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.

TABLE OF CONTENTS

FIGURES AND TABLES

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

EVIDENCE REPORT

INTRODUCTION

Prostate cancer is the most common malignancy in men with an estimated 268,490 new cases in the United States in 2022, and approximately 70% of these patients presented with localized disease.[1] Prostate cancer is also the most common malignancy seen within the Veterans Health Administration (VHA), accounting for approximately 12,500 new diagnoses per year.[2] Risk stratification of these patients to help define prognosis and guide treatment has been traditionally based on clinical features including prostate specific antigen (PSA) level, tumor staging, and biopsy results.[3] However, multiple options exist for treatment, and there is variability in patient outcomes not otherwise explained by currently recognized risk factors. To improve the determination of individual risk of adverse clinical outcomes, tissue-based genomic classifier tests have been developed. The goal of these tests is to refine the current clinically based classification schema and inform personalized recommendations for treatment intensity. Depending on the point during the clinical course of prostate cancer at which the genomic classifier test is obtained, it can guide treatment intensity—such as the decision to pursue active surveillance, surgery, or radiation with or without hormonal therapy—or the timing and extent of adjuvant treatment following prostatectomy. Ultimately, the hope is for these tests to inform patient-physician decision-making, improve patient outcomes, and reduce overtreatment.

Tissue-based genomic classifiers have been developed with the goal of refining prognosis and personalizing treatment intensity. Three of the currently commercially available genomic classifier tests are Decipher, Oncotype (formerly Oncotype DX GPS, hereafter referred to as Oncotype), and Prolaris (Appendix A). Each test provides a score based on the expression of an empirically derived panel of genes in a patient's biopsy or prostatectomy specimen that can be used to estimate the risk of important clinical outcomes.[4-6] While large prospective studies are underway to assess the use of at least 1 of these studies (Decipher) to guide treatment intensity,[7,8] results are not likely to be available for a decade or more. In the meantime, a review of what is currently known about genomic classifier tests in localized prostate cancer is needed to inform interim guidance for clinical care.

This systematic review examines (1) whether adding Decipher, Oncotype, or Prolaris tests to existing clinical risk models changes a patient's risk classification, (2) how the use of these tests impacts treatment choice and if use of these tests causes harm, and (3) what prognostic value is offered by these tests beyond existing clinical risk models. We also consider available evidence on associations between test results (*ie*, risk classifications) and patient-important clinical outcomes, particularly survival.

METHODS

TOPIC DEVELOPMENT

This topic was developed at the request of the National Radiation Oncology Program and the National Oncology Prostate Cancer Clinical Pathways team. Findings from this review will be used to inform the multidisciplinary national clinical pathway for prostate cancer used by all providers who diagnose, treat, and manage prostate cancer patients in the VA.

KEY QUESTIONS

The following key questions (KQs) were the focus of this review:

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?

The conceptual framework developed to guide the approach to this review is in Appendix C.

PROTOCOL

A preregistered protocol for this review can be found on the PROSPERO international prospective register of systematic reviews (http://www.crd.york.ac.uk/PROSPERO/; registration number CRD42022347950).

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.[9] 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. In addition, we hand-searched previous systematic reviews conducted on this or a related topic for potential included studies.

STUDY SELECTION

Eligibility Criteria

Studies identified through our primary search were classified independently based on title and abstract by 2 investigators for relevance to the KQs from our *a priori* established eligibility criteria. All citations classified for inclusion by at least 1 investigator were reviewed at the full-text review level. The citations designated for exclusion by 1 investigator at the title and abstract level underwent screening by a second investigator. If both investigators agreed on exclusion, the study was excluded. All articles meeting eligibility criteria were included for data abstraction. All results were tracked in an electronic database (for referencing, EndNote, Clarivate Analytics,

Philadelphia, PA; for data abstraction, DistillerSR, Evidence Partners Inc., Manotick, ON, Canada).

Table 1 describes the study eligibility criteria organized by PICOTS elements (population, intervention, comparator, outcome, timing, setting) and other criteria such as study design, language, and publication type. Note that while the Oncotype genomic classifier test was acquired by MDxHealth and is now named Genomic Prostate Score (GPS), we use the name Oncotype throughout this report because most of the articles reviewed used the former name.

Table 1. Eligibility Criteria

	Inclusion	Exclusion
Population	KQ1, 2A: Patients with localized prostate cancer who are seeking first-line definitive treatment KQ2B: Patients who have localized prostate cancer who have undergone radical prostatectomy considering post-surgical treatment intensity KQ3: Patients who have localized prostate cancer who have undergone definitive radiation or surgery	Patients with metastatic (*de novo* or recurrent) prostate cancer with either distant or non-regional nodal metastases
Index prognostic factor	Tissue-based, multigene expression classifiers, specifically: • Decipher • Oncotype (now known as genomic prostate score or GPS) • Prolaris (*eg,* cell cycle progression [CCP] molecular score) Tissue upon which testing is run can be from a diagnostic biopsy or prostatectomy Diagnostic test does not need to be run at the time of tissue acquisition	• Germline genetic testing • Next-generation sequencing • Other gene signatures (*eg,* Post-Operative Radiation Therapy Outcomes Score [PORTOS])
Comparator prognostic factors	Clinical feature-based prediction models (*eg*, AUA/NCCN, CAPRA-S) Prediction models must include the following minimum core set of clinical features: PSA, Gleason score, and clinical tumor (T) stage	Prognostic factors not meeting the minimal core set
Outcomes	KQ1: Changes in risk classification or reclassification, difference in classification KQ2: Proportion choosing active surveillance, change in management/treatment decision-making, addition of ADT to definitive radiation, receipt of adjuvant radiation with or without ADT	• Adverse pathology at prostatectomy • Patient experience of treatment decision (*eg*, decision conflict)

	Inclusion	Exclusion
	harms (*eg,* complications from unnecessary treatment) KQ3[a]: Biochemical recurrence-free survival, metastasis-free survival, prostate-cancer-specific mortality, overall survival, time-on surveillance	
Timing	Time test is measured: At any time point in clinical management but run on tissue obtained at diagnostic biopsy or prostatectomy Time outcomes assessed: • KQ1: At the time of first-line definitive treatment determination • KQ2: At the time of relevant treatment decision (intensity) and at least 12 months after receipt of treatment (harms) • KQ3: At least 3 years after test measurement	
Setting	Any setting	
Study design	Randomized trials (including pilot studies), pre-post studies, retrospective/prospective cohort studies (including observational data drawn from randomized trials), case-control studies, individual meta-analysis	• Case studies/series • Systematic reviews
Publication types	Full publication in a peer-reviewed journal	Letters, opinion pieces, editorials, reviews, dissertations, meeting abstracts, protocols without results
Years	2010 to April 20, 2022	

Notes. [a] Given the relatively recent development of these tests, longer term clinical outcomes data, such as overall survival, may not be available. As such, we also considered intermediate outcomes such as biochemical recurrence and metastases-free survival.
Abbreviations. AUA=American Urological Association; KQ=key question; PSA=prostate-specific antigen.

DATA ABSTRACTION AND ASSESSMENT

Data from published reports were abstracted into a customized DistillerSR 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 CHARMS-PF (the checklist for critical appraisal and data extraction for systematic reviews of prognostic factor studies).[10] Key characteristics abstracted included participant descriptors (*eg,* age, race, clinical status), test type (*eg,* Decipher, Prolaris, Oncotype), comparator, prognostic effect estimates, and outcomes. We attempted to identify when a cohort of patients from an individual institution was included in multiple analyses. When identified, we prioritized data from the larger or more inclusive cohort analysis.

In accordance with our *a priori* plan, we prioritized randomized trials, prospective cohort studies, cohorts with longer follow-up duration (>5 years), nested case-control studies, and validation or confirmatory studies over training cohorts or data used to establish a test, given the volume of identified literature. While cost outcomes were not relevant to the determination of eligibility, we had an *a priori* interest in cost outcomes related to the use of genomic classifier tests in the identified studies and these outcomes were collected when available.

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 QUIPS because it is a validated tool to assess quality in prognostic factor studies.[11] Domains included in QUIPS include study participation, study attrition, prognostic factor measurement, outcome measurement, adjustment for other prognostic factors, and statistical analysis and reporting. Based on a previously published approach, any study that was rated high in 1 or more domains or moderate for 3 or more was considered high ROB overall. Any study that was rated low ROB in all 6 domains or up to 1 moderate ROB was considered low ROB overall.[12] Studies that did not meet either of those conditions were considered moderate ROB overall. For KQ1 and KQ2, studies that did not otherwise address KQ3 as a prognostic outcomes study, we used ROB-2[13] for randomized trials and ROBINS-I for observational designs.[14] These criteria included adequacy of randomization and allocation concealment, comparability of groups at baseline, blinding, completeness of follow-up and differential loss to follow-up, whether incomplete data were addressed appropriately, validity of outcome measures, protection against contamination, selective outcomes reporting, and conflict of interest. We assigned a summary ROB rating to individual studies.

We paid attention to particular areas of potential bias among identified studies including the following: (1) exclusion of patient data due to inadequate tissue or incalculable test score beyond 20% to 30%, which we considered average sample loss; (2) evidence that tests were run by a lab other than the commercial lab for the specific test, as the tests themselves are proprietary and a non-commercial lab would not be able to provide the equivalent testing results; (3) inadequate adjustment for basic clinical features considered standard for clinical risk assessment, and (4) clarity and consistency in the selection of patient populations for analysis, especially when data were drawn from multiple health care systems or clinics.

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 over-read by a second given the volume of included studies for observational designs. Lastly, we audited ROB assessments for consistency across the included studies. These criteria included adequacy of randomization and allocation concealment, comparability of groups at baseline, blinding, completeness of follow-up and differential loss to follow-up, whether incomplete data were addressed appropriately, validity of outcome measures, protection against contamination, selective outcomes reporting, and conflict of interest. We assigned a summary ROB score to individual studies. 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 over-read by a second given the volume of included studies.

SYNTHESIS

We summarized the primary literature using data abstracted from eligible studies. Summary tables describe key study characteristics of the primary studies (*eg,* study design, patient demographics, and genomic test). We determined the feasibility of completing a quantitative synthesis (*ie,* meta-analysis) to estimate summary prognostic effects; feasibility decisions were based on the volume of relevant literature, types of effect measures reported, and completeness of results. We also considered similarity in patient treatment status (before or after first-line definitive treatment), test studied, follow-up duration, and definitions of risk classifications. 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, details of which are not publicly available, and therefore cannot be determined to be equivalent. Conceptual consistency was primarily based on clinical contextualization and consideration for which types of patient cohorts were similar enough for a combined analysis (*eg*, those who had received or not received definitive initial treatment for prostate cancer). For each key outcome, we reviewed operationalized definitions of the outcome variable and combined those that were sufficiently conceptually similar after discussion with clinical experts on our team (MB, DC).

For KQ1, we prioritized data reported at the individual level (*ie*, reporting of specific number or percentage of patients that changed from 1 specific risk level to another risk level). We summarized no change, higher, and lower reclassification across reported levels (*eg*, % changed from low to favorable intermediate plus, % changed from unfavorable intermediate to high for reclassification to a higher level). When individual-level data were not provided, we report summary % reclassification as given in the primary article. For KQ2, we similarly categorized treatment intensification versus de-intensification based on how it was reported by the included studies.

For KQ 3, we considered a minimum set of established prognostic factors (*ie,* PSA, Gleason score, and clinical tumor [T] stage) for adequate adjustment. However, some otherwise eligible studies did not include T stage as a component of the regression model. Studies were not excluded for this reason. Because the genomic classifier tests of interest can be reported as both a continuous variable and a categorical variable, we report both. We only aggregated outcomes when there were at least 3 studies with the same outcome, based on the rationale that 2 studies do not provide adequate evidence for summary effects. We grouped outcomes by time point of outcome measurement (*eg,* 3–5 years after test measurement vs 6–10 years) due to the current understanding of the natural history of prostate cancer. Specifically, we did not combine discrimination statistics (*eg*, AUC, c-index) at 5 and 10 years, as the natural history of prostate cancer is slow and there is no clear existence of a natural plateau for outcomes that would support combining such data. For time-to-metastasis or metastasis-free survival, we combined studies defining this outcome by distant and/or regional metastases. This decision was driven by the recognition that while the location of metastases can drive treatment decisions, attention to the location of metastases has evolved over time, including during the time span of many of the included studies. We have noted the definitions for key outcomes in our study characteristics table (Appendix E).

For survival and other time-to-event outcomes, we abstracted hazard ratios and corresponding 95% confidence intervals (CIs). We also abstracted calibration (*eg,* O:E ratio) and discrimination (*eg,* c-statistic, AUC) statistics for models with and without the addition of a genomic classifier.

Given the KQs guiding this review, we only abstracted adjusted prognostic effect estimates with the most adjusted analyses prioritized.

When studies reported multiple models using different approaches to adjusting for clinical risk factors (*eg*, NCCN vs CAPRA risk stratification models, individual clinical risk factors), we prioritized the use of models as follows:

- For patient cohorts with an intact prostate and who had *not* received definitive therapy, we prioritized models using NCCN risk categorization, followed by CAPRA, and then individual clinical features.
- For patients who received radical prostatectomy, we prioritized CAPRA-S, followed by models with individual clinical risk factors.

Random-effects models were used for meta-analyses, and for analyses with fewer than 20 studies, we used the Knapp-Hartung approach to better account for uncertainty in estimates of the amount of heterogeneity among studies.

We evaluated heterogeneity using visual inspection of forest plots and 95% prediction intervals. When prediction intervals encompassed values that substantially differed from the effect estimate (in magnitude, direction, or both), we concluded that there was substantial heterogeneity present. Potential sources of heterogeneity we planned to investigate included case-mix variation, study characteristics (*eg*, follow-up time), analytic approach, and risk of bias.

When a quantitative synthesis was not feasible, we summarized the data narratively. We gave more weight to the evidence from higher quality studies with more precise estimates of effect. 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. We analyzed potential reasons for inconsistency in prognostic effects across studies by evaluating differences in the study population, clinical status, comparator, and timing and definition of outcome variables.

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 were unable to consider meta-regression to explore quantitative or qualitative interactions of prespecified, potential effect modifiers including duration of follow-up due to lack of individual patient data. In addition, we narratively considered the representation of subgroups within identified studies in comparison with the VA population.

We assessed certainty of evidence using GRADE with consideration of guidance around adaptation for prognostic studies.[15] We limited GRADE ratings to outcomes for KQ3 due to the volume and comparability of relevant studies. In brief, this approach requires assessment of 4 domains: ROB, consistency, directness, and precision. Additional domains to be used when appropriate are dose-response association, impact of plausible residual confounders, strength of association (magnitude of effect), and publication bias. We assigned a summary rating of high, moderate, low, or very low strength of evidence based on consensus among 3 investigators (KG,

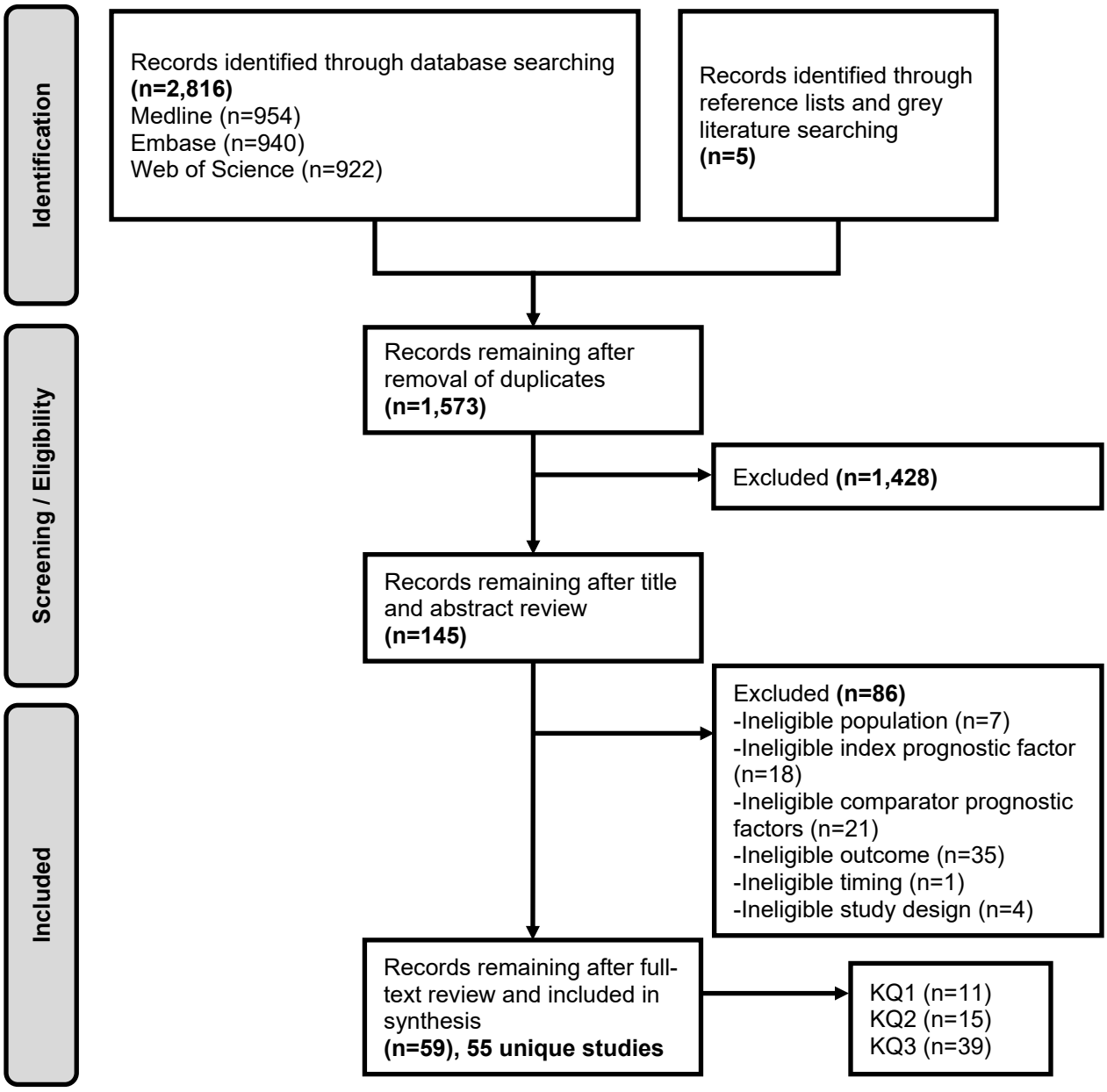
MB, AG). We did not downgrade for observational study designs as suggested for prognostic GRADE assessment.[15] Studies that included patient data from the 1980s or early 1990s were downgraded for indirectness because patients in these studies have limited comparability with patients receiving modern cancer screening and treatment.

RESULTS

LITERATURE FLOW

The literature flow diagram (Figure 1) summarizes the results of the study selection process (full list of excluded studies is in Appendix D).

Figure 1. Literature Flowchart



LITERATURE OVERVIEW

We identified 2,816 records through searches of MEDLINE (via Ovid), Embase, and Web of Science (see Figure 2). 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, there were 1,573 articles remaining in total. After applying inclusion and exclusion criteria to titles and abstracts, 145 articles remained for full-text review. Of these, 59 were included and retained for data abstraction. Of the 59 articles, 55 were identified as unique studies. They consisted of 1 randomized controlled trial, 1 secondary analysis of a randomized trial, 1 individual patient-level meta-analysis, 2 case-control studies, and 50 observational cohorts (8 prospective, 42 retrospective). The studies were conducted across North America, Europe, and Asia (United States, Britain, Finland, France, Germany, Italy, and China). Multiple studies had overlapping cohorts and it was not always possible to determine the exact degree of overlap. There were 4 studies conducted exclusively in the VA; however, 7 additional studies included a VA cohort.

Of the 55 studies, 11 related to KQ1, 15 to KQ2, and 39 to KQ3. Several studies had outcomes related to more than one KQ. For details of study characteristics, see Appendix E.

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

KEY QUESTION 1: 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?

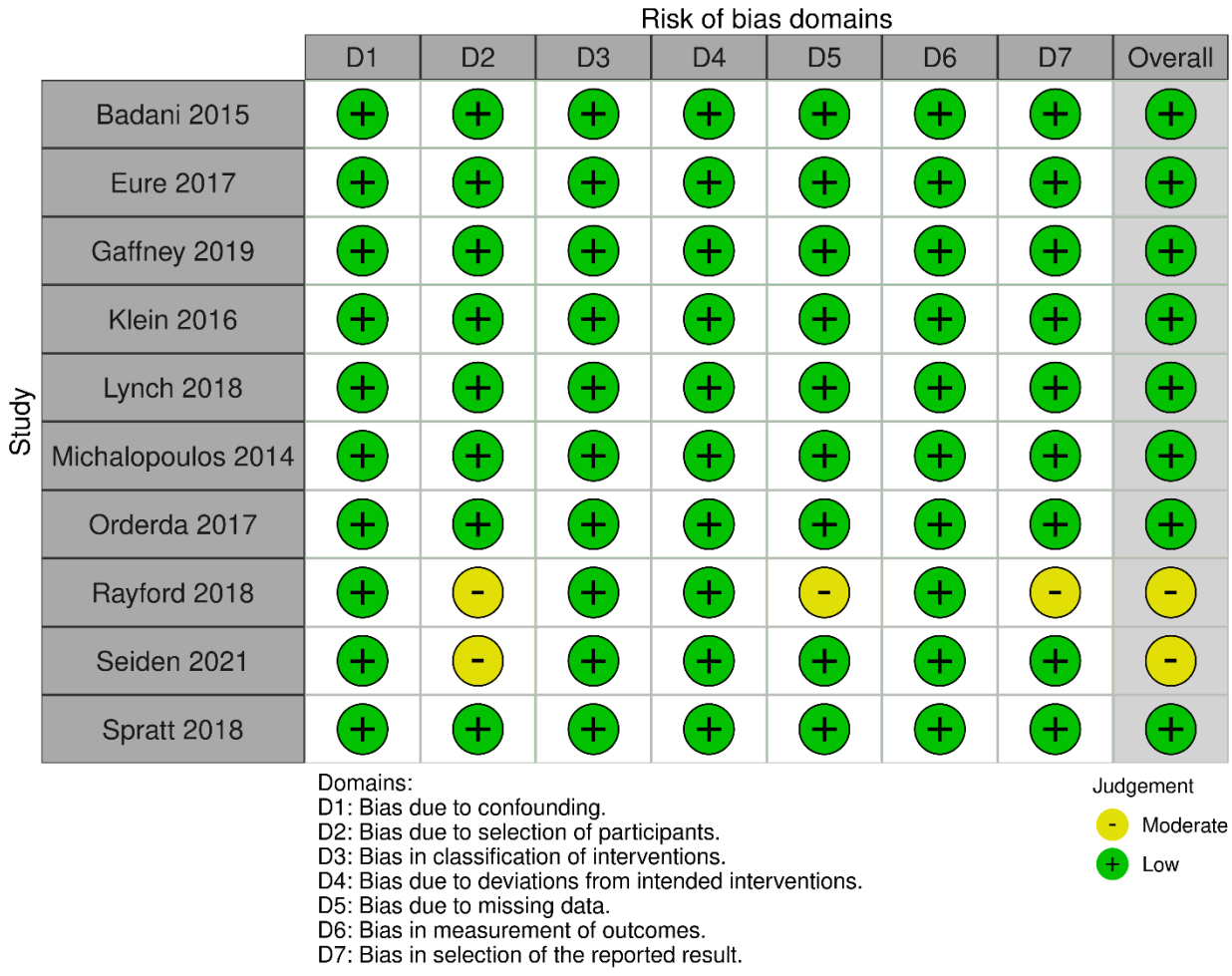
Key Points

- The majority of identified studies (6 of 11) that evaluated risk reclassification used the Oncotype genomic classifier run on biopsy specimens prior to definitive treatment.
- Prior to definitive treatment, Oncotype results did not change the clinical risk level in 37% to 81% of patients, reclassified 3.2% to 44% of patients to a higher risk, and reclassified 12% to 35.4% of patients to a lower risk. For Decipher, 1 large study reported no change in classification in 24% of patients, a higher classification in 35%, and a lower classification in 41%. The 1 small study using Prolaris had no change in classification of 58% of participants, higher reclassification in 25%, and lower in 17%.
- Among baseline intermediate-risk patients, Oncotype led to a lower reclassification in 4% to 28.8% of cases, higher classification in 2% to 69%, and no change in classification in 15% to 96%.
- Few studies examined differences in risk reclassification among minoritized racial groups.

Overall, 11 studies reported risk reclassification from baseline risk assessment with clinical features to genomic classifier testing.[16-26] Reclassification was reported either as a direct comparison of risk levels using the same risk labels or through integration of the genomic

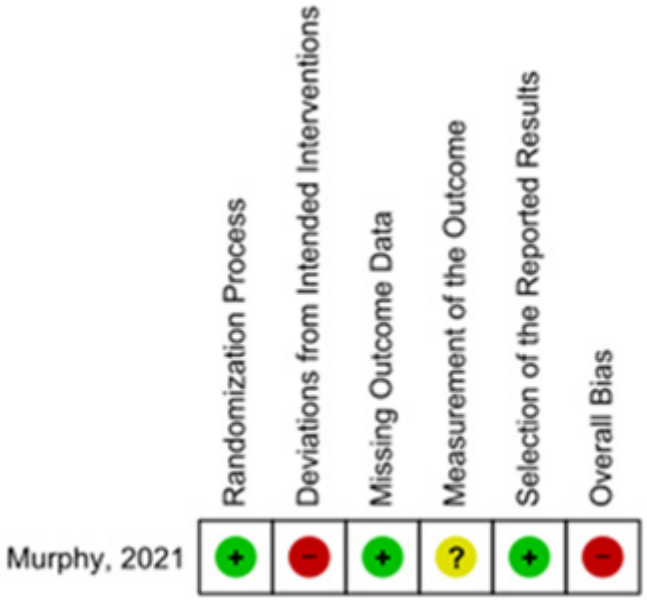
classifier test results into the clinical features risk assessment. One study reported risk reclassification among intervention arm patients in a randomized trial,[23] 4 in prospective cohort studies,[19,20,24,26] 5 in retrospective cohort studies,[16,17,21,22,25] and 1 in an ambidirectional study (a retrospective and prospective cohort).[18] Three studies used Decipher,[19,22,26] 6 used Oncotype,[16-18,20,23,24] and 2 used Prolaris.[21,25] 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.[19,21,22] Overall, among studies reporting time of data collection, the years over which the prospective studies were conducted ranged from 2012–2014, the retrospective studies drew data from years 2013–2020, and the ambidirectional study from 2013–2014 for the retrospective component and 2015–2016 for the prospective component. For clinical-based risk classification, 8 studies used NCCN criteria,[16-20,22-24] 1 AUA,[25] 1 EUA,[21] and 1 CAPRA-S.[26] Eight studies were found to have low ROB,[17-22,24,26] 2 moderate ROB,[16,25] and 2 high/serious ROB (designation term depending on instrument used).[18,23] 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 (Figures 2 and 3).

