Evidence Synthesis Program

Prognostic Value of Genomic Classifier Testing for Prostate Cancer: A Systematic Review

March 2023

U.S. Department of Veterans Affairs
Veterans Health Administration
Health Services Research & Development Service

Recommended citation: Boyer MJ, Carpenter D, Gingrich JR, et al. Prognostic Value of Genomic Classifier Testing for Prostate Cancer: A Systematic Review. Washington, DC: Evidence Synthesis Program, Health Services Research and Development Service, Office of Research and Development, Department of Veterans Affairs. VA ESP Project #09-010; 2023.

AUTHORS

Author roles, affiliations, and contributions to the present report (using the CRediT taxonomy) are summarized in the table below.

Author	Role and Affiliation	Report Contribution
Matthew J. Boyer, MD, PhD	Medical Instructor, Department of Radiation Oncology, Duke University School of Medicine Durham, NC	Conceptualization, Methodology, Investigation, Formal analysis, Writing – original draft, Writing – review & editing
David Carpenter, MD, MHS	Resident, Department of Radiation Oncology, Duke University School of Medicine Durham, NC	Conceptualization, Methodology, Investigation, Formal analysis, Writing – original draft, Writing – review & editing
Jeffrey R. Gingrich, MD, FACS	Staff Physician, Durham VA Medical Center Professor of Surgery, Urology, Duke University Durham, NC	Conceptualization, Methodology, Investigation, Writing – original draft, Writing – review & editing
Sudha Raman, PhD	Assistant Professor, Department of Population Health Sciences at Duke University School of Medicine Durham, NC	Conceptualization, Methodology, Investigation, Writing – original draft, Writing – review & editing
Deepika Sirohi, MD	Associate Professor, Department of Pathology, University of Utah and ARUP Laboratories Salt Lake City, UT	Conceptualization, Methodology, Investigation, Writing – original draft, Writing – review & editing
Amir Alishahi Tabriz, MD, PhD, MPH	Assistant Member, Department of Health Outcomes and Behavior Moffitt Cancer Center Tampa, FL	Conceptualization, Methodology, Investigation, Writing – original draft, Writing – review & editing
Alexis Rompre-Brodeur, MD, FRCSC	Assistant Professor of Surgery, Division of Urology, McGill University Montreal, Canada	Conceptualization, Methodology, Investigation, Writing – original draft, Writing – review & editing
Joseph Lunyera, MBChB	Medical Instructor, Department of Medicine, Duke University School of Medicine Durham, NC	Conceptualization, Methodology, Investigation, Writing – original draft, Writing – review & editing
Fahmin Basher, MD, PhD	Fellow, Hematology, Medical Oncology, Duke University Durham, NC	Conceptualization, Methodology, Investigation, Writing – original draft, Writing – review & editing

Author	Role and Affiliation	Report Contribution
Rhonda L. Bitting, MD	Staff Physician, Genitourinary Medical Oncology, Durham VA Medical Center Associate Professor of Medicine, Duke University Durham, NC	Conceptualization, Methodology, Investigation, Writing – original draft, Writing – review & editing
Andrzej Kosinski, PhD	Professor of Biostatistics & Bioinformatics, Duke University Durham, NC	Data curation, Methodology, Formal analysis, Visualization, Writing – review & editing
Sarah Cantrell, MLIS, AHIP	Associate Director for Research & Education, Duke University Medical Center Library & Archives, Duke University School of Medicine Durham, NC	Conceptualization, Methodology, Writing – review & editing
Adelaide M. Gordon, MPH	Project Coordinator, Durham Evidence Synthesis Program (ESP) Center Research Health Science Specialist, Durham Center of Innovation to Accelerate Discovery and Practice Transformation, Durham VA Health Care System Durham, NC	Conceptualization, Data curation, Methodology, Investigation, Project administration, Formal analysis, Writing – original draft, Writing – review & editing
Belinda Ear, MPH	Research Assistant, Durham ESP Center, Durham, NC Research Health Science Specialist, Durham Center of Innovation to Accelerate Discovery and Practice Transformation, Durham VA Health Care System Durham, NC	Conceptualization, Data curation, Methodology, Project administration, Writing – original draft
Jennifer M. Gierisch, PhD	Co-director, Durham ESP Center Core Investigator, Durham Center of Innovation to Accelerate Discovery and Practice Transformation, Durham VA Health Care System Associate Professor, Department of Population Health Sciences, Duke University School of Medicine Durham, NC	Conceptualization, Methodology, Investigation, Writing – review & editing
Morgan Jacobs, MPH	Research Assistant, Durham ESP Center Durham, NC	Data curation, Project administration, Visualization

