 5. Fifty patients had RP pathology and the reclassification results for these participants are discussed here; 21 (42%) met the definition of adverse pathology. The NCCN risk classification categorized 2 (4%) patients as very low risk, 34 (68%) as low risk, and 14 (28%) as low-intermediate risk. Twenty-three (46%) of patients were reclassified using GPS and the percentage with adverse pathology for the reclassification is shown in Table 9 as derived from data provided in the text. Confidence intervals were not provided.

One publication compiled results of three cohorts, two of which were used for test development, of contemporary (1997-2011) patients in a prostatectomy study (N=441; Cleveland Clinic database, 1987-2004), a biopsy study (N=167; Cleveland Clinic database, 1998-2007), and an independent clinical validation study cohort (N=395; mean age, 58 years; University of California, San Francisco Urologic Oncology Data Base, 1998-2011).[11]

Results from the clinical validation study and prostatectomy study provide information on the potential clinical validity of this test. The cohorts had a mix of low to low-intermediate clinical risk characteristics using National Comprehensive Cancer Network (NCCN) or American Urological Association (AUA) criteria. Patients included in the validation and prostatectomy studies would be considered (a) eligible for active surveillance based on clinical and pathologic findings and (b) representative of patients in contemporary clinical practice. However, all patients elected radical prostatectomy within 6 months of their initial diagnostic biopsies.

The clinical validation study was designed to evaluate the ability of Oncotype Dx® Prostate to predict tumor pathology in needle biopsy specimens. It was prospectively designed, used masked review of prostatectomy pathology results, and as such met the REMARK (Reporting Recommendations for Tumor Marker Prognostic Studies) guidelines for biomarker validation.[12] In the prostatectomy study, all patients with clinical recurrence (local recurrence
or distant metastasis) were selected, together with a random sample of those who did not recur, using a stratified cohort sampling method to construct a 1:3 ratio of recurrent to nonrecurrent patients. The prespecified primary end point of the validation study was the ability of the Genomic Prostate Score (GPS) to predict the likelihood of favorable pathology in the needle biopsy specimen. Favorable pathology was defined as freedom from high-grade or non-organ-confined disease. In the prostatectomy study, the ability of the GPS to further stratify patients within AUA groupings was related to clinical recurrence-free interval in regression-to-the-mean estimated survival curves.

The validation study results showed that the GPS could refine stratification of patients within specific NCCN criteria groupings, as summarized in Table 3. The proportions in Table 1 were estimated from a plot of GPS versus the percent likelihood of favorable pathology. These findings suggest that a lower GPS would reclassify the likelihood of favorable pathology (i.e., less biologically aggressive disease) upward (i.e., a potentially lower risk of progression), and vice versa within each clinical stratum. For example, among patients in the cohort classified by NCCN criteria as low risk, the mean likelihood of favorable pathology in a tumor biopsy was about 76%, with 24% then having unfavorable pathology. With the GPS, the estimated likelihood of favorable tumor pathology was broadened, ranging from 55% to 86%, conversely reflecting a 45% to 14% likelihood of adverse pathology, respectively.

Table 1. Reclassification of Prostate Cancer Risk Categories With Oncotype Dx® Prostate

<table>
<thead>
<tr>
<th>NCCN Risk Level</th>
<th>Estimated Mean Likelihood of Favorable Tumor Pathology</th>
<th>Estimated Corresponding GPS, Range</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>NCCN Criteria, %</td>
<td>GPS + NCCN, % Range</td>
</tr>
<tr>
<td>Very low</td>
<td>≈84</td>
<td>63-91</td>
</tr>
<tr>
<td>Low</td>
<td>≈76</td>
<td>55-86</td>
</tr>
<tr>
<td>Intermediate</td>
<td>≈56</td>
<td>29-75</td>
</tr>
</tbody>
</table>

GPS: Genomic Prostate Score; NCCN: National Comprehensive Cancer Network.

In effect, the risk of adverse tumor pathology indicated by the GPS could be nearly halved (24%-14%) at 1 extreme, or nearly doubled (24%-45%) at the other, but the actual number of patients correctly or incorrectly reclassified between all 3 categories cannot be ascertained from the data provided. The results suggested that the combination of GPS plus clinical criteria could reclassify patients on an individual basis within established clinical risk categories. However, whether these findings support a conclusion that the GPS could predict the biological aggressiveness of a tumor—hence its propensity to progress—based solely on the level of pathology in a biopsy specimen is unclear. Moreover, extrapolation of this evidence to a true active surveillance population, for which the majority in the study would be otherwise eligible, is difficult because all patients had elective radical prostatectomy within 6 months of diagnostic biopsy.