Figure 2. ROBINS-I Risk of Bias Assessment for KQ1 Studies



Notes. Some of these studies are included for both KQ1 and KQ2. ROB assessment may differ by outcome.

Figure 3. Risk of Bias Assessment for the Randomized Trial



Notes. Green plus denotes low risk of bias; red dash denotes high risk of bias; yellow question mark denotes some concerns about risk of bias.

Risk Reclassification Among Patients Prior to Definitive Treatment

Decipher

We identified 2 studies that examined the change in risk classification following Decipher testing (Figure 4 and Table 2).[19,22] One low ROB study[19] included data from 2 prospective cohorts. The first cohort included 4,960 patients diagnosed between 2014 and 2016 with NCCN clinical risk classification ranging from low to very high and who had a Decipher test run on radical prostatectomy tissue. The second cohort included 977 patients diagnosed in 2016 with a similar range of NCCN clinical risk classification and who underwent Decipher testing on pre-treatment biopsy tissue specimens. Reclassification results were similar across the 2 included cohorts comparing risk from the 4-tier NCCN risk groups (*ie*, low, favorable intermediate, unfavorable intermediate, high) to a 6-tier risk spectrum combining both clinical and genomic information, with 21% and 24% of cases without a classification change, 43% and 35% with a higher reclassification, and 36% and 41% with a lower reclassification. Of note, reclassification using a novel six-tier system that incorporated both NCCN risk and genomic risk found a 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%. A smaller low ROB retrospective study[22] including 57 men with predominantly low or intermediate risk disease by clinical assessment found half of cases had no change in risk classification while 30% were reclassified to a lower risk and 14% higher.

Oncotype

Six studies[16-18,20,23,24] with a total of 907 patients evaluated changes in risk classification after Oncotype genomic testing (Table 2). Identified studies included patients with NCCN classification of very low to intermediate risk. Overall, the median proportion without a change in risk classification was 69% (range from 37% to 81%), median percent reclassified to a higher

risk was 6% (range from 3.2% to 44%), and median percent reclassified to a lower risk was 18 (range from 12% to 35.4%).

The 1 randomized trial[23] examined change in risk among the 104 patients assigned to the intervention arm who had undergone Oncotype testing on biopsy specimens and who had baseline NCCN risk evenly distributed from very low to low intermediate risk (note that low intermediate was defined as favorable intermediate risk but excluding patients with Grade Group 2 and more than 3 positive scores and including patients with a PSA of 10–20 ng/mL if the PSA density was less than 0.15). Overall, in this study 40% of patients had no change in risk, 43% were reclassified as higher risk, and 17% were lower.

Two prospective studies evaluated Oncotype effect of risk reclassification, and each found that the majority of patients had no change in risk classification. First, a low ROB prospective study[24] of 158 patients from 3 urologic practices with risk assessment before and after testing reported no change in 61.4% with 3.2% reclassified lower and 35.4% higher. Second, a low ROB study[20] prospectively collected data on 258 newly diagnosed patients after institutional availability of genomic testing with Oncotype. The majority of patients (77.5%) had no change in risk classification while 5.8% were adjusted higher and 16.7% lower.

The other 3 Oncotype studies were retrospective. One study[17] evaluated 134 patients at a single institution in a low ROB study and found that the majority of patients had no change in classification and more patients were reclassified to a lower risk than changed to a higher risk category. One small, moderate ROB, single-institution cohort study of 63 patients[16] found 37% had no change in risk classification. Finally, 1 study[18] evaluated risk reclassification among 190 newly diagnosed Veterans from 6 VA facilities after Oncotype testing became available and found that 81% had no change in risk classification while 7% were adjusted higher and 12% lower.

Prolaris

One low ROB study[21] evaluated a small retrospective cohort study with 52 newly diagnosed patients from 2 academic hospitals who had Prolaris tests run on biopsy specimen (Table 2). Of this cohort with 29% of patients classified as high risk by European Association of Urology criteria, 58% were unchanged, 25% changed to a higher risk classification, and 17% lower.

Figure 4. Median Percent Reclassification by Genomic Classifier Test Type

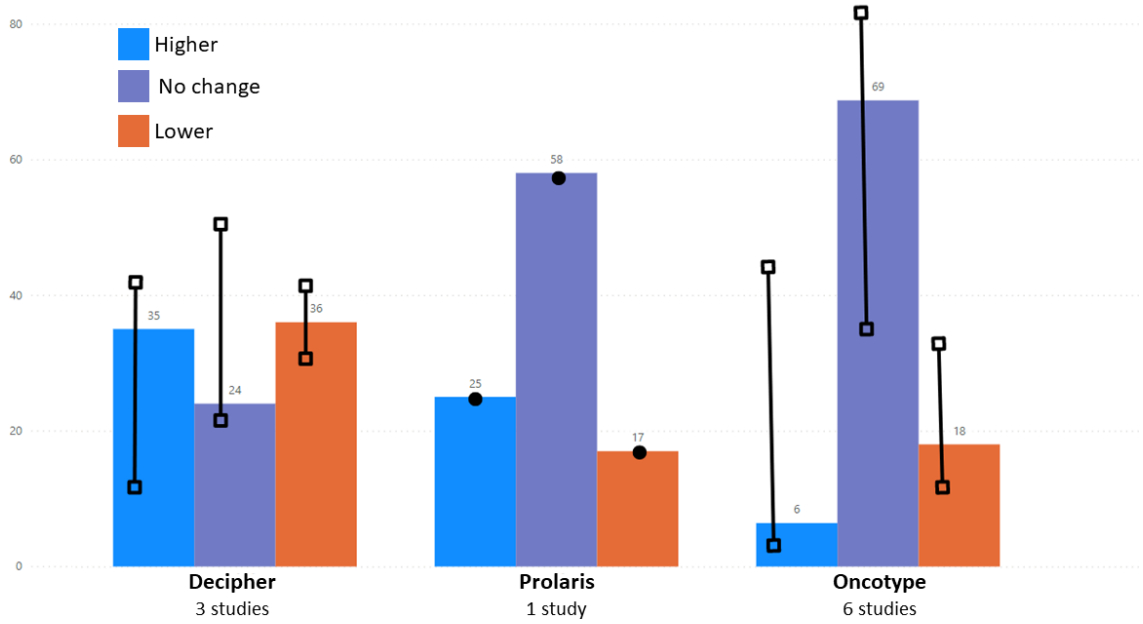


Table 2. Change in Risk Classification with Genomic Classification Testing Among Men Prior to Definitive Treatment

Study Total N Design Tissue Source	Clinical Characteristics Treatment Median Test Value	Pre-test Risk Classification	No Change in Risk N (%)	Reclassified Higher N (%)	Reclassified Lower N (%)
Decipher					
Spratt, 2018[19] [a,b] 4,960 Prospective cohort I RP specimen	PSA <200 Stage T1c, T3b, clinical N0 Median GC: NR	NCCN: Low: 203 (4.1%) Favorable int: 948 (19.1%) Unfavorable int: 634 (12.8%) High/very high: 3,145 (64.0%)	21%	43%	36%
Spratt, 2018[19] [a,b] 977 Prospective cohort II Biopsy specimen	PSA <200 Stage T1c, T3b, clinical N0 Median GC: NR	NCCN: Low: 315 (32.2) Favorable int: 198 (20.3) Unfavorable int: 284 (29.1) High/very high:180 (19.4)	24%	35%	41%
Klein, 2016[22] [c] 57 Retrospective, single institution Biopsy specimen	PSA >20 Stage pT3 or positive margin or GG ≥8 All later underwent RP Median GC: 0.38 (IQR 0.29 to 0.49)	NCCN: Low: 23 (40.4%) Int: 27 (47.4%) High: 4 (7.0%) Unknown: 3 (5.3%)	51%	14%	30%
Oncotype					
Murphy, 2021[23] 191 patients (104 intervention with test) ENACT Randomized trial Biopsy specimen	Newly diagnosed mean GPS within NCCN groups: very low 5 26.9, low 5 27.2, low intermediate 5 32.4	NCCN: Very low: 40 (38.5%) Low: 36 (34.6%) Low-int: 28 (26.9%)	40%	43%	17%

Study **Total N** **Design** **Tissue Source**	**Clinical Characteristics** **Treatment** **Median Test Value**	**Pre-test Risk Classification**	**No Change in Risk** **N (%)**	**Reclassified Higher** **N (%)**	**Reclassified Lower** **N (%)**
Lynch, 2018[18] 190 tested patients Comparative cohort before/after test availability, 6 VAMCs Biopsy specimen	Newly diagnosed Median GPS: 26.5 (range 0 to 61)	NCCN: Very low: 42 (22%) Low: 81 (43%) Int: 67 (35%)	81%	7%	12%
Seiden, 2021[16] 63 men managed with AS Retrospective, single Institution Biopsy specimen	GG 6,7 Managed with AS Median GPS 25 (IQR 19 to 4)	NCCN: Very low: 7 (11%) Low: 24 (38%) Favorable int: 31 (49%) Unfavorable int: 1 (2%)	37%	44%	19%
Eure, 2017[20] 258 (posttest) Comparative cohort before (retrospective) and after (prospective) institutional testing Biopsy specimen	Newly diagnosed Median GPS: NR	NCCN: Very low: 68 (26.4%) Low: 111 (43.0%) Int: 79 (30.6%)	77.5%	5.8%	16.7%
Badani, 2015b[24] 158, 3 urology practices Prospective before and after test (own patients) Biopsy specimen	Newly diagnosed Median GPS score: NR	NCCN: Very low: 35 (22.2) Low: 71 (44.9) Low-int: 52 (32.9)	61.4%	3.2%	35.4%
Gaffney, 2019[17] 134 Retrospective, single institution Biopsy specimen	All patients with GPS GG 6 = 87 (65%) 7 = 47 (35%) 32 later underwent definitive treatment	NCCN: Very low: 31 (23.1%) Low: 45 (33.6%) Int: 58 (43.3%)	76%	4%	20%

Study Total N Design Tissue Source	Clinical Characteristics Treatment Median Test Value	Pre-test Risk Classification	No Change in Risk N (%)	Reclassified Higher N (%)	Reclassified Lower N (%)
Prolaris					
Oderda, 2016[21] 52 Retrospective cohort, 2 academic hospitals Biopsy specimen	Newly diagnosed (All later underwent RP) Median CCP -0.15 (-1.7 to 1.4)	EAU: Low: 13 (25%) Int: 24 (46%) High: 15 (29%)	58%	25%	17%

Notes. [a] Spratt, 2018 included data for 2 separate cohorts; [b] Clinical data were pre-treatment; [c] Reclassification was unknown for 3 patients.
Abbreviations. CCP=cell cycle progression; EAU=European Association of Urology; GG=grade group; GPS=genomic prostate score; Int=intermediate; RP=radical prostatectomy.

Subgroups of Interest

By Baseline Risk Category

Eight studies provided, to varying extents, risk reclassification rates stratified by clinical risk assessment (Table 3).[16-18,20,22-24,26] No Prolaris studies included relevant subgroup analyses. In general, the majority of patients undergoing risk assessment with 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.

Of the 5 studies that reported changes among patients with a baseline clinical risk classification of *very low*, all used Oncotype testing and NCCN criteria for risk classification.[17,18,20,23,24] Sixty-six percent of patients identified to be at very low risk in the randomized trial remained at the same risk classification after testing,[23] while the other 4 observational studies found 88% to 100% remained at very low risk with and without the test.

Eight studies reported changes in risk classification among patients at baseline *low risk*, 6 with Oncotype,[16-18,20,23,24] and 2 with Decipher (3 cohorts total).[19,22] Among the Oncotype studies, 38% to 70% of patients were not reclassified, 4.2% to 29% were reclassified to a higher risk category, and 20% to 57.7% were reclassified to a lower risk category. One Decipher-based study reported 2 cohorts, 1 using biopsy specimen ($N = 4960$) and 1 using prostatectomy tissue ($N = 977$)[19]; among baseline low-risk patients 25.7% and 19.7% remained the same, 25.7% and 12.8% were reclassified higher, and 57.1% and 67.5% lower. The Decipher-based study of 57 patients[22] found 86.9% of patients were not reclassified, 13.1% reclassified as higher risk, and none were reclassified as lower risk.

The same 6 Oncotype-based studies reported reclassification among patients at baseline *intermediate* risk. Three studies[17,18,20] ($N = 895$) found no change among 78% to 96% of patients, reclassification to a higher risk level among 0% to 2%, and reclassification to a lower risk level among 4% to 21%. Two studies reported changes among patients defined as low-intermediate risk (defined above) at baseline; 38% of the 104 patients at low-intermediate risk in the randomized trial[23] reported no change. Twenty-nine percent of patients were reclassified to a higher risk level and 32% to lower risk level. The retrospective cohort[24] found no change among 71.2% of 158 patients, 0% higher, and 28.8% lower. In this study, low intermediate risk was defined as favorable intermediate risk but with 3 or fewer positive cores (or 33% or fewer) if Gleason Grade Group 2. The small retrospective cohort of 63 patients[16] reported that 65% of the 31 patients at baseline favorable intermediate risk were reclassified higher. The study with 2 cohorts reported changes across both favorable-intermediate and unfavorable-intermediate risk categories. For the favorable-intermediate group, 28.3% (biopsy cohort) and 21.0% (prostatectomy cohort) experienced no change, 27.3% and 22.3% higher, and 44.4% and 56.8% lower. For the unfavorable-intermediate individuals, 24.6% and 23.8% were not reclassified, 40.5% and 36.0% were higher, and 34.9% and 40.2% were lower. The other small Decipher-based study[22] reported risk reclassification among baseline intermediate risk patients with 9 of 27 (33.3%) having no change, 18.5% reclassified higher, and 48.1% lower.

Only the Decipher-based studies reported risk reclassification among high-risk patients. The large study with both a radical prostatectomy and biopsy tissue cohort reported similar findings.[19] Across both cohorts, 16.2% and 21.0% had no change in risk classification when incorporating genomic classifier risk scores, 64.4% and 51.85% higher (to a risk category of very

high), and 19.4% and 27.2% lower. The 4 patients in the study by Klein et al[22] who were classified as high risk at baseline were moved to a lower category.

Table 3. Reclassification by Baseline Clinical Risk Levels Prior to Definitive Treatment

Study Total N Design Tissue Source	Clinical Characteristics Treatment Median Test Value	Very Low	Low	Intermediate	High
Decipher					
Spratt, 2018[19] [a] 4,960 Prospective cohort I Radical prostatectomy	PSA <200 Stage T1c, T3b, clinical N0 Median GC: NR	—	No change: 40/203 (19.7%) Higher: 26/203 (12.8%) Lower: 137/203 (67.5%)	*Favorable Int* No change: 199/948 (21.0%) Higher: 211/948 (22.3%) Lower: 538/948 (56.8%) *Unfavorable Int* No change: 151/634 (23.8%) Higher: 228/634 (36.0%) Lower: 255/634 (40.2%)	No change: 668/3175 (21.0%) Higher: 1644/3175 (51.8%) Lower: 863/3175 (27.2%)
Spratt, 2018[19] [a] 977 Prospective cohort II Biopsy specimen	PSA <200 Stage cT1c-T3bN0 Median GC: NR	—	No change: 81/315 (25.7%) Higher: 54/315 (17.1%) Lower: 180/315 (57.1%)	*Favorable Int* No change: 56/198 (28.3%) Higher: 54/198 (27.3%) Lower: 88/198 (44.4%) *Unfavorable Int* No change: 70/284 (24.6%) Higher: 115/284 (40.5%)	No change: 29/180 (16.1%) Higher: 116/180 (64.4%) Lower: 35/180 (19.4%)

Study **Total N** **Design** **Tissue Source**	**Clinical Characteristics** **Treatment** **Median Test Value**	**Very Low**	**Low**	**Intermediate**	**High**
				Lower: 99/284 (34.9%)	
Klein, 2016[22] 57 Retrospective, single institution Biopsy specimen	PSA >20 Stage pT3 or positive margin or GG ≥8 All later underwent RP Median GC: 0.38 (IQR 0.29 to 0.49)	—	No change: 20 (86.9%) Higher: 3 (13.1%) Lower: none	No change: 9 (33.3%) Higher: 5 (18.5%) Lower: 13 (48.1%)	No change: 0 Higher: 0 Lower: 4 (100%)
Oncotype					
Murphy, 2021[23] 191 patients (104 intervention with test) ENACT Randomized trial Biopsy specimen	Newly diagnosed Median GPS: NR	No change: 19/29 (66%) Higher: 10 (34%) Lower: NA	No change: 13 (38%) Higher: 10 (29%) Lower: 11 (32%)	*Low-Int* No change: 4 (15%) Higher: 18 (69%) Lower: 4 (15%)	—
Lynch, 2018[18] 190 tested patients Comparative cohort before/after test availability, 6 VAMCs Biopsy specimen	Newly diagnosed Median GPS: 26.5 (range 0 to 61)	No change: 37 (88%) Higher: 5 (12%) Lower: N/A	No change: 57 (70%) Higher: 8 (10%) Lower: 16 (20%)	No change: 59 (88%) Higher: 0 Lower: 8 (12%)	—
Seiden, 2021[16] 63 men managed with AS Retrospective, single Institution Biopsy specimen	GG 6,7 Managed with AS Median GPS: 25 (IQR 19 to 4)	—	Higher: 7/24 (29%)	*Fav–Int* Higher: 20/31 (65%)	—

Study **Total N** **Design** **Tissue Source**	**Clinical Characteristics** **Treatment** **Median Test Value**	**Very Low**	**Low**	**Intermediate**	**High**
Eure, 2017[20] 258 (post test) Comparative cohort before (retrospective) and after (prospective) institutional testing Biopsy specimen	Newly diagnosed Median GPS: NR	No change: 61 (90%) Higher: 7 (10%) Lower: NA	No change: 63 (57%) Higher: 8 (7%) Lower: 40 (36%)	No change: 76 (96%) Higher: 0 Lower: 3 (4%)	—
Badani, 2015b[24] 158, 3 urology practices Prospective before and after test (own patients) Biopsy specimen	Newly diagnosed Median GPS: NR	N = 35 No change: 33 (94.3%) Higher: 2 (5.7%) Lower: NA	N = 71 No change: 27 (38.0%) Higher: 3 (4.2%) Lower: 41 (57.7%)	*Low-int*, N = 52 No change: 37 (71.2%) Higher: 0 Lower: 15 (28.8%)	—
Gaffney, 2019[17] 134 Retrospective, single institution Biopsy specimen	All patients with GPS GG 6 = 87 (65%) 7 = 47 (35%) 32 later underwent definitive treatment	No change: 31/31 (100%)	No change: 26/45 (58%) Higher: 4/45 (9%) Lower: 15/45 (33%)	No change: 45/58 (78%) Higher: 1/58 (2%) Lower: 12/58 (21%)	—
Prolaris (No Studies)					

Notes. [a] Spratt, 2018, included data for 2 separate cohorts.
Abbreviations. AS=active surveillance; Fav=favorable; GG=grade group; GPS=genomic prostate score; Int=intermediate; IQR=interquartile range; PSA=prostate-specific antigen.

By Race

Three studies included a focus on the effect of genomic testing on risk reclassification among a historically minoritized population, specifically Black men. A randomized trial that included a majority of Black men (140 of 200 participants or 70%) did not report reclassification by race after Oncotype testing.[23] A second moderate ROB study from a single-community urologic oncology practice[25] compared the impact of Prolaris test results on risk classification between 150 Black and 60 White men with prostate cancer. 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% higher compared with 50% lower and 10% higher for White men. Finally, a retrospective, single-institution cohort study[16] included 63 participants (all Black) (results in Table 3) and found 37% without a change in risk, 44% higher, and 19% lower after Oncotype testing.

Risk Reclassification Among Patients at the Time of Prostatectomy

One low ROB study (PRO-ACT)[26] assessed the impact of Decipher on community-based urologist adjuvant treatment decisions among high risk patients after radical prostatectomy, but also noted that among those tested, 65.4% CAPRA-S intermediate patients were reclassified to a low-risk level. No other studies evaluated the effect of the genomic classifier tests of interest on risk reclassification after prostatectomy.

KEY QUESTION 2: DOES TISSUE-BASED GENOMIC TESTING IMPACT THE CHOICE OF TREATMENT INTENSITY OR HARMS AMONG A) INDIVIDUALS WITH LOCALIZED PROSTATE CANCER BEFORE FIRST-LINE DEFINITIVE TREATMENT OR B) INDIVIDUALS WHO HAVE UNDERGONE RADICAL PROSTATECTOMY?

Key Points

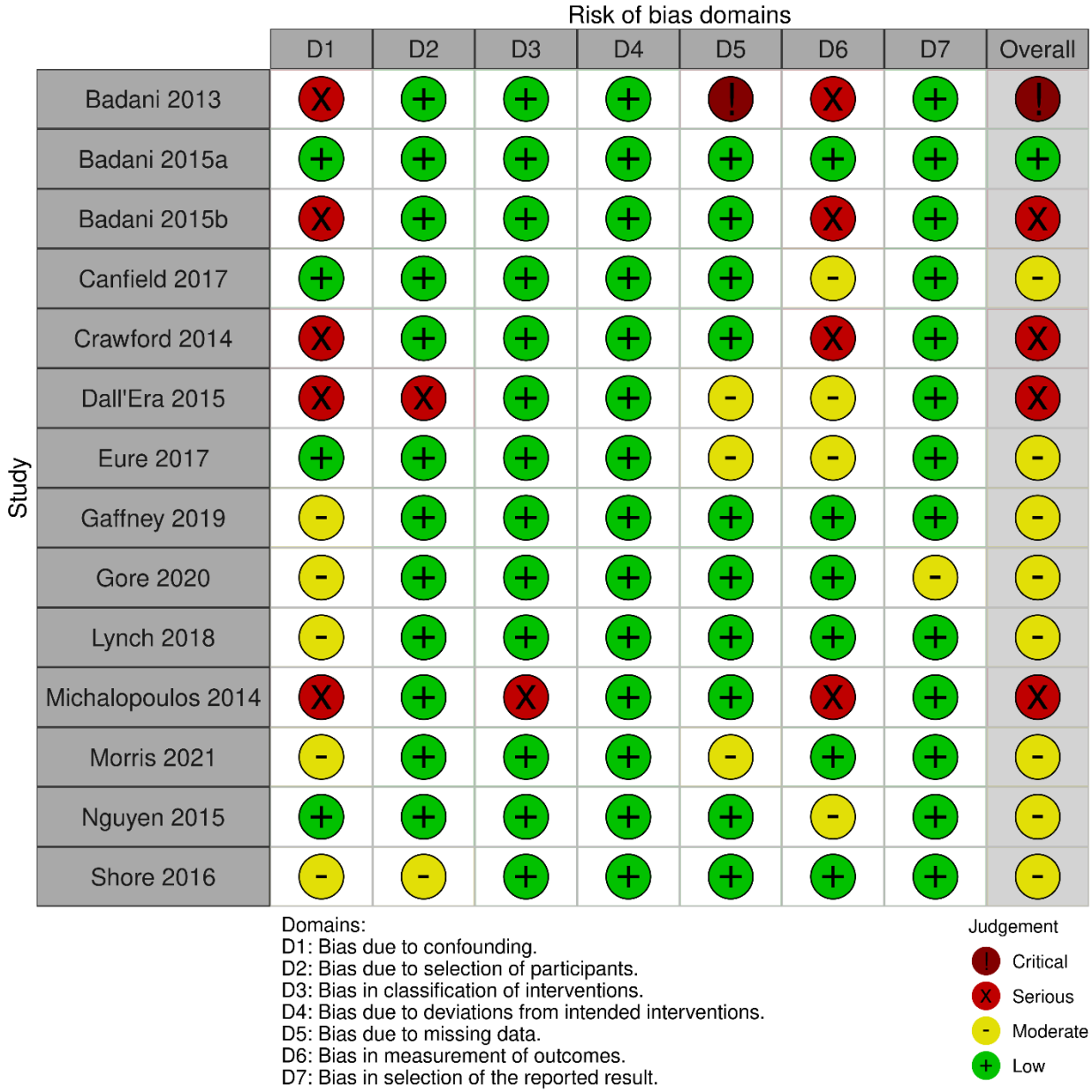
- Across 14 observational studies at the time of diagnosis and after prostatectomy, there was a pattern of changes to treatment recommendations after receipt of genomic classifier tests. The only available randomized trial, which incorporated Oncotype test results into treatment decisions, found no evidence of altered choice of treatment after receipt of the test.
- The impact of Oncotype or Prolaris results on treatment recommendations was evaluated only prior to first-line definitive treatment, while Decipher was used only evaluated after prostatectomy.
- Across the observational studies, rates of recommending active surveillance after an initial diagnosis were higher with genomic classifier use, though there was no clear pattern of adjuvant treatment or surveillance after prostatectomy based on test use.
- There was no clear pattern of treatment recommendations at the time of diagnosis attributable to genomic classifier testing.
- Patients classified as higher risk by Decipher at the time of prostatectomy were less likely to receive a recommendation of surveillance.

- Significant limitations of the evidence were that many patients were diagnosed and treated prior to the current era of prostate cancer management, and harms due to genomic classifier testing were not reported by any study.

