Comparative Effectiveness Review
Number 146

Effective Health Care Program

Therapies for Clinically Localized Prostate Cancer: Update of a 2008 Systematic Review

Executive Summary

Background

Prostate cancer is the most common nondermatologic cancer in men.[1,2] The American Cancer Society has estimated that 241,740 men were expected to receive a diagnosis of prostate cancer in 2012, and 28,170 were expected to die from the disease.[1] Approximately 90 percent of those who receive such a diagnosis have cancer confined to the prostate gland, which is the definition of clinically localized disease. Since 2004, the prostate cancer incidence rate has decreased by 2.7 percent annually among men 65 years of age or older and has remained steady among men younger than age 65.[1] The major risk factors for prostate cancer are advanced age, race and ethnicity (the highest incidence is in blacks), and family history.

Many cases of prostate cancer have a protracted course if left untreated. Many men die with prostate cancer rather than from it.[3] During its early stages, clinically localized prostate cancer is usually asymptomatic.[4] However, as the cancer grows, it may cause urinary problems such as blood in the urine, pain or a burning sensation during urination, a weak urine stream, inability to urinate, and frequent urination, especially at night. These presenting symptoms, along with a physical examination, prostate-specific antigen (PSA) levels, and biopsy, may be used to evaluate patients for the presence of prostate cancer.

Effective Health Care Program

The Effective Health Care Program was initiated in 2005 to provide valid evidence about the comparative effectiveness of different medical interventions. The object is to help consumers, health care providers, and others in making informed choices among treatment alternatives. Through its Comparative Effectiveness Reviews, the program supports systematic appraisals of existing scientific evidence regarding treatments for high-priority health conditions. It also promotes and generates new scientific evidence by identifying gaps in existing scientific evidence and supporting new research. The program puts special emphasis on translating findings into a variety of useful formats for different stakeholders, including consumers.

The full report and this summary are available at **www.effectivehealthcare.ahrq.gov/reports/final.cfm**.

AHRQ
Agency for Healthcare Research and Quality
Advancing Excellence in Health Care • www.ahrq.gov

Effective Health Care

The PSA test is used to measure blood levels of PSA, a protein produced by the prostate gland.[4] Elevated PSA levels may indicate the presence of prostate cancer, but elevations are also seen in conditions such as benign prostatic hyperplasia and prostatitis. Conversely, some patients with prostate cancer do not have elevated levels of PSA.[5] Moreover, the cutpoint separating a "normal" PSA level from an abnormal level also remains a subject of debate. In recent years, more frequent use of PSA testing has intensified concern about overdiagnosis of prostate cancer (i.e., detection of cancer that would have remained silent and caused the patient no illness throughout his lifetime).[2,4]

In May 2012, the U.S. Preventive Services Task Force recommended against PSA-based screening for prostate cancer in healthy men of all ages, concluding that the harms of screening outweigh the benefits (Grade D recommendation).[6] However, health care professionals and professional societies have continued to debate the merits of PSA-based screening. Potential benefits of regular PSA screening include early cancer detection and reduced mortality rates. Potential harms include anxiety related to abnormal results, pain, infection, bleeding from diagnostic biopsies, and morbidity from definitive treatment in men who may not need such treatment.[7-10] No organization (including the American Urological Association) currently recommends routine PSA-based screening.

Determining which men with clinically localized prostate cancer are most likely to benefit from interventions such as surgery and radiation could potentially improve the balance of benefits and harms, especially in those identified by screening. Current practice is to use tumor grade as the primary prognostic variable in patients with clinically localized prostate cancer.[2] After biopsy confirms the presence of the cancer, pathologists report tumor grade using the Gleason score, which ranges from 2 to 10.[4] Gleason 8 and higher tumors are considered the most aggressive, Gleason 7 tumors are considered somewhat less aggressive, and Gleason 6 or lower tumors are considered potentially indolent.[11]

A biopsy-based Gleason score may not always accurately reflect the real aggressiveness of the prostate cancer. Therefore, efforts are underway to identify more reliable prognostic factors. PSA, PSA kinetics (rate of rise in PSA over time and doubling time for PSA), and digital rectal examination are still very important when deciding treatment. Additionally, radiographic imaging in high-risk disease is valuable, along with other diagnostic assessments, before making definitive treatment decisions.

Staging is the process of assessing whether the cancer is confined to the prostate gland or has spread and the extent of the spread.[4] Staging of prostate cancer could be clinical (based on a digital rectal examination of the prostate gland, imaging tests, prostate biopsy, and laboratory tests) or pathological (based on surgery and examination of resected prostate tissue). The staging system currently used is the American Joint Committee on Cancer TNM classification.[4] TNM classification is based on the extent of primary tumor (T stages), whether cancer has spread to the adjacent lymph nodes (N stages), and any metastasis (M stages).[4,12] TNM categories are combined with the Gleason histologic score and PSA results (stage grouping) to determine the overall stage, commonly reported as stage I, IIA, IIB, III, or IV, with stage I being the least advanced and stage IV being the most advanced. In the absence of a Gleason histologic score, staging can be based on the TNM classification.

Another categorization—incorporating PSA levels, Gleason histologic score, and TNM stage—stratifies tumors into low, intermediate, and high risk: the concept reflects the likelihood of progressing with no treatment or recurring after early intervention. The levels are defined as follows:[4]

- Low risk (corresponding to stage I): a PSA level of 10 ng/mL or less, a Gleason score of 6 or less, and clinical stage T1c or T2a
- Intermediate risk (roughly corresponding to stage IIA): a PSA level of greater than 10 to 20 ng/mL, a Gleason score of 7, or clinical stage T2b but not qualifying for high risk
- High risk (roughly corresponding to stage IIB): a PSA level of greater than 20 ng/mL, a Gleason score of 8 or higher, or clinical stage T2c

This risk-assessment scheme, although commonly used, has significant limitations in assessing patients in the intermediate- and high-risk groups. A good example of a risk-assessment scheme developed and validated across populations is the University of California, San Francisco, Cancer of the Prostate Risk Assessment (CAPRA). The CAPRA is associated with both overall and cause-specific survival and can be used to predict disease recurrence and mortality after radical prostatectomy (RP).[13-16] These risk-assessment tools may be improved in the future with the use of biomarkers (e.g., actinin alpha 1, derlin 1).

Clinicians make pretreatment assessment of whether prostate cancer is localized by determining tumor stage, basing their decision on clinical examinations (e.g., digital

rectal examination, imaging and laboratory tests, prostate biopsy). According to a 2013 clinical practice guideline published by the National Comprehensive Cancer Network, clinically localized prostate cancer includes clinical stage T1–T3a, N0–X, and M0.[17] This expert opinion–based guideline further categorizes clinically localized disease based on the recurrence risk as follows:

- Very low recurrence risk: T1c, Gleason score ≤6, PSA <10 ng/mL, fewer than three prostate biopsy cores positive, ≤50 percent cancer in each core, PSA density <0.15 ng/mL/g
- Low recurrence risk: T1–T2a, Gleason score ≤6, PSA <10 ng/mL
- Intermediate recurrence risk: T2b–T2c or PSA 10–20 ng/mL or Gleason score 7
- High recurrence risk: T3a or Gleason score 8–10 or PSA >20 ng/mL

The focus of this report is clinically localized prostate cancer (T1–T3a). Locally advanced (T3b–T4), metastatic, and recurrent prostate cancer are outside the scope of this report.

Therapies for Clinically Localized Prostate Cancer

The primary goal of treating clinically localized prostate cancer is to target men most likely to need intervention to prevent disability or death while minimizing intervention-related complications. Frequently used treatment options include the following:

- RP, including laparoscopic or robotic-assisted prostatectomy
- External beam radiotherapy (EBRT), including conventional radiation, intensity-modulated radiation (IMRT), three-dimensional conformal radiation therapy (3D-CRT), stereotactic body radiation therapy, and proton beam therapy
- Interstitial brachytherapy (BT)
- Cryotherapy
- Observation or watchful waiting (WW); the two terms are used interchangeably throughout the report
- Active surveillance (AS)
- Hormonal therapy (e.g., androgen-deprivation therapy [ADT])
- High-intensity focused ultrasound (HIFU)

Choice of treatment options may be influenced by numerous factors. These include patient age and health at the time of diagnosis, life expectancy, and estimated likelihood of cancer progression without treatment; surgeon experience and preference; treatment-related convenience and costs; and potential for eradication and adverse effects (e.g., incontinence, sexual dysfunction).[4] Before choosing any intervention, the patient's overall health status should be assessed because it may influence response to therapy, severity of complications, and life expectancy.[4]

The National Cancer Institute and the Centers for Disease Control and Prevention sponsored a National Institutes of Health (NIH) State-of-the Science Conference in December 2011 to better understand the risks and benefits of AS and other observational management strategies for low-grade localized prostate cancer detected by PSA screening.[3] AS (with curative intent) usually includes hands-on followup in which PSA levels are checked, prostate biopsies may be repeated, and subsequent treatment is planned. The panel concluded that AS should be offered to patients with low-risk prostate cancer.[3]

The NIH panel used the term "watchful waiting" to describe a palliative observational strategy—that is, waiting for symptoms to appear and then intervening to manage the symptoms. In the 2008 Comparative Effectiveness Review "Comparative Effectiveness of Therapies for Clinically Localized Prostate Cancer," these two approaches were considered together.[18] In the literature, the distinction between AS (with curative intent) and other observational strategies (with palliative intent) has not always been clear; however, for this systematic review update we attempted to separate the two using the definitions proposed at the 2011 NIH State-of-the-Science Conference.[3]

Objectives of This Review

This report updates a 2008 systematic review conducted by the University of Minnesota Evidence-based Practice Center (EPC).[18] This update examines the same four Key Questions (KQs) as the original report and summarizes the more recent evidence comparing the relative effectiveness and safety of treatment options for clinically localized prostate cancer.