The prostatectomy study, although used to identify genes to include in the GPS, provided estimates of clinical recurrence rates stratified by AUA criteria (Epstein et al.), compared with rates after further stratification according to the GPS from the validation study. The survival curves for clinical recurrence reached a duration of nearly 18 years based on the dates individuals in the cohort were entered into the database (1987-2004). The restratifications are summarized in Table 2. The GPS groups are defined by tertiles defined in the overall study.
Table 2. Reclassification of Prostate Cancer 10-Year Clinical Recurrence Risk With Oncotype Dx® Prostate

<table>
<thead>
<tr>
<th>Overall 10-Year Risk, %</th>
<th>10-Year Risk, % (GPS Low Group)</th>
<th>10-Year Risk, % (GPS Intermediate Group)</th>
<th>10-Year Risk, % (GPS High Group)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(AUA Risk Level)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.4 Low</td>
<td>2.0</td>
<td>3.4</td>
<td>7.0</td>
</tr>
<tr>
<td>9.6 (Intermediate)</td>
<td>2.8</td>
<td>5.1</td>
<td>14.3</td>
</tr>
<tr>
<td>18.2 (high)</td>
<td>6.2</td>
<td>9.2</td>
<td>28.6</td>
</tr>
</tbody>
</table>

AUA: American Urological Association; GPS: Genomic Prostate Score.

In the NCCN intermediate group, for example, the 10-year recurrence rate among radical prostatectomy patients was 9.6%. When the GPS was used in the analysis, the 10-year recurrence rate fell to as low as 2.0% (71% reduction) among patients in the low GPS group and 5.1% (47% reduction) in the intermediate GPS group, but rose to 14.3% (49% increase) in the high GPS group. These data suggest the GPS can reclassify a patient’s risk of recurrence based on a specimen obtained at biopsy. However, the findings do not necessarily reflect a clinical scenario of predicting disease progression in untreated patients under active surveillance.

In summary, the evidence from Klein et al.[13] on clinical validity for Oncotype Dx® Prostate suggests the GPS can reclassify a patient’s risk of recurrence based on a specimen obtained at biopsy. However, whether these findings support a conclusion that the GPS could predict the biological aggressiveness of a tumor—hence its propensity to progress—based solely on the level of pathology in a biopsy specimen is unclear.

Clinical Utility

Klein et al. also reported a decision-curve analysis that they have proposed reflects the clinical utility of Oncotype Dx® Prostate.[13] The analysis investigated the predictive impact of the GPS in combination with the Cancer of the Prostate Risk Assessment (CAPRA) validated tool[14] versus the CAPRA score alone on the net benefit for the outcomes of patients with high-grade disease (Gleason >4+3), high-stage disease, and combined high-grade and high-stage disease. They reported that, over a range of threshold probabilities for implementing treatment, “incorporation of the GPS would be expected to lead to fewer treatments of patients who have favorable pathology at prostatectomy without increasing the number of patients with adverse pathology left untreated.” For example, at a threshold risk of 40% (eg, a man weighing the harms of prostatectomy versus benefit over active surveillance at 4:6) the test could identify 2 per 100 with highgrade or high-stage disease at a fixed false positive rate compared with using the CAPRA score alone. However, no confidence intervals were presented for the decision curve analysis. Thus, an individual patient could use the findings to assess his balance of benefits and harms (net benefit) when weighing the choice to proceed immediately to curative radical prostatectomy with its attendant adverse sequelae, or deciding to enter an active surveillance program. The latter would have an immediate benefit realized by forgoing radical prostatectomy, but perhaps would be associated with greater downstream risks of disease progression and subsequent therapies.

Klein’s decision-curve analyses suggest a potential ability of the combined GPS and CAPRA data to help patients make decisions based on relative risks associated with immediate treatment or deferred treatment (i.e., active surveillance). This would reflect the clinical utility of...
the test. However, it is difficult to ascribe possible clinical utility of Oncotype Dx® Prostate in active surveillance because all patients regardless of clinical criteria elected radical prostatectomy within 6 months of diagnostic biopsy. Moreover, the validity of using different degrees of tumor pathology as “markers” to extrapolate the risk of progression of a tumor in vivo is unclear.

**POLARIS®**

**Analytic Validity**

Although there is no reference standard for gene expression profiling tests, other measures of technical performance are relevant and include reproducibility, tissue-sample adequacy, potential batch effects, and test-set bias. In 2015 Warf et al. evaluated the precision of the Cell Cycle Progression (CCP) score using 6 formalin-fixed, paraffin-embedded (FFPE) biopsy (three replicate scores) and 12 FFPE RP (4-6 replicate scores) specimens. Overall precision was estimated from replicate samples, intended to reflect combined variation from tissue dissection through gene expression. Across replicate samples, the standard deviation of the CCP score was 0.1 (95% confidence interval [CI], 0.98 to 0.13). After eight weeks of sample storage, results were similar. In 2013, Myriad Genetics reported 95.3% of samples were adequate to produce a CCP score.

Information is available on the performance of the TaqMan array platform (Applied Biosystems) used in Prolaris® and Oncotype DX® Prostate from the MicroArray Quality Control (MAQC) project, thus providing an indirect evaluation of the analytical validity of Prolaris®. In the MAQC project, initiated and led by FDA scientists, expression data on 4 titration pools from 2 distinct reference RNA samples were generated at multiple test sites on 7 microarray-based and 3 alternative technology platforms including TaqMan. According to the investigators, the results provided a framework to assess the potential of array technologies as a tool to provide reliable gene expression data for clinical and regulatory purposes. The results showed very similar performance across platforms, with a median coefficient of variation of 5% to 15% for the quantitative signal and 80% to 95% concordance for the qualitative detection call between sample replicates.