Fifteen studies addressed the clinical utility, or impact of genomic classifier tests, on treatment intensity recommended and/or received: 5 for Decipher,[26-30] 7 for Oncotype,[17,18,20,23,24,31,32] and 3 for Prolaris.[33-35] For Decipher-based studies, all considered treatment intensity after prostatectomy. For Oncotype- and Prolaris-based studies, treatment intensity determination was after biopsy and prior to first-line definitive treatment. Study designs addressing this key question included retrospective examinations 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. For each specific genomic classifier test, at least 1 study considered treatment recommendations within risk-based subgroups. Two studies[17,22] reported subgroup analysis by race/ethnicity, both of which used Oncotype.[18,23] 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). The specific definition for treatment intensity varied by study such that some focused any interventional treatment versus observation[26,29,30,33,34] and others employed a more nuanced approach to treatment.[18,24,35] Across this group of 15 studies, 1 was found to have low ROB,[29] 8 moderate ROB,[17,18,20,27,28,31,33,35] 5 high/serious ROB,[23,24,26,32,34] and 1 critical ROB.[30] Common sources of potential bias include selection bias from providers choosing to order the test for included patients, reporting bias, outcome measurement approach, and missing data (Figure 5). Results for studies addressing this key question are listed in Table 4.

Next, we describe findings by tests used in the pre-definitive treatment setting followed by post-prostatectomy according to classifier test.

Figure 5. ROBINS-I Risk of Bias Assessment KQ2 Studies



Notes. Some of these studies are included for both KQ1 and KQ2. ROB assessment may differ by outcome.

Overall Treatment Recommendations

Pre-definitive First-line Treatment

Decipher

No studies evaluated the impact of Decipher test use on first-line treatment decisions.

Oncotype

Two studies evaluated the effect of the Oncotype genomic classifier test on newly diagnosed patients with very low to intermediate risk disease based on the NCCN classification.[18,24] First, a

prospective study[24] enrolled 158 newly diagnosed men from 3 clinical sites and collected urologist treatment recommendations before and after receiving of Oncotype test results. Overall, 26% of patients received a change of treatment recommendation after receipt of test results, with 16% of a lower intensity and 9% of a higher intensity, while 75% of treatment recommendations were unchanged. A second study with moderate ROB[18] compared patient management strategies across 6 VA medical centers among a retrospective cohort of 200 patients treated prior to the introduction of Oncotype testing and a separate prospective cohort of 190 patients with similar prostate cancer who agreed to undergo genomic classification testing. Among patients who received Oncotype test results, 16% received any change in treatment recommendation, with 4% and 12% receiving a decreased and increased treatment intensity recommendation, respectively (Table 4).

Prolaris

Three studies reported on the impact of the Prolaris genomic classifier tests on overall treatment recommendations[33-35]: 2 prospectively evaluated provider recommendations before and after receipt of test results,[34,35] and 1 compared cohorts of patients before and after the test became available as part of routine practice.[33] In a large, prospective registry (PROCEDE-1000)[35] of 1,206 patients with newly diagnosed prostate cancer whose providers (N = 124 from 21 states) completed 4 sequential questionnaires up to 6 months after ordering the test, nearly half of patients (95% CI [45.0, 50.6]) received a different treatment after Prolaris testing compared with the documented pre-test treatment recommendation. Of those who received a different treatment than initially recommended, 72.1% were a decrease in intensity versus 26.9% that increased. In addition, 17.6% of patients (213 of 1,206) experienced a change in treatment modality between non-interventional to interventional. The second prospective study[34] enrolled 305 patients at the time Prolaris was ordered on biopsy specimens and compared treatment recommendations at the time of test ordering to recommendations reported after test results were received. Overall, in a majority of cases providers reported a change in treatment recommendation from before to after receiving Prolaris test results (64.9%, 95% CI [59.4, 70.1%]), with 40% receiving a decreased treatment intensity recommendation and 24.9% an increased intensity recommendation. Finally, a retrospective study evaluated treatment recommendations for 150 men from a single urologic practice before and after the institutional availability of routine Prolaris testing.[33] Among men with Gleason 6 or 7 newly diagnosed prostate cancer, every unit increase in CCP score (Prolaris) had a greater odds of selecting definitive treatment than active surveillance (OR = 2.09, 95% CI [1.16, 3.94]); in comparison, each unit increase in CAPRA score had OR = 1.29, 95% CI [1.01, 1.66]).

Table 4. Impact on Overall Treatment Intensity by Genomic Classifier Test: Pre-definitive Treatment

Study Setting Total N Design	Population Clinical Characteristics	Any Change in Recommendations Pre- to Post-test				Other	Impact on Treatment Recommendation Change (OR, 95% CI)
Decipher (No Studies)							
Oncotype							
Lynch, 2018[18] 6 VAMCs 190 tested patients Comparative cohort before-after test availability	Newly diagnosed NCCN very low, low, and intermediate risk Median GPS: 26.5 (range: 0 to 61)	16%	4%	84%	12%	—	—
Badani, 2015b[24] 1 academic and 2 community-based urology practices 158 patients [a] Prospective pre-post treatment recommendation	Newly diagnosed NCCN very low, low, and intermediate risk Median GPS: 21 (IQR 13 to 32)	25%	15.8%	75.3%	8.8%	—	—
Prolaris							
Shore, 2016[35] 21 states 1206 patients 124 providers Prospective registry before/after test	Newly diagnosed Stage T1c: 72.1% Mean CCP: -0.7 (range -2.8 to 2.0)	47.8% (95% CI [45.0, 50.6)	Of those with change: 72.1%	—	Of those with change: 26.9%	—	OR = 1.027 change in treatment per 1 unit increase CCP score

Study Setting Total N Design	**Population Clinical Characteristics**	**Any Change in Recommendations Pre- to Post-test**	⬇	⊘	⬆	**Other**	**Impact on Treatment Recommendation Change (OR, 95% CI)**
Crawford, 2014[34] US-based practices 305 patients Prospective pre/post-test result	New diagnosis AUA risk: Low: 44.3% Inter: 42.9% High: 12.8% Median CCP: -0.71 (SD 0.83)	64.9% (95% CI [59.4%, 70.1%])	40%	35.1%	24.9%	—	—
Morris, 2021[33] 150 Retrospective comparative cohort before/after initiation testing protocol	New diagnosis Gleason on biopsy 6 or 7 Median CCP: -0.5 (IQR -0.9 to 0.0)	—	—	—	—	—	OR = 2.08 (1.16 to 3.94) per unit increase CCP score

Notes. [a] 175 enrolled, but 158 with evaluable data.
Abbreviations. CCP=cell cycle progression; IQR=interquartile range.

Post-prostatectomy

Decipher

Five studies (6 publications) were identified that addressed the impact of the Decipher genomic classifier test on treatment recommendations after prostatectomy (Table 5).[26-30,36] Two prospective studies (3 publications) evaluated the impact of Decipher genomic classifier test results on treatment recommendations before and after receipt of test results among providers in the context of treating their own patients.[26,27,36] Two articles[27,36] published results on analyses from the moderate ROB PRO-IMPACT Trial. PRO-IMPACT was a prospective study across 19 sites that evaluated treatment recommendations at the time the Decipher test was ordered versus after the results were received and then again 12 months after enrollment.[27,36] Among 242 patients eligible for either ART (non-organ-confined disease) or SRT (PSA ≥0.2 ng/ml post radical prostatectomy), treatment recommendations changed after Decipher testing for 17% and 30% of patients in the ART and SRT cohorts, respectively. Patients categorized as high risk had significantly greater odds of receiving ART at 12 months follow-up (OR = 2.99, 95% CI [1.3, 6.9]) and similarly higher odds of SRT (OR = 3.13, 95% CI [1.4, 7.1]) compared with patients categorized as not-high risk. Michalopoulos et al (PRO-ACT) similarly recruited 15 community-based urologists to provide treatment recommendations with and without Decipher results in the context of caring for patients after radical prostatectomy and who were eligible for adjuvant therapy.[26] In this serious ROB study, physician participants submitted a median of 6 patient tests. In approximately one-third of cases (30.8, 95% CI [23%, 39%]), urologists changed their treatment recommendation after Decipher test results were reviewed. This was also measured as the odds of a change in treatment recommendation with a Decipher test (OR = 4.04, 95% CI [2.36, 6.92]).

Three studies asked practicing radiation oncologists and/or urologists to provide treatment recommendations based on deidentified patient cases with and without GC test results.[28-30] The ASSESS-D study recruited 51 urologists from the AUA directory and had them review treatment plans with and without Decipher test results for 110 case histories from patients post radical prostatectomy who had an undetectable PSA (10 per urologist).[29] Thirty-one percent of case reviews had a change in treatment recommendation after genomic classifier test results. Those with a high genomic classifier score (vs low) had a significantly higher odds of experiencing a change in treatment recommendation (OR = 8.57, 95% CI [5.27, 14.26]). Nguyen et al evaluated differences in the impact of a genomic classifier test on post-prostatectomy adjuvant treatment recommendations between urologists and radiation oncologists in a moderate ROB study.[28] Decipher test results led to a change in treatment recommendations 35% of the time for radiation oncologists and 45% for urologists. Patients with high-risk Decipher results cared for by radiation oncologists had a significantly higher odds of any change in treatment recommendation than unchanged recommendations before to after testing (OR = 4.17, 95% CI [2.26, 7.70]); in comparison, urologist recommendations for patients with high-risk Decipher results had an even higher odds of any change in treatment recommendation (OR = 6.51, 95% CI [4.29, 9.88]). Finally, in a study at high ROB, the DECIDE study group[30] asked 21 urologists to give treatment recommendations after reviewing pathology reports with and without genomic classifier test results from deidentified high-risk, post-radical prostatectomy (RP) cases. Twelve cases involved patients undergoing consideration for adjuvant treatment and 12 for salvage treatment (each case was reviewed by multiple physicians). Overall, 43% (95% CI [37%, 49%]) of ART case reviews experienced a change in treatment recommendation, with the Decipher test results with 27%

(95% CI [19%, 35%]) having a decrease and 37% (95% CI [28%, 46%]) experiencing an increase in treatment intensification. Among SRT case reviews, 53% experienced changes in recommendations (95% CI [45%, 60%]) with 16% (95% CI [11%, 23%]) decreasing intensity and 61% increasing treatment intensity (95% CI [42%, 78%]).

Oncotype

No studies evaluated the impact of Oncotype test use on post-prostatectomy treatment decisions.

Prolaris

No studies evaluated the impact of Prolaris test use on post-prostatectomy treatment decisions.

Table 5. Impact on Overall Treatment Intensity by Genomic Classifier Test: Post-prostatectomy

Study Total N Design	Patient Characteristics	Impact on Treatment Decision	Other Outcomes
Decipher			
Gore, 2017[36] PRO-IMPACT study (X results) 19 sites 265 patients Prospective before-after test (own patients)	Post radical prostatectomy; non-organ confined prostate cancer or positive surgical margins (ART); or PSA increase or BCR [a] (SRT) Median GC score: ART group: 6.2% (IQR 0.5 to 44.2) SRT group: 6.5% (IQR 0.5 to 62.8)	Overall changes in treatment decisions after test results: ART: 27/150; 18% (95% CI [12%, 25%]) SRT: 37/115; 32% (95% CI [24%, 42%])	Odds decision to pursue ART/SRT after test results per 5% increase in GC score: ART: OR = 1.48 (95% CI [1.19, 1.85]; p < 0.001) SRT: OR = 1.30 (98% CI [1.03, 1.65]; p = 0.03)
Gore, 2020[27] PRO-IMPACT study 12-month follow up 19 sites 246 patients Prospective before-after test (own patients)	Post radical prostatectomy; non-organ confined prostate cancer or positive surgical margins (ART); or PSA increase or BCR [a] (SRT) Median GC score: ART group: 6.2% (IQR 0.5 to 44.2) SRT group: 6.5% (IQR 0.5 to 62.8)	Change from recommended pre-test to treatment administered at 12 months: ART: 31/140; 22% (95% CI [16%, 30%]) SRT: 25/106; 24% (95% CI [16%, 33%])	ART Treatment decision change with GC score high vs not-high risk: OR = 9.75 (95% CI [3.3, 28.0]; p < 0.001) Before test to 12 months follow-up (treatment received): OR = 2.99 (95% CI [1.3, 6.9]; p = 0.01) SRT; treatment decision change with GC score high vs not-high risk: OR = 8.02 (95% CI [2.9, 22]; p < 0.001) Before test to 12 months follow up (treatment received) OR = 3.13 (95% CI [1.4, 7.1]; p = 0.006)
Michalopoulos, 2014[26] 2014 PRO-ACT study	Post radical prostatectomy with T3 disease or positive SM	Change from treatment recommended before to after GC results:	Influence on treatment decision-making per 5 unit increase in GC score:

Study Total N Design	Patient Characteristics	Impact on Treatment Decision	Other Outcomes
15 urologists 146 patients Prospective before-after test (own patients)	Median GC 4.2% (range 1.3% to 41.5%)	30.8% (95% CI [23%, 89%]) Any treatment to observation: 42.5% (95% CI [27%, 59%]) Observation to any treatment: 17.6% (11% to 26%)	OR = 4.04 (95% CI [2.36, 6.92]; p < 0.0001)
Badani, 2015a[29] 110 cases 51 urologists Deidentified case history review with and without test	Post radical prostatectomy with undetectable PSA Median GC 3.85 (min, max: 1.2, 33.4) % reclassified: NR	Change from treatment recommended before to after GC results: 31% (95% CI [27%, 35%]) Change from any treatment to observation: 38% (95% CI [32%, 45%]) Change from observation to any treatment: 16% (95% CI [12%, 20%])	Impact on treatment decision-making MVA with GC score low vs high: OR = 8.57 (95% CI [5.27, 14.26]; p < 0.001)
Badani, 2013[30] 12 patient cases (ART) 12 patient cases (SRT) 21 urologists from 18 sites Deidentified case history review with and without test	Post radical prostatectomy with adverse pathology Median GC: NR	Change from treatment recommended before to after GC results: ART: 43% (95% CI [37%, 49%]) SRT: 53% (95% CI [45%, 60%])	Change from treatment to observation (decrease): ART: 27% (95% CI [19%, 35%]) SRT: 16% (95% CI [11%, 23%]) Change from observation to treatment (increase): ART: 37% (95% CI [28%, 46%]) SRT: 61% (95% CI [42%, 78%])
Nguyen, 2015[28] 11 patient cases 26 radiation oncologists	Post prostatectomy	Change from treatment recommended before to after GC results: Radiation oncologists: 35%	Impact on treatment recommendations MVA with GC high vs not high): Radiation oncologists OR = 4.17; (95% CI [2.26, 7.70])

Study Total N Design	Patient Characteristics	Impact on Treatment Decision	Other Outcomes
20 urologists Deidentified case history review with and without test	D'Amico risk (n): 2 – Low 4 – Intermediate 5 – High Median GC: NR	Urologists: 45%	Urologists: OR = 6.51; (95% CI [4.29, 9.88]
Oncotype (No Studies)			
Prolaris (No Studies)			

Notes. [a] BCR defined as PSA ≥0.2 ng/ml with a confirmatory reading.
Abbreviations. ART=adjuvant radiotherapy; GC=genomic classifier; MVA=Multivariate Analysis; OR=odds ratio; SRT=salvage radiotherapy.

Specific Treatment Choice

Pre-definitive Treatment

Decipher

No studies evaluated the impact of Decipher test results on specific treatment choice among patients prior to first-line definitive treatment.

Oncotype

We identified 6 studies (total 10,338 patients), including 1 randomized trial,[23] 1 prospective before-and-after receipt of test study,[24] and 4 retrospective or combination retrospective/prospective comparative cohort studies (Table 6).[18,20,31,32] All studies examined the occurrence of active surveillance among men with newly diagnosed prostate cancer who had intermediate or low-risk disease. In all the observational studies, rates of active surveillance were higher among those patients who had received oncotype test results; however, in the 1 trial, the rate of active surveillance was lower in the study arm that received the oncotype test results.

The only randomized trial (Engaging Newly Diagnosed Men about Cancer treatment options, ENACT)[23] evaluated the impact of receiving genomic testing on treatment decisions; specifically, 200 patients across 3 institutions (including 1 VA) with newly diagnosed NCCN very low- to intermediate-risk prostate cancer were randomized to receive standard counseling with or without the results of Oncotype genomic testing and were then evaluated the effect on uptake of AS. Eighty-eight percent of patients in the arm without the Oncotype test were recommended to receive active surveillance at the second treatment visits compared to 77% in the arm that did receive the Oncotype test. Overall, the intention-to-treat analysis of the initial choice of treatment (AS, surgery or radiation, undecided) found no significant difference between groups ($p = 0.067$), though there was a lower odds of AS use with testing though likely underpowered (OR = 0.49; 95% CI [0.22, 1.09]). Of note, 38 (43%) participants who underwent Oncotype testing were moved to a higher-risk group and 15 (17%) were moved to a lower-risk group.

One Oncotype-based prospective study enrolled 158 newly diagnosed men from 3 clinical sites with NCCN very low-, low-, or low-intermediate-risk prostate cancer and collected urologist treatment recommendations before and after receiving test results.[24] They reported a 24% relative increase in AS recommendations after review of test results from 41% to 51%. Twenty-four of 38 patients with risk reclassification with Oncotype result had a change in treatment recommendation, all concordant with Oncotype test results. Another observational study reported a pre-specified interim analysis comparing treatment recommendation patterns before (retrospective by chart review) and after the introduction of genomic classifier testing (prospective) with Oncotype at multiple U.S. community-based urology practices.[20] Across all risk categories, more patients were managed with AS in the tested group (74%) versus the untested group (62%), while persistence of AS at 1 year was 55% among those who were tested versus 34% who had not undergone testing (relative difference = 62%). A large retrospective cohort study used the Optum research database to identify 8920 men with low-risk cancer and reviewed their electronic medical records and administrative claims data from 2013 to 2016.[31] They considered the role of Oncotype testing and/or MRI imaging (for our purposes, we only considered those who had undergone Oncotype testing [$N = 300$] versus those who had neither testing nor MRI [$N = 7446$]). Of those who had undergone testing with Oncotype at 6 months,

89% had no observed therapy (labeled as AS) versus 84% at 12 months; in comparison, at 6 months those without testing or MRI only 60% had no observed therapy versus 56% at 12 months. Overall, patients who had Oncotype testing had a 31.2% higher occurrence of no observed therapy than those without (95% CI [22.6, 39.7]). A VA-based study compared retrospective data on patient management strategies among 200 patients from 6 medical centers treated prior to the introduction of Oncotype testing to a prospective cohort of 190 patients who agreed to undergo genomic classification testing.[18] Sixty-two percent of untested patients and 74% of tested patients were on active surveillance at 6 months (12% absolute difference; 19% relative difference). Of note, 26 of patients who received Oncotype testing had known Agent Orange exposure. Finally, a retrospective cohort study in which 15 urologists who had ordered at least 4 Oncotype tests provided data on 87 patients previously treated without genomic testing and 124 patients with Oncotype testing, all of whom had either Gleason score 3+3 or low volume 3+4 disease.[32] They reported an absolute increase of 11% in the use of AS with patients who received Oncotype testing versus those who did not (61 vs 50%; p =0.110) and a similar difference in AS received (67% vs 43%).

Prolaris

No tests evaluated the impact of Prolaris test results on specific treatment choice among patients prior to first-line definitive treatment.

Table 6. Impact on Active Surveillance Use by Genomic Classifier Test

Study Total N Design	Patient Characteristics	Active Surveillance: Without Test	Active Surveillance: With Test	Other Outcomes
Decipher (No Studies)				
Oncotype				
Murphy, 2021[23] ENACT study 191 patients Randomized trial	Newly diagnosed NCCN favorable intermediate or below Median GPS score: NR	At 2nd treatment visit: 88%	At 2nd treatment visit: 77%	Odds of choosing AS with test vs without: OR = 0.49 (95% CI [0.22, 1.09])
Badani, 2015[24] 158 patients 3 clinical sites Prospective before-after test (own patients)	Newly diagnosed NCCN low-intermediate or lower Median GPS: 21 (IQR range: 13, 32) 38 patients (24%) with risk reclassification post-GPS	At time test ordered: 41%	After receipt of test results: 51%	—
Eure, 2017[20] 247 (before) 258 (after) Comparative cohort before (retrospective) and after (prospective) institutional testing	Newly diagnosed NCCN intermediate or lower Median GPS: NR 23% pts had risk reclassification	40%	62%	AS persistence at 1 year: With GPS test: 89% Without GPS test: 86%
Lynch, 2018[18] 200 (2013-2014) 190 (2015-2016) 6 VA HCS Retrospective cohorts before and after institutional testing	Newly diagnosed NCCN intermediate risk or lower Median GPS (tested): 26.5 (range 0-61)	At 6 months: 62%	At 6 months: 74%	—

Study Total N Design	Patient Characteristics	Active Surveillance: Without Test	Active Surveillance: With Test	Other Outcomes
	Risk reclassification: Lower: 12% Higher: 7%			
Canfield, 2017[31] 300 GPS only 7446 no GPS/MRI Retrospective, comparative cohort before-after testing availability	Newly diagnosed AUA low risk Median GPS: NR Reclassification: NR	No observed therapy At 6 months: 60% At 12 months: 56%	No observed therapy At 6 months: 89% At 12 months: 84%	At 6 months – 31% higher use of no observed therapy among tested than non-tested (22.6, 39.7; p<0.001)
Dall'Era, 2015[32] 87 without testing 124 with testing 15 urologists Retrospective	Newly diagnosed NCCN Intermediate risk or lower Median GPS: NR % reclassification: NR	Recommended [a]: 30/60 = 50% Received: 43%	Recommended: 69/114 = 61% Received: 67%	Absolute increase in recommended AS: 11% Relative increase in AS recommendation: 22% Absolute increase in AS received: 14% Relative increase in AS received: 56%
Prolaris (No Studies)				

Notes. [a] Active surveillance (AS) defined as "active surveillance" or "watchful waiting" as noted in patient's medical record.

Post-radical Prostatectomy

Decipher

Two studies (3 publications) evaluated the impact of Decipher test results on specific treatment recommendations (Table 7).[27,29,36] As noted above, Gore et al published 2 studies from 1 trial that enrolled patients who were eligible for both ART and SRT.[27,36] They found that 88.7% of the ART patients were recommended to have observation before Decipher test results compared to 79% after and from 60% to 51% SRT patients recommended for observation with Decipher test results. The second study found the same percentage receiving a recommendation for observation before to after test results were received, although, as noted above, an overall change in treatment recommendations of 30.8%.[29] Of note, 42.5% who were initially recommended to receive ART were changed to observation.

Table 7. Impact on Observation After Prostatectomy by Genomic Classifier Test

Study Total N Design	Clinical Characteristics	Before Test Results	After Test Results
Decipher			
Gore, 2017[36] PRO-IMPACT 19 sites 265 patients Prospective before-after test (own patients)	Post radical prostatectomy; non-organ confined prostate cancer or positive surgical margins (ART); or PSA increase or BCR [a] (SRT) Median GC score: ART group: 6.2% (IQR 0.5 to 44.2) SRT group: 6.5% (IQR 0.5 to 62.8)	Percentage observation: ART: 88.7% SRT: 58.3%	Percentage observation ART: 79% SRT: 51%
Michalopoulos, 2014[26] PRO-ACT study 15 urologists 146 patients Prospective before-after test (own patients)	Post radical prostatectomy with T3 disease or positive SM Median GC 4.2% (Range 1.3% to 41.5%)	ART: 27.4% Observation: 69.9% Other: 2.7%	ART: 27.4% [a] Observation: 71%
Oncotype (No Studies)			
Prolaris (No Studies)			

Notes. [a] As noted above, treatment recommendations were revised for 30.8% of patients.
Abbreviations. ART=adjuvant radiotherapy; BCR=biochemical recurrence; GC=genomic classifier; IQR=interquartile range; PSA=prostate-specific antigen; SRT=salvage radiotherapy.

Harms

None of the included studies reported on outcomes related to patient harm due to genomic classifier testing. Despite this, acute harms due to the test itself would be expected to be minimal,

if any, due to the genomic classifier being run on available tissue specimens. Long-term harms, namely inferior clinical outcomes, due to modified treatment in the absence of the proven predictive ability of the genomic classifier could be present though were not addressed.

Subgroups

We examined subgroup analyses of *a priori* prioritized characteristics, specifically race and clinical risk categorization.

Race

Two studies reported the impact of genomic classifier tests (both used Oncotype) on treatment recommendations by race.[18,23] The ENACT trial[23] recruited 191 participants, the majority of whom were from historically underrepresented racial/ethnic populations, with 70% Black, 12.5% Hispanic or Latino, 1% Asian, and the remaining White. Genomic testing was hypothesized to increase adoption of AS including among Black men, but the study found comparable rates of AS adoption regardless of race (Hispanic, Latino, and Asian patients were pooled due to low sample sizes). Odds ratios for each group suggest that genomic testing increased the odds of undergoing AS, but ratios in all groups were nonsignificant. Lack of significance may be attributable to the small sample size of race/ethnicity subgroups, which likely limited statistical power. The second study[18] considered the impact of Oncotype testing on selection of AS across groups of White, Black, and "other" Veterans. Black Veterans who underwent Oncotype testing had a higher percentage selecting AS (80%) compared to untested Black Veterans (66%). A similar pattern was noted among White Veterans, with an absolute increase of 11% selecting AS after testing, and 20% absolute increase among Veterans identified as "other" race after testing. There was a significance difference with p-value < 0.01 across all 3 categories.

Risk

The impact of genomic classifier test results on treatment recommendations in patients with different risk classifications was explored in the identified studies in 2 ways: among studies that evaluated the effect of receipt of test results by the risk estimate from the test itself and by baseline clinical risk classification (*eg*, NCCN risk categorization).

Effect of Tests on Treatment Recommendations at First-line Treatment Decision

Among the studies evaluating the effect of tests on treatment recommendations at the time of first-line treatment choice, 4 Oncotype-based studies[18,20,23,24] and 1 Prolaris-based study[34] reported on effect stratified by baseline clinical risk classification. The randomized trial by Murphy et al found no statistically significant association between baseline risk classification of NCCN low or low intermediate and choice of active surveillance as initial treatment.[23] Across the other observational studies, there was no clear pattern of which baseline clinical risk population more often was described to have a change in treatment plan (Table 8).