Key Questions and Scope

Key Questions

The KQs are as follows:

Key Question 1: What are the comparative risks and benefits of the following therapies for clinically localized prostate cancer?

a. Radical prostatectomy, including open (retropubic and perineal) and laparoscopic (with or without robotic assistance) approaches
b. External beam radiation therapy, including standard therapy and therapies designed to decrease exposure to normal tissues such as three-dimensional conformal radiotherapy, intensity-modulated radiation therapy, proton beam therapy, and stereotactic body radiation therapy
c. Interstitial brachytherapy
d. Cryotherapy
e. Watchful waiting
f. Active surveillance
g. Hormonal therapy
h. High-intensity focused ultrasound

Key Question 2: How do specific patient characteristics (e.g., age, race/ethnicity, presence or absence of comorbid illness, preferences such as tradeoff of treatment-related adverse effects vs. potential for disease progression) affect the outcomes of these therapies overall and differentially?

Key Question 3: How do provider/hospital characteristics (e.g., geographic region, case volume, learning curve) affect outcomes of these therapies overall and differentially?

Key Question 4: How do tumor characteristics (e.g., Gleason score, tumor volume, screen-detected vs. clinically detected tumors, PSA levels) affect the outcomes of these therapies overall and differentially?

Scope

An analytic framework showing the populations, interventions, comparators, outcomes, timing, and setting (PICOTS) in diagram form is shown in Figure 1 of the full report.

Population: KQs 1–4: The population comprised men considered to have clinically localized prostate cancer (T1–T3a, N0–X, M0–X), regardless of age, histologic grade, or PSA level. Studies were excluded if more than 15 percent of men with disease stage higher than T3a were enrolled and data were not reported separately for men with T1, T2, and/or T3a prostate cancer.

Interventions: For KQs 1–4, we included treatment options for men with clinically localized prostate cancer: RP (including retropubic, perineal, laparoscopic, robotic assisted), EBRT (including conventional radiation, IMRT, 3D-CRT, proton beam, and stereotactic body radiation therapy), interstitial BT, cryotherapy, WW, AS, hormonal therapy, and HIFU.

Comparators: Comparators were any interventions of interest listed above.

Outcomes: The primary outcome is overall mortality or survival. Additional outcomes include prostate cancer–specific mortality or survival, biochemical (PSA) progression, metastatic and/or clinical progression-free survival, health status, and quality of life (QOL). We focused primarily on common and severe adverse events of treatment, including bowel, bladder, and sexual dysfunction, as well as harms from biopsy such as bleeding and nosocomial infections. For KQ 3, we focus on RP compared with other interventions in association with provider location, case volume, and affiliation with academic centers.

Timing: Duration of followup was appropriate for the outcome under consideration.

Settings: All settings were considered.

Methods

Search Strategy

Medical Librarians in the ECRI Institute–Penn Medicine EPC Information Center performed literature searches following established systematic review protocols. We searched the following databases using controlled vocabulary and text words: Embase®, MEDLINE®, PubMed®, and the Cochrane Library from January 1, 2007, through March 7, 2014.

Study Selection

We used the same study selection criteria as in the 2008 report. For KQs 1, 2, and 4, we included randomized trials only if the randomized treatment allocation was based on men with clinically localized disease and if clinical outcomes were reported for T1, T2, and T3a disease separately from T3b and T4 disease. We also included large nonrandomized comparative studies (N ≥500) that controlled for potentially confounding variables. For KQ 3, we included multicenter studies that compared RP with another treatment of interest, enrolled 500 or more patients, used appropriate statistical techniques to control for potentially confounding variables, and examined the effect of provider characteristics on survival of patients with localized prostate cancer.

Data Extraction and Management

We used the DistillerSR® (Evidence Partners, Inc., Ottawa, Ontario, Canada) Web-based systematic review software for abstract screening. One team member extracted data directly into a Word document and a second team member

reviewed the extractions. The data extracted included study, patient, tumor, and intervention characteristics and predefined outcomes. We calculated standard errors, regression coefficients, and 95% confidence intervals (CIs) from reported means, standard deviations, and sample size when provided and appropriate, if not already done in the original study.[19] Also, because of the possibility of subjective interpretation, we judged the risk-of-bias items in duplicate. We resolved all discrepancies through discussion. Multiple publications of the same study (e.g., publications reporting subgroups, other outcomes, longer followup) were identified by examining author affiliations, study designs, enrollment criteria, and enrollment dates. Multiple publications were used only when each publication had unique data not reported in the most comprehensive and recent publication.

Risk-of-Bias Assessment of Individual Studies

Because of the possibility of subjective interpretation, two researchers assessed methodologic risk of bias for each study and resolved discrepancies by consensus. When consensus could not be reached, a third researcher adjudicated.

We assessed the risk of bias by following the guidelines in the chapter "Assessing the Risk of Bias of Individual Studies When Comparing Medical Interventions" in the "Methods Guide for Effectiveness and Comparative Effectiveness Reviews."[20] This involved evaluating several items such as randomization, allocation concealment, intention-to-treat-analysis, and completeness of followup. Additionally, we assessed fidelity to the protocol to address performance bias and blinding of outcome assessors to address detection bias when outcomes were subjective.

To be considered as having low risk of bias, the study must have met all the following conditions: randomization or pseudorandomization (e.g., using instrumental variable analysis) of study participants to treatment groups, concealment of allocation, data analysis based on the intention-to-treat-principle, an outcome that was objective if outcome assessors were not blinded or blinding of outcome assessors was not reported, a difference of 15 percent or less in the length of followup for the comparison groups, data for more than 85 percent of enrolled patients provided at the timepoint of interest, and no clear indication of lack of fidelity to the protocol.

To be considered as having high risk of bias, the study must have met at least one of the following criteria: trial did not randomly or pseudorandomly (i.e., using instrumental variables) assign patients to study groups and did not blind outcome assessors, trial had a difference of 15 percent or more in the length of followup for comparison groups, or trial stated that there was not good fidelity to the protocol. To be considered as having medium risk of bias, the study met neither the criteria for low risk of bias nor the criteria for high risk of bias.

Data Synthesis

Because of the differences in study designs, treatments, patient and tumor characteristics, and reporting of outcomes, the 2008 report did not pool studies for KQs 1, 2, and 4. For the same reason, we performed only qualitative analysis in this update.

Because randomized controlled trials (RCTs) and nonrandomized comparative studies differed substantially in average risk of bias, we performed separate qualitative analyses and present results separately for these study designs. The findings from the RCTs and nonrandomized comparative studies were included in our discussion and formed the basis of our overall conclusion. We further stratified the results from the RCTs based on comparisons across and within primary treatment categories.

Generally, we report summaries of effectiveness and adverse event outcomes with ranges according to treatment option, tumor characteristics, and group sample size. For KQ 1, we summarize and discuss comparative risks, benefits, and outcomes of therapies. For KQ 2, we summarize how patient characteristics affect outcomes. For KQ 4, we summarize how tumor characteristics affect outcomes. For KQ3, we were unable to identify any studies that met our inclusion criteria.

Strength–of-Evidence Grading

We provided evidence grades for the following patient-oriented outcomes: overall mortality or survival, prostate cancer–specific survival, progression to metastases, and QOL. We assessed strength of evidence by following the guidelines from the article "Grading the Strength of a Body of Evidence When Comparing Medical Interventions" by Owens and colleagues.[21] We graded the strength of evidence based on the following domains: risk of bias (low, medium, or high), consistency (consistent, inconsistent, or unknown/not applicable), directness (direct or indirect), and precision (precise or imprecise). Two independent graders assessed each domain, and differences were resolved by consensus.