**Clinical Validity: Needle Biopsy, Conservative Management**

In 2015, Cuzick et al examined three U.K. cancer registries from 1990 to 2003 to identify men with prostate cancer who were conservatively managed following needle biopsy, with follow-up through December 2012. Men were excluded if they had undergone RP or radiation therapy within 6 months of diagnosis. A combination of the CCP and Cancer of the Prostate Risk Assessment (CAPRA) scores was used to predict prostate cancer death. There were 989 men who fit eligibility criteria; CCP scores were calculable for 761 (77%) and combined CCP and clinical variables were available for 585 (59%). Median age at diagnosis was 70.8 years and median follow-up was 9.5 years. The prostate cancer mortality rate was 17% (n=100), with 29% (n=168) dying from competing causes. Higher CCP scores were associated with increased 10-year risk of prostate cancer mortality: 7% (CCP score <0), 15% (CCP score 0-1), 36% (CCP score 1-2), 59% (CCP score >2). A 1-unit increase in CCP was associated with a crude hazard ratio (HR) for death of 2.08 (95% CI, 1.76 to 2.46) and when adjusted for CAPRA
score yielded a HR of 1.76 (95% CI, 1.47 to 2.14). For the combined CAPRA/CCP score, the HR for 10-year prostate cancer mortality increased to 2.17 (95% CI: 1.83 to 2.57). The c-statistic for the CAPRA score was 0.74; adding the CCP score increased the C statistic to 0.78 (no confidence intervals for the AUC were reported). Treatment changes after 6 months were documented in only part of 1 of the 3 cohorts; at 24 months, 45% of the men in this cohort had undergone radiotherapy or prostatectomy. Therefore, the potential effect of treatment changes on prognostic estimates is uncertain.

The original peer-reviewed evidence on the clinical validity of Prolaris® comprises a retrospective cohort (n=349) culled from 6 cancer registries in Great Britain.[18] The investigators assert the CCP score alone was more prognostic than either PSA titer or Gleason score for tumor-specific mortality at 10-year follow-up. Although the patients may be similar to those of a modern U.S. cohort, comparability is unclear from the single publication that is available. Furthermore, the study is limited by the use of archived biopsy specimens, with attendant issues of reproducibility and test reliability.

Cuzick et al. reported this original validation study of Prolaris® to determine its prognostic value for prostate cancer death in a conservatively managed needle biopsy cohort.[18] The authors did not state whether this study adhered to the PRoBE (prospective-specimen-collection, retrospective-blinded evaluation) criteria suggested by Pepe and colleagues for an adequate biomarker validation study.[19] They noted that the cell cycle expression data were read blind to all other data, which conformed to the criteria; however, patients were identified retrospectively from tumor registries and there were no case-control subjects, which does not conform to the criteria.

Patients with clinically localized prostate cancer diagnosed by needle biopsy between 1990 through 1996 were identified in 6 registries. Additional inclusion criteria included age younger than 76 years at diagnosis, available baseline prostate-specific antigen (PSA) measurement, and conservative management.[20] Exclusion criteria included radical prostatectomy, death, evidence of metastatic disease within six months of diagnosis, or hormone therapy prior to diagnostic biopsy. A cell cycle progression (CCP) score consisting of expression levels of 31 predefined cell cycle progression genes and 15 housekeeper genes was generated using TaqMan low-density arrays. The values of each of the 31 CCP genes were normalized by subtraction of the average of up to 15 nonfailed housekeeper genes for that replicate.

Of 776 patients diagnosed by needle biopsy, 349 (79%) produced a CCP score and had complete baseline and follow-up information. The median potential follow-up time was 11.8 years during which a total of 90 deaths from prostate cancer occurred within 2799 person-years of actual follow-up. The main assessment of the study was a univariate analysis of the association between death from prostate cancer and the CCP score. A further predefined assessment of the added prognostic information after adjustment for the baseline variables was also undertaken. The primary end point was time to death from prostate cancer. A number of covariates were evaluated: centrally reviewed Gleason primary grade and score; baseline PSA value; clinical stage; extent of disease (percent of positive cores); age at diagnosis; Ki-67 immunohistochemistry; and initial treatment. The results are shown in Table 3.
Table 3. Univariate and Multivariate Analysis for Death From Prostate Cancer in the Cuzick 2012 Validation Study