Table 8. Test Effect on First-line Treatment Decisions by Baseline Clinical Risk Determination

Study Total N Design	Clinical Characteristics	Outcomes
Decipher (No Studies)		
Oncotype		
Murphy, 2021[23] ENACT study 191 patients Randomized trial	Newly diagnosed NCCN favorable intermediate or below Median GPS score: NR	Association intervention arm (test vs no-test) with treatment choice of active surveillance: • Low: OR = 0.28 (0.05 to 1.50) • Low intermediate: OR for active surveillance = 0.32 (0.10 to 1.08)
Lynch, 2018[18] 200 (2013-2014) 190 (2015-2016) 6 VA health care systems Retrospective cohorts before-after institutional testing	Newly diagnosed NCCN intermediate risk or lower Median GPS (tested): 26.5 (range 0–61) Risk reclassification: Lower :12% Higher: 7%	Treatment recommendation changes most common among NCCN *intermediate* risk patients: • 5% decreased intensity • 22% increased intensity No statistically significant different in use of active surveillance across NCCN risk groups between tested and untested cohorts (p = 0.20)
Eure, 2017[20] 247 (before) 258 (after) Comparative cohort before (retrospective) and after (prospective) institutional testing	Newly diagnosed NCCN intermediate or lower Median GPS: NR 23% pts had risk reclassification	Overall change in management plan: • NCCN very low: 16% • NCCN low: 28% • NCCN intermediate: 23% Choice of active surveillance management: • NCCN very low: Untested: 57% Tested: 88% • NCCN low: Untested: 43% Tested 74% • NCCN intermediate Untested: 19% Tested: 23% Changes were “directionally consistent with GPS predicted risk”
Badani, 2015[24] 158 pts 3 clinical sites Prospective before-after test (own patients)	Newly diagnosed NCCN low-intermediate or lower Median GPS: 21 (IQR range: 13, 32)	Decrease in treatment intensity by NCCN risk: • Very low: 1/35 (2.8%) • Low: 21/71 (29.6%) • Low intermediate: 3/52 (5.8%) Increase in treatment intensity by NCCN risk: • Very low: 3/35 (8.6%)

Study Total N Design	Clinical Characteristics	Outcomes
	38 pts (24%) with risk reclassification post-GPS	• Low: 5/71 (7.0%) • Low intermediate: 6/52 (11.5%)
Prolaris		
Crawford, 2014[34] 305 patients Prospective pre/post-test result	New diagnosis AUA risk Low: 44.3% Intermediate: 42.9% High: 12.8% Median CCP: -0.71 +/-0.83	Overall change in treatment recommendations by AUA risk level: • Low: 31.8% 24.4% interventional to non-interventional 7.4% non-interventional to interventional • Intermediate: 29% 16.% intervention to non-interventional 12.2% non-interventional to interventional • High: 33.3% 15.4% interventional to non-interventional 17.9% non-interventional to interventional

Abbreviations. AUA=American Urological Association; CCP=cell cycle progression; GPS=genomic prostate score.

Test Effect on Treatment Recommendations after Radical Prostatectomy by Test Risk Prediction

Decipher

Five studies (6 articles) reported changes to treatment recommendations for patients at the time of radical prostatectomy by test-based risk assessment, all of which were Decipher-based studies (Table 9).[26,27,29,30,36,37] Overall, patients with genomic classifier or Decipher test results indicating higher risk received lower rates of recommendation for observation post-prostatectomy.

Table 9. Test Effect on Treatment Recommendations After Radical Prostatectomy by Test Risk Prediction

Study Total N Design	Clinical Characteristics	Observation Treatment Recommendation Pre-test	Genomic Classifier Risk Stratification		
Decipher					
Gore, 2017[36] PRO-IMPACT study 19 sites 265 patients Prospective before-after test (own patients)	Post radical prostatectomy; non-organ confined prostate cancer or positive surgical margins (ART); or PSA increase or BCR[a] (SRT) Median GC score: ART: 6.2% (IQR 0.5 to 44.2) SRT: 6.5% (IQR 0.5 to 62.8)	ART Observation recommended: 88% SRT Observation recommended: 60%	**Low** ART Observation recommended after test results: 60/63 = 95% After 12 months:48/140 = 76% SRT Observation recommended post test: 24/34 = 71% After 12 months: 16/34 = 47%	**Intermediate** ART Observation recommended after test results: 24/33 = 73% After 12 months: 23/33 = 70% SRT Observation recommended after test results: 17/28 = 61% After 12 months: 12/28 = 43%	**High** ART Observation recommended after test results: 27/44 = 61% After 12 months: 27/44 = 61% SRT Observation recommended after test results: 13/44 = 30% After 12 months: 12/44 = 27%
Badani, 2015[29] 110 cases 51 urologists Deidentified case history review with and without test	Post radical prostatectomy with undetectable PSA Median GC 3.85 (min, max: 1.2, 33.4) % reclassified: NR	Observation recommended: 57%	**Low** Observation: 80.7%	**High** Observation: 34.6%	
Badani, 2013[30] 12 patient cases (ART) 12 patient cases (SRT) 21 urologists from 18 sites	Post radical prostatectomy with adverse pathology Median GC: NR	Observation: 47.5%	Observation: 79%	Observation: 8%	

Study **Total N** **Design**	**Clinical Characteristics**	**Observation Treatment Recommendation** **Pre-test**	**Genomic Classifier Risk Stratification**	
Deidentified case history review with and without test				
Michalopoulos, 2014[26] PRO-ACT 15 urologists 146 patients Prospective before-after test (own patients)	Post radical prostatectomy with T3 disease or positive SM Median GC: 4.2% (Range: 1.3 to 41.5%)		**Low** • Decrease: 15 (17.1%) • No change: 72 (81.8%) • Increase: 1 (1.1%)	**High**[b] • Decrease: 4 (7.6%) • No change: 28 (52.8%) • Increase: 21 (39.6%)
Shahait, 2021[37] 398 2 prospective cohorts	Post radical prostatectomy with adverse pathology Median GC: 0.593%	NR	HR = 4.28 (2.81 to 6.50) higher chance of receiving secondary therapy with high GC vs low/intermediate GC	
Oncotype (No Studies)				
Prolaris (No Studies)				

Notes. [a] High/low determination based on Decipher predicted risk relative to average risk for original study population; [b] Receipt of treatment recommendations higher among patients with high-risk Decipher scores vs low risk (p<0.001).
Abbreviations. ART=adjuvant radiotherapy; BCR=biochemical recurrence; GC=genomic classifier; HR=hazard ratio; IQR=interquartile range; PSA=prostate-specific antigen; SRT=salvage radiotherapy.

KEY QUESTION 3: AMONG PATIENTS WITH LOCALIZED PROSTATE CANCER, WHAT IS THE PROGNOSTIC EFFECT OF TISSUE-BASED GENOMIC TESTS AFTER ADJUSTING FOR EXISTING PROGNOSTIC CLINICAL FEATURES ON KEY CLINICAL OUTCOMES (*eg*, BIOCHEMICAL RECURRENCE-FREE SURVIVAL, METASTASES-FREE SURVIVAL) FOLLOWING DEFINITIVE TREATMENT?

Key Points

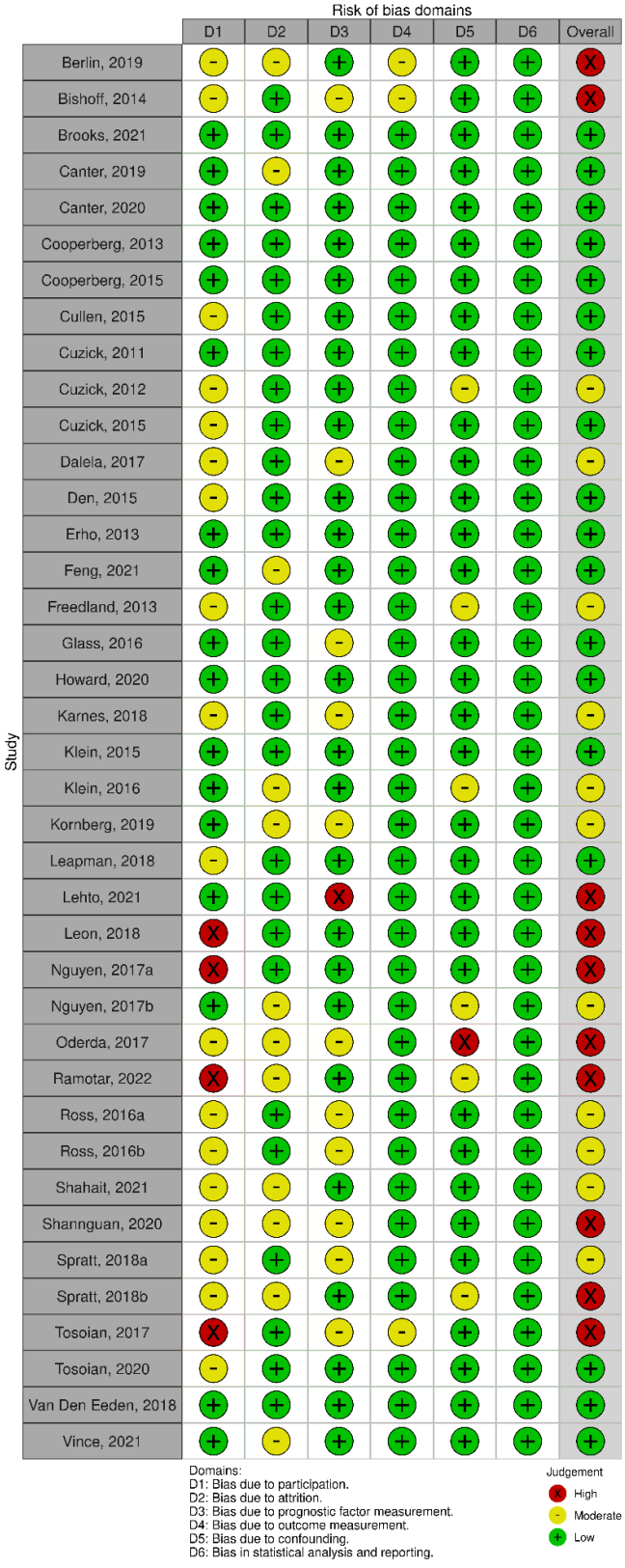
- While 39 studies addressed the prognostic ability of genomic classifiers, the clinical classification schemes they were compared to and outcomes assessed varied greatly.
- Only 2 of the 39 studies included prospectively collected cohorts, and only 9 of the 39 studies included patients treated with definitive radiation.
- Patients in these studies were diagnosed from the 1980s to the mid-2010s, a long period that saw many advancements in the diagnosis, treatment, and follow-up of patients with prostate cancer. Despite this, genomic classifiers show a consistent albeit modest improvement in prognosis when compared to clinical models.
- For biochemical recurrence, the Decipher summary hazard ratio (HR) was 1.20 (95% CI [1.00, 1.43]), the Oncotype HRs ranged from 1.10 to 2.73, and the Prolaris summary hazard ratio for BCR was 1.44 (95% CI [1.28, 1.62]).
- For development of metastases, the HR ranged from 1.17 to 2.05 for Decipher, 2.24 to 2.34 for Oncotype, and 2.03 to 4.19 for Prolaris.
- For prostate-cancer-specific mortality, the HRs for Decipher ranged from 1.39 to 1.81, the range for Oncotype was 2.30 to 2.69, and the summary HR for Prolaris was 1.722 (95% CI [1.58, 1.87]).

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.[4,6,19,21,22,37-71] Across studies, there was substantial variability in the clinical risk-classification models, outcome of interest, and statistical measure used to assess the impact of the genomic classifier. Seven studies compared the prognostic ability of the genomic classifier 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; however, in the study that included all 3 tests, tissue sample processing and analysis was performed by the institution at which the patients were treated as opposed to the company that developed the test.[40] Twenty-four studies ran the genomic classifier on prostatectomy tissue,[4,22,37,39-46,51,52,54,57,59-61,63,64,66,68,71] 20 on biopsy tissue,[6,19,21,22,38,42,43,46-50,53,55,58,62,65,67,69,71] and 5 on a combination of the two.[22,42,43,46,71] At least 26 studies included patients diagnosed prior to 2000,[4,6,19,22,39-41,43,44,46,52-57,59,61-65,67-69,71] and at least 9 included patients diagnosed prior

to 1990.[4,6,22,39,44,54,55,64,69] One study did not report the timeframe from which the patients were drawn, while another described patients as diagnosed prior to 2017.[42,50] 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.[19,38,43,46,48,49,55,58,67] Two studies did not report the treatments received.[62,69]

Common risks of bias among included studies for this KQ 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. Eighteen studies were found to have low ROB,[4,6,38,39,41,43,44,46,48,50,53,60,62-66,68] 11 moderate ROB,[22,37,47,52,54,57-59,61,67,69] and 10 high ROB[19,21,40,42,45,49,51,56,71] (Figure 6). Of note, 17 studies appear to have been sponsored or co-authored by the commercial companies with rights to the genomic classifier tests under study.

Figure 6. Risk of Bias Assessment for Prognostic Studies



Next, we discuss KQ3 results by outcome and genomic classifier studied.

Biochemical Recurrence

Decipher

Four studies evaluated the additional benefit of the Decipher score in predicting biochemical recurrence (BCR).[42,49,58,60] All 4 were retrospective, 1 had low ROB,[60] 1 moderate ROB,[58] and 2 high ROB.[42,49] For Decipher, the summary estimate HR for BCR across 3 studies ($N = 445$) was 1.20 (95% CI [1.00, 1.43]; 95% prediction interval [PI] [1.00, 1.43) (Figure 7), indicating a 20% increase in the risk of BCR with a higher Decipher score when clinical classification schemes are also considered. Two studies evaluated patients undergoing radiation and 2 evaluated patients post-prostatectomy.

A low ROB study evaluated 224 men with high-risk pathologic features after prostatectomy.[60] In a model that included age and CAPRA-S scores, 0.1 unit increases in the Decipher score correlated with a significantly increased risk of BCR with a HR of 1.17 (95% CI [1.04, 1.33]) (Table 10).[60] However, in the same model, 1 unit increases in the CAPRA-S scores predicted a similar increase in risk of BCR with a HR of 1.14 (95% CI [1.01, 1.29]).[60] The AUC remained in the range generally considered to be poor with a non-significant increase in the AUC from 0.64 (0.56 to 0.63) to 0.69 (0.61 to 0.76) after the addition of the Decipher score to the CAPRA-S score for discrimination of BCR at 10 years (Figure 8).[60]

In a moderate ROB study of 100 men with either intermediate- or high-risk prostate cancer, there was no evidence of an improvement in prognostic ability per 0.1 increase in Decipher score in either a model with NCCN (HR = 1.16, 95% CI [0.96, 1.141]) or CAPRA (HR = 1.08, 95% CI [0.89, 1.32]) classification schemes.[58]

In a high ROB retrospective study of 121 men with intermediate-risk prostate cancer who underwent dose-escalated radiation therapy alone without ADT,[49] Decipher score as a continuous variable was significantly associated with an increased risk of BCR (HR = 1.36, 95% CI [1.09, 1.71]) while unfavorable versus favorable intermediate risk classification was not. Additionally, even though Decipher's AUC for BCR at 5 years (0.78; 95% CI [0.59, 0.91]) was not significantly different from the NCCN classification (0.56; 95% CI [0.43, 0.66]), a combined NCCN and Decipher model improved the AUC for BCR to 0.85 (95% CI [0.73, 1.00]). Concerns for potential ROB came from a lack of clarity of participation by a potentially eligible pool of patients and a lack of information about missing data.

The other high ROB study evaluated the prognostic ability of the Decipher score in 81 patients who underwent prostatectomy and post-operative radiation.[42] In a model with multiple pathologic features and the Decipher score as a categorical value, only receipt of salvage versus adjuvant therapy and the Decipher score were significantly associated with risk of BCR. Compared to a high Decipher score, low (HR = 0.32; 95% CI [0.13, 0.75]) and intermediate (HR = 0.4; 95% CI [0.18, 0.89]) Decipher scores were associated with a lower risk of BCR. This study also reported an acceptable AUC of 0.742 (95% CI [0.643, 0.84]) for a model with pathologic features and the Decipher score; however, it is unclear how much Decipher added value here as neither an AUC for pathologic features nor Decipher score alone was provided. ROB assessment for this study was primarily driven by a lack of information around the source

populations for the cohorts, the recruitment time period, or how many samples were excluded due to insufficient or inadequate tissue.

Oncotype

Three retrospective studies, including 2 low ROB studies and 1 moderate ROB study, evaluated the prognostic ability of the Oncotype score for BCR. All studies used biopsy tissue from patients who underwent prostatectomy.[47,53,65] The study-specific HRs ranged from 1.10 (95% CI [1.10, 1.21]) to 2.7 (95% CI [1.84, 3.96]). These 3 studies were not combined in a meta-analysis due to underlying conceptual heterogeneity.

In a cohort of 402 men, a model accounting for NCCN risk classification showed an increase in BCR risk for every increase in 20 units of the Oncotype score with an HR of 2.73 (95% CI [1.84, 3.96]).[65] In this model, NCCN risk grouping was not a significant predictor of BCR.

In a retrospective study of 257 men treated from 1995 to 2010, the Oncotype score per 20 units had significant increased risk in BCR with HRs of 2.11, 2.41, and 2.30 in models containing NCCN, AUA, or CAPRA classification schemes, respectively.[53] NCCN and CAPRA remained significant in their respective models, while AUA did not. There was a corresponding increase in the AUC for BCR from NCCN alone to NCCN with Oncotype (0.59 to 0.68); however, it remained below generally acceptable levels of discrimination and confidence intervals were not presented to assess significance.

Finally, in a retrospective of 215 men who underwent prostatectomy following a course of active surveillance, Oncotype was associated with an increased risk of BCR per 5 unit increase in a model that included the CAPRA score (HR = 1.10; 95% CI [1.00, 1.21]).[47] This study was found to be at moderate risk of bias due to concerns related to study attrition and specificity of prognostic factor measurement.

Prolaris

Nine retrospective studies (3 low ROB, 1 moderate ROB, and 5 high ROB) evaluated the ability of the Prolaris or cell cycle progression score (CCP) to predict BCR. The summary effect estimate across these studies showed an increased risk of BCR with increasing Prolaris score with an HR of 1.44 (95% CI [1.28, 1.62]; 95% PI [1.28, 1.62]). Eight studies were performed in patients who underwent prostatectomy,[4,21,45,50,51,56,68,71] and 1 study in patients who underwent definitive radiation,[67] with biopsy and prostatectomy tissue each analyzed in 5 of the 9 studies.

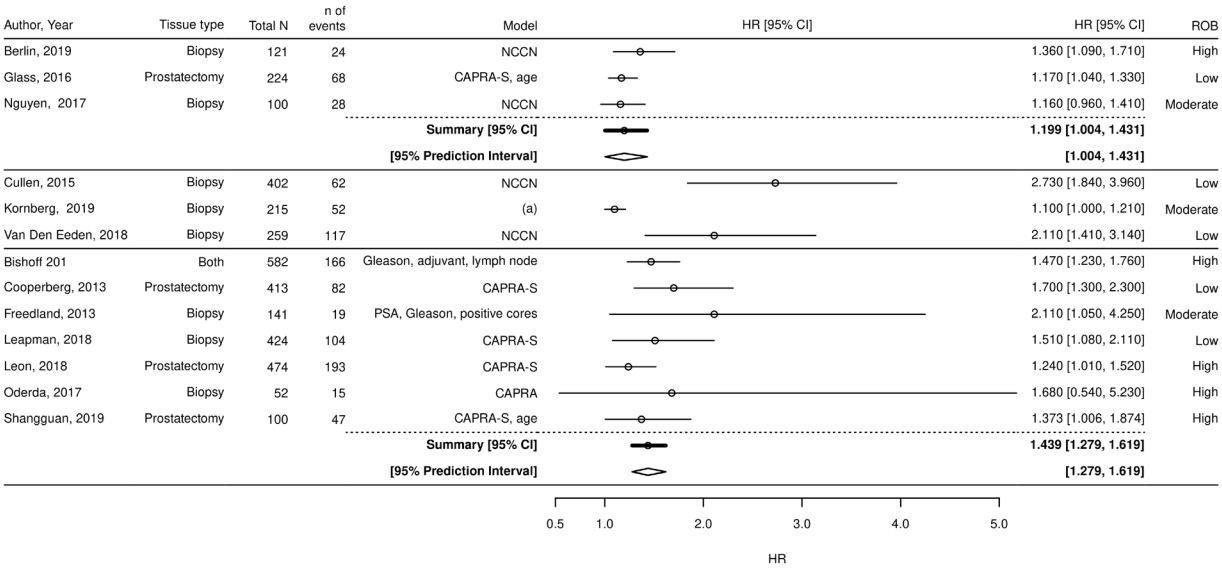
One low ROB study evaluated BCR post-prostatectomy among 246 men and found an HR of 1.7 (95% CI [1.3, 2.3]) for CCP as a continuous variable in a model with CAPRA-S. In this study, CCP as a categorical variable from 0 to 1 (HR = 5.2, 95% CI [1.2, 21.7]) or greater than 1 (HR = 9.5, 95% CI [2.0, 45.2]) were significant; however, CCP as a categorical variable from -1 to 0 was not (HR = 3.4, 95% CI [0.8, 14.1]) in a model incorporating the CAPRA-S score.[68] The Prolaris score was also predictive of BCR in another low ROB study of 424 men who underwent prostatectomy prior to 2017.[50] When incorporating CAPRA-S, CCP had an HR of 1.51 (95% CI [1.08, 2.11]) for BCR.[50] This corresponded with an acceptable AUC for the CCP and CAPRA-S of 0.72 (95% CI was not reported nor was AUC for CCP or CAPRA-S alone). The third low ROB study demonstrated an HR of 1.74 (95% CI [1.39, 2.17]) for the CCP per unit increase in a

cohort of 336 men who underwent prostatectomy when included in a model with other pathologic characteristics.[4]

The other 5 studies reporting the relationship between CCP and BCR among men post-prostatectomy were found to be at high ROB primarily due to lack of clarity about participation among potentially eligible patients, study attrition, and non-standard or unclear prognostic factor measurement. The first included 100 men post-prostatectomy and found an HR of 1.373 (95% CI [1.006, 1.874]) in a model containing CCP per unit increase and CAPRA score. They also demonstrated HRs for high versus low CCP of 10.912 (95% CI [3.0, 39.7]) and 7.481 (2.1, 26.4) for intermediate versus low CCP with notably wide confidence intervals.[45] This corresponded to a C-index for prediction of BCR of 0.77 (95% CI [0.69, 0.85]) which was not significantly greater than of the CAPRA-S score (0.71; 95% CI [0.63, 0.79]) or CCP (0.74; 95% CI [0.66, 0.83]) alone.[45] In a second retrospective cohort of 474 men from the 2000s, HRs of 1.24 (95% CI [1.01, 1.52]) and 1.28 (95% CI [1.03, 1.59]) were found in models including CAPRA-S or multiple pathologic characteristics, respectively, with the latter including Ki-67 and PTEN expression.[51] In the third study of 236 patients (76 Veterans) with low-risk prostate cancer who underwent prostatectomy, the HR for BCR with CCP as a continuous variable was 1.41 (95% CI [1.02, 1.96]) in MVA incorporating CAPRA score.[56] In this study, similar albeit poor AUCs for BCR were observed at 5 (0.662) and 10 (0.65) years post-prostatectomy for models with CCP and CAPRA compared to CAPRA alone at 5 (0.557) and 10 (0.542) years. A fourth retrospective study including only 52 patients postprostatectomy reported a non-statistically significant HR for CCP in a model with CAPRA (1.68; 95% CI [0.54, 5.23]); however, the AUC for the CCP and CAPRA combined was 0.86 (95% CI or AUC for CCP or CAPRA alone not reported).[21] Finally, the fifth cohort including 582 patients postprostatectomy (176 Veterans) demonstrated an HR of 1.47 (95% CI [1.23, 1.76]) when incorporating pathologic features in MVA model for BCR.[71]

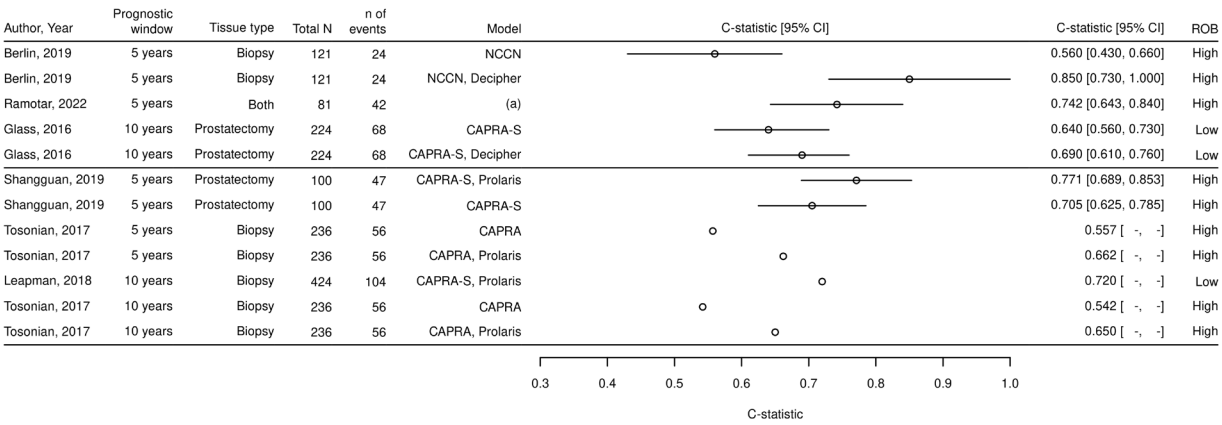
The 1 study assessing BCR following definitive radiation included 141Veterans treated between 1991 and 2006. In a model with pathologic features and concurrent ADT use, there was an increased risk of BCR with increasing Prolaris score with a HR of 2.11 (95% CI [1.05, 4.25]).[67] In this moderate ROB study, a small increase in the AUC was observed from 0.78 with clinical features alone to 0.80 with clinical features and CCP score, bringing the AUC into what is generally considered excellent discrimination.