Author	Role and Affiliation	Report Contribution
Karen M. Goldstein, MD, MSPH	Co-director, Durham ESP Center Core Investigator, Durham Center of Innovation to Accelerate Discovery and Practice Transformation, Durham VA Health Care System Staff Physician, Durham VA Medical Center Associate Professor, Department of Medicine, Division of General Internal Medicine, Duke University Durham, NC	Conceptualization, Data curation, Methodology, Supervision, Investigation, Formal analysis, Visualization, Writing – original draft, Writing – review & editing

This report was prepared by the Evidence Synthesis Program Center located at the **Durham VA Medical Center**, directed by Jennifer M. Gierisch, PhD, MPH, and Karen M. Goldstein, MD, MSPH, and funded by the Department of Veterans Affairs, Veterans Health Administration, Health Services Research and Development.

PREFACE

The VA Evidence Synthesis Program (ESP) was established in 2007 to provide timely and accurate syntheses of targeted health care topics of importance to clinicians, managers, and policymakers as they work to improve the health and health care of Veterans. These reports help:

- Develop clinical policies informed by evidence;
- Implement effective services to improve patient outcomes and to support VA clinical practice guidelines and performance measures; and
- Set the direction for future research to address gaps in clinical knowledge.

The program comprises four ESP Centers across the US and a Coordinating Center located in Portland, Oregon. Center Directors are VA clinicians and recognized leaders in the field of evidence synthesis with close ties to the AHRQ Evidence-based Practice Center Program. The Coordinating Center was created to manage program operations, ensure methodological consistency and quality of products, interface with stakeholders, and address urgent evidence needs. To ensure responsiveness to the needs of decision-makers, the program is governed by a Steering Committee composed of health system leadership and researchers. The program solicits nominations for review topics several times a year via the program website.

The present report was developed in response to a request from the National Radiation Oncology Program and the National Oncology Prostate Cancer Clinical Pathways team. The scope was further developed with input from Operational Partners (below), the ESP Coordinating Center, the review team, and the technical expert panel (TEP). The ESP consulted several technical and content experts in designing the research questions and review methodology. In seeking broad expertise and perspectives, divergent and conflicting opinions are common and perceived as healthy scientific discourse that results in a thoughtful, relevant systematic review. Ultimately, however, research questions, design, methodologic approaches, and/or conclusions of the review may not necessarily represent the views of individual technical and content experts.

ACKNOWLEDGMENTS

The authors are grateful to Liz Wing for editorial and citation management support, Katherine Van Loon for citation management support, and the following individuals for their contributions to this project:

Operational Partners

Operational partners are system-level stakeholders who help ensure relevance of the review topic to the VA, contribute to the development of and approve final project scope and timeframe for completion, provide feedback on the draft report, and provide consultation on strategies for dissemination of the report to the field and relevant groups.

Edward Obedian, MD
Chief, Radiation Oncology
Bronx VA Hospital

Michael G. Chang, MD
Chief, Radiation Oncology
Richmond VA Medical Center

Maria Kelly, MD
Executive Director
National Radiation Oncology Program

Technical Expert Panel

To ensure robust, scientifically relevant work, the TEP guides topic refinement; provides input on key questions and eligibility criteria, advising on substantive issues or possibly overlooked areas of research; assures VA relevance; and provides feedback on work in progress. TEP members are listed below:

Daniel Spratt, MD
Chair
University Hospitals Cleveland Medical Center
Department of Radiation Oncology

Jeremy Shelton, MD, MSHS
Director of Surgical Informatics/Chief of Telesurgery
West Los Angeles VA Medical Center
National Surgery Office

Andrew J. Armstrong, MD, ScM, FACP
Professor of Medicine, Surgery, Pharmacology and Cancer Biology Director of Research
Duke Cancer Institute Center for Prostate and
Urologic Cancers Divisions of Medical Oncology and Urology Duke University

Eric Klein, MD
Chairman
Glickman Urological & Kidney Institute Taussig Cancer Institute,
Cleveland Clinic

Peer Reviewers

The Coordinating Center sought input from external peer reviewers to review the draft report and provide feedback on the objectives, scope, methods used, perception of bias, and omitted evidence (see Appendix F for disposition of comments). Peer reviewers must disclose any relevant financial or non-financial conflicts of interest. Because of their unique clinical or content expertise, individuals with potential conflicts may be retained. The Coordinating Center works to balance, manage, or mitigate any potential nonfinancial conflicts of interest identified.