<table>
<thead>
<tr>
<th>Variable</th>
<th>N</th>
<th>Univariate Hazard Ratio (95% CI)</th>
<th>Multivariate Hazard Ratio (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-unit increase in CCP score</td>
<td>349</td>
<td>2.02 (1.62 to 2.53)</td>
<td>1.65 (1.31 to 2.09)</td>
</tr>
<tr>
<td><strong>Gleason score</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;7</td>
<td>106</td>
<td>0.46 (0.25 to 0.86)</td>
<td>0.61 (0.32 to 1.16)</td>
</tr>
<tr>
<td>7</td>
<td>152</td>
<td>Referent</td>
<td>Referent</td>
</tr>
<tr>
<td>&gt;7</td>
<td>91</td>
<td>2.70 (1.72 to 4.23)</td>
<td>1.90 (1.18 to 3.07)</td>
</tr>
<tr>
<td><strong>log (1+PSA)/(ng/mL)</strong></td>
<td>349</td>
<td>1.70 (1.31 to 2.20)</td>
<td>1.37 (1.05 to 1.79)</td>
</tr>
<tr>
<td><strong>Proportion of positive cores</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;50%</td>
<td>69</td>
<td>0.50 (0.22 to 1.12)</td>
<td></td>
</tr>
<tr>
<td>50 to &lt;100%</td>
<td>106</td>
<td>Referent</td>
<td></td>
</tr>
<tr>
<td>100%</td>
<td>160</td>
<td>1.66 (1.01 to 2.73)</td>
<td></td>
</tr>
<tr>
<td><strong>Age at diagnosis (y)</strong></td>
<td>349</td>
<td>1.00 (0.96 to 1.04)</td>
<td></td>
</tr>
<tr>
<td><strong>Clinical stage</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T1</td>
<td>38</td>
<td>0.75 (0.32 to 1.75)</td>
<td></td>
</tr>
<tr>
<td>T2</td>
<td>106</td>
<td>Referent</td>
<td></td>
</tr>
<tr>
<td>T3</td>
<td>43</td>
<td>1.74 (0.90 to 3.38)</td>
<td></td>
</tr>
<tr>
<td><strong>Hormone use</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>200</td>
<td>Referent</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>149</td>
<td>1.97 (1.30 to 2.98)</td>
<td></td>
</tr>
</tbody>
</table>

CI: confidence interval

The median CCP score was 1.03 (IQ range, 0.41-1.74). The primary univariate analysis suggested that a 1-unit increase in CCP score was associated with a 2-fold increase in the risk of dying from prostate cancer. In preplanned multivariate analyses, extent of disease, age, clinical stage, and use of hormones had no statistically significant effect on risk; only the Gleason score and PSA remained in the final model. Further exploratory multivariate modeling to produce a combined score, including CCP, Gleason score, and PSA level, suggested a strong, predominant nonlinear influence of the CCP score in predicting the risk of death from prostate cancer (p=0.008).

Cuzick and colleagues suggested this combined score provided additional discriminatory information to help identify low-risk patients who could be safely managed by active surveillance. For example, among patients with a Gleason score of 6, for whom uncertainty existed as to the appropriate management approach, the predicted 10-year prostate cancer death rate ranged from 5.1% to 20.9% based on Gleason score and PSA; the range when assessed against the combined CCP, Gleason, and PSA score was 3.5% to 41%. However, the authors cautioned that because death rates were rare in this group, larger cohorts would be required to fully assess the value of the CCP combined score. Measures that would suggest improved discriminatory ability, eg., AUC or reclassification, were not reported. Evidence was not provided that the test could correctly reclassify men initially at high risk to lower risk to avoid overtreatment, or conversely those initially at low risk to high risk to avoid undertreatment.

Table 4 shows Kaplan-Meier analyses of 10-year risk of prostate cancer death stratified by CCP score groupings. Cuzick et al. reported no significance tests for the estimates. Nor did they explain the apparent substantial difference in mortality rates among patients in the 0 ≤
CCP ≤ 2 grouping (range, 19.3-21.1%) and those in the 2 < CCP ≤ 3 and > 3 groupings (range, 48.2-74.9%). The difference may simply reflect clinical criteria, for example, proportions of lower compared with higher Gleason grade cancers, respectively.

Table 4. Kaplan-Meier Estimates of Prostate Cancer Death at 10 Years According to CCP Score Groupings in the Cuzick 2012 Validation Study

<table>
<thead>
<tr>
<th>CCP Score Group</th>
<th>N</th>
<th>10-Year Death Rate (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CCP ≤ 0</td>
<td>36</td>
<td>19.3</td>
</tr>
<tr>
<td>0 &lt; CCP ≤ 1</td>
<td>133</td>
<td>19.8</td>
</tr>
<tr>
<td>1 &lt; CCP ≤ 2</td>
<td>114</td>
<td>21.1</td>
</tr>
<tr>
<td>2 &lt; CCP ≤ 3</td>
<td>50</td>
<td>48.2</td>
</tr>
<tr>
<td>&gt; 3</td>
<td>16</td>
<td>74.9</td>
</tr>
</tbody>
</table>

Clinical Validity: Posttreatment (radical prostatectomy and external beam radiation therapy)

In 2014 Bishoff et al.[21] examined the prognostic ability of the CCP score in 3 cohorts: Martini Clinic (n=283, simulated biopsies from FFPE RP specimen), Durham Veterans Affairs Medical Center (n=176, diagnostic biopsies) and Intermountain Healthcare (n=123, diagnostic biopsies). The combined analysis included all 582 patients. Gleason scores were 7 or lower in 93% of men. In the combined cohorts, a unit increase in the CCP score increased the adjusted HR for BCR by 1.47 (95% CI, 1.23 to 1.76). Metastatic events (n=12) were too few to draw conclusions. Although the CCP score was associated with increased risk of BCR, the analyses do not allow examining whether the CCP score provides improved discrimination over clinicopathologic variables.