We assigned the strength of evidence an overall grade of high, moderate, low, or insufficient, as outlined by Owens and colleagues.[21] Briefly, a high grade reflects high confidence that the effect estimate lies close to the true effect; a moderate grade reflects moderate confidence; a

low grade reflects limited confidence; and an insufficient grade reflects either no evidence, inability to estimate an effect, or no confidence in the effect estimate. The decision to grade an evidence base as insufficient rather than low often reflected an imprecise effect estimate (a non–statistically significant effect with 95% CIs wide enough to allow the possibility of a significant benefit for one treatment compared with another) in an evidence base with only one or two studies. However, we also graded as insufficient evidence from a single study with medium risk of bias or from fewer than three consistent studies with high risk of bias, even when findings were direct and precise. Because multiple factors other than treatment can influence apparent differences between interventions, we placed a high value on replication of findings, even more so for studies with high risk of bias. Further explanation of this conservative approach to evidence grading appears in the Discussion.

When evidence came from subgroup analyses (KQs 2 and 4), we lowered the strength-of-evidence grade by one level. For example, when the strength of evidence for a primary analysis in KQ 1 was low, strength of evidence for subgroup analyses from the same studies was considered insufficient. We adopted this approach because subgroup analyses were usually underpowered to detect differences between treatments and sometimes not prespecified at the beginning of the study. In general, subgroup analyses should be considered as hypothesis generating rather than definitive analyses.

Applicability

Applicability assessment refers to how generalizable findings from this report are to other populations and settings. We assessed applicability by following the guidelines in the article "Assessing the Applicability of Studies When Comparing Medical Interventions" by Atkins and colleagues.[22] The applicability of the evidence involves the following five aspects: patients, interventions, comparisons, outcomes, and settings.[22]

We addressed factors relevant to the applicability of the evidence by evaluating patient selection in both observational studies and clinical trials. We considered the primary biology and epidemiology (grade and stage of the prostate cancer) and the present-day clinical practice setting. The typical interventions, comparisons, outcomes (e.g., overall mortality, prostate cancer–specific survival), and settings of care were also used to specify more clearly the most applicable study characteristics (i.e., most typical of care for patients with localized prostate cancer in the United States).

Peer Review and Publication

Peer reviewers were invited to provide written comments on the draft report based on their clinical, content, or methodological expertise. The EPC considered peer review comments on the preliminary draft of the report in preparation of the final report. The dispositions of the peer review comments are documented and will be published 3 months after the publication of the evidence report.

Results

Evidence Base

Our searches of the literature identified 5,210 potentially relevant articles. We excluded 1,508 articles by reviewing the titles, 3,420 by reviewing the abstracts, and 221 by reviewing the full-length articles. Figure 2 in the full report is a flow chart that describes in detail the exclusion process and the reasons for exclusion at each review level. The remaining 61 publications, describing 52 unique studies, made up the evidence base for this review.

All 52 studies met the inclusion criteria for review for KQ 1. Thirteen of these studies also met the inclusion criteria for KQ 2, and 20 of them met the inclusion criteria for KQ 4. Studies that addressed KQ 1 reported data for patient-oriented outcome measures such as overall survival, all-cause mortality, prostate cancer–specific mortality, QOL, and adverse events. Evidence addressing KQ 2 or 4 came solely from subgroup analyses of some larger studies that addressed KQ 1. Although these subgroup analyses reported data on overall survival, all-cause mortality, or prostate cancer–specific mortality for specific patient subgroups, they did not report adverse events that occurred in these subgroups.

KQ 1: Comparative Risks and Benefits of Therapies for Clinically Localized Prostate Cancer

Eight RCTs in 16 publications addressed comparative risks and benefits for various therapies. Our risk-of-bias assessments for the eight trials appear in Table C-1 of Appendix C. Of these eight RCTs, seven were categorized as medium risk of bias for all outcomes excluding the QOL outcome. One study received a rating of low risk of bias.[23] Because QOL is subjectively interpreted, studies that did not blind outcome assessors received a lower rating for this outcome.

Table A summarizes our findings from RCTs on the major health outcomes for KQ 1. These outcomes include overall survival, all-cause mortality, prostate cancer–specific mortality, QOL, and progression to metastases, for which we assessed the strength of evidence. For the

comparison of RP versus WW, the Scandinavian Prostate Cancer Group-4 (SPCG-4) trial reported that all-cause and prostate cancer–specific mortality at the end of the 15-year followup period favored RP, but the strength of evidence was insufficient. Both the Prostate Cancer Intervention Versus Observation Trial (PIVOT) and SPCG-4 studies reported data on all-cause and prostate cancer–specific mortality at the end of the 12-year followup period, but we found that the evidence on these outcomes at this timepoint was insufficient to draw any conclusion (based mostly on imprecision in the statistically nonsignificant effect sizes). However, both trials found that progression to metastases was significantly lower among patients in the RP group than in the WW group; the strength of evidence was moderate due to consistent and precise findings in medium risk-of-bias trials. The evidence on other patient-oriented outcomes based on the two trials is insufficient to permit conclusions.

For the comparison of 3D-CRT alone versus 3D-CRT combined with ADT,[23] data on overall survival, all-cause mortality, and prostate cancer–specific mortality reported in the trial favor the combined treatments. Although a single trial, the study was precise with a low risk of bias, which allowed a low strength-of-evidence grade. For the comparison of EBRT alone versus EBRT combined with ADT, data on overall survival, all-cause mortality, and prostate cancer–specific mortality reported in the trial favor the combined treatments with an insufficient strength-of-evidence grade.

Table A. Summary of the main findings from randomized controlled trials for Key Question 1

Comparison and Outcome	Evidence Base	Findings	Risk of Bias	Consistency	Directness and Precision	SOE Grade
RP vs. WW, all-cause mortality	2 trials SPCG-4[24-26] PIVOT[27] (N = 1,426)	SPCG-4: Favors RP at 15 years. ARR, 6.6%; 95% CI, -1.3% to 14.5%. Cumulative incidence: 46.1% vs. 52.7%; RR, 0.75; 95% CI, 0.61 to 0.92 No significant difference between the interventions at 12 years. ARR, 7.1%; 95% CI, -0.5 to 14.7%. Cumulative incidence: 32.7% vs. 39.8% (137 vs. 156 deaths); RR, 0.82; 95% CI, 0.65 to 1.03 PIVOT: No significant difference between the interventions at 12 years. ARR, 2.9%; 95% CI, -4.1% to 10.3% (171 [47.0%] vs. 183 [49.9%] deaths); HR, 0.88; 95% CI, 0.71 to 1.08	Medium	Consistent	Direct Imprecise	Insufficient
RP vs. WW, PCSM	2 trials SPCG-4[24-26] PIVOT[27] (N = 1,426)	SPCG-4: Favors RP at 12 and 15 years. ARR, 6.1%; 95% CI, 0.2% to 12.0%. Cumulative incidence: 14.6% vs. 20.7%; RR, 0.62; 95% CI, 0.44 to 0.87 PIVOT: No significant difference between the interventions. ARR, 2.6%; 95% CI, -1.1 to 6.5 (21 [5.8%] vs. 31 [8.7%] deaths); HR, 0.63; 95% CI, 0.36 to 1.09	Medium	Inconsistent	Direct Imprecise	Insufficient
RP vs. WW, QOL	1 trial SPCG-4[24-26] (N = 695)	No significant difference between the interventions at median followup of 12.2 years	High	Consistency unknown (single study)	Direct Imprecise	Insufficient

Table A. Summary of the main findings from randomized controlled trials for Key Question 1 (continued)