ABBREVIATIONS TABLE

ART	Adjuvant radiotherapy
AS	Active surveillance
AUA	American Urological Association
AUC	Area under the curve
BCR	Biochemical recurrence
CCP	Cell cycle progression
CHARMS-PF	Checklist for critical appraisal and data extraction for systematic reviews of prognostic factor studies
CI	Confidence interval
COE	Certainty of evidence
EAU	European Association of Urology
GC	Genomic classifier
GG	Grade group
GPS	Genomic Prostate Score test
HR	Hazard ratio
IQR	Interquartile range
KQ	Key question
MRI	Magnetic resonance imaging
NCCN	National Comprehensive Cancer Network
PICOTS	Population, intervention, comparator, outcome, timing, setting
PORTOS	Post-operative Radiation Therapy Outcomes Score
PSA	Prostate-specific antigen
QUIPS	Quality In Prognosis Studies
ROB	Risk of bias
ROBINS-I	Risk of Bias in Nonrandomized Studies of Interventions
RP	Radical prostatectomy
SRT	Salvage radiotherapy
VHA	Veterans Health Administration

EXECUTIVE SUMMARY

Key Findings

- Eleven studies (6,953 patients) reported risk reclassification either as a direct comparison or through integration of the genomic classifier test results with a clinical feature-based risk assessment.
- While there was a wide range of impact on risk reclassification across studies, we found that there was no change in risk classification for a majority of patients across test types: Decipher (21% to 51%; 3 studies), Oncotype (37% to 81%; 6 studies), and Prolaris (58%; 1 study).
- Across 14 observational studies (2,561 patient cases) at the time of diagnosis and after prostatectomy, there was a pattern of changes to treatment recommendations after receipt of genomic classifier tests. However, in the single randomized trial (191 patients) evaluating the impact of the incorporation of Oncotype test results into treatment decisions, there was no statistically significant effect.
- Prolaris or Oncotype usage prior to definitive treatment led to a change in management 16% to 65% of the time, while Decipher studies noted that a higher risk score increased odds of influencing treatment post-prostatectomy.
- Overall, we found that these tests do provide modest additional prognostic information with respect to biochemical recurrence, development of metastatic disease, and prostate cancer–specific mortality; however, the certainty of this evidence was very low for Oncotype and Prolaris and low for Decipher.
- Significant limitations of the evidence include that it is largely drawn from patients diagnosed and treated prior to the current era of prostate cancer management and that no harms due to genomic classifier testing were reported by any study.

INTRODUCTION

Prostate cancer is the most common malignancy in men, with an estimated 268,490 new cases in the United States in 2022, and 12,500 new diagnoses annually within the Veterans Health Administration (VHA). A major challenge for prostate cancer management is identifying patients who would benefit from treatment and tailoring the intensity of that treatment to personalized risk assessments. Risk stratification traditionally has been based on readily available clinical features; however, multiple options exist for treatment, and there is variability in patient outcomes not otherwise explained by currently recognized risk factors. Individualized prognosis beyond clinically based risk stratification schemas could inform patient-physician decision-making, reduce unnecessary overtreatment, and improve patient outcomes. A relatively recent advancement in prostate cancer risk stratification is the development of commercially available, tissue-based genomic classifiers. This systematic review addresses the impact of 3 commercial genomic classifier tests—Decipher, Oncotype DX GPS (now named Genomic Prostate Score but

referred to in this report as Oncotype), and Prolaris—on risk classification, treatment choice and harms, and the prognostic ability of these tests beyond the clinical features of patients diagnosed with or treated for localized prostate cancer. (See Appendix A for guidelines for the 3 tests.)

Thus, this review sought to address 3 key questions (KQs):

KQ1: Among individuals with localized prostate cancer who are considering first-line definitive treatment, does the addition of a tissue-based genomic test to existing clinical risk models impact risk classification?