Figure 7. Hazard Ratio Forest Plot for Biochemical Recurrence by Test Type (Decipher, Oncotype, Prolaris)[a]



Notes. [a] Model includes CAPRA, PSA, age, tissue source (confirmatory vs diagnostic biopsy), clinical institution (UCSF vs other), genomic prostate score testing (clinical care vs research).

Figure 8. C-statistic Forest Plot for Biochemical Recurrence by Test Type (Decipher, Prolaris)[a]



Notes. [a] Model includes tumor stage (pT3-4 vs pT2), PSA pre-PORT, surgical margins (positive vs negative), ISUP Grade Group (2,3,4-5 vs 1), PORT modality (salvage vs adjuvant), intraductal carcinoma and cribriform architecture (positive vs negative).

Table 10. Studies Reporting Biochemical Recurrence

Study Total N Design Risk of Bias (ROB)	Clinical Characteristics Tissue Source Treatment Study Duration	Model Variables	Results (95% CI)
Decipher			
Berlin, 2019[49] 121 Retro High ROB	NCCN intermediate risk Biopsy Dose-escalated, image-guided RT without ADT 2005-2011	NCCN classification *unfavorable vs favorable intermediate risk*	HR = 1.36 (1.09 to 1.71)
		NCCN	AUC (5 years) Clinical features 0.56 (0.43 to 0.66) Clinical features and test 0.85 (0.73 to 1.00)
Ramotar, 2022[42] 81 Retro High ROB	Post "maximal local therapies" (RP and PORT) with pathology slides available for review Biopsy, RP RP and RT NR	Tumor stage p*T3-4 vs pT2* PSA pre-PORT Surgical margins *positive vs negative* ISUP Grade Group *2,3,4-5 vs 1* PORT modality *salvage vs adjuvant* IDC/CA *positive vs negative*	Intermediate Decipher vs high HR = 0.4 (0.1 to, 0.89) Low Decipher vs high HR = 0.32 (0.13 to 0.75)
		Tumor stage p*T3-4 vs pT2* PSA pre-PORT Surgical margins *positive vs negative* ISUP Grade Group *2,3,4-5 vs 1* PORT modality *salvage vs adjuvant* IDC/CA *positive vs negative*	AUC Clinical features and test 0.742 (0.643 to 0.84)
Nguyen, 2017b[58] 100 Retro Moderate ROB	NCCN intermediate and high risk treated with RT and ADT Biopsy RT and ADT	NCCN *High vs intermediate risk*	HR = 1.16 (0.96 to 1.41)

Study **Total N** **Design** **Risk of Bias (ROB)**	**Clinical Characteristics** **Tissue Source** **Treatment** **Study Duration**	**Model Variables**	**Results (95% CI)**
Glass, 2016[60] 224 Retro Low ROB	RP with high-risk pre-op features (PSA > 20 or Gleason score ≥ 8) pT3, or +SM RP RP 2001-2013	CAPRA-S Age at diagnosis	HR = 1.17 (1.04 to 1.33)
		CAPRA-S	AUC Clinical features 0.64 (0.56 to 0.73) Clinical features and test 0.69 (0.61 to 0.76)
Oncotype			
Van Den Eeden, 2018[53] 259 Retro Low ROB	RP within 12 months of diagnosis Biopsy RP 1995-2010	NCCN *High vs low and very low* *Intermediate vs low and very low*	HR = 2.11 (1.41 to 3.14)
Kornberg, 2019[47] 215 Retro Moderate ROB	Active surveillance patients who had RP ≥ 6 months after starting. Patients had organ confined Gleason 3 + 3 or low volume 3 + 4 prostate cancer with PSA < 20 and CAPRA score < 6 Biopsy RP 2001-2016	CAPRA, Age at diagnosis PSA density at time of genomic test Tissue source *Confirmatory vs diagnostic biopsy* Clinical institution *UCSF vs other* GPS testing *Clinical care vs research*	HR = 1.10 (1.00 to 1.21)
Cullen, 2015[65] 402 Retro Low ROB	Biopsy Gleason score 6 or 7, PSA ≤ 20, ≤ cT2, and RP ≤ 6 months after diagnosis Biopsy RP 1990-2011	NCCN *Low vs very low* *Intermediate vs very low*	HR = 2.73 (1.84 to 3.96)
		NCCN	AUC Clinical features only 0.59 (NR) Clinical features and test 0.68 (NR)

Study Total N Design Risk of Bias (ROB)	Clinical Characteristics Tissue Source Treatment Study Duration	Model Variables	Results (95% CI)
Prolaris			
Cuzick, 2011[4] 336 Retro Low ROB	RP without neoadjuvant therapy RP RP 1985-1995	Log(1+baseline PSA) Gleason score *7, >7 vs <7* Pathological stage Surgical margins	HR = 1.74 (1.39 to 2.17)
Shangguan, 2019[45] 100 Retro High ROB	Adverse pathology (SVI, ECE, positive surgical margins) after RP RP RP 2010-2014	CAPRA score Age	HR = 1.373 (1.006 to 1.874)
		CAPRA score Age	High Prolaris vs low HR = 10.912 (3.0 to 39.691)
		CAPRA score Age	Intermediate Prolaris vs low HR = 7.481 (2.118 to 26.425)
		CAPRA score	AUC Clinical features 0.705 (0.625 to 0.785) Clinical features and test 0.771 (0.689 to 0.853)
Leapman, 2018[50] 424 Retro Low ROB	Clinically localized, treated with RP Biopsy RP Prior to 2017	CAPRA-S	HR = 1.51 (1.08 to 2.11)
		CAPRA-S	AUC Clinical features and test 0.72 (NR)
Leon, 2018[51] 474 Retro High ROB	Treated with RP RP RP 2000-2007	CAPRA-S score (other regression components not clear)	HR = 1.24 (1.01 to 1.52)
Tosoian, 2017[56]	RP for Gleason score ≤6	CAPRA	HR = 1.41 (1.02 to 1.96)

Study Total N Design Risk of Bias (ROB)	Clinical Characteristics Tissue Source Treatment Study Duration	Model Variables	Results (95% CI)
236 Retro High ROB	Biopsy RP 1994-2006	CAPRA	AUC (5 years) Clinical features 0.557 (NR) Clinical features and test 0.662 (NR)
		CAPRA	AUC (10 years) clinical features 0.542 (NR) Clinical features and test 0.65 (NR)
Oderda, 2017[21] 52 Retro High ROB	Treated with RP Biopsy RP 2013-2015	CAPRA	HR = 1.68 (0.54 to 5.23)
			AUC Clinical test and features 0.86 (NR)
Freedland, 2013[67] 141 Retro Moderate ROB	Treated with RT Biopsy RT 1991-2006	Log(1 + PSA) Gleason score *7, >7 vs <7* Percent positive cores Concurrent ADT	HR = 2.11 (1.05 to 4.25)
		Log(1 + PSA) Gleason score Percent positive cores Concurrent ADT	AUC Clinical features 0.78 (NR) Clinical features and test 0.80 (NR)
Bishoff, 2014[71] 582 Retro High ROB	Clinically localized, treated with RP Biopsy, RP RP 1994-2006	Log(1 + PSA) Gleason score *7, >7 vs <7* Percent positive cores Adjuvant treatment Age at diagnosis	HR = 1.47 (1.23 to 1.76)
Cooperberg, 2013[68]	Treated with RP without adjuvant or neoadjuvant	CAPRA-S	HR = 1.7 (1.3 to 2.3)

Study Total N Design Risk of Bias (ROB)	Clinical Characteristics Tissue Source Treatment Study Duration	Model Variables	Results (95% CI)
413 Retro Low ROB	therapy and with >5 years follow-up RP RP 1994-2011	CAPRA-S	CCP Score > -1 to 0 vs ≤ -1 HR = 3.4 (0.8 to 14.1)
		CAPRA-S	CCP Score > 0 to 1 vs ≤ -1 HR = 5.2 (1.2 to 21.7)
		CAPRA-S	CCP Score > 1 vs ≤ -1 HR = 9.5 (2.0 to 45.2)

Abbreviations. ECE=extra-capsular extension; PSA=prostate specific antigen; SVI=seminal vesicle involvement.

Metastases

Decipher

Sixteen studies with 3,587 participants addressed the ability of the Decipher score to predict metastases following definitive treatment of prostate cancer, including 15 retrospective studies and 1 secondary analysis of a prospective, randomized trial.[6,19,22,40,41,43,44,49,52,55,58,59,61,63,64] Across 9 studies (N = 2,139), the summary effect estimate showed an increase in risk of metastases with continuous increase in Decipher score with an HR of 1.32 (95% CI [1.22, 1.44]; 95% PI [1.15, 1.52]) (Figure 9). Notably, these 16 studies drew on patients diagnosed from 1987 and 2016, over which time management of prostate cancer evolved. Six studies were found to have low ROB, 5 moderate ROB, and 5 high ROB. The number of metastatic events in these studies were low, ranging from 5 to 104, leading to large variability in reported findings.

The first low ROB study was an ancillary analysis of data from 352 men treated on RTOG 9601, a phase III prospective randomized trial evaluating the addition of 2 years of ADT to post-prostatectomy radiation. Patients eligible for this study had a rising PSA to 0.2 to 4.0 following prostatectomy with pathology showing either T2N0 disease with positive margins or T3N0 disease.[41] This study demonstrated an HR of 1.17 (95% CI [1.05 to 1.32]) for metastases for Decipher as a continuous variable in a model including clinical and pathological characteristics as well as treatment received (ADT or placebo) as part of the phase III study. When considered as a categorical variable, Decipher high versus low scores had an HR of 1.74 (95% CI [1.08, 2.84]) accounting for the same variables used in the model above (Table 11). Of note, this report was underpowered to detect a statistically significant interaction between the Decipher score and the effect of ADT and therefore did not demonstrate the Decipher score as a predictive biomarker for ADT use.

A second low ROB study included 405 Veterans with prostate cancer and employed multiple clinical classification schemes while evaluating Decipher as both a continuous variable (per 0.1 units) and as a categorical variable.[43] In models with Decipher as a continuous variable, the HRs for metastases were 1.34 (95% CI [1.19, 1.50]) and 1.33 (95% CI [1.19, 1.48]) with incorporation of NCCN and CAPRA classification schemes, respectively. As a categorical

variable, HRs for metastases were 2.95 (95% CI [1.75, 4.98]) and 3.09 (95% CI [1.88, 5.06]) comparing high (>0.6) and low (<0.45) Decipher scores in models with NCCN and CAPRA classifications, respectively (Figure 10). In addition, the AUC for metastases at 5 years increased with the addition of the Decipher score to almost acceptable discrimination; specifically, the AUC increased from 0.46 (95% CI [0.38, 0.53]) to 0.67 (95% CI [0.60, 0.75]) and from 0.59 (95% CI [0.50, 0.67]) to 0.71 (95% CI [0.65, 0.78]) when including the Decipher score with NCCN and CAPRA models, respectively (Figure 11).

In another low ROB study of 548 Veterans who had undergone prostatectomy, Decipher as a categorical variable showed an HR of 9.60 (95% CI [3.51, 32]) for prediction of metastases when comparing high to low scores and 6.51 (95% CI [2.33, 21.8]) for intermediate to low scores.[44] The number of events in this cohort was 37, leading to the large confidence intervals. In the fourth low ROB study, a nested case control design was used to create 2 cohorts of men post-prostatectomy: one including patients with either no evidence of recurrence or biochemical recurrence only and one with patients with clinical metastases.[6] The authors then demonstrated an increased odds of metastases with every 10% increase in Decipher score in a model incorporating multiple pathologic and clinical factors (OR = 1.36, 95% CI [1.16, 1.60]). In addition, they reported an AUC of 0.74 for a model with Decipher and clinical features as compared to an AUC of 0.69 for clinical features alone, although CI were not reported. In another low ROB study of 188 patients limited to those who received post-prostatectomy radiation, models with CAPRA-S showed similar HR for metastases with Decipher of 1.69 (95% CI [1.24, 2.31]) and a significant increase in the AUC from 0.66 (95% CI [0.56, 0.78]) to 0.85 (95% CI [0.79, 0.93]).[63] Finally, Klein et al reported ORs for the Decipher score of 1.43 and 1.48 for metastases with the continuous Decipher score in models with CAPRA-S or clinical features, respectively.[64] Including Decipher with clinical features in this study led to a non-significant increase in the AUC to 0.78 (95% CI [0.68, 0.89]) compared to 0.72 (95% CI [0.6, 0.84]) with clinical features alone.

Four of the 5 moderate ROB studies evaluated the additive predictive value of Decipher for metastases reported similar findings. One retrospective study of 260 men observed an HR of 1.32 (95 CI [1.17, 1.51]) for metastases with Decipher by 0.1 unit increase in a model with CAPRA-S.[61] In addition, the AUC for metastases at 10 years increased from 0.77 (95% CI [0.69, 0.85]) for CAPRA-S alone to 0.87 (95% CI [0.77, 0.94]) for CAPRA-S and Decipher score, although this was not significant. A second retrospective study including 422 patients with adverse pathology on radical prostatectomy[59] included CAPRA-S and radiation treatment type in their model and reported a HR of 1.28 (1.08, 1.52) for metastases with Decipher. A third study of 100 intermediate and high-risk patients who underwent radiation combined with ADT found that the Decipher score had an HR of 1.37 (95% CI [1.06, 1.78]) in a model with NCCN.[58] In a fourth small study of 57 men with high-risk clinical or pathologic features, the Decipher score had an HR of 1.64 (95% CI [1.11, 2.42]) for metastasis development in a model including the CAPRA-S score.[22] In this same study, the addition of the Decipher score to NCCN classification led to an excellent AUC for metastases at 0.88 (95% CI [0.76, 0.96]) versus 0.75 (95% CI [0.64, 0.87]) for NCCN classification alone. The fifth moderate ROB study included 150 men with a persistent PSA following prostatectomy and demonstrated similar HRs with wide confidence intervals for metastases when Decipher as a categorical variable of high versus low and intermediate when included in a model with CAPRA-S (HR 8.72; 95% CI [2.25, 39.8]) or clinical features (HR 5.61; 95% CI [1.48, 22.7]).[52] The latter corresponded with a non-significant increase in AUC

from 0.69 (95% CI [0.41, 0.89]) with clinical features alone to 0.83 (95% CI 0.70, 1.00) with clinical features and the Decipher score.

The remaining 5 studies were retrospective, high ROB studies due to variable levels of concern related to inability to determine proportion of participating eligible patients, non-standard prognostic factor measurement, and study attrition. One study included 121 intermediate-risk patients who underwent radiation alone as definitive treatment and demonstrated an HR of 2.05 (95% CI [1.24, 4.24]) for metastases in a MVA model including NCCN classification.[49] When NCCN classification was replaced in the model by clinical features, a similar HR was obtained (2.07 95% CI [1.17, 5.24]). AUC improved numerically in this study by adding the Decipher score to clinical features alone (0.89 vs 0.86, respectively).

The remaining 4 high ROB studies all included patients who underwent prostatectomy, with 3 that also included patients treated with definitive radiation. In the 1 study with both prostatectomy and radiation patients, a model containing both clinical features and treatment received demonstrated an increased risk of metastases associated with Decipher with an HR of 1.39 (95% CI [1.09, 1.8]).[55] Additionally, this study of 235 patients showed that the incorporation of Decipher score into a model with NCCN improved the AUC for prediction of metastases from 0.66 (95% CI [0.53, 0.77]) to 0.74 (95% CI [0.66, 0.82]), although this increase was not significant. A validation study of 235 patients revealed an AUC of 0.84 (95% CI [0.61, 0.93]) for a combination of clinical and genomic characteristics, compared to 0.68 (95% CI [0.64, 0.73)] for NCCN classification alone.[19] In addition, their clinical-genomic classification scheme of low, intermediate, and high risk showed HRs with wide confidence intervals of 22.3 (95% CI [2.9, 2,863.8]) and 61.6 (95% CI [8.1, 7,914.9]) for intermediate and high risk as compared to low risk when stratified by treating institution and adjusted for treatment received. In a study limited to 160 patients with Gleason Grade Group 2-4 prostate cancer, the predictive ability for metastases for all 3 genomic classifier tests was assessed; specifically, with the addition of Decipher, the AUC increased from 0.55 (95% CI [0.5, 0.6]) to 0.74 (95% CI [0.69, 0.78]).[40]

Oncotype

Three retrospective studies (2 low ROB and 1 high ROB) evaluated the prognostic ability of the Oncotype score to predict metastases in men who underwent prostatectomy.[39,40,53] The 2 low ROB studies reported similar findings supporting an modest additive value of Oncotype test per 20 units when combined with standard clinical features or risk schemas. The first, a low ROB study, showed an HR of 2.34 (95% CI [1.42, 3.86]) for Oncotype by 20 units in a model with NCCN classification among 259 patients who received radical prostatectomy within 12 months of diagnosis.[53] Similar results for the Oncotype score were seen with models including AUA classification (HR 2.51 95% CI [1.49, 4.23]) and CAPRA score (HR 2.63 95% CI [1.58, 4.36]). This corresponded to an increase in AUC from 0.66 with NCCN classification alone to 0.75 with NCCN and Oncotype features, although 95% CI were not provided to assess significance. The second low ROB cohort included 428 patients treated with prostatectomy between 1987 and 2004 and found that increases in Oncotype score per 20 units were associated with an increased risk of metastases in a model with clinical features (HR = 2.24; 95% CI [1.49, 3.53]), and AUC increased from 0.772 to 0.824 when incorporating Oncotype with clinical features compared to clinical features alone.[39] Finally, in the same high ROB study that assessed the additive

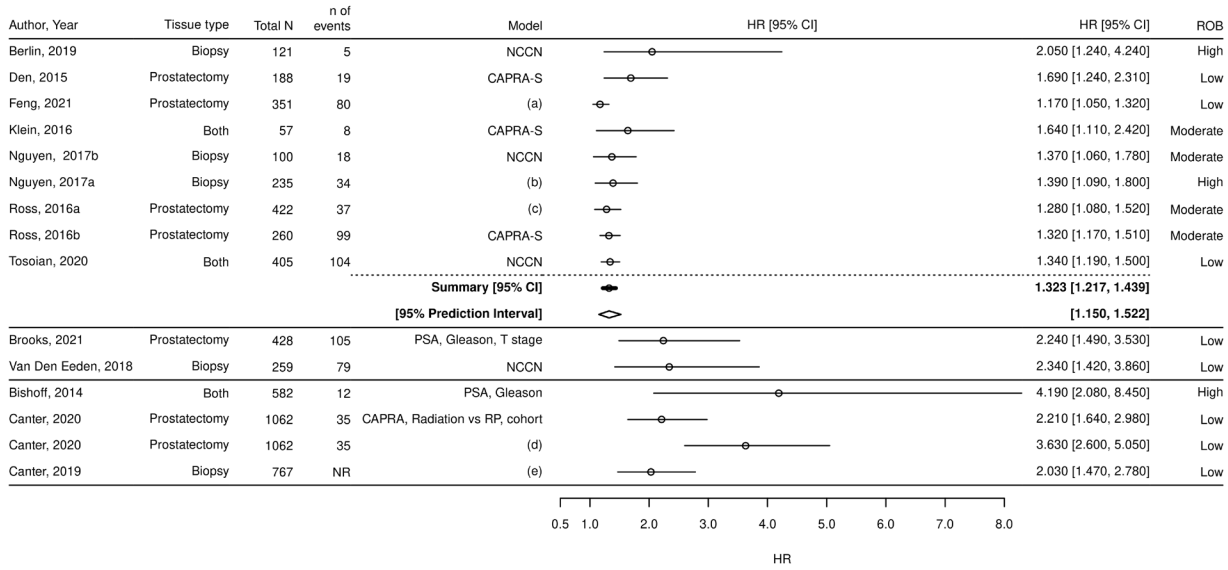
prognostic value of all 3 tests in 1 cohort of patients, the Oncotype score had an AUC of 0.65 (95% CI [0.6, 0.7]) compared to 0.55 (95% CI [0.5, 0.6]) for clinical features alone.[40]

Prolaris

Four retrospective studies addressed the additive prognostic value of the Prolaris test for metastases.[40,46,48,71] The 2 low ROB studies with the same first author (Canter) reported similarly sized increases in metastatic risk associated with increases in Prolaris score. The first (2020) included 1062 patients from 4 institutions (131 Veterans) who had undergone prostatectomy or radiation therapy with or without ADT. In a model that included the CAPRA score and treatment received, CCP had an HR of 2.21 (95% CI [1.64, 2.98]) for metastases.[46] This increased risk translated to an improvement in AUC from 0.86 to 0.89 when the CCP score was combined with the CAPRA score. The second study by the same author (2019) included a cohort of 767 men, some of whom were included in the previously noted Canter paper, who underwent either definitive radiation or prostatectomy.[48] In a model containing the CAPRA score, treatment received, and race, the HR for the per unit increase in CCP was 2.03 (95% CI [1.47, 2.78]). The AUC for CAPRA only increased from 0.88 to 0.9 with the addition of CCP.

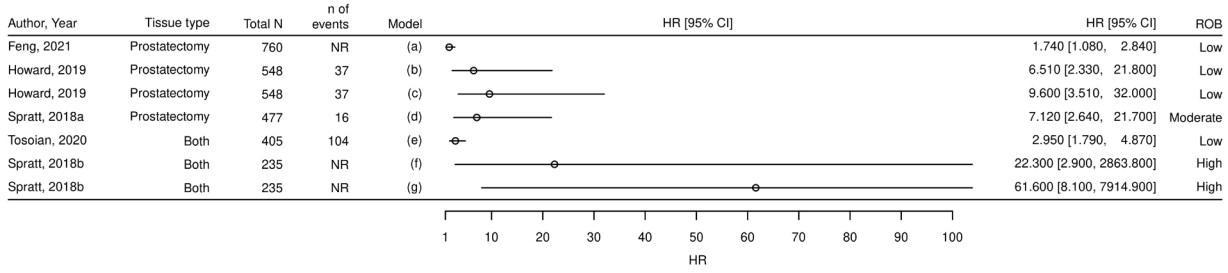
The other 2 studies evaluating the additive predictive value of Prolaris for metastases had high ROB but reported findings in the same direction and similar magnitude. First, 1 study ($N = 582$) that included patients post-prostatectomy was found to have multiple potential sources of risk of bias including study participation and measurement of both prognostic factors and outcomes. In a model that included clinical features, the HR for the CCP score was found to be 4.19 (95% CI [2.08, 8.45]).[71] Finally, the other high ROB study was the same study which evaluated AUC for all 3 relevant genomic classifier tests and showed an AUC of 0.73 (95% CI [0.69, 0.78]) for the CCP score compared to 0.55 (95% CI [0.5, 0.6]) for clinical features alone.[40]

Figure 9. Hazard Ratio Forest Plot for Metastasis by Test Type (Decipher, Oncotype, Prolaris)



Notes. [a] Model includes age (≥65 vs <65), race (Black vs non-Black) Gleason score (8-10 vs ≤7), T stage (pT3 vs pT2), PSA, positive surgical margins, PSA nadir status (non-nadir vs nadir <0.5), ADT vs placebo; [b] Model includes age, log2 (PSA), grade group, clinical stage, first-line treatment RP, first-line treatment RT ADT; [c] Model includes CAPRA-S, treatment (adjuvant radiation vs minimal residual disease salvage radiation, salvage radiation, no radiation); [d] CAPRA, treatment institutional cohort; [e] CAPRA, ancestry (Black vs non-Black), primary treatment.

Figure 10. Hazard Ratio for Categorical Studies Reporting Metastasis by Test Type (Decipher)



Notes. [a] Treatment, age, Black vs non-Black, Gleason, T score, PSA, margin status, nadir, Decipher (high vs low); [b] CAPRA-S, age, Black vs non-Black, Decipher (intermediate vs low); [c] CAPRA-S, age, Black men vs non Black men, Decipher (high vs low risk); [d] CAPRA-S, PSA, Decipher (high vs low/intermediate); [e] Age, PSA, Grade Group, T-stage, Decipher (high vs low); [f] Clinical-genomic risk grouping NCCN + Decipher (intermediate vs low); [g] Clinical-genomic risk grouping NCCN + Decipher (high vs low).