Comparison and Outcome	Evidence Base	Findings	Risk of Bias	Consistency	Directness and Precision	SOE Grade
RP vs. WW, QOL (urinary leakage)	2 trials SPCG-4[24-26] PIVOT[27] (N = 1,426)	Favors WW for urinary leakage (2–4 years) SPCG-4: OR, 2.3; 95% CI, 1.6 to 3.2 PIVOT: RR, 2.69; 95% CI, 1.61 to 4.51	Mediuma	Consistent	Direct Precise	Low[a]
RP vs. WW, QOL (erectile dysfunction at 4 years)	2 trials SPCG-4[24-26] PIVOT[27] (N = 1,426)	SPCG-4: No significant difference between interventions for erectile dysfunction at 4 years PIVOT: RR, 1.84; 95% CI, 1.59 to 2.11. Favors WW at 2 years	Medium	Inconsistent	Direct Imprecise	Insufficient
RP vs. WW, QOL (bowel dysfunction)	1 trial PIVOT[27] (N = 731)	No significant difference between interventions for bowel dysfunction	Medium	Consistency unknown (single study)	Direct Imprecise	Insufficient
RP vs. WW, progression to metastases	2 trials SPCG-4[24-26] PIVOT[27] (N = 1,426)	Favors RP SPCG-4: RR, 0.65; 95% CI, 0.47 to 0.88 PIVOT: HR, 0.40; 95% CI, 0.22 to 0.70	Medium	Consistent	Direct Precise	Moderate
RALRP vs. LRP, QOL (urinary continence, erectile function)	1 trial[28] (N = 120)	Favors RALRP at 1 year Urinary continence: 95% vs. 83.3%; p = 0.042 Erectile function: 80% vs. 54.2%; p = 0.02	High	Consistency unknown (single study)	Direct Precise	Insufficient
RRP vs. BT, QOL	1 trial[29] (N = 200)	No significant difference between the interventions at 5-year followup	High	Consistency unknown (single study)	Direct Imprecise	Insufficient
RPP vs. RRP, QOL (urinary continence, erectile function)	1 trial[30] (N = 200)	Favors RRP for erectile function (60% vs. 42%; p = 0.032) at 2 years; no significant between-group difference in urinary continence	High	Consistency unknown (single study)	Direct Precise (erectile function) Imprecise (urinary continence)	Insufficient
3D-CRT vs. 3D-CRT plus ADT, overall survival	1 trial[23] (N = 206)	Favors 3D-CRT plus ADT at median 7.6-year followup HR, 3.0; 95% CI, 1.5 to 6.4 (44 vs. 30 deaths)	Low	Consistency unknown (single study)	Direct Precise	Low
3D-CRT vs. 3D-CRT plus ADT, all-cause mortality	1 trial[23] (N = 206)	Favors 3D-CRT plus ADT at median 7.6-year followup HR, 1.8; 95% CI, 1.1 to 2.9	Low	Consistency unknown (single study)	Direct Precise	Low

Table A. Summary of the main findings from randomized controlled trials for Key Question 1 (continued)

Comparison and Outcome	Evidence Base	Findings	Risk of Bias	Consistency	Directness and Precision	SOE Grade
3D-CRT vs. 3D-CRT plus ADT, PCSM	1 trial[23] (N = 206)	Favors 3D-CRT plus ADT at median 7.6-year followup HR, 4.1; 95% CI, 1.4 to 12.14 (14 vs. 4 deaths)	Low	Consistency unknown (single study)	Direct Precise	Low
EBRT vs. EBRT plus ADT, overall survival	1 trial[31] (N = 1,979)	Favors EBRT plus ADT at median 9.1-year followup HR, 1.17; 95% CI, 1.10 to 1.35 (57% vs. 62% survival rate)	Medium	Consistency unknown (single study)	Direct Precise	Insufficient
EBRT vs. EBRT plus ADT, PCSM	1 trial[31] (N = 1,979)	Favors EBRT plus ADT at median 9.1-year followup HR, 1.87; 95% CI, 1.27 to 2.74 (8 vs. 4 deaths)	Medium	Consistency unknown (single study)	Direct Precise	Insufficient
EBRT vs. EBRT plus ADT, QOL (sexual function)	1 trial[31] (N = 1,979)	Favors EBRT at 1 year OR, 1.72; 95% CI, 1.17 to 2.52; p = 0.004	High	Consistency unknown (single study)	Direct Precise	Insufficient
EBRT vs. cryotherapy, overall survival	1 trial[32] (N = 244)	No significant difference between interventions at 5 years. Difference, 1.2 (95% CI, 6.8–9.2)	Medium	Consistency unknown (single study)	Direct Imprecise	Insufficient
EBRT vs. cryotherapy, PCSM	1 trial[32] (N = 244)	No significant difference between interventions at 5 years. Difference, 0.3 (95% CI, 4.8–5.4)	Medium	Consistency unknown (single study)	Direct Imprecise	Insufficient
EBRT vs. cryotherapy, QOL (urinary function)	1 trial[33] (N = 244)	Favors cryotherapy (p value was statistically significant) at 3 years	High	Consistency unknown (single study)	Direct Precise	Insufficient
EBRT vs. cryotherapy, QOL (bowel function)	1 trial[33] (N = 244)	No significant difference between interventions at 3 years	High	Consistency unknown (single study)	Direct Imprecise	Insufficient
EBRT vs. cryotherapy, QOL (sexual function)	1 trial[33] (N = 244)	Favors EBRT (p-value was statistically significant) at 3 years	High	Consistency unknown (single study)	Direct Precise	Insufficient

[a]The evidence base for this outcome contained 1 medium and 1 high risk-of-bias study; because of this borderline between medium and high risk, the strength of evidence was lowered from moderate to low.
Note: For the interpretation of SOE grading, see definitions of evidence grades in the Methods section under Strength-of-Evidence Grading.
3D-CRT = 3-dimensional conformal radiation therapy; ADT = androgen-deprivation therapy; ARR = absolute risk reduction; BT = brachytherapy; CI = confidence interval; EBRT = external beam radiation therapy; HR = hazard ratio; LRP = laparoscopic radical prostatectomy; OR = odds ratio; PCSM = prostate cancer–specific mortality; PIVOT = Prostate Intervention Versus Observation Trial; QOL = quality of life; RALRP = robotic-assisted laparoscopic radical prostatectomy; RP = radical prostatectomy; RPP = radical perineal prostatectomy; RRP = radical retropubic prostatectomy; RR = relative risk; SOE = strength of evidence; SPCG 4 = Scandinavian Prostate Cancer Group 4; WW = watchful waiting.

Of 44 nonrandomized comparative studies included, we categorized 41 as high risk of bias for all reported outcomes. (See Table 10 in the full report for risk-of-bias assessment criteria and Table C-2 of Appendix C for individual study assessments.) We categorized the three remaining studies as medium risk of bias because all used instrumental variable analysis, which effectively "pseudorandomizes" patients into different groups and can account for both measured and unmeasured confounders.[34]

Table B summarizes our findings from nonrandomized comparative studies on overall survival, overall mortality, prostate cancer–specific mortality, or QOL for each treatment comparison and outcome with evidence from at least three nonrandomized comparative studies. (See the Results section in the full report for a full description of evidence for all comparisons.) Although the majority of studies had a high risk of bias, the evidence base for all-cause mortality and prostate cancer–specific mortality for the comparison of RP and EBRT included six studies with consistent and precise findings that provide low strength of evidence favoring RP. For all other comparisons/outcomes, the strength of evidence was insufficient.

The definition and severity of adverse events varied greatly across the studies. Adverse events such as urinary incontinence and erectile dysfunction were mostly reported among men who underwent RP. Adverse events such as genitourinary toxicity, gastrointestinal toxicity, and erectile dysfunction were reported among men who received radiation therapy.

Table B. Summary of the main findings from nonrandomized comparative studies for Key Question 1

Comparison and Outcome	Evidence Base	Findings	Risk of Bias	Consistency	Directness and Precision	SOE Grade
RP vs. EBRT, all-cause mortality	6 studies[35-40] (N = 22,771)	Favors RP Five of 6 studies found that overall mortality was significantly lower after RP (followup, 3–15 years)	High	Consistent	Direct Precise	Low
RP vs. EBRT, PCSM	6 studies[35,37-41] (N = 23,301)	Favors RP All 6 studies found that PCSM was significantly lower after RP (followup 3–15 years)	High	Consistent	Direct Precise	Low
RP vs. BT, PCSM	3 studies[35,39,42] (N = 22,337)	Outcomes between groups did not differ significantly in any study	High	Consistent	Direct Imprecise	Insufficient
RP vs. observation, all-cause mortality	4 studies[34,36,40,43] (N = 131,114)	Favors RP with multivariable regression or propensity score analyses, but 1 study using instrumental variable analysis did not find a significant between-group difference	High	Inconsistent	Direct Imprecise	Insufficient
RP vs. observation, PCSM	3 studies[34,40,43] (N = 63,219)	Favors RP with multivariable regression or propensity score analyses, but 1 study using instrumental variable analysis did not find a significant between-group difference	High	Inconsistent	Direct Imprecise	Insufficient
RALRP vs. RRP, QOL	3 studies[44-46] (N = 2,108)	In 1 study, RALRP was associated with greater problems with incontinence. The 2 treatment groups did not differ in sexual dysfunction Two studies found no between-group differences for continence or sexual function	High	Inconsistent for continence; consistent for sexual dysfunction	Direct Imprecise	Insufficient

BT = brachytherapy; EBRT = external beam radiation therapy; PCSM = prostate cancer–specific mortality; QOL = quality of life; RALRP = robotic-assisted laparoscopic radical prostatectomy; RP = radical prostatectomy; RRP = radical retropubic prostatectomy; SOE = strength of evidence.