KQ2: Does tissue-based genomic testing impact the choice of treatment intensity or harms:

A. Among individuals with localized prostate cancer before first-line definitive treatment?

B. Among individuals who have undergone radical prostatectomy?

KQ3: Among patients with localized prostate cancer, what is the incremental prognostic effect of tissue-based genomic tests beyond existing prognostic clinical features on key clinical outcomes (*eg*, biochemical recurrence-free survival, metastases-free survival) following definitive treatment?

METHODS

Data Sources and Searches

The MEDLINE (via Ovid), Embase (via Elsevier), and Web of Science (via Clarivate) databases were searched from 2010 to the present using a combination of database-specific controlled vocabulary terms and keywords searched in the titles and abstracts related to prostate cancer and genomic tests. An experienced medical librarian devised and conducted the search, with input on keywords from the other authors. The search strategies were peer reviewed by another librarian using a modified PRESS checklist. The original searches were conducted on April 20, 2022. Case reports, editorials, letters, comments, and conference abstracts were excluded from the search, as were animal-only studies. The full, reproducible search strategies are in Appendix B.

Study Selection

Studies identified through our primary search were classified independently at both title and abstract and full text. All articles meeting our *a priori* eligibility criteria were included for data abstraction. We specifically sought to include articles evaluating these tests among patients with localized prostate cancer who are seeking first-line definitive treatment or those with localized prostate cancer who have undergone radical prostatectomy considering post-surgical treatment intensity. Articles addressing prognostic ability had to include either an established clinical feature–based prediction model or a minimum core set of clinical features (*eg*, prostate-specific antigen [PSA], Gleason score, and clinical tumor stage).

Data Abstraction and Assessment

Data from published reports were abstracted into a customized database by 1 reviewer and over-read by a second reviewer. Disagreements were resolved by consensus or by obtaining a third reviewer's opinion when consensus was not reached. We extracted data elements from included

studies, including descriptors to assess applicability, quality elements, population characteristics, intervention details, and outcomes including prognostic effect estimates and adverse events. Our extraction process was guided by the checklist for critical appraisal and data extraction for systematic reviews of prognostic factor studies (CHARMS-PF).

For risk of bias (ROB) assessment, we selected the appropriate tool relevant to the included study design. For studies solely or primarily relevant to KQ3 or prognostic outcomes, we used the Quality In Prognosis Studies (QUIPS) tool. For studies that did not otherwise address prognostic outcomes, we used RoB-2 for randomized trials and Risk of Bias in Nonrandomized Studies of Interventions (ROBINS-I)for observational designs. We completed ROB in duplicate for 25% of included studies. Because we found sufficient agreement, the remaining included studies were assessed for ROB by 1 investigator and overread by a second, given the volume of included studies for observational designs. Last, we audited ROB assessments for consistency across the included studies. Areas of concern for potential bias included exclusion of patient data due to inadequate tissue or incalculable test score, tests run by a lab other than the commercial lab for the specific test, inadequate adjustment to account for standard clinical risk assessment features, and consistency and clarity of source patient populations.

Synthesis

We summarized key study characteristics of the primary literature using data abstracted from eligible studies. When feasible based on the volume of relevant literature, types of effect measures reported, and completeness of results, we completed a quantitative synthesis (*ie,* meta-analysis) to estimate summary prognostic effects. Effect estimates were grouped by outcome, statistical effect measure, treatment status of the patient population (before or after first-line definitive treatment), genomic classifier test studied, and follow-up duration. We did not combine outcomes across the 3 types of commercial genomic classifier tests, as each test evaluates the expression profile of distinct gene panels that differ across the tests. Because the genomic classifier tests of interest can be reported as both a continuous variable and a categorical variable, we report both. Time-to-event outcomes are summarized as hazard ratios, and dichotomous outcomes are summarized with risk ratios or odds ratios. We also sought to identify studies reporting calibration (*eg,* O:E ratio) and discrimination (*eg,* c-statistic, AUC) statistics for models with and without the addition of a genomic classifier. Random-effects models were used, as was the Knapp-Hartung approach to adjust the standard errors of the estimated coefficients. We evaluated for statistical heterogeneity using visual inspection and 95% prediction intervals. When a quantitative synthesis was not feasible, we summarized the data narratively. A narrative synthesis focused on documenting and identifying patterns of effect of incremental benefit of genomic classifier tests after consideration of existing clinical prognostic factors.