Figure 11. C-statistic Forest Plot for Metastasis and Decipher

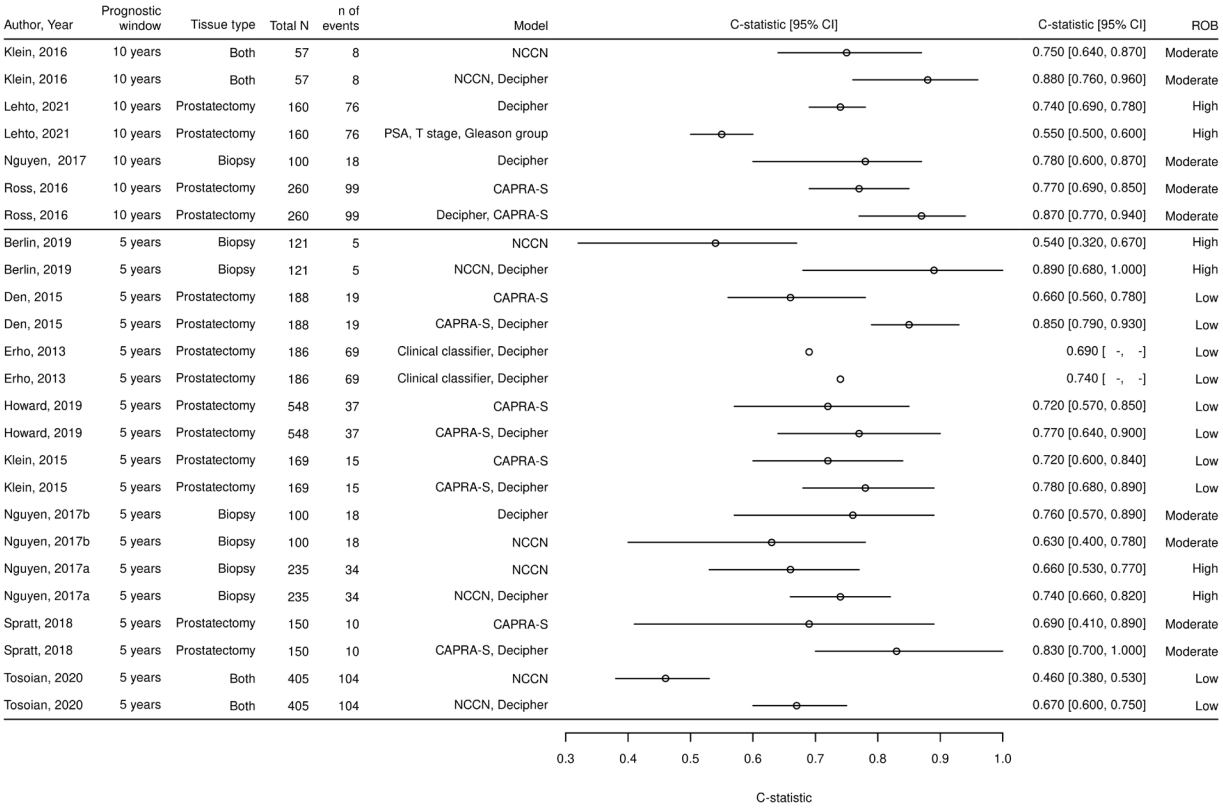


Table 11. Studies Reporting Metastases

Study Total N Design Risk of Bias (ROB)	Clinical Characteristics Tissue Source Treatment Study Duration	Model Variables	Results (95% CI)
Decipher			
Erho, 2013[6] 186 RCC Low ROB	Treated with RP with no evidence of recurrence, BCR only, or metastasis within 5 years of RP RP RP 1987-2001	Pre-op PSA Pathologic Gleason score ≥8 SVI Tumor volume Lymph node involvement Positive surgical margins ECE	OR = 1.36 (1.16, 1.60)
		Pre-op PSA Pathologic Gleason score ≥8 SVI Tumor volume Lymph node involvement Positive surgical margins ECE	AUC Clinical features 0.69 (NR) Clinical features and test 0.74 (NR)
Feng, 2021[41] 351 Retrospective ancillary study of a phase III trial Low ROB	Treated with RP and PLND with pT2N0M0 and positive surgical margins or pT3N0M0, and PSA of 0.2-4 at least 8 weeks after surgery KPS ≥ 80 No prior therapy other than short period ADT No liver disease Life expectancy of ≥10 years RP RP and RT ±2 years ADT 1998-2003	Age: ≥*65 vs <65* Race: *Black vs non-Black* Gleason score: *8-10 vs ≤7* T stage: *pT3 vs pT2* PSA at trial entry Positive surgical margins PSA nadir status *Non-nadir vs nadir (<0.5)* ADT vs placebo	HR = 1.17 (1.05, 1.32)
		Age: ≥ *65 vs <65* Race: *Black vs non-Black* Gleason score: *8-10 vs ≤ 7* T stage: *pT3 vs pT2* PSA at trial entry Positive surgical margins PSA nadir status *Non-nadir vs nadir (<0.5)* ADT vs placebo	Decipher High/intermediate vs low HR = 1.74 CI (1.08, 2.84)

Study Total N Design Risk of Bias (ROB)	Clinical Characteristics Tissue Source Treatment Study Duration	Model Variables	Results (95% CI)
Tosoian, 2020[43] 405 Retro Low ROB	cT3a-T4, or PSA >20, or Gleason Grade Group 4-5 with no neoadjuvant ADT or evidence of nodal disease prior to RP Biopsy, RP RP or RT+ADT 1995-2005	Age PSA Grade Group *4,5 vs 1-3* T stage *T2, T3/4 vs T1*	Decipher High vs low HR = 2.95 (1.79, 4.87) Intermediate vs low 1.43 (0.80, 2.53)
		NCCN *Very high vs high risk*	1.34 (1.19, 1.50)
		NCCN *Very high vs high risk*	Decipher High vs low HR = 2.95 (1.75, 4.98) Intermediate vs Low 1.56 (0.87, 2.80)
		NCCN	AUC Clinical features 0.46 (0.38, 0.53) Clinical features and test 0.67 (0.60, 0.75)
Berlin, 2019[49] 121 Retro High ROB	NCCN intermediate risk, treated with curative intent DE-IGRT without ADT Biopsy 2005-2011	NCCN *unfavorable vs favorable intermediate risk*	HR = 2.05 (1.24, 4.24)
		Age Pre-diagnostic PSA T stage *cT2b/c vs cT1/2a* ISUP grade *3 vs 1 and 2* Percent positive cores *≥ 50 vs <50*	HR = 2.07 (1.17, 5.24)
		NCCN	AUC Clinical features 0.54 (0.32, 0.67) Clinical features and test 0.89 (0.68, 1.00)
Nguyen, 2017a[55] 235 Retro High ROB	NCCN intermediate or high risk treated with RT with or without ADT or prostate cancer with adverse pathology on RP Biopsy RP or RT with or without ADT 1987-2014	Patient's age Log2 (PSA) Grade Group Clinical stage First-line treatment RP First-line treatment RT ADT	HR = 1.39 (1.09, 1.8)
		NCCN	AUC Clinical features 0.66 (0.53, 0.77) Clinical features and test 0.74 (0.66, 0.82)

Study Total N Design Risk of Bias (ROB)	Clinical Characteristics Tissue Source Treatment Study Duration	Model Variables	Results (95% CI)
Nguyen, 2017b[58] 100 Retro Moderate ROB	NCCN intermediate and high risk, treated with RT and ADT Biopsy RT+ADT 2001-2013	NCCN high	HR = 1.37 (1.06, 1.78)
Klein, 2016[22] 57 Retro Moderate ROB	Treated with RP with either pre-op PSA >20, pT3 or positive margins, or pathologic Gleason score ≥ 8 Bx, RP RP 1987-2008	CAPRA-S	HR = 1.64 (1.11, 2.42)
Ross, 2016a [59] 422 Retro Moderate ROB	Adverse pathology patients treated with RP and adjuvant RT, RT for minimal PSA, RT with higher PSA recurrence compared to patients with no RT 1990-2010	CAPRA-S, treatment (adjuvant radiation vs minimal residual disease salvage radiation, salvage radiation, no radiation)	HR = 1.28 (1.08, 1.52)
Ross, 2016b[61] 260 Retro Moderate ROB	Treated with RP with CAPRA-S score ≥ 3, pathologic Gleason score ≥ 7, and post-RP PSA nadir < 0.2 RP RP 1992-2010	CAPRA-S	HR = 1.32 (1.17, 1.51)
		CAPRA-S	AUC Clinical features 0.77 (0.69, 0.85) Clinical features and test 0.87 (0.77, 0.94)
Den, 2015[63] 188 Retro Low ROB	pT3 and/or positive surgical margins at RP treated with PORT RP RP+RT 1990-2009	CAPRA-S score	HR = 1.69 (1.24, 2.31)
		CAPRA-S score	AUC Clinical features 0.66 (0.56, 0.78) Clinical features and test 0.85 (0.79, 0.93)
Howard, 2020[44] 548 Retro Low ROB	Treated with RP with either pT3a, positive margins, SVI, or had PORT RP RP with or without RT 1989-2016	Age at RP CAPRA-S High vs low/intermediate Race: *Black vs non-Black*	Decipher Intermediate vs low HR = 6.51 (2.33, 21.8) Decipher High vs low HR = 9.60 (3.51, 32.0)
			AUC Clinical features 0.72 (0.57, 0.85) Clinical features and test 0.77 (0.64, 0.90)

Study Total N Design Risk of Bias (ROB)	Clinical Characteristics Tissue Source Treatment Study Duration	Model Variables	Results (95% CI)
Spratt, 2018a[52] 150 Retro Moderate ROB	Treated with RP with persistently detectable PSA RP RP 1990-2015	CAPRA-S Post-operative PSA	Decipher High vs low/intermediate HR = 7.12 (2.64, 21.7)
		Pre-op PSA Pathologic grade groups 4 and 5, 3 vs. 1 and 2 Positive margins Pathologic T stage T3b and T4, T3a vs T2	Decipher High vs low/intermediate HR = 5.61 (1.48, 22.7)
		CAPRA-S	AUC Clinical features 0.69 (0.41, 0.89) Clinical features and test 0.83 (0.70, 1.00)
Klein, 2015[64] 169 RCC Low ROB	Treated with RP with either pre-op PSA >20 pT3 or positive margins or Gleason score ≥ 8 and pN0, undetectable post-RP PSA, no neoadjuvant or adjuvant therapy, and a minimum of 5-yr follow-up for those who remained metastasis free RP RP 1987-2008	CAPRA-S	OR = 1.43 (1.07, 1.91)
		CAPRA-S	AUC Clinical features 0.72 (0.6, 0.84) Clinical features and test 0.78 (0.68, 0.89)
Spratt, 2018b[19] 235 Retro High ROB	PSA <200, cT1c-T3b, and cN0 Biopsy RP or RT with or without ADT 1995-2005	Treatment, stratified by institution	Clinical-genomic risk grouping Intermediate vs low HR = 22.3 (2.9, 2,863.8)
		Treatment, stratified by institution	Clinical-genomic risk grouping High vs low HR = 61.6 (8.1, 7914.9)
		6 tier clinical-genomic risk groups (10 years)	AUC 0.84 (0.61, 0.93)
Lehto, 2021[40] 160 RCC High ROB	Treated with RP with Gleason 3+4, 4+3, 4+4 and AJCC 8th ed Stage II-II without neo-adjuvant treatment RP RP 1992-2015	PSA T Stage Grade group	AUC Clinical features 0.55 (0.5, 0.6) Test only 0.74 (0.69, 0.78)

Study Total N Design Risk of Bias (ROB)	Clinical Characteristics Tissue Source Treatment Study Duration	Model Variables	Results (95% CI)
Oncotype			
Van Den Eeden, 2018[53] 259 Retro Low ROB	RP within 12 months of diagnosis Biopsy RP 1995-2005	NCCN *High vs low and very low* Intermediate vs *low and very low*	HR = 2.34 (1.42, 3.86)
		CAPRA	HR = 2.63 (1.58, 4.36)
		NCCN	AUC Clinical features 0.66 (NR) Clinical features and test 0.75 (NR)
Brooks, 2021[39] 428 Retro Low ROB	Treated with RP RP RP 1987-2004	Log2(PSA) Grade high vs low T Stage high vs low	HR = 2.24 (1.49, 3.53)
		Log2(PSA) High grade High T stage	AUC Clinical features 0.772 (NR) Clinical features and test 0.824 (NR)
Lehto, 2021[40] 160 RCC High ROB	Treated with RP with Gleason 3+4, 4+3, 4+4 and AJCC 8th ed Stage II-II without neo-adjuvant treatment RP RP 1992-2015	PSA T Stage Grade group	AUC Clinical features 0.55 (0.5, 0.6) Test only 0.65 (0.6, 0.7)
Prolaris			
Canter, 2020[46] 1062 Retro Low ROB	Patients from 4 different institutions Biopsy or RP RP or RT with or without ADT 1994-2006	CAPRA, Treatment Institutional cohort	HR = 2.21 (1.64, 2.98)
		Treatment Institutional cohort	CCR score (CAPRA and Prolaris scores) HR = 3.63 (2.60, 5.05)
		CAPRA	AUC Clinical features 0.857 (NR) CCR 0.894 (NR)

Study Total N Design Risk of Bias (ROB)	Clinical Characteristics Tissue Source Treatment Study Duration	Model Variables	Results (95% CI)
Bishoff, 2014[71] 582 Retro High ROB	Clinically localize, treated with RP Biopsy, RP RP 1994-2006	Log(1 + PSA) Gleason Score 7,>7 vs <7 Percent positive cores Adjuvant treatment Age at diagnosis	HR = 4.19 (2.08, 8.45)
Canter, 2019[48] 767 Retro Low ROB	Prostate adenocarcinoma with PSA < 100, cT1-T3M0 that did not undergo TURP, cryosurgery, or laser ablation Biopsy RP, RT with or without ADT, ADT or none 2006-2011	CAPRA Ancestry: *Black vs non-Black* Primary treatment	HR = 2.03 (1.47, 2.78)
		CAPRA	AUC Clinical features 0.88 (NR) Clinical features and test 0.90 (NR)
Lehto, 2021[40] 160 RCC High ROB	Treated with RP with Gleason 3+4, 4+3, 4+4 and AJCC 8th ed Stage II-II without neo-adjuvant treatment RP RP 1992-2015	PSA Stage Grade group	AUC Clinical features 0.55 (0.5, 0.6) Test only 0.73 (0.69, 0.78)

Abbreviations. ADT=androgen deprivation therapy; AJCC=American Joint Committee on Cancer; AUC=area under the curve; BCR=biochemical recurrence; CAPRA=Cancer of the Prostate Risk Assessment; ECE=extra-capsular extension; NCCN=National Comprehensive Cancer Network; HR=hazard ratio; PLND=pelvic lymph node dissection; PORT=post-operative radiation therapy; PSA=prostate-specific antigen; RP=radical prostatectomy; RT=radiation therapy; SVI=seminal vesicle involvement.

Prostate-cancer-specific Mortality

Decipher

Five studies (3 low ROB,[41,44,66] 1 moderate ROB,[54] and 1 high ROB[40]) addressed the impact of the Decipher score on the prediction of prostate-cancer-specific mortality in addition to standard clinical or pathologic features.[40,41,44,54,66] Two low ROB studies (N = 538) examined the additive benefit of Decipher with this outcome among patients post-prostatectomy and reported similar HRs of 1.81 (95% CI [1.48, 2.25]) and 1.39 (95% CI [1.20, 1.63]) (Figure 12 and Table 12). The third low ROB study was an exclusively Veteran cohort (N = 548) who underwent prostatectomy between 1989 and 2016. In models with CAPRA-S and race (Black vs non-Black), Decipher was rounded to have HRs of 25.5 (95% CI [2.84, 3,365]) and 56.0 (95% CI [6.82, 7,297]) for intermediate and high risk Decipher scores, compared to low risk, respectively.[44] Of note, this low ROB study had an event rate of 12 for PCSM, leading to large CIs. This corresponded with a non-significant increase in AUC from 0.81 (95% CI [0.63, 0.95]) with CAPRA-S alone to 0.856 (95% CI [0.71, 0.98]) with CAPRA-S and Decipher (Figure 13). Finally, the ancillary study of RTOG 9601 discussed previously assessed prostate-cancer-specific mortality in men following prostatectomy. Models including both clinical and pathologic features demonstrated HRs of 1.39

(95% CI [1.20, 1.63]) and 2.94 (95% CI [1.57, 5.81]) with Decipher as a continuous and categorical variable (high or intermediate versus low), respectively.[41] Of note, the study by Feng et al also showed the prognostic ability of the Decipher score for overall survival, with an HR of 1.17 (95% CI [1.06, 1.29]) in a model similar to those for prostate-cancer-specific mortality.[41]

The 1 moderate ROB study included 561 men who underwent prostatectomy (113 of whom were Veterans). In a model including CAPRA-S, increases of 0.1 unit in Decipher score had a greater odds of prostate-cancer-specific mortality (OR = 1.34, 95% CI [1.2, 1.5]).[54] This was associated with a non-significant increase in AUC from 0.73 (95% CI [0.68, 0.78]) to 0.76 (95% CI [0.71, 0.82]) when incorporating Decipher with CAPRA-S. Concerns for potential ROB in this study were moderate for both study participation by eligible individuals and prognostic factor measurement.

In the same high ROB study by Lehto et al discussed above which evaluated AUCs for all 3 tests, the AUC for prostate-cancer-specific mortality did significantly increase from 0.55 (95% CI [0.49, 0.6]) with clinical features to 0.72 (95% CI [0.66, 0.77]) with inclusion of the Decipher score.[40]

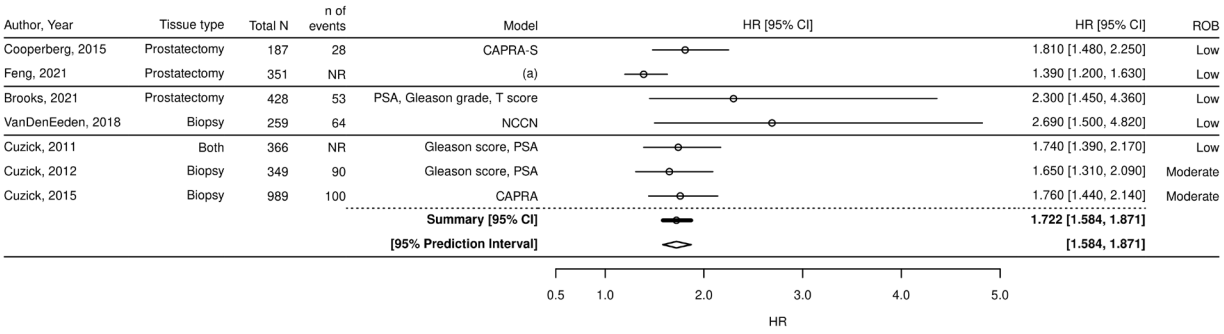
Oncotype

Three studies, all retrospective (2 low ROB and 1 high ROB) and in men who underwent prostatectomy, evaluated the Oncotype score's additive prognostic value for prostate-cancer-specific mortality.[39,40,53] Each showed an increase in the AUC: from 0.71 with NCCN classification alone to 0.81 with the addition of the Oncotype score in Van den Eeden et al,[53] from 0.762 to 0.822 (95% CIs not reported) in Brooks et al,[39] and from 0.55 (95% CI [0.49, 0.6]) to 0.69 (95% CI [0.63, 0.74]) with the addition of the Oncotype score to models with clinical features in the high ROB study by Lehto et al.[40] The HRs for prostate-cancer-specific mortality were reported in 2 of the publications, 2.69 (95% CI [1.50, 4.82]) in a model with NCCN in Van den Eeden et al[53] and 2.30 (95% CI [1.45, 4.36]) in a model with clinical features in Brooks et al.[39]

Prolaris

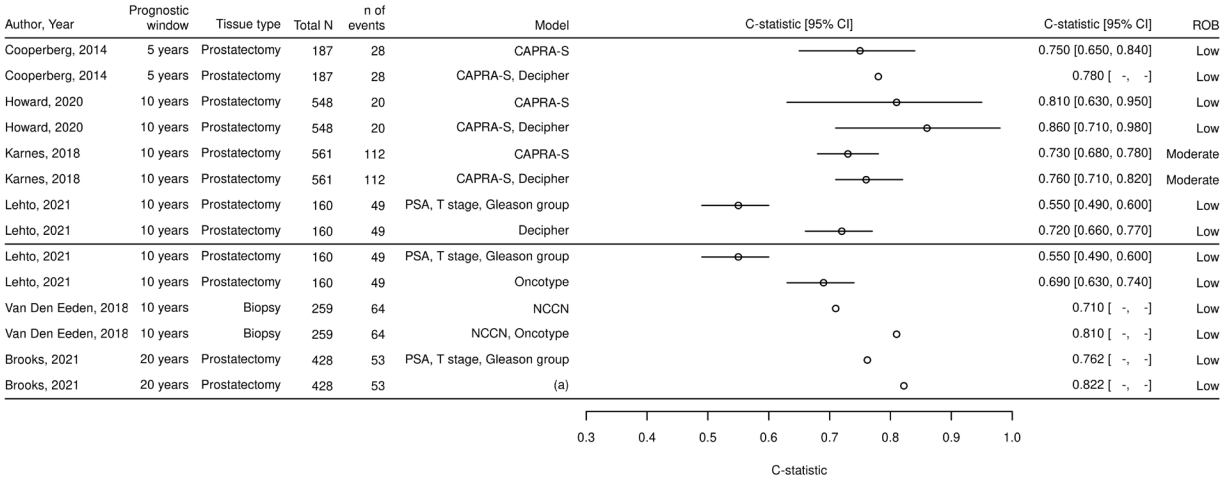
Four studies assessed the additive prognostic value for prostate-cancer-specific mortality by the Prolaris score or CCP.[4,40,62,69] Three Prolaris studies (N = 1,675) contributed to a meta-analysis of the additive prognostic effect and reported a summary HR of 1.72 (95% CI [1.58, 1.87]; 95% PI [1.58, 1.87]). Three of these were by Cuzick et al.[4,62,69] The first was a low ROB study published in 2011, and included a cohort of 337 men diagnosed by TURP; within this cohort the CCP score had an HR for prostate-cancer-specific mortality of 2.57 (95% CI [1.93, 3.43]) when included in a model with clinical features.[4] The second moderate ROB study by Cuzick et al, from 2012, evaluated CCP scores derived from biopsy specimens in a similar manner to the first, showing an HR of 1.65 (95% CI [1.31, 2.09]) for prostate-cancer-specific mortality in a model with clinical features.[69] The final Cuzick et al report from 2015 (low ROB), in a separate cohort of patients with biopsy specimens from the 2012 manuscript, demonstrated an HR of 1.76 (95% CI [1.44, 2.14]) in a model with CAPRA. In addition, the AUC was reported to increase from 0.74 to 0.78 with the addition of the CCP score to the CAPRA score.[62] A significant increase in AUC was also reported by Lehto et al, from 0.55 (95% [CI 0.49, 0.60]) to 0.66 (95% CI [0.61, 0.74]), with the addition of the CCP score to clinical features in a cohort of 160 men who underwent prostatectomy.[40]

Figure 12. Hazard Ratio Forest Plot for Prostate-cancer-specific Mortality by Test Type (Decipher, Oncotype, Prolaris)[a]



Notes. [a] Model includes age (≥65 vs <65), Black men vs non Black men, Gleason score (8-10 vs ≤7), T stage (pT3 vs pT2), PSA at trial entry, positive surgical margins, PSA nadir status (non-nadir vs nadir <0.5), ADT vs placebo.

Figure 13. C-statistic Forest Plot for Prostate-cancer-specific Mortality by Test Type (Decipher, Oncotype)



Notes. [a] Model includes PSA, T stage, Gleason group, Oncotype.

Table 12. Studies Reporting Prostate-cancer-specific Mortality

Study Total N Design ROB	Clinical Characteristics Tissue Source Treatment Study Duration	Model Variables	Results (95% CI)
Decipher			
Karnes, 2018[54] 561 Retro Moderate ROB	Treated with RP with pT3, pN1, positive margins, or Gleason score >7 RP RP 1987-2010	CAPRA-S	OR = 1.34 (1.20, 1.50)
		CAPRA-S	AUC Clinical features 0.73 (0.68, 0.78) Clinical features and test 0.76 (0.71, 0.82)
Feng, 2021[41] 351 Ancillary study of a phase III trial Low ROB	Treated with RP and PLND with pT2N0M0 and positive surgical margins or pT3N0M0, and PSA of 0.2-4 at least 8 weeks after surgery KPS ≥ 80 No prior therapy other than short period ADT No liver disease Life expectancy of ≥ 10 years RP RP and RT ±2 years ADT 1998-2003	Age: ≥ 65 vs <65 Race: Black vs non-Black Gleason score: 8-10 vs ≤7 T stage: pT3 vs pT2 PSA at trial entry Positive surgical margins PSA nadir status Non-nadir vs nadir (<0.5) ADT vs placebo	HR = 1.39 (1.20, 1.63)
		Age: ≥65 vs <65 Race: Black vs non-Black Gleason score: 8-10 vs ≤7 T stage: pT3 vs pT2 PSA at trial entry Positive surgical margins PSA nadir status Non-nadir vs nadir (<0.5) ADT vs placebo	Decipher High/intermediate vs low HR = 2.94 (1.57, 5.81)
Cooperberg, 2015[66] 187 Retro Low ROB	Treated with RP with pre-op PSA >20, Gleason score ≥ 8 or pT3b without neo-adjuvant therapy, metastatic disease or lack of a PSA nadir post-RP RP RP 2000-2006	CAPRA-S Adjuvant therapy (RT or ADT)	HR = 1.81 (1.48, 2.25)
		CAPRA-S	AUC Clinical features 0.75 (0.65, 0.84) Clinical features and test 0.78 (NR)

Study **Total N** **Design** **ROB**	**Clinical Characteristics** **Tissue Source** **Treatment** **Study Duration**	**Model Variables**	**Results (95% CI)**
Howard, 2020[44] 548 Retro Low ROB	Treated with RP with either pT3a, positive margins, SVI, or had PORT RP RP with or without RT 1989-2016	Age at RP CAPRA-S High vs low/intermediate Race: Black vs non-Black	Decipher Intermediate vs low HR = 25.5 (2.84, 3365) Decipher High vs low HR = 56.0 (6.82, 7297)
		CAPRA-S	AUC Clinical features 0.81 (0.63, 0.95) Clinical features and test 0.86 (0.71, 0.98)
Lehto, 2021[40] 160 RCC High ROB	Treated with RP with Gleason 3+4, 4+3, 4+4 and AJCC 8th ed Stage II-II without neo-adjuvant treatment RP RP 1992-2015	PSA T Stage Grade group	AUC Clinical features 0.55 (0.49, 0.6) Test only 0.72 (0.66, 0.77)
Oncotype			
Van Den Eeden, 2018[53] 259 Retro Low ROB	RP within 12 months of diagnosis Biopsy RP 1995-2005	NCCN	HR = 2.69 (1.50, 4.82)
		NCCN	AUC Clinical features 0.71 (NR) Clinical features and test 0.81(NR)
Brooks, 2021[39] 428 Retro Low ROB	Treated with RP RP RP 1987-2004	Log2(PSA) Grade *high vs low* T stage *high vs low*	HR = 2.30 (1.45, 4.36)
		Log2(PSA) High grade High T stage	AUC Clinical features 0.762 (NR) Clinical features and test 0.822 (NR)

Study Total N Design ROB	Clinical Characteristics Tissue Source Treatment Study Duration	Model Variables	Results (95% CI)
Lehto, 2021[40] 160 RCC High ROB	Treated with RP with Gleason 3+4, 4+3, 4+4 and AJCC 8th ed Stage II-II without neo-adjuvant treatment RP RP 1992-2015	PSA T stage Grade group	AUC Clinical features 0.55 (0.49, 0.6) Test only 0.69 (0.63, 0.74)
Prolaris			
Cuzick 2015[62] 989 Retro Low ROB	Clinically localized, with age <76 and excluding metastatic disease or PSA >100 or treatment within 6 months of diagnosis Biopsy NR 1990-2003	CAPRA	HR = 1.76 (1.44, 2.14)
		CAPRA	AUC Clinical features 0.74 (NR) Clinical features and test 0.78 (NR)
Cuzick, 2012[69] 349 Retro Moderate ROB	Clinically localized, with age <76 and excluding metastatic disease or treatment within 6 months of diagnosis Biopsy NR 1990-1996	Log(1+PSA) Gleason score <7, >7 vs 7	HR = 1.65 (1.31, 2.09)
Cuzick, 2011[4] 366 Retro Low ROB	RP without neoadjuvant therapy RP RP 1985-1995	Log(1+baseline PSA) Gleason score 7,>7 vs <7	HR = 1.74 (1.39, 2.17)
Canter, 2019[48] 767 Retro Low ROB	Prostate adenocarcinoma with PSA < 100, cT1-T3M0 that did not undergo TURP, cryosurgery, or laser ablation Biopsy RP, RT with or without ADT, ADT or none 2006-2011	CAPRA	AUC Clinical features 0.91 (NR) Clinical features and test 0.94 (NR)
Lehto, 2021[40] 160 RCC High ROB	Treated with RP with Gleason 3+4, 4+3, 4+4 and AJCC 8th ed Stage II-II without neo-adjuvant treatment RP RP 1992-2015	PSA Stage Grade group	AUC Clinical features 0.55 (0.49, 0.6) Test only 0.66 (0.61, 0.74)

Other Reported Outcomes

Decipher

Three additional studies assessed alternative or composite endpoints with Decipher testing (Table 13).[37,38,57,60] A low ROB study of 241 patients treated with definitive radiation or prostatectomy from the Decipher GRID registry were evaluated for time-to-treatment failure defined as biochemical recurrence or initiation of salvage therapy after definitive treatment.[38] The HR for time to treatment failure was 2.98 (95% CI [1.22, 7.29]) in a model containing NCCN risk classification and other clinical features. Of interest, Vince et al is the only study addressing KQ3 that includes patients undergoing genomic classifier testing at the time of diagnosis and not a cohort aggregated from banked tissue specimens.[38]

In the low ROB study by Glass et al discussed previously, clinical recurrence was assessed, and although noted to be distinct from biochemical recurrence, was not clearly defined.[60] Regardless, in a model with CAPRA-S Decipher did show a significant HR for clinical recurrence of 1.48 (95% CI [1.09, 2.01]) with a non-significant increase in the AUC from 0.73 (95% CI [0.49, 0.95]) for clinical features alone to 0.84 (95% CI [0.7, 0.96]) with clinical features and the Decipher score. A second, moderate ROB study also assessed the time to clinical recurrence with Decipher.[57] In this study, time to clinical recurrence was a composite endpoint of biopsy proven prostate bed recurrence or development of regional or distant metastatic disease on imaging. Among 512 men (including 104 Veterans) who had undergone prostatectomy, Decipher as a categorical variable was found to have HRs of 1.40 (95% [CI 0.7, 2.74]) and 2.93 (95% CI [1.58, 5.55]) for intermediate and high risk Decipher scores, respectively, in a model with clinical features. This corresponded to a non-significant increase in the AUC from 0.79 (95% CI [0.73, 0.86]) to 0.85 (95% CI [0.80, 0.89]) when adding in Decipher results.

