

# An Evidence Review of Active Surveillance in Men With Localized Prostate Cancer

## Executive Summary

### Background<sup>a</sup>

In 2011, more than 240,000 men are projected to be diagnosed with prostate cancer, and 33,000 are projected to die from the disease in the United States. In the United States, most cases of prostate cancer are detected via prostate-specific antigen (PSA) screening. The cancer is usually localized, and most tumors have low histological grades and low Gleason scores. Indeed, more than half of prostate cancers detected by PSA screening are expected to be early-stage, low-risk tumors. Such cancers are an infrequent cause of death, and those affected are more likely to die of unrelated causes.

A number of immediate active treatment options are available for localized prostate cancer. Most commonly, radical prostatectomy (RP) or radiation therapy (RT), with or without androgen deprivation therapy (ADT), are offered with curative intent. However, the clinical benefit of immediate therapy with curative intent has not yet been demonstrated for localized prostate cancer in a PSA-screened population. It is likely that a large number of men are receiving active treatment with curative intent without much likelihood of obtaining any clinical benefit due to the slow progression of many prostate tumors. Both surgical and radiation treatments result in significant short- and long-term adverse events, including impotence, urinary dysfunction, and other complications. Thus, determination of the

### Evidence-based Practice Program

The Agency for Healthcare Research and Quality (AHRQ), through its Evidence-based Practice Centers (EPCs), sponsors the development of evidence reports and technology assessments to assist public- and private-sector organizations in their efforts to improve the quality of health care in the United States. The reports and assessments provide organizations with comprehensive, science-based information on common, costly medical conditions and new health care technologies. The EPCs systematically review the relevant scientific literature on topics assigned to them by AHRQ and conduct additional analyses when appropriate prior to developing their reports and assessments.

AHRQ expects that the EPC evidence reports and technology assessments will inform individual health plans, providers, and purchasers as well as the health care system as a whole by providing important information to help improve health care quality.

The full report and this summary are available at [www.ahrq.gov/clinic/epcix.htm](http://www.ahrq.gov/clinic/epcix.htm).

<sup>a</sup>Please refer to the reference list in the full report for a full documentation of statements contained in the Executive Summary.



appropriate management strategy for early-stage, low-risk prostate cancer is an important public health concern.

Active surveillance (AS) and watchful waiting (WW) are two observational followup strategies that forgo immediate therapy in patients with prostate cancer, with the goal of minimizing the morbidities and costs of immediate active treatment for men who may never develop cancer-related symptoms or who are interested in palliative treatments only. AS is curative in intent, and WW is palliative. AS is appropriate in men with disease believed to be indolent and therefore may not require therapy. Because prediction tools are imperfect, these men are monitored closely and treated with curative intent at signs of progression or patient choice. In this way, the considerable adverse effects of treatment are at best avoided, and at least deferred. This approach is to be distinguished from men for whom treatment is deemed inappropriate because of comorbidity; for these men, WW is generally considered, as it offers the option of palliative therapy upon symptomatic disease progression. AS often entails a multifactorial followup of patients—monitoring of PSA values, digital rectal examinations (DRE), prostate imaging, and periodic prostate biopsies—while WW is commonly a relatively passive strategy—with interventions triggered by symptoms. However, there is a continuum of aggressiveness of followup for both AS and WW, as practiced in the community. It should be noted that even though the two terms are used commonly in the scientific literature, the attended intents (curative vs. palliative) of these approaches are not always made clear. Furthermore, many analyses or databases combine AS, WW, and noncurative interventions like primary ADT in their analyses, making it impossible to ferret out issues specifically related to AS.

Immediate active treatment has tradeoffs, including the harms of short- and long-term complications from curative treatments and the benefits of potential reductions in long-term morbidity and mortality. Thus, AS and other observational management strategies may be considered by men who are more interested in avoiding the risks of curative treatment. Therefore, it is important to clarify the appropriate eligibility criteria and followup protocols for these observational strategies that could minimize both unnecessary early curative treatments and avoidable prostate cancer symptoms and deaths. Of course, this assumes that AS is as effective as (or no worse than) immediate curative treatments in an appropriate subgroup of men diagnosed

with prostate cancer. This, however, remains to be proven. Furthermore, some men may be uncomfortable with observational management and feel a strong need to “do something,” and thus AS may be rarely offered, chosen, or adhered to. Therefore, the factors affecting these actions also warrant further investigation.

The National Cancer Institute and the Centers for Disease Control and Prevention are sponsoring a National Institutes of Health (NIH) State-of-the-Science Conference in December 2011 to examine these and other essential issues regarding the role of AS (as opposed to immediate curative intent therapy) in the management of early-stage, low-risk prostate cancer. The NIH has tasked the Agency for Healthcare Research and Quality (AHRQ) Evidence-based Practice Center (EPC) Program to provide the present review of evidence for use in this conference.