Nominators for this review expressed an *a priori* interest in differences in the effect of these genomic classifier tests by key subpopulations, specifically race/ethnicity and risk classification at the time of genomic classifier test sample collection. Given that this is a patient-level characteristic, we sought to identify analyses conducted within the primary literature that identified effect modification (*eg,* subgroup analyses, regression model explanatory variables).

We assessed certainty of evidence using GRADE with consideration of guidance around adaptation for prognostic studies. We limited GRADE ratings to outcomes for KQ3 due to the volume and comparability of relevant studies.

RESULTS

Results of Literature Search

We identified 2,816 records through searches of MEDLINE (via Ovid), Embase (via Elsevier), and Web of Science (via Clarivate). An additional 5 articles were identified through hand searching and reviewing bibliographies of relevant review articles for a total of 2,821 articles. After removing duplicates, 1,573 articles remained. After applying inclusion and exclusion criteria to titles and abstracts, 145 articles remained for full-text review. Of these, 59 articles (55 unique studies) were included and retained for data abstraction. They consisted of 1 randomized controlled trial, 1 secondary analysis of a randomized controlled trial, 1 individual patient-level meta-analysis, 2 case-control studies, and 50 observational cohorts (8 prospective, 42 retrospective). Multiple studies had overlapping cohorts. There were 4 studies conducted solely in the VA; however, 7 additional studies included a VA cohort.

In the results section below for each KQ, we report results by genomic classifier test type in alphabetical order for consistency (*eg*, Decipher, Oncotype, Prolaris).

Summary of Results for Key Questions

KQ1

Eleven studies reported risk reclassification between baseline risk assessment with clinical features and after genomic classifier testing. Reclassification was reported either as a direct comparison of risk levels using the same risk labels or through integration of the genomic classifier test results into the clinical features risk assessment. Three studies used Decipher, 6 used Oncotype, and 2 used Prolaris. The majority of patients in these 11 studies fell into intermediate or lower baseline clinical risk classification, with only 3 studies including patients at high risk. Overall, the years over which the studies drew data ranged from 2012 to 2020. Most used NCCN clinical-based risk classification criteria as a comparator. Eight studies were found to have low ROB, 2 moderate ROB, and 2 high or serious ROB. Common sources of ROB for articles relevant to this KQ included missing patient-level data due to inadequate or poor-quality sample, lack of information about which patients were receiving the test, and inadequate control of confounders.

First, we considered reclassification among patients prior to definitive treatment. We found 1 large study (2 cohorts of 4,960 and 977 patients from radical prostatectomy and biopsy tissue, respectively) using Decipher that reported no change in risk classification in 21% and 24% of patients, a higher classification in 43% and 35%, and a lower classification in 36% and 41% when considering the 4-tier NCCN risk groups. Of note, reclassification using a novel six-tier system that incorporated both NCCN risk and genomic risk found a found greater proportion reclassified with no change in 33.3%, reclassification to a lower risk level in 27.7%, and to a higher level in 38.9%. One small study ($N = 57$) reported no change in risk classification in 50% of patients, a higher classification in 14%, and a lower classification in 10% (3% or 5% unknown). The 6 Oncotype-based studies included a total of 907 patients with NCCN risk very low to intermediate. The test did not change risk level in 37% to 81% of patients, reclassified to higher risk among 3.2% to 44%, and reclassified to a lower risk among 12% to 35.4%. One of the Oncotype studies was conducted with 177 VA patients from 6 facilities and found that 81% had no change in risk classification while 7% were adjusted higher and 12% lower. The 1 small

study ($N = 52$) using Prolaris had no change in 58% of participants, higher reclassification in 25%, and lower in 17%. There were minimal data examining rates of risk reclassification with genomic classifier testing at the time of prostatectomy.

We also considered reclassification across subgroups including by baseline risk and among minoritized racial groups. Eight studies provided data on reclassification by baseline risk (none used Prolaris). In general, the majority of patients undergoing risk assessment with evaluated genomic classifier tests did not change risk levels; however, this appears to be more consistent among patients at very low risk based on clinical features. Among intermediate-risk patients, Oncotype testing seems to more often reclassify risk to lower categories when reclassified. Specifically, among baseline intermediate-risk patients, Oncotype led to lower reclassification in 20% to 57.7% of cases, higher in 7% to 42%, and no change in 38% to 70%. Only the Decipher-based studies reported risk reclassification among high-risk patients. The 1 large Decipher-based study reported similar findings across the biopsy and prostatectomy cohorts with 16.15% and 21.0% unchanged, respectively, 64.4% and 51.85% higher (to very high), and 19.4% and 27.2% lower.