In 2013 Myriad-funded study, Freedland et al. evaluated the CCP score’s ability to predict biochemical recurrence (BCR) in a cohort of men treated with external beam radiation therapy (EBRT).[22] The CCP score was derived retrospectively from diagnostic biopsy specimens of men diagnosed with prostate cancer from 1991 to 2006 (n=141). The primary outcome assessed was time from EBRT to BCR. In a multivariable analysis with Gleason score, PSA, percent positive cores, and androgen deprivation therapy, the hazard ratio (HR) was 2.11 for a one-unit increase in CCP score (equivalent to a doubling of gene expression) (p=0.034), indicating that CCP provides prognostic information that is not provided by standard clinical parameters. At ten years post-EBRT, the CCP score was associated with prostate cancer specific mortality (p-value = 0.013). The limitations of this study include small size of the cohort, small number of treatment failures (only 19 patients [13%] had BCR), and short follow-up time. The authors conceded that “definitive conclusions regarding time dependency will require additional studies”.

In 2013 Cooperberg et al.[11] sought to evaluate the CCP score in a RP cohort and the incremental improvement over the Cancer of the Prostate Risk Assessment Postsurgical (CAPRA-S) score for predicting BCR employing a prospective-retrospective design (conforming to a PROBE study design). A prognostic model was developed from the RP cohort described by Cuzick et al (2011).[23] The validation cohort was obtained from patients identified from the University of California, San Francisco (UCSF) Urologic Oncology Database. Tissue sufficient to obtain a CCP score was available for 413 men (69% of the 600 eligible samples).
Both UCSF and Myriad Genetics performed statistical analyses. In the validation cohort, 95% had Gleason scores of 7 or lower, 16% of samples had positive margins, 4% had seminal vesicle invasion, and 23% had extracapsular extension. BCR occurred in 82 men (19.9%). The unadjusted HR for BCR increased by 2.1 (95% CI, 1.6 to 2.9) per unit increase in CCP score. A predictive model for the combined CCP/CAPRA-S developed in the Cuzick et al (2011) [23] RP cohort applied to the UCSF cohort obtained an AUC for BCR with CAPRA-S alone of 0.73 increasing to 0.77 for the combined CCP/CAPRA-S.

Cuzick et al (2011) [23] examined the potential use of the Prolaris® CCP test combined with a clinical score following RP, using a retrospective cohort and the prospective-retrospective design for archived samples. The study also included a cohort of men with localized prostate cancer detected from specimens obtained during transurethral resection of the prostate, which is not a population of interest here, and so has not been described. Men conservatively managed post RP between 1985 and 1995 were identified from a tumor registry (n=366 with CCP scores, Scott and White Clinic, in Texas). The primary end point was time to biochemical recurrence (BCR) and the secondary end point was prostate cancer death. Myriad Genetics assessed CCP scores blindly. The median age of patients was 68 years and the median follow-up 9.4 years. Gleason scores were 7 or lower in 96%, but margins were positive in 68%. Cancers were clinically staged as T3 in 34%; following RP, 64% was judged pathologic stage T3. CCP score was associated with BCR (adjusted HR=1.77; 95% CI, 1.40 to 2.22). Analyses of prostate cancer deaths in the RP cohort were problematic, owing to only 12 (3%) deaths. The clinical score included PSA, stage, positive surgical margins, and Gleason score. The model was optimized using stepwise variable selection (eg, a development model). The AUC for BCR within 5 years in the RP cohort was 0.825 for the clinical score and 0.842 for the combined clinical/CCP score. The discriminatory ability of the clinical score is of note. Although the CCP increased the AUC by 2%, whether that improvement might be clinically useful is unclear lacking reclassification or examination of net benefit.

Clinical Utility

One large prospective registry study, funded by Myriad, was recently published that evaluated the impact of the CCP test on treatment decision making for patients newly diagnosed with prostate cancer. [24] Patients (n=1206) with newly diagnosed prostate adenocarcinoma had the CCP test performed on initial prostate biopsy tissue. Changes in treatment decision making was tracked using the answers provided by physicians in sequential surveys relative to initial therapy recommendations (before cell cycle progression). The CCP test caused a change in actual treatment in 47.8% of patients, 72.1% of which were reductions and 26.9% of which were increases in treatment. For each clinical risk category there was a significant change in treatment modality (intervention vs nonintervention) before vs after CCP testing (p=0.0002). This study did not report any changes in patient-important outcomes, such as biochemical recurrence, cancer-specific survival or long term survival. Although this study reported a change of management in a modest percentage of patients, there was no evidence that these changes in management lead to clinically important improvements in health outcomes.