## Objectives

The objective of this report is to summarize the existing literature regarding the role of AS in the management of early-stage, low-risk prostate cancer. Both the report and the corresponding NIH State-of-the-Science conference are a part of the NIH Consensus Development Program, the purpose of which is to evaluate the scientific evidence on a particular topic and develop a consensus statement that advances research in that area. This statement is developed by an independent panel that is assembled for the conference. The panel on AS will hear the scientific data, including the findings of the present evidence review, and will then use that information to compose their statement. Additional information about the NIH Consensus Development Program can be found at: <http://consensus.nih.gov/>.

The Conference planning committee crafted the Key Questions to be addressed at the conference, and the EPC was charged with systematically reviewing the relevant literature to address them. Key Question 1 pertains to temporal trends in the natural history of prostate cancer in the United States. Key Question 2 relates to the definitions of observational management strategies (i.e., those involving no active treatment) for prostate cancer used in the published literature. Key Question 3 relates to the factors that influence the offer or acceptance of or adherence to AS. Key Question 4 pertains to the comparative effectiveness of AS versus active treatments for localized prostate cancer. Key Question 5 addresses recommendations for future

research on observational management strategies for localized prostate cancer. The exact wordings of the Key Questions provided to the EPC for systematic review are as follows:

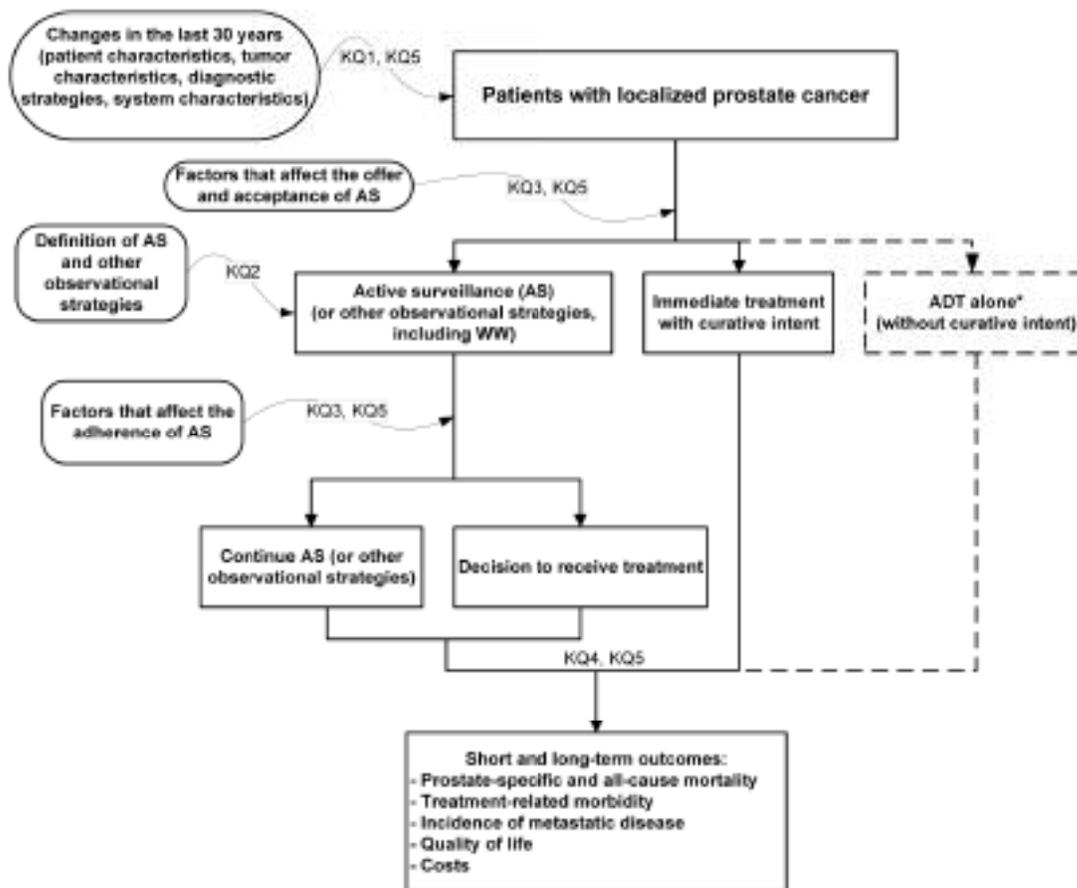
## Key Questions

1. How have the patient population and the natural history of prostate cancer diagnosed in the United States changed in the last 30 years?
  - a. Patient Characteristics
    - i. Age
    - ii. Comorbidity
    - iii. Race/ethnicity
  - b. Tumor Characteristics
    - i. Stage
    - ii. Tumor volume
    - iii. Gleason score
    - iv. PSA
  - c. Diagnostic Strategies
    - i. Biopsy Frequency
    - ii. # cores
    - iii. Histopathologic grading changes
  - d. System Characteristics
    - i. Differences in geographical access
2. How are active surveillance and other observational management strategies defined?
  - a. Common Metrics
    - i. Age
    - ii. Gleason
    - iii. # cores
    - iv. % cores
    - v. PSA (velocity, doubling time)
    - vi. Imaging
    - vii. Behavioral indicators
  - b. Followup Protocols
    - i. Gleason
    - ii. # cores
    - iii. % cores
    - iv. PSA
    - v. Imaging
    - vi. Behavioral indicators
3. What factors affect the offer of, acceptance, and adherence to active surveillance?
  - a. Physician Factors
    - i. Primary care
    - ii. Diagnosing physician
    - iii. Consultant – second opinion
    - iv. Clinical factors
  - b. Patient Factors
    - i. Family involvement
    - ii. Personal preferences
    - iii. Risk perceptions
    - iv. Family history
    - v. Social support
  - c. Delivery System
    - i. Economic incentives and disincentives
      1. Insurance Type (HMO, military, private)
      2. Availability of technology
    - ii. Geographic location
      1. Small area variation
      2. Regional variation
      3. Urban vs. rural
    - iii. Academic centers vs. private practice
  - d. Communication Strategies
    - i. Risk assessment, predictive models
    - ii. Decisionmaking tools and aids
4. What are the comparative short- and long-term outcomes of active surveillance versus immediate treatment with curative intent for localized prostate cancer?
  - a. Prostate-specific and all cause mortality
  - b. Morbidity of primary treatment decision
  - c. Incidence of metastatic disease
  - d. Quality of life
  - e. Costs
5. What are the research needs regarding active surveillance (or watchful waiting) in localized prostate cancer?

## Analytic Framework

To guide this systematic review and facilitate the interpretation of Key Questions, we developed an analytic framework (Figure A) that depicts the logical progression and interconnection of all five Key Questions for this report.