One study compared reclassification between Black and White men with prostate cancer ($N = 150$ and $N = 60$, respectively) in a single community urologic oncology practice. Overall, there was no change in risk for 43% of Black men compared with 40% of White men, while 33% of Black men were reassigned to a lower risk and 24% to a higher risk compared with 50% lower risk for White men and 10% higher risk.

KQ2

Fifteen studies addressed the impact of genomic classifier tests on treatment intensity recommended and/or received: 5 for Decipher, 7 for Oncotype, and 3 for Prolaris. The impact of Oncotype or Prolaris results on treatment recommendations was evaluated only prior to first-line definitive treatment, while Decipher was evaluated only after prostatectomy. Study designs addressing KQ2 included retrospective examination of documented treatment recommendations before and after receipt of test results, prospective collection of provider recommendations before and after test results, and deidentified case reviews by providers with and without test results. Of note, outcomes for impact on treatment are reported in multiple ways across studies, including overall change, increase/decrease in treatment recommendation, and change in specific treatment recommended (*eg*, active surveillance [AS]). The specific definition for treatment intensity varied by study such that some focused on any interventional treatment versus observation and others focused on specific types of definitive treatment. Across this group of 15 studies, 1 was found to have low ROB, 8 moderate ROB, 5 high/serious ROB, and 1 critical ROB. Common sources of potential bias include having the providers chose which patients for whom to order the test, reporting bias, outcome measurement approach, and missing data.

Across 14 observational studies at the time of diagnosis and after prostatectomy, there was a pattern of treatment recommendations changing after receipt of genomic classifier tests. However, in the single identified randomized trial which evaluated the impact of the incorporation of Oncotype test results into treatment decisions, there was no evidence of altered choice of treatment after receipt of the test. Based on studies using Prolaris or Oncotype, use of a genomic classifier test prior to definitive treatment led to a change 16% to 64.9% of the time, reduced intensity 4% to 40% of the time, and increased intensity 8.8% to 24.9% of the time. Post-prostatectomy clinical utility was reported 2 ways. First, the influence of the genomic

classifier test on treatment decision-making was estimated to have an odds ranging from 4.04 to 8.57 (based on high vs not-high risk test results, or per 5% increase in score). Odds of pursuing adjuvant radiation therapy ranged from 1.48 (per 5% increase in score) and 9.75 (high vs not-high), while odds of pursuing salvage radiation therapy were 1.30 (per 5% increase) and 8.02 (high vs not-high). Harm as a result of genomic classifier testing was not reported by any study.

Impact of receipt of genomic classifier testing was also reported for specific treatment choices. Across included observational studies, rates of recommending AS after an initial diagnosis were higher after receipt of genomic classifier (range 51%–89%), though there was no clear pattern for adjuvant treatment or surveillance after prostatectomy. There was also no clear pattern between clinical risk classification and the effect of genomic classifier testing on treatment recommendation at the time of diagnosis.

When considering the effect of race on clinical utility, the 1 randomized trial ($N = 191$) included a prespecified hypothesis that use of genomic testing would increase adoption of AS, including among Black men. They found no evidence of a statistically significant difference in AS uptake by race across the following groups. A second study using 2 retrospective cohorts from 6 VA health care systems found a 14% absolute difference in use of AS among black Veterans between untested (66%) and tested (80%) groups compared with an 11% absolute difference among White Veterans (61% vs 72%) with a p value of <0.01.

KQ3

Thirty-nine studies including more than 10,000 patients addressed the utility of adding or incorporating genomic classifiers into clinical risk classification schemes to enhance prognostic accuracy across multiple disease outcomes. There was substantial variability in the clinical risk classification employed, outcome definitions, and statistical measures used to assess the impact of genomic classifiers. Seven studies evaluated test prognostic ability in addition to NCCN risk classification, 22 to CAPRA or CAPRA-S, 1 to AUA, and 24 to a combination of clinical features unique to the study, with a plurality of studies reporting multiple comparisons across clinical risk classification schemes. Sixteen studies investigated biochemical recurrence, 20 the rate of metastases, and 10 prostate-cancer-specific mortality, all of which were retrospective in design. Five studies included composite endpoints, of which 2 were prospective and the remaining 3 retrospective. Twenty-two studies employed Decipher, 5 Oncotype, and 14 Prolaris, with 1 study investigating all 3 genomic classifiers. Twenty-four studies ran the genomic classifier on prostatectomy tissue, 20 on biopsy tissue, and 5 on a combination of the two. Twenty-six studies included patients diagnosed prior to 2000 and 9 included patients diagnosed prior to 1990. The majority of studies, 34, included patients who underwent prostatectomy as their initial treatment. Nine studies included patients who were treated with definitive radiation, with only 3 studies including patients that solely received definitive radiation. Two studies did not report the treatments received.