KQ 2: Specific Patient Characteristics Affecting Outcomes of the Therapies

We identified four RCTs and nine nonrandomized comparative studies that addressed the impact of significant patient characteristics on outcomes. Two RCTs comparing RP and WW and another two RCTs comparing EBRT alone and EBRT plus ADT performed subgroup analysis according to patient characteristics. In the PIVOT trial,[27] investigators reported no differences in all-cause mortality and prostate cancer–specific mortality between RP and WW when patients were stratified according to age. In contrast, investigators in the SPCG-4 trial[24] reported that the advantages of RP over WW in all-cause mortality, prostate cancer–specific mortality, and progression to metastases were statistically significant for patients younger than 65 years of age but not for the older patient group. The SPCG-4 trial investigators noted that the findings of the subgroup analyses should be interpreted with caution because these analyses may misleadingly dismiss differences because of a lack of power.[24]

One study reported that 3D-CRT plus ADT was associated with significantly lower 8-year all-cause mortality compared with 3D-CRT alone for patients with no comorbidity or a minimal comorbidity score. However, for patients with a moderate or severe comorbidity score, all-cause mortality did not differ significantly between the two treatments. For reasons described in the Methods section, all subgroup analyses were considered inconclusive, with insufficient strength of evidence.

Table C summarizes our findings on overall survival, overall mortality, prostate cancer–specific mortality, or QOL from the randomized trials that addressed KQ 2. Results for nonrandomized comparative studies can be found in the Results section of the full report.

Table C. Summary of the main findings from randomized controlled trials for Key Question 2

Comparison	Outcome	Evidence Base	Patient Characteristics by Which Data Were Stratified	Findings	SOE Grade
RP vs. WW	All-cause mortality, PCSM, and progression to metastases at 15 year followup	1 trial SPCG-4[24-26] (N = 695)	Age	There was a significant reduction in all-cause mortality, PCSM, and progression to metastases in the younger than 65 years age category but not in the 65 years or older category.	Insufficient for patient subgroup
RP vs. WW	All-cause mortality and PCSM at 12 years	1 trial PIVOT[27] (N = 731)	Age, race, self-reported performance status	No significant difference between interventions in either younger than 65 years or 65 or older age group, race (white, black, and other), or performance (score 0 or 1–4) category.	Insufficient for patient subgroup
3D-CRT vs. 3D-CRT plus ADT	All-cause mortality at 8 years	1 trial[23] (N = 206)	Comorbidity scores	Among patients with no or minimal comorbidity, all-cause mortality was higher for the EBRT-alone group than for the EBRT plus ADT group. Among men with moderate or severe comorbidity, all-cause mortality was not significantly different between the 2 treatment groups.	Insufficient for patient subgroup

Table C. Summary of the main findings from randomized controlled trials for Key Question 2 (continued)

Comparison	Outcome	Evidence Base	Patient Characteristics by Which Data Were Stratified	Findings	SOE Grade
EBRT vs. EBRT plus ADT	Overall survival, PCSM	1 trial[31] (N = 1,979)	Age, race	Age group was unrelated to survival. EBRT plus ADT was associated with a significantly lower PCSM than EBRT alone among men older than 70 years of age, but not among men 70 years of age or younger. EBRT plus ADT was also associated with significantly greater overall survival and significantly lower PCSM among white patients but not among black patients.	Insufficient for patient subgroup

3D-CRT = 3-dimensional conformal radiation therapy; ADT = androgen-deprivation therapy; EBRT = external beam radiation therapy; PCSM = prostate cancer–specific mortality; PIVOT = Prostate Intervention Versus Observation Trial; RP = radical prostatectomy; SOE = strength of evidence; SPCG 4 = Scandinavian Prostate Cancer Group-4; WW = watchful waiting.

KQ 3: Provider/Hospital Characteristics Affecting Outcomes of the Therapies

We did not identify any comparative study directly examining how provider or hospital characteristics influence the effectiveness of different treatments. As a result, this review does not add new information on this KQ beyond that from the 2008 report. The 2008 report found that results from national administrative databases and surveys suggested that provider/hospital characteristics—including RP procedure volume, physician specialty, and geographic region—affect outcomes. Screening practices can influence the characteristics of patients receiving diagnoses and tumors detected. Screening practices and treatment choices varied by physician specialty and across U.S. regions. Given the diverse readership of this report, we would also like to note a landmark U.S. Government Accountability Office report that found a growing concern that financial incentives (a provider characteristic) may continue to drive treatment selection and costs.[47]

KQ 4: Tumor Characteristics Affecting Outcomes of the Therapies

We identified 4 RCTs and 16 nonrandomized comparative studies that addressed the effect of tumor characteristics. Two RCTs compared RP and WW; another RCT compared EBRT alone and EBRT plus ADT and performed subgroup analysis according to tumor characteristics. In the PIVOT trial,[27] investigators reported that RP did not reduce all-cause mortality and prostate cancer–specific mortality among men with PSA levels of less than 10 ng/mL but resulted in a significant reduction in all-cause mortality (but not prostate cancer–specific mortality) among men with PSA levels higher than 10 ng/mL. In contrast, investigators in the SPCG-4 trial[24] reported that the PSA level (<10 vs. ≥10 ng/mL) did not alter RP's effect in reducing all cause mortality or prostate cancer–specific mortality. However, the tumor stage differed in these trials. In PIVOT almost 45 percent of the men had T2 prostate cancer, whereas in the SPCG-4 study the figure was almost 75 percent.

In another trial, adding short-term ADT to EBRT led to significantly higher overall survival and lower prostate cancer–specific mortality among patients with intermediate-risk prostate cancer, but not among patients with high- or low-risk prostate cancer, compared with EBRT alone. For reasons described in the Methods section, all subgroup analyses were considered inconclusive, with insufficient strength of evidence.

Table D summarizes our findings on overall survival, overall mortality, prostate cancer–specific mortality, or a global QOL score from the RCTs that addressed KQ 4. Results for nonrandomized comparative studies can be found in the Results section of the full report; all findings had insufficient strength of evidence.

Table D. Summary of the main findings from randomized controlled trials for Key Question 4

Comparison	Outcome	Evidence Base	Tumor Characteristics by Which Data Were Stratified	Findings	SOE Grade
RP vs. WW	All-cause mortality and PCSM at median followup of 10 years	1 trial PIVOT[27] (N =731)	PSA levels	No reduction in all-cause mortality among men with PSA levels of ≤10 ng/mL treated with RP compared with WW. All-cause mortality (but not PCSM) was reduced by 13.2% among men with PSA levels of >10 ng/mL who were treated with RP compared with WW.	Insufficient for patient subgroup
RP vs. WW	All-cause mortality at 15 year followup	1 trial SPCG-4[24-26] (N = 695)	PSA levels	No reduction in all-cause mortality among men with PSA levels of <10 ng/mL or ≥10 ng/mL treated with RP compared with WW at 15 year followup.	Insufficient for patient subgroup
RP vs. WW	All-cause mortality at 15 year followup	1 trial SPCG-4[24-26] (N = 695)	Gleason score	No reduction in all-cause mortality among men with Gleason score <7 or ≥7 treated with RP compared with WW at 15-year followup.	Insufficient for patient subgroup
RP vs. WW	All-cause mortality and PCSM at median followup of 10 years	1 trial PIVOT[27] (N = 731)	Risk level based on PSA levels, Gleason score, or tumor stage	There was a 31% relative reduction in all-cause mortality among men with intermediate tumor risk treated with RP compared with WW. There was a significant reduction in PCSM among men with PSA >10 ng/mL and men with high-risk tumors who were treated with RP compared with WW.	Insufficient for patient subgroup
RP vs. WW	All-cause mortality and distant metastases at 15-year followup	1 trial SPCG-4[24-26] (N = 695)	Risk level based on PSA levels, Gleason score, or a WHO grade of 1	There were significant absolute between-group reductions of 13.2% for all-cause mortality and 11.4% for distant metastases among men with low-risk tumors who were treated with RP compared with those in WW at 15-year followup.	Insufficient for patient subgroup
EBRT vs. EBRT plus ADT	Overall survival and PCSM at 10 years	1 trial[31] (N = 1,979)	Risk level based on PSA levels, Gleason score, or tumor stage	Among men with intermediate-risk tumors, overall survival was increased to 60% in the EBRT plus ADT group compared with 54% in the EBRT-alone group. Among men with low-risk tumors, overall survival was increased to 67% in the EBRT plus ADT group compared with 60% in the EBRT-alone group. There was no reduction in PCSM among men with low-risk tumors who were treated with EBRT alone compared with EBRT plus ADT	Insufficient for patient subgroup

ADT = androgen deprivation therapy; EBRT = external beam radiation therapy; PCSM = prostate cancer–specific mortality; PIVOT = Prostate Intervention Versus Observation Trial; PSA = prostate-specific antigen; RP = radical prostatectomy; SOE = strength of evidence; SPCG 4 = Scandinavian Prostate Cancer Group-4; WHO = World Health Organization; WW = watchful waiting.