Two retrospective survey studies that assessed the potential impact of Prolaris® on physicians' treatment decisions. [25,26] The authors of each study have suggested their findings support the "clinical utility" of the test, based on whether the results would lead to a change in treatment.
Although this information may be useful in assessing the potential test uptake by urologists, it does not demonstrate clinical utility in clinical settings. In a decision-curve analysis, Cooperberg et al.\cite{11} found the CAPRA-S score superior to CCP alone (as well as treat-none or treat-all strategies) in men postprostatectomy. A combined CCP/CAPRA-S predictor appeared only slightly better than CAPRA-S alone for thresholds of approximately 30% or more. For example, at a threshold of 30% (ie, meaning a man would value the harm-to-benefit of treatment such as radiotherapy as 3:7), the combined CCP/CAPRA-S would detect about two more men per 100 likely to experience BCR if the false-positive rate was fixed. However, the lack of confidence intervals for the decision-curve analysis, together with the small difference, is consistent with an uncertain net benefit obtained by adding CCP to the CAPRA-S score.

**Systematic Review**

In 2016, results of a systematic review and meta-analysis supported by the manufacturer were reported.\cite{27} Published and unpublished studies of prognostic validity or clinical utility of CCP testing were eligible for inclusion. Seven published studies were identified; 5 were clinical validity studies. Two were reviewed in the previous paragraphs and the remaining validity studies will be reviewed in a subsequent section on post-RP management. The other 2 “utility” studies are discussed in the following section. Two validity studies reported outcomes for disease-specific mortality\cite{18,23} but of the 2 only the Cuzick et al (2012) included newly diagnosed patients, so the pooled outcome is not of relevance in this section.

**DECIPHER®**

**Analytic Validity**

Published data on the analytic validity of the Decipher test consists of one study, which was performed on surgical resection specimens from patients with prostate cancer identified to be in a postsurgery high-risk population. The Decipher test platform was performed in formalin-fixed, paraffin-embedded (FFPE) tissue to assess the differential expression in the discovery, validation and clinical application.\cite{28} Matched FFPE and unfixed fresh-frozen specimens from paired tumor and normal samples from kidney, lung and colon were compared and the microarray signals derived from the degraded RNA extracted from FFPE specimens was found to be highly analogous to the signals from the RNA in the fresh frozen specimens.

**Clinical Validity**

The clinical validity of the Decipher test genomic classifier (GC) has been reported in studies to predict metastasis, mortality or BCR after RP in patients with postoperative high-risk features like pathologic stage T2 with positive margins, pathologic stage T3 disease or a rising PSA.\cite{29-37} All studies were conducted from registry data. The development study was a nested case-control design,\cite{39} four were case-cohort studies, and four used retrospective cohorts. Owing to apparent overlap in samples, the number of unique patients in the studies is difficult to ascertain. Seven studies were supported by GenomeDx, which offers the Decipher test; all studies identified multiple authors as company employees.
Four studies,\(^{32-35}\) including the test (validation) sample from the development study, examined men observed following radical prostatectomy and undergoing adjuvant or salvage radiotherapy. Median follow-up periods ranged from 6.4 to 16.9 years. The distributions of Gleason scores in the studies varied—from 24.3% to 49.3% with 8 or higher and 0.4% to 15.1% with 6 or lower. Extracapsular extension of the tumor ranged from 42.7% and 72.3% of men across the studies.

In 2016, Klein et al. evaluated the ability of the Decipher genomic classifier in predicting metastasis from the prostate needle biopsy diagnostic tumor tissue from 56 men.\(^{37}\) Median followup time was eight years. In that time, eight patients metastasized and three died of PCa. Decipher plus NCCN model had an improved c-index of 0.88 (95% confidence interval [CI] 0.77-0.96) compared to NCCN alone (c-index 0.75; 95% CI 0.64-0.87). Using the Cox multivariable analysis, Decipher was the only significant predictor of metastasis when adjusting for age, preoperative PSA and biopsy Gleason score, with a hazard ratio of 1.72 per 10% increase; 95% CI 1.07-2.81; \(p=0.02\).

In 2015, Ross et al.\(^{38}\) assessed the prognostic accuracy for metastasis through 10 years, excluding men receiving any adjuvant therapy following radical prostatectomy over median follow-up periods of 7.8 and nine years. The investigators reported a 6.5% 5-year cumulative incidence of metastases in men with GC scores of 0.45 or lower, compared with 30.3% in those with scores higher than 0.60. The AUCs for development of metastases was 0.76 for the GC. In addition, it was found that combining the GC with the best clinicopathologic tool improve the AUC. The study did not include a “standard” reclassification table, but did report 10-year cumulative incidence of metastases stratified by GC and CAPRA-S. The GC appeared to discriminate within CAPRA-S categories, but appeared to add little to a score greater than 5.