Last, time to secondary therapy was reported in a moderate ROB study among patients treated with radical prostatectomy.[37] In a model including age, PSA, pathological grade group, positive surgical margins, extraprostatic extension, and seminal vesicle invasion, the HR for Decipher was 1.46 (1.34 to 1.66).

Oncotype

No studies with Oncotype were identified that evaluated endpoints other than BCR, metastases, or prostate-cancer-specific mortality.

Prolaris

One low ROB retrospective study employing Prolaris among 424 patients treated with radical prostatectomy also reported a composite endpoint of metastasis or prostate-cancer-specific mortality.[50] In a model including CAPRA-S and CCP scores, the HR for Prolaris was 2.15 (95% CI [1.36, 3.39]) for this composite endpoint.

Table 13. Studies Reporting Other Outcomes

Study Total N Design Risk of Bias (ROB)	Clinical Characteristics Tissue Source Treatment Study Duration	Model Variables	Results
Decipher			
Time to Treatment Failure			
Vince, 2021[38] 241 Prospective Low ROB	Clinically localized, underwent testing as part of routine clinical care and were able to be matched with Decipher GRID registry Biopsy RP or RT 2015-2019	NCCN Age Log(PSA) Log(prostate volume) BMI Percent positive cores	HR = 2.98 (1.22 to 7.29)
Time to Secondary Therapy			
Shahait, 2021[37] 398 Prospective Moderate ROB	Treated with RP RP RP 2013-2018	Age Log2(pre-op PSA) Pathological GG 4-5 vs 1-3 Positive surgical margins EPE SVI	HR = 1.49 (1.34 to 1.66)
Time to Clinical Recurrence			
Dalela, 2017[57] 512 Retrospective Moderate ROB	Treated with RP with ≥pT3a, positive margins, and/or lymph node invasion who achieved a PSA nadir after RP RP RP 1990-2010	Log2(PSA) T stage: pT3a, pT3b-4 vs pT2 Pathologic Gleason score: 8-10 vs ≤7 Lymph node invasion Surgical margins Adjuvant RT Adjuvant ADT	Decipher Intermediate vs low risk HR = 1.40 (0.7, 2.74) High vs low risk HR = 2.93 (1.58 to 5.55)
		Log2(PSA) T stage pT3a, pT3b-4 vs pT2	AUC Clinical features 0.79 (0.73 to 0.86) Clinical features and test 0.85 (0.8 to 0.89)

Study Total N Design Risk of Bias (ROB)	Clinical Characteristics Tissue Source Treatment Study Duration	Model Variables	Results
		Pathologic Gleason score: 8-10 vs ≤7 Lymph node invasion Surgical margins Adjuvant RT Adjuvant ADT	
Glass, 2016[60] 224 Retrospective Low ROB	RP with high-risk pre-op features (PSA > 20 or GS ≥ 8) pT3, or +SM RP RP 2001-2013	CAPRA-S Age at diagnosis	HR = 1.48 (1.09 to 2.01)
		CAPRA-S	AUC Clinical features 0.73 (0.49 to 0.95) Clinical features and test 0.84 (0.7 to 0.96)
Oncotype (No Studies)			

Prolaris			
Metastasis or Prostate-cancer-specific Mortality			
Leapman, 2018[50] 424 Retrospective Low ROB	Treated with RP Biopsy RP Prior to 2017	CAPRA-S	HR = 2.15 (1.36 to 3.39)
		CAPRA-S	AUC Clinical features and test 0.81 (NR)

DISCUSSION

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 impact on risk reclassification reported across studies, there was no clear pattern in these changes across tests. We did find that there was no change in risk classification for a majority of patients apart from a potentially greater rate of reclassification among those at intermediate risk by clinical features. Despite the large proportion of patients without a change, across the identified studies there were still clinically meaningful proportions of included patients who experienced a change in risk assessment that could contribute to important changes in treatment. Of note, most of the data on risk reclassification have been generated with the Oncotype test and were almost exclusively related to risk assessment at the time of initial diagnosis. 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 this was not found in the single randomized trial. 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. Last, we found that these tests do seem to provide additional prognostic information with respect to biochemical recurrence, development of metastatic disease, and prostate-cancer-specific mortality; we have the most certainty of this effect with Decipher compared to the other 2 tests. The value of that additional prognostic information is limited by these findings that 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. Of note, we did not find any evidence of acute harms of the tests studied, although there is likely limited harm as the test does not require new tissue acquisition and does not identify or disclose genetic risk applicable to patient family members.

While not specified in our KQs, an outcome of interest for the nominators of this topic was evidence related to the cost and economic value of genomic tests in the management of prostate cancer. Unfortunately, none of the identified studies reported cost-related outcomes. One recent systematic review on health economic evidence for both liquid and tissue-based molecular tests provides us with some cost-related information. Four of the 22 studies included by Degeling et al were relevant to the tests of concern in this review.[72] They reported that 3 of the 4 studies found that these tests led to increased overall costs but concluded that they were cost effective when considering improved clinical outcomes; 1 study found cost savings for low-risk populations and increased costs among intermediate-risk patients. Lobo et al provide some guidance in this area using model simulation based on individualized decision analysis to estimate additional quality-adjusted life years (QALYs) based on genomic classifier test risk estimates in patients post-prostatectomy.[73] Specifically, they found an additional 0.07 QALYs with use of genomic classifier testing.

CERTAINTY OF EVIDENCE FOR KEY OUTCOMES

To provide context for the findings described in this report, we conducted certainty of evidence (COE) ratings for those outcomes with adequate volume and comparability of relevant studies. These ratings reflect the degree of confidence we have for the summary findings. We made our COE assessments by genomic classifier test type across the KQ3 outcomes of biochemical

recurrence (BCR), metastases, and prostate-cancer-specific mortality (Table 14). Overall, we noted that while the effect estimates were consistent in showing a clinically relevant additive benefit of the genomic tests, our confidence assessments were frequently downgraded because most identified studies used older data that have limited relevance to modern clinical practice (indirectness).

For Decipher, we have low COE that this test provides additional prognostic information for risk of BCR, metastases, and prostate-cancer-specific mortality. For BCR, this determination was limited by ROB across the 4 relevant studies and imprecision of effect estimates. For metastases, while 15 observational studies contributed data to this effect estimate, our assessment was downgraded for ROB and serious indirectness, as much of the contributing patient data from the 1980s to 1990s reflect a different era of management standards. For prostate-cancer-specific mortality, our assessment of the findings from 5 studies was downgraded due to indirectness as noted for metastases and the imprecision of the effect estimate. Additional prospective studies with data drawn from current era of prostate cancer management could change this assessment.

For Oncotype, we have very low COE across all 3 outcomes. Indirectness due to temporal source of the data was noted for each outcome. In addition, for BCR, we downgraded for imprecision and inconsistency across the 3 relevant studies. For metastases and prostate-cancer-specific mortality, the same 3 studies contributed data, and we downgraded both of these outcomes for ROB and imprecision. Additional studies, especially prospective, conducted in the current management era would affect this assessment.

For Prolaris, we have very low COE across all 3 outcomes. Our assessment for all 3 outcomes was downgraded for ROB, indirectness, imprecision, and concerns for potential publication bias (BCR and metastases only). Of note, only 4 studies contributed relevant data. Additional studies, especially prospective studies conducted in the current management era would affect this assessment.

Table 14. Certainty of Evidence for Genomic Tests and Biochemical Recurrence, Metastasis, and Prostate-cancer-specific Mortality

Outcome	Number of Studies	Findings	Certainty of Evidence (Rational)
Decipher			
Biochemical recurrence	4 observational studies (525 patients)	HR range (0.32 to 1.36) AUC range clinical features (0.56, 0.64) AUC range clinical features and genomic test (0.69 to 0.85)	**Low certainty** (Downgraded for serious risk of bias and serious imprecision)
Metastases	15 observational studies (3,165 patients)	HR range (1.17 to 61.6) OR range (1.36, to 1.48) AUC range clinical features (0.46 to 0.88) AUC range clinical features and genomic test (0.67 to 0.89)	**Low certainty** (Downgraded for serious risk of bias and serious indirectness)
Prostate-cancer-specific mortality	5 observational studies (1,807 patients)	HR range (1.39 to 56.0) OR range (1.20) AUC range clinical features (0.55 to 0.81) AUC range clinical features and genomic test (0.71 to 0.78)	**Low certainty** (Downgraded for serious indirectness and serious imprecision)
Oncotype			
Biochemical recurrence	3 observational studies (876 patients)	HR range (1.10 to 2.73) AUC range clinical features (0.59) AUC range clinical features and genomic test (0.68)	**Very low certainty** (Downgraded for serious inconsistency, serious indirectness, and serious imprecision)
Metastases	3 observational studies (793 patients)	HR range (2.24 to 2.63) AUC range clinical features (0.55 to 0.77) AUC range clinical features and genomic test (0.65 to 0.824)	**Very low certainty** (Downgraded for serious risk of bias, serious indirectness, and serious imprecision)
Prostate-cancer-specific mortality	3 observational studies (847 patients)	HR range (2.69, 2.30)	**Very low certainty** (Downgraded for serious risk of bias, serious

Outcome	Number of Studies	Findings	Certainty of Evidence (Rational)
		AUC range clinical features (0.55 to 0.762) AUC range clinical features and genomic test (0.69 to 0.822)	indirectness, serious imprecision)
Prolaris			
Biochemical recurrence	9 observational studies (2,758 patients)	HR range (1.24 to 10.9) AUC range clinical features (0.542 to 0.78) AUC range clinical features and genomic test (0.65 to 0.86)	**Very low certainty** (Downgraded for very serious risk of bias, serious indirectness, serious imprecision, and suspected publication bias)
Metastases	4 observational studies (2,571 patients)	HR range (2.05 to 4.19) AUC range clinical features (0.55 to 0.894) AUC range clinical features and genomic test (0.90) Test only (0.73)	**Very low certainty** (Downgraded for serious risk of bias, serious inconsistency, serious indirectness, and serious imprecision)
Prostate-cancer-specific mortality	3 observational studies (1,989 patients)	HR range (1.65 to 2.57) AUC range clinical features (0.74, 0.55) AUC range clinical features and genomic test (0.78) AUC test only (0.66)	**Very low certainty** (Downgraded for serious risk of bias, serious inconsistency, very serious indirectness, and serious imprecision)

Abbreviations. AUC=area under the curve; OR=odds ratio; HR=hazard ratio.

CLINICAL AND POLICY IMPLICATIONS

Identifying patients who would benefit from treatment and tailoring the intensity of that treatment to an individual patient remains a major challenge for prostate cancer management despite multiple existing risk stratification tools. Previously developed risk stratification (or prognostic) tools utilize combinations of readily available clinical parameters including tumor grade, PSA level, and clinical stage to provide some assessment of individual patient prognosis; however, these tools have limited prognostic ability. The purpose of this review has been to determine whether 3 commercially available genomic classifiers can refine prognosis assessment and lead to more personalized management of patients with prostate cancer.

KQ1

Among individuals with localized prostate cancer who are considering first-line definitive treatment, does genomic testing impact risk stratification? At present, the most commonly used clinical risk stratification incorporates clinical tumor stage, Gleason score or grade group classification, as well as serum PSA. The rationale for pursuing individualized tumor characterization through genomic classification testing is that it may further refine risk stratification. The majority of studies addressing KQ1 utilized the Oncotype genomic classifier, were retrospective in design, and most frequently utilized NCCN risk stratification as a comparator. For patients with very low to intermediate risk disease, more often than not, there was no significant change in risk stratification compared to the clinical classification scheme. In this patient population, genomic testing may only serve to potentially reassure these patients that active surveillance remains a reasonable option for them. Similarly, for those already treated with prostatectomy, most patients were classified into the same or lower risk classification between clinical risk factors and genomic classifier testing. For patients classified pre-treatment as intermediate risk, Oncotype testing could potentially reassure those who are averse to treatment that active surveillance remains a reasonable option. There were limited available data on change in risk stratification by race; however, what we did identify did not appear to be significantly influenced by race, providing some reassurance that these results appear applicable across different races.

KQ2

How has genomic testing impacted the choice of treatment intensity to date? Across observational studies, when testing occurred at the time of diagnosis, there have been more frequent recommendations for active surveillance. This would be in line with the finding that most patients tend to be classified at similar or lower risk levels after genomic testing. However, in the single randomized trial evaluating the incorporation of Oncotype testing into treatment decisions, there was no statistically significant effect of receipt of testing on treatment choice. Although there was a tendency to recommend adjuvant therapy when patients were classified as higher risk by Decipher at the time of prostatectomy, there was no clear pattern of adjuvant treatment or surveillance after prostatectomy based on genomic classifier use. Because adjuvant therapies such as radiation with or without hormonal therapy can be associated with increased long-term side effects and appear equivalent in outcomes to early salvage radiation at the time of biochemical recurrence for a large portion of patients, further investigation into use in this setting to identify those at most need of additional therapy would seem warranted. We note that evidence of change in management after testing primarily reflects current practice patterns surrounding test use and is only helpful if occurring in the context of a test with evidence of acceptable predictive ability.

KQ3

Finally, among individuals who have undergone definitive treatment for localized prostate cancer, can genomic testing of the pre-treatment biopsy or radical prostatectomy specimen refine prognosis in terms of biochemical recurrence-free survival, metastases-free survival, and prostate-cancer-specific mortality beyond clinical classification schemes? Overall, genomic classifiers showed meaningful additive value in prognostic ability beyond the previously available clinical models and by improvements in discriminatory characteristics.[74] However, only 2 of the 39 studies included prospectively collected cohorts, and only 9 of the 39 studies

included patients treated with definitive radiation therapy. In addition, patients in these studies were diagnosed and managed over several decades, while screening patterns, pathologic grading and biopsy methodology, radiological staging utilizing pelvic MRI, treatment recommendations and techniques, and follow-up of patients with prostate cancer have evolved significantly. Even more contemporary changes, including staging with PET imaging and artificial-intelligence-generated classification schemes, warrant consideration as the prognostic utility of genomic classifiers continues to be assessed. In addition, while these studies demonstrated improved prognostic ability of the genomic classifiers over clinical models, they were not designed to predict response to therapy, and therefore these findings are insufficient to guide intensification or de-escalation of treatment despite the demonstration of change in treatment intensity shown in KQ2.

PRIOR SYSTEMATIC REVIEWS

KQ1

Only 1 prior systematic review explicitly discussed any of the genomic classifiers of interest in reference to risk reclassification when compared to existing clinical risk models.[75] Olleik et al reviewed 4 studies reporting reclassification with use of Decipher, 1 with Oncotype, and 3 with Prolaris, mostly in the post-radical prostatectomy setting. Six of the articles in the Olleik review were included in this report, and we additionally included 6 studies related to Oncotype. Estimates of frequency of reclassification as described by Olleik et al displayed a similar variability across studies but were not described by baseline risk classification as we have done.

KQ2

Three recent systematic reviews evaluated the extent to which tissue-based genomic testing impacts the choice of treatment intensity.[75-77] All published since 2017, these reviews include a total of 16 articles, all of which except 3 were eligible for our review. All 3 reviews used a formalized risk of bias tool to assess article quality. The 3 reviews included articles about Prolaris in low- and intermediate-risk localized prostate cancer and summarize these studies about change in treatment before and after Prolaris CCP results as ranging between 48% and 65%, including a portion where interventional treatment was reduced (37% to 72%) or increased (23% to 27%). Two reviews suggested similar findings (*eg*, a 21% to 24% decrease in interventional treatment). The Decipher test was examined in only 1 review (which included 4 studies about post-prostatectomy, all of which were included in our review), and authors summarized that the test's clinical utility was reflected in a change of post-prostatectomy treatment recommendations in 31% to 51% of the time, with 16% to 42% changing from any to no treatment.[75]

Our report also includes a randomized trial from 2021, which, interestingly, found no change in treatment in men randomized to receipt of genomic testing with Oncotype during treatment decision-making. We also explored testing timing (post-initial biopsy and post-prostatectomy), more detailed treatment changes as a result of genomic testing, and treatment changes by race subgroups.

KQ3

Six recent systematic reviews examined 1 or more of our 3 outcomes of interest[76,78-82] relevant to the prognostic effect of tissue-based genomic tests after adjusting for existing prognostic clinical features following definitive treatment.

Two systematic reviews examined only the Decipher test.[80,82] The remaining reviews included 17[79] and 21[76] studies about the utility of adding or incorporating genomic classifiers into clinical risk classification schemes to enhance prognostic accuracy across various disease outcomes. Our study improves upon this recent work, as we have included 39 studies, including the majority of the articles about all 3 genomic classifier tests from the last review,[76] with an additional 12 articles from after 2019, beyond the end of the Fine et al search period.

In terms of substantive findings, each review has noted that genomic classifier tests have been able to improve on prediction of clinical outcomes compared to clinical features alone. For example, Jairath et al found a consistent independent association between the Decipher test results and various endpoints, including BCR, metastases, and prostate-cancer-specific mortality.[82] For all 3 KQs, our review focused on Decipher, Oncotype, and Prolaris; however, some of the prior reviews included other tests, and few included all 3 tests. A few early reviews summarized the results of this rapidly changing field, but were not systematic reviews[81] or did not include a formal risk of bias assessment.[78,81]

Overall, our findings are largely consistent with prior reviews in that these tests provide additive information to existing clinical risk stratification tools related to reclassification, may change the treatment plan or actual treatment for some prostate cancer patients, and better predict clinical outcomes. Our review adds to these prior reviews with a significantly increased number of studies, more recent studies, formal risk of bias assessment for all included studies, and exploration of test effects by key subgroups.

LIMITATIONS

There are limitations to our review. While we developed an *a priori* protocol outlining populations of interest and a standardized approach to searching and evaluating the literature, we limited this evaluation to 3 commercially available genomic tests. Thus, our findings do not apply to other tests of genomic markers in the context of prostate cancer. In addition, we limited our search to publications after January 1, 2010, so any earlier publications related to the development of genomic tests of interest would not be captured.

We did not consider outcomes specific to the time of surgery such as adverse pathology, nor did we consider any literature related to use of these tests among patients with metastatic disease. Of note, there is ongoing investigation exploring the relationship between specific histologic distinctions, radiomics, and classification schemes derived by artificial intelligence in relation to prognosis; however, our review did not incorporate these burgeoning areas of scientific inquiry. We also note that there are current ongoing studies to assess the ability of genomic classifier tests in predicting response to treatment options, which was not within the scope of this review although at least 1 of the included studies reported some preliminary findings related to this question.[41] In an attempt to capture the breadth of existing evidence, we included studies with notable aspects of heterogeneity, as noted below, which likely reduces our ability to identify a

narrow effect estimate which could be applied to clinical practice. Finally, we were unable to compare across genomic test types as no identified literature provided a direct comparison.

Publication Bias

Regarding the identified relevant literature, we consider important limitations by potential source of bias. Much of the identified literature for KQ3 was conducted as retrospective cohort studies from individual or grouped institutional data from previously treated patients; many of these were from the same institutions (as noted by multiple linked studies evaluating the same cohort data).[4,22,27,36,56,64,69,71] Many of the included studies may overlap substantially, although in some cases, the amount of overlap is unclear.[52,57,59,61,63,77] It is unknown if other institutions conducted retrospective analyses of their own patient populations, which may have had different practice patterns, including patterns in which different patients underwent genomics tests. In addition, many of the included studies were supported by the companies that developed one of the 3 genomic tests of interest. It is notable that the bulk of the outcomes data for KQ3 for the Oncotype and Prolaris tests were among earlier endpoints such as biochemical recurrence, while Decipher-based studies were the predominant test studied for data with later or "harder" outcomes such as prostate-cancer-specific mortality. If studies for the other 2 tests of interest have been conducted with longer-term outcomes, we were unable to identify them in the published literature. It is possible that studies of Oncotype and Prolaris with longer-term outcomes have not been completed yet. We note that there are 4 Prolaris studies registered in clinicaltrials.gov that appear to have completed data collection but are without peer-reviewed publications or posted results[83-85] or were terminated due to poor enrollment.[86]

Study Quality

There are notable concerns related to study quality of the identified literature. For KQ2, study designs used to determine the impact of genomic classifier testing on treatment recommendations were primarily observational, with case reviews in abstract from actual clinical care or provider self-report of care recommended before and after receiving test results; we found only 1 randomized trial addressing KQ2.

For KQ3, many studies were retrospective and could not control for practice patterns. In addition, many studies employed genomic classifier tests run on stored biopsy or prostatectomy tissue, some of which could be up to 30 years old at the time of analysis. Thus, rates of unusable specimens due to inadequate tissue sample were often substantial. It is suspected that older samples and those from patients with earlier or more favorable stage cancers, with lower cancer burden, may have been more likely to have samples inadequate for genomic testing, which could bias the results to find a greater relationship between test scores and worse prognosis. We considered a loss of 20% to 30% to be acceptable given the context for these studies[87]; loss of sample at higher levels was a common cause of downgrading for risk of bias. Follow-up among studies pertinent to KQ3 was limited (the majority was less than 10 years) relative to the natural history of prostate cancer. Additional common sources of bias across included studies was lack of clarity around how the sampled population in retrospective studies were identified and sampled and potential confounding due to factors driving which patients received the test.

Heterogeneity

Potential sources of heterogeneity in effects include study design (*eg*, cohort, nested case-control, case-control), use of different clinical risk classification systems as comparators (*eg*, NCCN, CAPRA, AUA), different approaches to primary definitive treatment (*eg*, radical prostatectomy vs primary radiation therapy with or without hormonal treatment), and different definitions of the outcomes of interest (*eg*, various ways of measuring changes in treatment intensity, as well as biochemical recurrence). In addition, screening patterns have changed, and effective treatment options for prostate cancer have improved significantly over the last few decades; yet this literature draws from patient care provided between 1985 and 2019, which introduces significant heterogeneity. Screening patterns may also vary among the clinical settings where these patients were selected. In particular, while clinical risk prediction tools and genomic classifier tests use similar language for risk determination, different cut points are used and are based on different data input. There are other likely sources of heterogeneity that were not reported, such as potential differences in pathology practice patterns, potential differences in clinical practice patterns around the extent of procedures used to identify pathologic node presence (*eg*, was a lymph node dissection conducted), and the tendency of practitioners across institutions to order genomic tests at all and for specific types of patients during routine clinical practice. Lastly, the follow-up window described in the KQ3 studies varied and the long-term outcomes (*ie*, metastasis and PCSM) are rare events.