Discussion

Key Findings and Strength of Evidence

Extended followup data from SPCG-4 and the recently published findings from the PIVOT trial add to our understanding of the effects of RP versus WW or observation in subgroups. However, neither study compared RP with active surveillance. The strength of evidence from the SPCG-4 and PIVOT trials is graded as insufficient for all-cause mortality and prostate cancer–specific mortality at 12 or 15 years (meaning that the evidence does not permit a conclusion). However, both trials reported consistent findings regarding a significant reduction in progression to metastases in the RP group compared with the WW group. This consistency, combined with medium risk of bias and precision, means that the strength of evidence is moderate for this outcome. The 2008 report similarly showed a significant reduction in incidence of distant metastases in the RP group compared with the WW group based on 10-year followup of SPCG-4[18] but did not have evidence from PIVOT to support this finding.

We did not perform a meta-analysis on these outcomes, primarily because of differences between the two trials in enrolled patient populations. Compared with the SPCG-4 trial, the PIVOT enrolled a higher percentage of men with nonpalpable tumors (T1c, 50% vs. 12%) and with low PSA values.[26] The SPCG-4 trial used an eligibility criterion of T1 or T2 stage; however, given the lack of widespread PSA screening in the early portion of the study, these tumors are at higher risk of being understaged by digital rectal examination than PSA-screened tumors in the PIVOT. The two trials also differed in their protocol for the observation arms. Both trials reported similar hazard ratios for prostate cancer–specific mortality, but the hazard ratio for all-cause mortality was higher in the PIVOT than the SPCG-4 trial. This suggests that prostate cancer deaths in the PIVOT may have been diluted by deaths from other causes or competing risks. This conjecture, in turn, suggests that the underlying health of men in the two RCTs was different and poses the question of whether the PIVOT data can apply to a healthy cohort. Furthermore, in the PIVOT study, the median survival was assumed to be 15 years in the original study design and 10 years in the updated design. The PIVOT investigators failed to accrue their targeted enrollment of 2,000 patients to surgery or observation.

In our review, we were unable to draw any conclusions about the effect of various treatments on global QOL. Therefore, it is unclear how patients as a whole will balance the tradeoff between the potential benefit in long-term survival and the potential harms (e.g., urinary incontinence, sexual dysfunction) associated with the treatments. Ultimately, personal preferences and values play a significant role in this decisionmaking. This may be particularly true for patients with life expectancies of less than about 15 years.

This review and the 2008 report both attempted to evaluate whether a particular patient group (in terms of age, race, general health status, and various tumor risk factors) might benefit more than another group from compared interventions. Addressing this question would help patients and clinicians make better informed treatment decisions. The SPCG-4 trial reviewed in the 2008 report performed subgroup analysis by age and had already found that survival benefits of RP compared with WW may be limited to men younger than 65 years of age.[48]

The evidence reviewed in this update does not provide any consistent conclusion on this issue. For example, the SPCG-4 trial found that RP led to significantly lower all-cause and cancer-specific mortality compared with WW among patients younger than 65 years of age but not among the older patient group.[27] However, the PIVOT study did not have the same finding regarding age.[24] The PIVOT trial found that RP did not reduce all-cause or cancer-specific mortality among men with PSA levels of 10 ng/mL or less but resulted in a significant reduction among men with PSA of more than 10 ng/mL. However, this finding is not confirmed by the SPCG-4 trial, which found that overall mortality was reduced by RP regardless of PSA level. Despite these differences, the two trials also show some overlap in findings (reduced mortality with RP) for the subgroup of patients with PSA of more than 10 ng/mL. Nevertheless, inconsistency remains in the evidence. The subgroup analyses might have misleadingly dismissed differences because of the lack of statistical power.[24] Therefore, clear guidance regarding the appropriate patient population for RP is difficult to establish. Four observational studies that used multivariable or propensity score analyses to adjust for known confounding factors found a lower overall mortality risk with RP than with WW,[34,36,40,43] but when one of these studies also performed an instrumental variable analysis (which adjusts for known and unknown confounding factors), no significant between-group difference was observed.[34] Given that the patient population in this latter study was derived from a database of patients 65 years or older, the findings in this analysis are comparable to those of the SPCG-4 trial[24-26] for patients aged 65 years or older.

This current review also evaluated RCTs that compared EBRT alone versus EBRT combined with ADT31 and 3D-CRT alone versus 3D-CRT combined with ADT.23 The evidence based on both RCTs[23,31] suggests that the results for overall survival and prostate cancer–specific mortality favored the combined treatments, although only one RCT[23] met the threshold for low strength of evidence. However, in both studies, the dose of radiation therapy was lower than is currently known to be effective. These findings are similar to the findings of two RCTs summarized in the 2008 report.[18] The subgroup analysis in one RCT[23] also suggests that the advantage of 3D-CRT combined with ADT may occur only among patients with no comorbidity or a minimal comorbidity score for the outcome all-cause mortality. The evidence in another RCT[31] suggests that the advantage of EBRT combined with ADT may occur only among white patients for the outcome of overall survival and among white patients and men older than 70 years of age for the outcome of prostate cancer–specific mortality. For both outcomes, the study found a significant benefit for combined therapy among patients with intermediate-risk prostate cancer, but not among patients with high- or low-risk prostate cancer. In this study, the length of ADT (only 4 months) might have been too short for patients with high-risk disease. Therefore, although it appears that men with intermediate-risk prostate cancer may benefit from 4 to 6 months of ADT, this study could not adequately address either of the study endpoints in the cases in which longer term ADT may be needed. Moreover, treating low-risk patients with EBRT plus ADT would be considered substantial overtreatment by most national clinical practice guidelines. For these reasons, this evidence is weak and requires further validation by new studies before it can be used to form clinical guidance for choosing appropriate cases for the treatments.

For a single treatment comparison, we were able to draw a conclusion from observational evidence based on six studies of high risk of bias but with consistent findings. RP was favored over EBRT for both all-cause mortality[35-39] and prostate cancer–specific mortality with low strength of evidence.[35,37-41] However, we note that radiation dosage was not reported in some studies and a proportion of patients received a lower dose than what is currently considered effective. Furthermore, despite attempts to adjust for known confounders, observational studies are vulnerable to bias from unknown confounding factors. Therefore, RCTs are needed to address this comparison.

Similarly, the evidence for other treatment comparisons covered in the current review needs further validation, particularly via rigorously designed RCTs, to form a more reliable foundation for making clinical recommendations.

As noted in the Methods section, we chose a conservative approach when grading strength of evidence in this report, because multiple factors other than treatment can influence apparent differences in clinical outcomes between interventions observed in these studies. Accordingly, we placed a high value on replication of findings and believe that if the evidence was based on a single RCT, it should be considered sufficient evidence (low strength) only if that RCT had precise findings and was rated as low risk of bias. For studies rated as having high risk of bias, we set a higher bar and required at least three studies with consistent and precise findings. End-users of this report can reasonably choose to set a less conservative bar when making clinical or policy decisions.

Applicability

The evidence-based conclusions are applicable only to the types of patients enrolled in the studies underlying those conclusions, the types of clinical settings in which the studies were conducted, the types of interventions being compared, and the particular outcomes and followup periods reported. Table 37 in the full report summarizes factors that may restrict the applicability of the findings from the RCTs discussed in the previous section.

Although the restrictions on the applicability of the conclusions may vary across the evidence bases for different treatment comparisons, some restrictions may be common to most of these evidence bases. All but one of the RCTs in this review recruited their patients before 2002. Since then, the treatment options compared in many studies have greatly evolved. For example, open surgery was the main treatment technique for RP in the reviewed RCTs. However, in recent years, robotic-assisted surgery has become the dominant technique for RP in the United States. Similarly, for EBRT, BT, and other treatments, advances in technologies and knowledge may allow currently available treatments to better target the cancer, thereby improving the effectiveness and tolerance of treatments. Evidence based on dated medical techniques may not be applicable in current practice.

Additionally, patients studied in the RCTs included in this review may have a different risk profile from patients currently receiving a diagnosis of prostate cancer. Risk profiles may affect the findings of treatment comparisons, although we did not reach any definitive conclusions from the evidence reviewed for KQs 2 and 4 because of the lack of statistical power for detecting between-intervention

differences in the subgroup analyses. Ten to 15 years ago, prostate cancers were primarily detected by digital rectal examination or tissue specimens obtained during transurethral resection of the prostate for treating benign prostatic obstruction. Currently, the vast majority of prostate cancers detected in the United States are found by PSA testing. Men often start to receive PSA tests in their 40s and continue taking the test on a regular basis until their 80s. As a result, patients with an established diagnosis can be younger and have a more confined cancer than those studied in the reviewed RCTs, which further restricts the applicability of the reviewed evidence. Because of intensified concern about overdiagnosis of prostate cancer in recent years, the way to use PSA testing for screening prostate cancer and the criteria for establishing an abnormal PSA test result may continue to change. Patient and tumor characteristics of men with prostate cancer in the future are likely to be different from those of men diagnosed in the past as well as those of men diagnosed today.