In 2015, Den et al. reported on the use of the Decipher genomic classifier (GC) to provide prognostic and predictive information into the development of metastases in men receiving post-RP RT (either 3-dimensional conformal or IMRT).\(^{30}\) Genomic classifier scores were calculated from 188 men who were identified within the GenomeDx prostate cancer database with pathologic stage T3 or margin-positive prostate cancer and had received post-RP RT at 1 of 2 academic centers between 1990 and 2009. The primary endpoint was metastatic disease (regional or distant) documented on computed tomography or bone scan. Adjuvant versus salvage RT was defined by PSA levels of 0.2 ng/mL or less and more than 0.2 ng/mL before initiation of RT. The clinical characteristics of eligible patients included 72% of men with extraprostatic extension, 35% with seminal vesicle invasion, and 78% with positive surgical margins. Twenty-one percent of patients had a Gleason score of 8 or more. Fifty-one percent of patients received adjuvant RT (89% within 12 months of RP) and overall, patients received RT at a median of 5 months after RP (range, 1-160 months). Thirty percent of patients received hormonal therapy with RT. Median follow-up after RP and RT was 10 and 8 years, respectively. Cumulative incidence of metastatic disease at 5 years after RT for low, average and high GC scores was 0%, 9% and 29% (\(p=0.002\)). In a multivariate analysis, GC and pre-RP PSA were independent predictors of metastasis (both \(p<0.01\)). In the low GC score group (score <0.4) there was no difference in cumulative incidence of metastasis compared with patients who received adjuvant or salvage RT (\(p=0.79\)), however, for patients with higher GC scores (\(\geq 0.4\)), the cumulative incidence of metastasis at 5 years was 6% for patients treated with adjuvant RT compared to 23% treated with salvage RT (\(p<0.01\)). The authors concluded
that patients with low GC scores are best treated with salvage RT and those with high GC scores with adjuvant RT.

In 2014 Klein et al. evaluated whether use of the Decipher genomic classifier (GC) test improved accuracy in predicting metastasis within 5 years following radical prostatectomy (rapid metastasis [RM]). Participants included 169 patients who underwent radical prostatectomy between 1987 and 2008, of which 15 were RM and 154 were non-RM controls. Metastasis developed between 1.7 and 3.3 years (median 2.3 years). Test performance was evaluated both individually and in combination with clinical risk factors. After adjusting for clinical factors, Decipher was a significant predictor of RM (OR 1.48; p=0.018). Compared to the Stephenson model, the CAPRA-S, and previously reported biomarkers, Decipher had the highest concordance index (c-index), with the highest c-index achieved with integration of Decipher into the Stephenson nomogram.

Karnes et al. prospectively created a randomly selected subcohort from the same initial 1010 post-prostatectomy patients in the Cooperberg et al. study. Patients with metastasis at diagnosis or with any prior treatment for prostate cancer were excluded. A randomly selected subcohort was created, with genomic data was available for 219 patients. Following radical prostatectomy the rates of biochemical recurrence (BCR) at 3 years was 35% and metastasis at 5 years was 6%. Median genomic classifier scores were consistently higher in patients with metastases at last follow-up (mean 6.7 years). Median genomic classifier scores also increased with higher Gleason scores. The authors concluded that the higher net benefit of genomic-based classifiers suggested increased specificity (i.e., lower false positives) compared with clinical-only risk models. Because patients with intermediate risk tumors may progress to advanced disease, the authors recommended further study of genomic classifiers in randomized datasets to determine whether genomic classifier scores from diagnostic biopsy specimens can predict metastasis as well as postoperative specimens. A possible limitation of this study was that nearly 15% of patients were node-positive and 45% received adjuvant therapy. Whether the genomic classifier predicted benefit from local (i.e., radiation) or systemic (e.g., hormone) therapies could not be determined because patients were not randomized to these treatments.

In 2014, Den et al reported that in within a Decipher low-risk group that was treated post-RP with RT, there was no difference in oncologic outcomes (either biochemical failure or metastasis) whether they received adjuvant or salvage RT. For the men classified as high-risk by Decipher, a median 4-year PSA-free survival advantage was observed in the patients that received adjuvant versus salvage RT. Of these men classified as high-risk by GC, those who received adjuvant radiation had a 3% cumulative incidence of metastases as compared with 23% incidence of metastasis by 8 years in those who delayed treatment and received salvage radiation.

**Clinical Utility**

Several studies have compared physician’s treatment recommendations before and after receiving GC results. Because the studies did not include information on outcomes and clinical validity has not been established, it is not known whether these treatment decisions represent a clinical improvement in management.
Ross et al (2016) reported results of a retrospective, comparative study of RT after RP for 422 men with pT3 disease or positive margins.[43] The men were from 4 cohorts previously described (Karnes 2013; Den 2014; Ross 2016; Freedland 2016). The 4 treatment groups were adjuvant RT (n=111), minimal residual disease salvage RT (n=70), salvage RT (n=83), and no RT (n=157). The primary end point was metastasis. Thirty-seven men developed metastasis and the median follow-up was 8 years. Both CAPRA-S (HR=1.39; 95% CI, 1.18 to 1.62) and Decipher (HR=1.28; 95% CI, 1.08 to 1.52) were independently associated with metastasis in multivariable analysis. There was no evidence that treatment effect was dependent on genomic risk (interaction p=0.16 for CAPRA-S, p=0.39 for Decipher), Men with low CAPRA-S or low Decipher scores had a low risk of metastatic events regardless of treatment selection and men with high CAPRA-S or Decipher scores benefitted from adjuvant RT compared to the other treatments.