Eighteen studies were found to have low overall ROB, 11 moderate ROB, and 10 high ROB. Of note, 17 studies appear to have been sponsored or coauthored by the commercial companies with rights to the genomic classifier tests under study. Common causes of ROB among included studies for KQ3 include exclusion of potentially eligible participants due to insufficient tissue sample or tissue quality to run the genomic classifier test, exclusion of patients lost to follow-up or who might have had adverse outcomes in other health systems, inadequate adjusting for confounders in analysis; limited duration of follow-up, and lack of details about missing data.

Less common was having the genomic classifier test run by a lab other than the commercial lab for the specific test type.

Findings by Outcome and Test Type

Next, we describe the findings by outcome and then by genomic classifier test type. We focus on the effect estimates included in forest plots and meta-analyses when appropriate. Additional study findings not combined due to methods-based or conceptual heterogeneity, findings related to C-statistics, and uncommonly reported outcomes are described in detail in the full report.

Risk of biochemical recurrence

First, we considered the additive prognostic value of the tests for risk of biochemical recurrence. For Decipher, there was a 20% increase in the risk of BCR for patients with a higher Decipher score in models that included standard clinical classification factors with a summary estimate hazard ratio (HR) across 3 studies ($N = 445$) of 1.20 (95% CI [1.00, 1.43]; 95% prediction interval [PI] [1.00, 1.43]). One study included patients post-prostatectomy and 2 included patients after definitive radiation therapy. For Oncotype, the study-specific HR ranged from 1.10 (95% CI [1.10, 1.21) to 2.7 (1.84, 3.96) across 3 studies not combined in a meta-analysis due to underlying conceptual heterogeneity—2 of which had wide CIs and all of which were patients post-prostatectomy. For Prolaris, the summary estimate HR across 7 studies ($N = 2,186$) was 1.44 (95% CI [1.28, 1.62]; 95% PI [1.28, 1.62]). All but 1 of the 7 studies were post-prostatectomy, and 1 included patients treated with radiation therapy. Event rates across identified studies ranged from 15 to 193.

Risk of metastatic disease

Second, we considered the additive prognostic value of the tests for risk of metastases. For Decipher, there was a 32% increase in the risk of metastatic disease for patients with higher Decipher scores across 9 studies considering the test as a continuous outcome ($N = 2,139$); the summary HR was 1.32 (95% CI [1.22, 1.44]; 95% PI [1.15, 1.52]). Of these 9 studies, 4 included patients post-prostatectomy, 2 after radiation therapy, 1 after either radiation or prostatectomy, and 1 after a combination of therapies. When considered as a categorical outcome, the effect estimates were wide ranging and had very broad confidence intervals due to a limited number of metastatic events in the studies. There were only 2 studies that evaluated this outcome for the Oncotype test, and they reported greater risk estimates with the HRs from 2 contributing studies in patients post-prostatectomy ($N = 687$) being 2.24 (95% CI [1.49, 3.53]) and 2.34 (95% CI [1.42, 3.86]). For Prolaris, 1 study (n=582) that included patients post-prostatectomy and 2 that included patients post-prostatectomy or after radiation therapy with or without ADT reported HR ranging from 2.03 (95% CI [1.47, 2.78]) to 4.19 (95% CI [2.08, 8.45]).