Finally, we note that even in well-designed RCTs that found an apparent advantage of one intervention over another, subgroup analyses raise the possibility that not all patients in the target population will derive equal or even any benefit from the treatment with the best average outcome. This is of particular importance given the potential morbidities associated with prostate surgery and radiation therapies, which may be avoided if a more conservative intervention such as active surveillance is deemed appropriate.

Research Gaps

A fundamental research gap involves the development of better methods for staging prostate cancer that is detectable but not metastatic. With current technology, such staging is not straightforward, and choosing treatment based on stage for patients whose prostate cancer is detectable but not metastatic will be difficult until more precise imaging and diagnostic methods are available.

To further address this review's KQs, additional RCTs are needed. In Table G-1 and Table G-2 in Appendix G of the full report, we summarize nine ongoing clinical trials. Ideally, future RCTs should (1) recruit patients with PSA-detected prostate cancer; (2) compare patient-focused outcomes (e.g., all-cause and cancer-specific mortalities, QOL) between treatment options, including AS and techniques used in current practice, and be designed with a long followup. These RCTs should use standardized or validated patient outcome measures, have adequate power to detect significant treatment effects, and define patient subgroups of interest a priori. They should also enroll patients who are representative of current clinical practice using similar enrollment criteria that would allow comparison of the patients' outcomes across studies.

RCTs have had challenges achieving target enrollments for comparing different treatment options. For example, the PIVOT investigators did not achieve their stated target enrollment of 2,000 patients. This suggests that comparative effectiveness research to guide treatment decisions will likely require well-designed observational studies as well.

Observational studies with better design and conduct (e.g., cancer registries and large prospective population-based cohort studies, use of propensity score or instrumental variables, use of validated QOL measures) may provide useful evidence, particularly in cases in which large differences in outcomes might exist. Observational studies may help estimate treatment effectiveness in high-priority patient and tumor subgroups that have not been adequately addressed in RCTs. Findings from observational studies may also help in generating hypotheses and designing better RCTs. We noted and reported that some observational studies conflicted in findings based on analytic methods employed (e.g., instrumental variable analysis vs. propensity scoring vs. multivariable regression analysis). Most of the existing evidence from nonrandomized comparative studies comes with treatment-selection biases.

We did not identify any studies that compared AS with current treatment therapies. Because WW or observation is not AS, more studies are needed to assess the effectiveness of AS. These studies might necessitate adequate consideration of multiparametric magnetic resonance imaging as a tool to enhance observation or AS. Additional research comparing observation or AS with any early intervention is warranted to avoid potential overdiagnosis and overtreatment in men with PSA-detected cancer (especially low PSA/low-risk disease, but possibly intermediate PSA/intermediate-risk disease as well). Future RCTs that compare early intervention versus AS or other early interventions should target patients with higher PSA/higher risk disease, given that the benefits in this group remain uncertain.

Furthermore, because prostate cancer is a significant cause of mortality among men, a research need remains for better prognostic surrogate markers to predict the risk of recurrence among patients with clinically localized prostate cancer.

Finally, some studies discussed in this report suggest that outcomes of surgery and radiation are influenced by center and surgeon case volume and expertise. However, most of these studies did not provide information about practice of care that could have influenced the results. Future studies are needed to fill this gap.

Conclusions

Overall, the body of evidence for treating prostate cancer continues to evolve, but the evidence for most treatment comparisons is largely inadequate to determine comparative risks and benefits. Although limited evidence appears to favor surgery over WW or EBRT and favors radiotherapy plus ADT over radiotherapy alone, the patients most likely to benefit and the applicability of these study findings to contemporary patients and practice remain uncertain. More RCTs and better designed observational studies that reflect contemporary practice and can control for many of the known/unknown confounding factors that can affect long-term outcomes may be needed to evaluate comparative risks and benefits of therapies for clinically localized prostate cancer. We also believe that an urgent need exists to provide clinicians an improved way to categorize patients with prostate cancer into different groups based on associated risk factors. All treatments available for clinically localized prostate cancer can cause bothersome complications, including sexual, urinary, and bowel dysfunction. Patients should be informed and actively involved in the decisionmaking process and consider the benefits and harms of the various treatments.

References

1. American Cancer Society. Cancer Facts & Figures 2012. Atlanta: American Cancer Society; 2012. www.cancer.org/acs/groups/content/@epidemiologysurveilance/documents/document/acspc-031941.pdf.
2. American Urological Association Education and Research, Inc. Prostate-Specific Antigen Best Practice Statement: 2009 Update. Linthicum, MD: American Urological Association Education and Research, Inc.; 2009.
3. Ganz PA, Barry JM, Burke W, et al. National Institutes of Health State-of-the-Science Conference statement: role of active surveillance in the management of men with localized prostate cancer. NIH Consens State Sci Statements. 2011 Dec 5-7;28(1): 1-27. PMID: 23392076.
4. Thompson I, Thrasher JB, Aug G, et al. Guideline for the management of clinically localized prostate cancer: 2007 update. American Urological Association Education and Research, Inc.; 2007. www.auanet.org/guidelines.
5. Thompson IM, Goodman PJ, Tangen CM, et al. The influence of finasteride on the development of prostate cancer. N Engl J Med. 2003 Jul 17;349(3):215-24. PMID: 12824459.
6. Moyer VA, U.S. Preventive Services Task Force. Screening for prostate cancer: U.S. Preventive Services Task Force recommendation statement. Ann Intern Med. 2012 Jul 17;157(2):120-34. PMID: 22801674.
7. Brett AS, Ablin RJ. Prostate-cancer screening--what the U.S. Preventive Services Task Force left out. N Engl J Med. 2011 Nov 24;365(21):1949-51. PMID: 22029759.
8. Fowler FJ Jr, Barry MJ, Walker-Corkery B, et al. The impact of a suspicious prostate biopsy on patients' psychological, socio-behavioral, and medical care outcomes. J Gen Intern Med. 2006 Jul;21(7):715-21. PMID: 16808772.
9. McNaughton-Collins M, Fowler FJ Jr, Caubet JF, et al. Psychological effects of a suspicious prostate cancer screening test followed by a benign biopsy result. Am J Med. 2004 Nov 15;117(10):719-25. PMID: 15541320.
10. Ransohoff DF, McNaughton Collins M, Fowler FJ. Why is prostate cancer screening so common when the evidence is so uncertain? A system without negative feedback. Am J Med. 2002 Dec 1;113(8):663-7. PMID: 12505117.
11. Humphrey PA. Gleason grading and prognostic factors in carcinoma of the prostate. Mod Pathol. 2004 Mar;17(3):292-306. PMID: 14976540.
12. American Cancer Society. What's New in Prostate Cancer Research? September 4, 2012. www.cancer.org/cancer/prostatecancer/detailedguide/prostate-cancer-new-research. Accessed December 13, 2012.
13. Meurs P, Galvin R, Fanning DM, et al. Prognostic value of the CAPRA clinical prediction rule: a systematic review and meta-analysis. BJU Int. 2013 Mar;111(3):427-36. Epub 2012 Aug 9. PMID: 22882877.
14. Halverson S, Schipper M, Blas K, et al. The Cancer of the Prostate Risk Assessment (CAPRA) in patients treated with external beam radiation therapy: evaluation and optimization in patients at higher risk of relapse. Radiother Oncol. 2011 Dec;101(3):513-20. PMID: 21703711.
15. 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. J Clin Oncol. 2013 Apr 10;31(11):1428-34. PMID: 23460710.
16. Punnen S, Freedland SJ, Presti JC Jr, et al. Multi-institutional validation of the CAPRA-S score to predict disease recurrence and mortality after radical prostatectomy. Eur Urol. 2014 Jun;65(6):1171-7. Epub 2013 Apr 8. PMID: 23587869.
17. National Comprehensive Cancer Network. NCCN Clinical Practice Guidelines in Oncology: Prostate Cancer. V4.2013 [slide set]. Fort Washington, PA: National Comprehensive Cancer Network; July 26, 2013. www.nccn.org/professionals/physician_gls/pdf/prostate.pdf.

18. Wilt TJ, Shamliyan T, Taylor B, et al. Comparative Effectiveness of Therapies for Clinically Localized Prostate Cancer. Comparative Effectiveness Review No. 13. (Prepared by Minnesota Evidence-based Practice Center under Contract No. 290-02-0009.) Rockville, MD: Agency for Healthcare Research and Quality; February 2008. www.effectivehealthcare.ahrq.gov/reports/final.cfm. Also available at: www.ncbi.nlm.nih.gov/books/NBK43147/. PMID: 20704036.