Lobo et al (2015) reported an individualized decision analysis comparing the GC to “usual care” using data from the cohorts in Karnes et al (2013) and Den et al (2014).[44] The usual care probabilities of receiving each treatment were derived from the published literature. A 6% threshold for the GC score was used for GC-based treatment. Using the cohort from Karnes et al (2013), the estimated 10-year probability of metastasis or death was 0.32 (95% CI, 0.32 to 0.33) for usual care compared to 0.31 (95% CI, 0.30 to 0.32) for GC-based treatment. In the cohort from Den et al (2014), the estimated 10-year probability of metastasis or death was 0.28 (95% CI, 0.27 to 0.29) for usual care compared to 0.26 (95% CI, 0.25 to 0.27) for GC-based treatment.

In a retrospective review of 110 patients with pT3 disease (71%) or positive surgical margins (PSMs) (63%) following radical prostatectomy.[39] US board certified urologists (n=107) were invited to provide adjuvant treatment recommendations for 10 cases randomly drawn from the pool of case histories. These recommendations were based on clinical variable only; GC test results were not provided. After these recommendations were completed, the urologists were asked to make recommendations with GC test results provided for the same patients, with case histories randomly re-ordered to minimize recall bias. Of the 107 urologists invited to participate, 51 took part in the study, providing 530 recommendations without GC test results. The overall change in recommendations when GC test results were added to clinical information was 31%). Of the 303 patients recommended for observation based on clinical information only, 38 (13%; 95% CI 9-17%) were recommended for radiation therapy and 9 (3%; 95% CI 1-6%) for radiation plus hormone therapies when GC results were made available. Of 193 patients recommended for radiation therapy initially, when GC results were available 77 (40%; 95% CI: 33-47%) were changed to observation. Therapy intensity was also highly correlated with higher risk according to the GC test. This study had a number of limitations which the authors noted as typical for early-stage evaluations of biomarker technologies. These limitations included the recommendations being based on case histories from chart reviews, the 51 participating urologists may not be representative of all urologists treating prostate cancer, patient health status and pathological information was not available to the urologists, and the effect on health outcomes of any changes in treatment were not studied.
In 2014, Michalopoulos et al. assessed the effect of the GC test on urologists' decisions regarding treatment of men with high-risk disease post-RP.[40] Participating urologists were from 15 community practices who had ordered the GC test for 146 prostate cancer patients with either pathologic stage T3 or positive surgical margins post-RP. The urologists were asked to provide their treatment recommendations before and after receiving the GC test report. Prior to availability of the GC test result, treatment recommendations were based on Gleason score and CAPRA-S risk: 40 (27.4%) were recommended to undergo adjuvant therapy, 102 (69.9%) close observation and 4 (2.7%) “other.” Using the GC risk score, 61.6% and 38.4% of patients were identified as low- and high-risk of metastasis, respectively. More than 60% of high-risk patients were reclassified as low risk after the GC test results. Overall, adjuvant treatment recommendations were modified for 30.8% (95% CI, 23% to 39%) of patients. With the GC test results, 42.5% of patients who were initially recommended for adjuvant therapy were subsequently recommended observation. The GC test score also influenced the intensity of the treatment recommendation, with about 40% of patients classified as high-risk by GC score recommended more intense therapy versus 1.1% of those deemed low-risk by GC score. Limitations to the study included that treatment recommendations were submitted electronically and did not track the actual treatment administered, it was not possible to assess patient influence on the decision-making process, the association between GC test results and treatment recommendations was determined using “early adopters” of the test, and all participants were community-based physicians whose treatment recommendations may differ from those of academic centers.

PRACTICE GUIDELINE SUMMARY

NATIONAL COMPREHENSIVE CANCER NETWORK

The NCCN guidelines Version 1.2017 for prostate cancer do not include gene expression profile analysis in their recommendations, though tests are listed in the discussion on tissue-based molecular assays. The discussion section notes that the clinical utility has not been established in prospective RCTs or comparative effectiveness studies.[45]

SUMMARY

There is not enough research to show that gene expression profile testing improves health outcomes in patients with prostate cancer. Additionally, there are no practice guidelines based on research which recommend the use of these tests for the management of prostate cancer. Therefore, gene expression analysis for prostate cancer management is considered investigational.

REFERENCES


# CODES

<table>
<thead>
<tr>
<th>Codes</th>
<th>Number</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>CPT</td>
<td>0005U</td>
<td>Oncology (prostate) gene expression profile by real-time RT-PCR of 3 genes (ERG, PCA3, and SPDEF), urine, algorithm reported as risk score</td>
</tr>
<tr>
<td></td>
<td>81599</td>
<td>Unlisted multianalyte assay with algorithmic analysis</td>
</tr>
<tr>
<td>HCPCS</td>
<td>None</td>
<td></td>
</tr>
</tbody>
</table>

*Date of Origin: January 2014*