Risk of prostate-cancer-specific mortality

We also considered the additive prognostic value of the tests for risk of prostate-cancer-specific mortality. Overall, fewer studies reported this outcome. Risk estimates were of similar magnitude and direction for Decipher and Prolaris, while in studies evaluating Oncotype reporting slightly greater hazard rations, wider confidence intervals were noted. Two studies ($N = 538$) examined the additive benefit of Decipher with this outcome among patients post-prostatectomy and reported HR of 1.81 (95% CI [1.48, 2.25]) and 1.39 (95% CI [1.20, 1.63]). Median duration of follow-up in these studies was 7 and 13 years. For Oncotype, 2 studies ($N =$

687) followed post-prostatectomy patients for a median of 9.8 and 15.5 years. They reported HRs of 2.30 (95% CI [1.45, 4.36]) and 2.69 (95% CI [1.50, 4.82]). Three Prolaris studies (*N* = 1,675) contributed to a meta-analysis of the additive prognostic effect and reported a summary HR of 1.72 (95% CI [1.58, 1.87]; 95% PI [1.58 to 1.87]). Duration of follow-up in these 3 studies ranged from 9.5 to 11.8 years.

DISCUSSION

Key Findings and Strength of Evidence

We evaluated the impact of 3 genomic classifier tests—Decipher, Oncotype, and Prolaris—on risk reclassification, treatment recommendations, and key clinical outcomes among patients with prostate cancer at the time of diagnosis and after definitive initial treatment. While there was a wide range of impacts on risk reclassification across studies, there was no clear pattern. We note that while there was no change in risk classification for a majority of patients, that still meant a clinically meaningful proportion across identified studies experienced a change that could influence important treatment changes. A lack of change was most consistent among patients at very low risk based on clinical features. With respect to the clinical utility of these tests, we found that providers do change their treatment recommendations after receipt of test results in observational studies, although note that these changes in practice occurred in the presence of a lack of established ability for these tests to predict treatment response. The single randomized trial identified did not find a statistically significant change in treatment after receipt of Oncotype test results. Evidence around clinical utility was distinct by test type and timeframe such that Oncotype and Prolaris were studied only at initial diagnosis and Decipher only after prostatectomy.

Overall, we found that these tests seem to provide modest additional prognostic information over existing clinical risk prediction schemas with respect to biochemical recurrence, development of metastatic disease, and prostate-cancer-specific mortality. For Decipher, which had the largest number of studies, we have low certainty of evidence (COE) that this test provides additional prognostic information for risk of biochemical recurrence, metastases, and prostate-cancer-specific mortality. For Oncotype and Prolaris, we have very low COE across all 3 outcomes. We note that while the effect estimates were consistent in showing a small, but potentially clinically relevant additive benefit of the genomic tests, our confidence assessments were frequently downgraded for issues related to indirectness, reflecting the era from which the data were drawn, imprecision of the estimates, and inconsistency.

Applicability

The evidence supporting the benefit of additional prognostic information afforded by genomic classifier tests in the context of prostate cancer is limited by the fact that these findings largely stem from patients diagnosed and treated prior to the current era of prostate cancer management defined by advanced screening practices as well as evolution in pathologic assessment, staging, and treatment modalities. Regarding relevance to the VA, several of the identified studies included VA-sourced data. Across all included studies, the patient populations were similar based on patient characteristics (*eg,* age, comorbidities) to the Veteran population such that these findings are expected to be generalizable to the VA clinical setting.

Future Research

To strengthen the body of evidence for the KQs outlined in this report, we suggest the following study design and analytic considerations. First, for studies addressing KQ3, prospective cohort studies with sufficient follow-up (*eg*, 15 years) would be ideal, while randomized trials would be preferred to determine if these tests are indeed predictive of treatment outcome. Greater certainty may be supported through additional studies evaluating Oncotype and Prolaris in patients after definitive treatment and evaluating Decipher in patients prior to definitive treatment. In addition, studies offering direct comparisons across these tests could inform determinations of comparative value. Finally, harms from use of these tests should be reported as an important outcome. Any future studies adding to this body of literature should provide explicit descriptions of the source of cohort data (especially when there is potential or apparent overlap across publications), outline attrition rates from cohort populations due to inadequate tissue samples or test results, and employ a standardized and broadly accepted set of core potential confounding measures for analytic adjustment.

Conclusions

Genomic classifier tests offer the potential to improve prognostic assessment for patients with prostate cancer and to provide critical information for patient-provider deliberations on key management decisions. While there is some evidence elucidating when such tests may lead to a change in risk classification and supporting tendency to spur a change in management, the key data needed to inform the value of these tests lies in their ability to accurately predict the risk of key long-term clinical outcomes that are relevant to patients. Definitive evidence of the prognostic ability of these tests is still needed from current management-era data. In the meantime, providers and their patients can take note that genomic classifier tests appear to provide some additional prognostic benefits that could offer value when treatment decisions are uncertain.