19. Dawson B, Trapp RG. Basic & Clinical Biostatistics (LANGE Basic Science). McGraw-Hill/Appleton & Lange; 2004.

20. Viswanathan M, Ansari MT, Berkman ND, et al. Assessing the risk of bias of individual studies in systematic reviews of health care interventions. AHRQ Publication No. 12-EHC047-EF. In: Methods Guide for Effectiveness and Comparative Effectiveness Reviews. AHRQ Publication No. 10(14)-EHC063-EF. Rockville, MD: Agency for Healthcare Research and Quality. January 2014. Chapters available at www.effectivehealthcare.ahrq.gov.

21. Owens DK, Lohr KN, Atkins D, et al. Grading the strength of a body of evidence when comparing medical interventions-Agency for Healthcare Research and Quality and the Effective Health Care Program. J Clin Epidemiol. 2010 May;63(5):513-23. PMID: 19595577.

22. Atkins D, Chang S, Gartlehner G, et al. Assessing applicability when comparing medical interventions: Agency for Healthcare Research and Quality and the Effective Health Care Program. J Clin Epidemiol. 2011 Nov;64(11):1198-207. Epub 2011 Apr 3. PMID: 21463926.

23. D'Amico AV, Chen MH, Renshaw AA, et al. Androgen suppression and radiation vs radiation alone for prostate cancer: a randomized trial. JAMA. 2008 Jan 23;299(3):289-95. PMID: 18212313.

24. Bill-Axelson A, Holmberg L, Ruutu M, et al. Radical prostatectomy versus watchful waiting in early prostate cancer. N Engl J Med. 2011 May 5;364(18):1708-17. PMID: 21542742.

25. Holmberg L, Bill-Axelson A, Steineck G, et al. Results from the Scandinavian Prostate Cancer Group Trial Number 4: a randomized controlled trial of radical prostatectomy versus watchful waiting. J Natl Cancer Inst Monogr. 2012 Dec;2012(45):230-3. PMID: 23271778.

26. Bill-Axelson A, Holmberg L, Filen F, et al. Radical prostatectomy versus watchful waiting in localized prostate cancer: the Scandinavian Prostate Cancer Group-4 randomized trial. J Natl Cancer Inst. 2008 Aug 20;100(16):1144-54. PMID: 18695132.

27. Wilt TJ, Brawer MK, Jones KM, et al. Radical prostatectomy versus observation for localized prostate cancer. N Engl J Med. 2012 Jul 19;367(3):203-13. PMID: 22808955.

28. Porpiglia F, Morra I, Lucci Chiarissi M, et al. Randomised controlled trial comparing laparoscopic and robot-assisted radical prostatectomy. Eur Urol. 2012 Jul 20;63(4):606-14. PMID: 22840353.

29. Giberti C, Chiono L, Gallo F, et al. Radical retropubic prostatectomy versus brachytherapy for low-risk prostatic cancer: a prospective study. World J Urol. 2009 Oct;27(5):607-12. PMID: 19455340.

30. Martis G, Diana M, Ombres M, et al. Retropubic versus perineal radical prostatectomy in early prostate cancer: eight-year experience. J Surg Oncol. 2007 May 1;95(6):513-8. PMID: 17226809.

31. Jones CU, Hunt D, McGowan DG, et al. Radiotherapy and short-term androgen deprivation for localized prostate cancer. N Engl J Med. 2011 Jul 14;365(2):107-18. PMID: 21751904.

32. Donnelly BJ, Saliken JC, Brasher PM, et al. A randomized trial of external beam radiotherapy versus cryoablation in patients with localized prostate cancer. Cancer. 2010 Jan 15;116(2):323-30. PMID: 19937954.

33. Robinson JW, Donnelly BJ, Siever JE, et al. A randomized trial of external beam radiotherapy versus cryoablation in patients with localized prostate cancer: quality of life outcomes. Cancer. 2009 Oct 15;115(20):4695-704. PMID: 19691092.

34. Hadley J, Yabroff KR, Barrett MJ, et al. Comparative effectiveness of prostate cancer treatments: evaluating statistical adjustments for confounding in observational data. J Natl Cancer Inst. 2010 Dec 1;102(23):1780-93. PMID: 20944078.

35. Nepple KG, Stephenson AJ, Kallogjeri D, et al. Mortality after prostate cancer treatment with radical prostatectomy, external-beam radiation therapy, or brachytherapy in men without comorbidity. Eur Urol. 2013 Sep;64(3):372-8. Epub 2013 Mar 13. PMID: 23506834.

36. Rice KR, Colombo ML, Wingate J, et al. Low risk prostate cancer in men >/= 70 years old: to treat or not to treat. Urol Oncol. 2013 Aug;31(6):755-60. Epub 2011 Aug 26. PMID: 21872499.

37. Cooperberg MR, Vickers AJ, Broering JM, et al. Comparative risk-adjusted mortality outcomes after primary surgery, radiotherapy, or androgen-deprivation therapy for localized prostate cancer. Cancer. 2010 Nov 15;116(22):5226-34. PMID: 20690197.

38. Hoffman RM, Koyama T, Fan KH, et al. Mortality after radical prostatectomy or external beam radiotherapy for localized prostate cancer. J Natl Cancer Inst. 2013 May 15;105(10):711-8. PMID: 23615689.

39. Mukherjee S, Reddy CA, Ciezki JP, et al. Risk of developing myelodysplastic syndromes in prostate cancer patients definitively treated with radiation. J Natl Cancer Inst. 2014 Mar;106(3):djt462. Epub 2014 Feb 27. PMID: 24577815.

40. Albertsen PC, Hanley JA, Penson DF, et al. 13-year outcomes following treatment for clinically localized prostate cancer in a population based cohort. J Urol. 2007 Mar;177(3):932-6. PMID: 17296379.

41. Degroot JM, Brundage MD, Lam M, et al. Prostate cancer-specific survival differences in patients treated by radical prostatectomy versus curative radiotherapy. Can Urol Assoc J. 2013 May;7 (5-6):E299-305. PMID: 23766831.

42. Kibel AS, Ciezki JP, Klein EA, et al. Survival among men with clinically localized prostate cancer treated with radical prostatectomy or radiation therapy in the prostate specific antigen era. J Urol. 2012 Apr;187(4):1259-65. PMID: 22335870.

43. Abdollah F, Sun M, Schmitges J, et al. Cancer-specific and other-cause mortality after radical prostatectomy versus observation in patients with prostate cancer: competing-risks analysis of a large North American population-based cohort. Eur Urol. 2011 Nov;60(5):920-30. PMID: 21741762.

44. Barry MJ, Gallagher PM, Skinner JS, et al. Adverse effects of robotic-assisted laparoscopic versus open retropubic radical prostatectomy among a nationwide random sample of Medicare-age men. J Clin Oncol. 2012 Feb 10;30(5):513-8. PMID: 22215756.

45. Krambeck AE, DiMarco DS, Rangel LJ, et al. Radical prostatectomy for prostatic adenocarcinoma: a matched comparison of open retropubic and robot-assisted techniques. BJU Int. 2009 Feb;103(4):448-53. PMID: 18778350.

46. Alemozaffar M, Sanda M, Yecies D, et al. Benchmarks for operative outcomes of robotic and open radical prostatectomy: results from the Health Professionals Follow-Up Study. Eur Urol. 2014 Feb 11. Epub ahead of print. PMID: 24582327.

47. U.S. Government Accountability Office. Medicare: Higher Use of Costly Prostate Cancer Treatment by Providers Who Self-Refer Warrants Scrutiny. Report to Congressional Requesters [GAO-13-525]. Washington: U.S. Government Accountability Office; July 2013. www.gao.gov/assets/660/656026.pdf.

48. Bill-Axelson A, Holmberg L, Ruutu M, et al. Radical prostatectomy versus watchful waiting in early prostate cancer. N Engl J Med. 2005 May 12;352(19):1977-84. PMID: 15888698.

Full Report

This executive summary is part of the following document: Sun F, Oyesanmi O, Fontanarosa J, Reston J, Guzzo T, Schoelles K. Therapies for Clinically Localized Prostate Cancer: Update of a 2008 Systematic Review. Comparative Effectiveness Review No. 146. (Prepared by the ECRI Institute–Penn Medicine Evidence-based Practice Center under Contract No. 290-2007-10063.) AHRQ Publication No. 15-EHC004-EF. Rockville, MD: Agency for Healthcare Research and Quality; December 2014. www.effectivehealthcare.ahrq.gov/reports/final.cfm.

AHRQ Pub. No. 15-EHC004-1-EF
December 2014