Applicability to the VA Population

Several of the identified studies included Veterans[18,23,43,44,46,52,54,56,57,59,67,71] diagnosed and/or treated within the VA health care system (VAHCS), and 1 study used Veteran-only data from the VAHCS. Across all included studies, the patient populations were reasonably 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. It is relevant that deployed Vietnam-era Veterans have potential exposure to Agent Orange, which is suspected to confer an increased risk of prostate cancer. While some patients included in non-VA-based cohorts may have had this exposure while not being identified as such, it is likely that this additional risk of a specific Veteran population is underappreciated based on our findings. Of note, 1 study included subgroup analyses by exposure to Agent Orange.[18]

FUTURE RESEARCH

To inform future work in this area, we consider the PICOTs framework (Table 15). For studies addressing KQ1 or KQ3, prospective cohort studies with sufficient follow-up would be preferable study designs, while randomized trials would be preferred for KQ2.

Table 15. Evidence Gaps for Effects of Genomic Tests on Key Outcomes

PICOTS Domain	Evidence Gap/Area for Future Exploration
Population	• Veterans with known Agent Orange exposure • Patients from historically marginalized populations • Patients with family history of early-onset prostate cancer • Patients with high risk germline mutations for prostate cancer
Intervention	• Indication for treatment of patients on active surveillance

PICOTS Domain	Evidence Gap/Area for Future Exploration
	• Oncotype/Prolaris in patients after definitive treatment • Decipher in patients prior to definitive treatment • Identifying candidates for focal therapy
Comparator	• Alternative genomic tests of interest for direct comparison • Direct comparison between Decipher, Oncotype, Prolaris, or other genomic tests • A broadly accepted test of core measures for adjustment in analyses, and/or 1 clinical risk classification system for each key clinical population
Outcomes	• Long-term outcomes such as prostate-cancer-specific mortality, overall mortality • Harms (*eg*, overtreatment, undertreatment) • Cost
Timing	• Follow-up beyond 5 years, ideally up to 15 years

We also note that there was minimal evidence analyzing the effect of genomic testing for prostate cancer on outcomes of interest among subpopulations including historically minoritized racial/ethnic groups. Black men experience more aggressive prostate cancer and suffer a higher mortality rate compared to White men.[88] It is important to recognize that race is a social construct and is not based on underlying genetics; however, ancestral genomic patterns of risk could be transferred through inherited genetics. Of note, recent findings suggest that race-based differences in prostate cancer outcomes are nullified by the receipt of standardized treatment and equal access to care.[89] Future studies examining treatment intensification after genomic testing could stratify by race/ethnicity of patients to explore and identify discriminatory practice patterns (The article by Rayford et al gives an example of such an approach.[25])

Ongoing Work

To project forthcoming evidence from currently active studies in this area, we conducted a rapid review of clinical trials.gov to identify studies in active recruitment, those not yet recruiting, or those that were closed but which do not yet have evidence of related publications. We have listed the identified ongoing studies relevant to each KQ in Appendix F. Importantly, these all are prospective studies. Note that this list is not exhaustive, as observational studies are routinely not registered in clinical trials.gov, and it may not capture work in this area being conducted in other countries. There is also evidence of analyses in earlier stages of dissemination (*eg,* conference abstracts[90]) which, if published in a peer-reviewed publication, could be eligible for future systematic reviews. Finally, there have been at least 3 additional relevant manuscripts published and 1 released as a preprint since our search date[91-94]; of note, the results of these more recent articles are generally consistent with our findings.

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 how frequently providers are changing treatment recommendations based on test reports, the key data needed to inform the value of these genomic classifier tests lies with their ability to accurately predict 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.

REFERENCES

1. Siegel RL, Miller KD, Fuchs HE, et al. Cancer statistics, 2022. *CA Cancer J Clin*. 2022;72(1):7-33.
2. Zullig LL, Jackson GL, Dorn RA, et al. Cancer incidence among patients of the U.S. Veterans Affairs Health Care System. *Mil Med*. 2012;177(6):693-701.
3. D'Amico AV, Whittington R, Malkowicz SB, et al. Biochemical outcome after radical prostatectomy, external beam radiation therapy, or interstitial radiation therapy for clinically localized prostate cancer. *JAMA*. 1998;280(11):969-74.
4. Cuzick J, Swanson GP, Fisher G, et al. Prognostic value of an RNA expression signature derived from cell cycle proliferation genes in patients with prostate cancer: a retrospective study. *Lancet Oncology*. 2011;12(3):245-55.
5. Knezevic D, Goddard AD, Natraj N, et al. Analytical validation of the Oncotype DX prostate cancer assay - a clinical RT-PCR assay optimized for prostate needle biopsies. *BMC Genomics*. 2013;14:690.
6. Erho N, Crisan A, Vergara IA, et al. Discovery and validation of a prostate cancer genomic classifier that predicts early metastasis following radical prostatectomy. *PLoS One*. 2013;8(6):e66855.
7. Two Studies for Patients With High Risk Prostate Cancer Testing Less Intense Treatment for Patients With a Low Gene Risk Score and Testing a More Intense Treatment for Patients With a High Gene Risk Score, The PREDICT-RT Trial. Available at: https://clinicaltrials.gov/ct2/show/NCT04513717. Accessed January 16, 2023.
8. Two Studies for Patients With Unfavorable Intermediate Risk Prostate Cancer Testing Less Intense Treatment for Patients With a Low Gene Risk Score and Testing a More Intense Treatment for Patients With a Higher Gene Risk Score. Available at: https://clinicaltrials.gov/ct2/show/NCT05050084. Accessed January 16, 2023.
9. McGowan J, Sampson M, Salzwedel DM, et al. PRESS Peer Review of Electronic Search Strategies: 2015 Guideline Statement. *J Clin Epidemiol*. 2016;75:40-6.
10. Riley RD, Moons KGM, Snell KIE, et al. A guide to systematic review and meta-analysis of prognostic factor studies. *Bmj*. 2019;364:k4597.
11. Hayden JA, van der Windt DA, Cartwright JL, et al. Assessing bias in studies of prognostic factors. *Ann Intern Med*. 2013;158(4):280-6.
12. Grooten WJA, Tseli E, Äng BO, et al. Elaborating on the assessment of the risk of bias in prognostic studies in pain rehabilitation using QUIPS-aspects of interrater agreement. *Diagn Progn Res*. 2019;3:5.
13. Sterne JAC, Savović J, Page MJ, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. *Bmj*. 2019;366:l4898.
14. Sterne JA, Hernán MA, Reeves BC, et al. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. *Bmj*. 2016;355:i4919.
15. Iorio A, Spencer FA, Falavigna M, et al. Use of GRADE for assessment of evidence about prognosis: rating confidence in estimates of event rates in broad categories of patients. *Bmj*. 2015;350:h870.
16. Seiden B, Weng S, Sun N, et al. NCCN Risk Reclassification in Black Men with Low and Intermediate Risk Prostate Cancer After Genomic Testing. *Urology*. 2021;22:22.
17. Gaffney C, Golan R, Cantu MD, et al. The Clinical Utility of the Genomic Prostate Score in Men with Very Low to Intermediate Risk Prostate Cancer. *Journal of Urology*. 2019;202(1):96-101.

18. Lynch JA, Rothney MP, Salup RR, et al. Improving risk stratification among veterans diagnosed with prostate cancer: impact of the 17-gene prostate score assay. *American Journal of Managed Care*. 2018;24(1 Suppl):S4-S10.
19. Spratt DE, Zhang J, Santiago-Jimenez M, et al. Development and Validation of a Novel Integrated Clinical-Genomic Risk Group Classification for Localized Prostate Cancer. *Journal of Clinical Oncology*. 2018;36(6):581-590.
20. Eure G, Germany R, Given R, et al. Use of a 17-Gene Prognostic Assay in Contemporary Urologic Practice: Results of an Interim Analysis in an Observational Cohort. *Urology*. 2017;107:67-75.
21. Oderda M, Cozzi G, Daniele L, et al. Cell-cycle Progression-score Might Improve the Current Risk Assessment in Newly Diagnosed Prostate Cancer Patients. *Urology*. 2017;102:73-78.
22. Klein EA, Haddad Z, Yousefi K, et al. Decipher Genomic Classifier Measured on Prostate Biopsy Predicts Metastasis Risk. *Urology*. 2016;90:148-52.
23. Murphy AB, Abern MR, Liu L, et al. Impact of a genomic test on treatment decision in a predominantly African American population with favorable-risk prostate cancer: A randomized trial. *Journal of Clinical Oncology*. 2021;39(15):1660-1670.
24. Badani KK, Kemeter MJ, Febbo PG, et al. The Impact of a Biopsy Based 17-Gene Genomic Prostate Score on Treatment Recommendations in Men with Newly Diagnosed Clinically Prostate Cancer Who are Candidates for Active Surveillance. *Urology Practice*. 2015;2(4):181-189.
25. Rayford W, Greenberger M, Bradley RV. Improving risk stratification in a community-based African American population using cell cycle progression score. *Translational Andrology & Urology*. 2018;7(Suppl 4):S384-S391.
26. Michalopoulos SN, Kella N, Payne R, et al. Influence of a genomic classifier on post-operative treatment decisions in high-risk prostate cancer patients: results from the PRO-ACT study. *Current Medical Research & Opinion*. 2014;30(8):1547-56.
27. Gore JL, du Plessis M, Zhang J, et al. Clinical Utility of a Genomic Classifier in Men Undergoing Radical Prostatectomy: The PRO-IMPACT Trial. *Practical Radiation Oncology*. 2020;10(2):e82-e90.
28. Nguyen PL, Shin H, Yousefi K, et al. Impact of a Genomic Classifier of Metastatic Risk on Postprostatectomy Treatment Recommendations by Radiation Oncologists and Urologists. *Urology*. 2015;86(1):35-40.
29. Badani KK, Thompson DJ, Brown G, et al. Effect of a genomic classifier test on clinical practice decisions for patients with high-risk prostate cancer after surgery. *BJU International*. 2015;115(3):419-29.
30. Badani K, Thompson DJ, Buerki C, et al. Impact of a genomic classifier of metastatic risk on postoperative treatment recommendations for prostate cancer patients: a report from the DECIDE study group. *Oncotarget*. 2013;4(4):600-9.
31. Canfield S, Kemeter MJ, Hornberger J, et al. Active Surveillance Use Among a Low-risk Prostate Cancer Population in a Large US Payer System: 17-Gene Genomic Prostate Score Versus Other Risk Stratification Methods. *Reviews in Urology*. 2017;19(4):203-212.
32. Dall'Era MA, Maddala T, Polychronopoulos L, et al. Utility of the Oncotype DX® Prostate Cancer Assay in Clinical Practice for Treatment Selection in Men Newly Diagnosed with Prostate Cancer: A Retrospective Chart Review Analysis. *Urology Practice*. 2015;2(6):343-348.

33. Morris DS, Woods JS, Edwards B, et al. Prognostic capabilities and clinical utility of cell cycle progression testing, prostate imaging reporting and data system, version 2, and clinicopathologic data in management of localized prostate cancer. *Urologic Oncology*. 2021;39(6):366.e19-366.e28.
34. Crawford ED, Scholz MC, Kar AJ, et al. Cell cycle progression score and treatment decisions in prostate cancer: results from an ongoing registry. *Current Medical Research & Opinion*. 2014;30(6):1025-31.
35. Shore ND, Kella N, Moran B, et al. Impact of the Cell Cycle Progression Test on Physician and Patient Treatment Selection for Localized Prostate Cancer. *J Urol*. 2016;195(3):612-8.
36. Gore JL, du Plessis M, Santiago-Jimenez M, et al. Decipher test impacts decision making among patients considering adjuvant and salvage treatment after radical prostatectomy: Interim results from the Multicenter Prospective PRO-IMPACT study. *Cancer*. 2017;123(15):2850-2859.
37. Shahait M, Liu VYT, Vapiwala N, et al. Impact of Decipher on use of post-operative radiotherapy: Individual patient analysis of two prospective registries. *BJUI Compass*. 2021;2(4):267-274.
38. Vince RA, Jr., Jiang R, Qi J, et al. Impact of Decipher Biopsy testing on clinical outcomes in localized prostate cancer in a prospective statewide collaborative. *Prostate Cancer & Prostatic Diseases*. 2021;20:20.
39. Brooks MA, Thomas L, Magi-Galluzzi C, et al. GPS Assay Association With Long-Term Cancer Outcomes: Twenty-Year Risk of Distant Metastasis and Prostate Cancer-Specific Mortality. *JCO Precision Oncology*. 2021;5.
40. Lehto TK, Sturenberg C, Malen A, et al. Transcript analysis of commercial prostate cancer risk stratification panels in hard-to-predict grade group 2-4 prostate cancers. *Prostate*. 2021;81(7):368-376.
41. Feng FY, Huang HC, Spratt DE, et al. Validation of a 22-Gene Genomic Classifier in Patients With Recurrent Prostate Cancer: An Ancillary Study of the NRG/RTOG 9601 Randomized Clinical Trial. *JAMA Oncology*. 2021;7(4):544-552.
42. Ramotar M, Chua MLK, Truong H, et al. Subpathologies and genomic classifier for treatment individualization of post-prostatectomy radiotherapy. *Urologic Oncology*. 2022;40(1):5.e1-5.e13.
43. Tosoian JJ, Birer SR, Jeffrey Karnes R, et al. Performance of clinicopathologic models in men with high risk localized prostate cancer: impact of a 22-gene genomic classifier. *Prostate Cancer & Prostatic Diseases*. 2020;23(4):646-653.
44. Howard LE, Zhang J, Fishbane N, et al. Validation of a genomic classifier for prediction of metastasis and prostate cancer-specific mortality in African-American men following radical prostatectomy in an equal access healthcare setting. *Prostate Cancer & Prostatic Diseases*. 2020;23(3):419-428.
45. Shangguan X, Qian H, Jiang Z, et al. Cell cycle progression score improves risk stratification in prostate cancer patients with adverse pathology after radical prostatectomy. *Journal of Cancer Research & Clinical Oncology*. 2020;146(3):687-694.
46. Canter DJ, Freedland S, Rajamani S, et al. Analysis of the prognostic utility of the cell cycle progression (CCP) score generated from needle biopsy in men treated with definitive therapy. *Prostate Cancer & Prostatic Diseases*. 2020;23(1):102-107.
47. Kornberg Z, Cooperberg MR, Cowan JE, et al. A 17-Gene Genomic Prostate Score as a Predictor of Adverse Pathology in Men on Active Surveillance. *Journal of Urology*. 2019;202(4):702-709.

48. Canter DJ, Reid J, Latsis M, et al. Comparison of the Prognostic Utility of the Cell Cycle Progression Score for Predicting Clinical Outcomes in African American and Non-African American Men with Localized Prostate Cancer. *European Urology*. 2019;75(3):515-522.
49. Berlin A, Murgic J, Hosni A, et al. Genomic Classifier for Guiding Treatment of Intermediate-Risk Prostate Cancers to Dose-Escalated Image Guided Radiation Therapy Without Hormone Therapy. *International Journal of Radiation Oncology, Biology, Physics*. 2019;103(1):84-91.
50. Leapman MS, Nguyen HG, Cowan JE, et al. Comparing Prognostic Utility of a Single-marker Immunohistochemistry Approach with Commercial Gene Expression Profiling Following Radical Prostatectomy. *European Urology*. 2018;74(5):668-675.
51. Leon P, Cancel-Tassin G, Drouin S, et al. Comparison of cell cycle progression score with two immunohistochemical markers (PTEN and Ki-67) for predicting outcome in prostate cancer after radical prostatectomy. *World Journal of Urology*. 2018;36(9):1495-1500.
52. Spratt DE, Dai DLY, Den RB, et al. Performance of a Prostate Cancer Genomic Classifier in Predicting Metastasis in Men with Prostate-specific Antigen Persistence Postprostatectomy. *European Urology*. 2018;74(1):107-114.
53. Van Den Eeden SK, Lu R, Zhang N, et al. A Biopsy-based 17-gene Genomic Prostate Score as a Predictor of Metastases and Prostate Cancer Death in Surgically Treated Men with Clinically Localized Disease. *European Urology*. 2018;73(1):129-138.
54. Karnes RJ, Choeurng V, Ross AE, et al. Validation of a Genomic Risk Classifier to Predict Prostate Cancer-specific Mortality in Men with Adverse Pathologic Features. *European Urology*. 2018;73(2):168-175.
55. Nguyen PL, Haddad Z, Ross AE, et al. Ability of a Genomic Classifier to Predict Metastasis and Prostate Cancer-specific Mortality after Radiation or Surgery based on Needle Biopsy Specimens. *European Urology*. 2017;72(5):845-852.
56. Tosoian JJ, Chappidi MR, Bishoff JT, et al. Prognostic utility of biopsy-derived cell cycle progression score in patients with National Comprehensive Cancer Network low-risk prostate cancer undergoing radical prostatectomy: implications for treatment guidance. *BJU International*. 2017;120(6):808-814.
57. Dalela D, Santiago-Jimenez M, Yousefi K, et al. Genomic Classifier Augments the Role of Pathological Features in Identifying Optimal Candidates for Adjuvant Radiation Therapy in Patients With Prostate Cancer: Development and Internal Validation of a Multivariable Prognostic Model. *Journal of Clinical Oncology*. 2017;35(18):1982-1990.
58. Nguyen PL, Martin NE, Choeurng V, et al. Utilization of biopsy-based genomic classifier to predict distant metastasis after definitive radiation and short-course ADT for intermediate and high-risk prostate cancer. *Prostate Cancer & Prostatic Diseases*. 2017;20(2):186-192.
59. Ross AE, Den RB, Yousefi K, et al. Efficacy of post-operative radiation in a prostatectomy cohort adjusted for clinical and genomic risk. *Prostate Cancer & Prostatic Diseases*. 2016;19(3):277-82.
60. Glass AG, Leo MC, Haddad Z, et al. Validation of a Genomic Classifier for Predicting Post-Prostatectomy Recurrence in a Community Based Health Care Setting. *Journal of Urology*. 2016;195(6):1748-53.
61. Ross AE, Johnson MH, Yousefi K, et al. Tissue-based Genomics Augments Post-prostatectomy Risk Stratification in a Natural History Cohort of Intermediate- and High-Risk Men. *European Urology*. 2016;69(1):157-65.

62. Cuzick J, Stone S, Fisher G, et al. Validation of an RNA cell cycle progression score for predicting death from prostate cancer in a conservatively managed needle biopsy cohort. *British Journal of Cancer*. 2015;113(3):382-9.
63. Den RB, Yousefi K, Trabulsi EJ, et al. Genomic classifier identifies men with adverse pathology after radical prostatectomy who benefit from adjuvant radiation therapy. *Journal of Clinical Oncology*. 2015;33(8):944-51.
64. Klein EA, Yousefi K, Haddad Z, et al. A genomic classifier improves prediction of metastatic disease within 5 years after surgery in node-negative high-risk prostate cancer patients managed by radical prostatectomy without adjuvant therapy. *European Urology*. 2015;67(4):778-86.
65. Cullen J, Rosner IL, Brand TC, et al. A Biopsy-based 17-gene Genomic Prostate Score Predicts Recurrence After Radical Prostatectomy and Adverse Surgical Pathology in a Racially Diverse Population of Men with Clinically Low- and Intermediate-risk Prostate Cancer. *European Urology*. 2015;68(1):123-31.
66. Cooperberg MR, Davicioni E, Crisan A, et al. Combined value of validated clinical and genomic risk stratification tools for predicting prostate cancer mortality in a high-risk prostatectomy cohort. *European Urology*. 2015;67(2):326-33.
67. Freedland SJ, Gerber L, Reid J, et al. Prognostic utility of cell cycle progression score in men with prostate cancer after primary external beam radiation therapy. *International Journal of Radiation Oncology, Biology, Physics*. 2013;86(5):848-53.
68. Cooperberg MR, Simko JP, Cowan JE, et al. Validation of a cell-cycle progression gene panel to improve risk stratification in a contemporary prostatectomy cohort. *Journal of Clinical Oncology*. 2013;31(11):1428-34.
69. Cuzick J, Berney DM, Fisher G, et al. Prognostic value of a cell cycle progression signature for prostate cancer death in a conservatively managed needle biopsy cohort. *British Journal of Cancer*. 2012;106(6):1095-9.
70. Bauman G, Breau RH, Kamel-Reid S, et al. Ontario health technology assessment series: Prolaris cell cycle progression test for localized prostate cancer: A health technology assessment. 2017;17(6):1-75.
71. Bishoff JT, Freedland SJ, Gerber L, et al. Prognostic utility of the cell cycle progression score generated from biopsy in men treated with prostatectomy. *J Urol*. 2014;192(2):409-14.
72. Degeling K, Pereira-Salgado A, Corcoran NM, et al. Health Economic Evidence for Liquid- and Tissue-based Molecular Tests that Inform Decisions on Prostate Biopsies and Treatment of Localised Prostate Cancer: A Systematic Review. *European Urology Open Science*. 2021;27:77-87.
73. Lobo JM, Dicker AP, Buerki C, et al. Evaluating the clinical impact of a genomic classifier in prostate cancer using individualized decision analysis. *PLoS ONE [Electronic Resource]*. 2015;10(3):e0116866.
74. Caetano SJ, Sonpavde G, Pond GR. C-statistic: A brief explanation of its construction, interpretation and limitations. *Eur J Cancer*. 2018;90:130-132.
75. Olleik G, Kassouf W, Aprikian A, et al. Evaluation of New Tests and Interventions for Prostate Cancer Management: A Systematic Review. *Journal of the National Comprehensive Cancer Network*. 2018;16(11):1340-1351.
76. Fine ND, LaPolla F, Epstein M, et al. Genomic Classifiers for Treatment Selection in Newly Diagnosed Prostate Cancer. *BJU International*. 2019;04:04.

77. Health Quality Ontario. Prolaris Cell Cycle Progression Test for Localized Prostate Cancer: A Health Technology Assessment. *Ontario Health Technology Assessment Series*. 2017;17(6):1-75.
78. Bostrom PJ, Bjartell AS, Catto JW, et al. Genomic Predictors of Outcome in Prostate Cancer. *European Urology*. 2015;68(6):1033-44.
79. Lamy PJ, Allory Y, Gauchez AS, et al. Prognostic Biomarkers Used for Localised Prostate Cancer Management: A Systematic Review. *European Urology Focus*. 2018;4(6):790-803.
80. Spratt DE, Zumsteg ZS, Feng FY, et al. Translational and clinical implications of the genetic landscape of prostate cancer. *Nature Reviews Clinical Oncology*. 2016;13(10):597-610.
81. Carrion DM, Rivas JG, Alvarez-Maestro M, et al. BIOMARKERS IN PROSTATE CANCER MANAGEMENT. IS THERE SOMETHING NEW? *Archivos Espanoles De Urologia*. 2019;72(2):105-115.
82. Jairath NK, Dal Pra A, Vince R, Jr., et al. A Systematic Review of the Evidence for the Decipher Genomic Classifier in Prostate Cancer. *European Urology*. 2021;79(3):374-383.
83. Prolaris Enhanced Risk Stratification - an ecONomic and clinicAL Evaluation (PERSONAL). Available at: https://clinicaltrials.gov/ct2/show/study/NCT03851211. Accessed January 16, 2023.
84. Clinical Outcomes in Men With Prostate Cancer Who Selected Active Surveillance Using Prolaris® Testing (URO-009Low). Available at: https://clinicaltrials.gov/ct2/show/results/NCT03511235. Accessed January 16, 2023.
85. Shore ND, Boczko J, Kella N, et al. Impact of CCP test on personalizing treatment decisions: Results from a prospective registry of newly diagnosed prostate cancer patients. *Journal of Clinical Oncology*. 2015;33(7_suppl):63-63.
86. Registry to Measure the Impact of Adding Genomic Testing. Available at: https://clinicaltrials.gov/ct2/show/results/NCT02454595. Accessed January 16, 2023.
87. Mikhaylenko DS, Sergienko SA, Alekseev BY, et al. Basic characteristics and features of the molecular genetic test systems designed for non-invasive diagnostics and prognosis of prostate cancer and bladder cancer. *Onkourologiya*. 2019;15(4):18-29.
88. Lillard JW, Jr., Moses KA, Mahal BA, et al. Racial disparities in Black men with prostate cancer: A literature review. *Cancer*. 2022;128(21):3787-3795.
89. Dess RT, Hartman HE, Mahal BA, et al. Association of Black Race With Prostate Cancer–Specific and Other-Cause Mortality. *JAMA Oncology*. 2019;5(7):975-983.
90. Spratt DE, Huang H-C, Michalski JM, et al. Validation of the performance of the Decipher biopsy genomic classifier in intermediate-risk prostate cancer on the phase III randomized trial NRG Oncology/RTOG 0126. *Journal of Clinical Oncology*. 2022;40(6_suppl):269-269.
91. Nguyen PL, Huang HR, Spratt DE, et al. Analysis of a Biopsy-Based Genomic Classifier in High-Risk Prostate Cancer: Meta-Analysis of the NRG Oncology/Radiation Therapy Oncology Group 9202, 9413, and 9902 Phase 3 Randomized Trials. *Int J Radiat Oncol Biol Phys*. 2022.
92. Dal Pra A, Ghadjar P, Hayoz S, et al. Validation of the Decipher genomic classifier in patients receiving salvage radiotherapy without hormone therapy after radical prostatectomy - an ancillary study of the SAKK 09/10 randomized clinical trial. *Ann Oncol*. 2022;33(9):950-958.

93. Janes JL, Boyer MJ, Bennett JP, et al. The 17-Gene Genomic Prostate Score Test Is Prognostic for Outcomes After Primary External Beam Radiation Therapy in Men With Clinically Localized Prostate Cancer. *Int J Radiat Oncol Biol Phys*. 2023;115(1):120-131.
94. Attard G, Parry M, Grist E, et al. Clinical testing of transcriptome-wide expression profiles in high-risk localized and metastatic prostate cancer starting androgen deprivation therapy: an ancillary study of the STAMPEDE abiraterone Phase 3 trial. *Res Sq*. 2023.
95. Spratt DE, Yousefi K, Deheshi S, et al. Individual Patient-Level Meta-Analysis of the Performance of the Decipher Genomic Classifier in High-Risk Men After Prostatectomy to Predict Development of Metastatic Disease. *Journal of Clinical Oncology*. 2017;35(18):1991-1998.