**Figure A. Analytic framework that depicts the five Key Questions that examine the role of active surveillance in the management of men with clinically localized prostate cancer**



ADT = androgen deprivation therapy; AS = active surveillance; KQ = Key Question.

\* ADT alone is a management strategy that is not reviewed in this report.

## Methods

The EPC convened a group of experts in the epidemiology and treatment of prostate cancer to form a Technical Expert Panel, which provided clinical and methodological expertise in interpreting the Key Questions, identifying important issues, and defining parameters for the review of evidence. In addition, input from these experts was sought when questions arose regarding the scope of the review.

### Literature Searches, Eligibility Criteria, and Screening

Multiple literature searches were performed in MEDLINE from inception to August 2011. We searched for recent systematic reviews, and subsequently conducted separate but overlapping searches for each of the first four Key Questions. We used search terms related to prostate cancer, active surveillance, watchful waiting, expectant management, and other related management strategies. We also searched for studies of specific databases, including SEER (Surveillance Epidemiology and End Results) and CaPSURE (Cancer of the Prostate Strategic Urologic Research Endeavor). For Key Question 4, we relied on previous systematic reviews on prostate cancer conducted for the AHRQ EPC program. Searches were supplemented with studies recommended to us by the Technical Expert Panel, reference lists of eligible primary studies and relevant review articles, and targeted searches for economic evaluations. We did not include unpublished data.

Below are the study eligibility criteria we used for the first four Key Questions (no specific literature search was performed for Key Question 5):

**Key Question 1:** Studies of large U.S.-based databases of patients with prostate cancer with time-trend data (reporting changes over a range of years) between 1980 and 2011. Studies must have had a sample size of at least 1000 patients. We also reviewed the latest version of the American Cancer Society Cancer Statistics report<sup>1</sup>, a recent SEER Survival Monograph<sup>2</sup>, and data available on the SEER Web site.<sup>b</sup>

**Key Question 2:** Studies of any design that reported protocols and management strategies for patients receiving observational management (i.e., no immediate curative treatment). We included both studies where the goal of

observation was to identify disease progression indicative of the need for curative treatments, and studies where the goal of observation was to determine the need for palliative treatments.

**Key Question 3:** Three types of studies were included. Firstly, we included studies that used quantitative methods to analyze databases or cohorts of patients to elucidate predictors of the offer or acceptance of or adherence to observational management strategies (including AS and WW). We excluded studies that analyzed ADT together with observational management strategies. We required multivariable analyses adjusting for a minimum of age and tumor stage (if the analysis was not limited to localized cancer) or using a propensity score. Secondly, we included studies using qualitative research methods (e.g., focus groups or surveys) to obtain information on factors that affect the offer or acceptance of or adherence to AS or WW. Eligible studies must have used a predefined approach to collect information. Thirdly, we also searched for experimental studies evaluating the effect of tools, such as decision aids, on the offer or acceptance of or adherence to AS (however, no such studies were found).

**Key Question 4:** We included randomized and nonrandomized, prospective or retrospective longitudinal comparative studies performed in a multicenter setting. Nonrandomized studies must have used multivariable or other methods to adjust for possible confounding, specifically for age and tumor stage, to warrant inclusion. The population of interest was men with clinically localized prostate cancer (T1-T2), without known lymph nodes (N0-X) or metastases (M0-X). No more than 20 percent of the study sample could exhibit more advanced disease. Studies had to compare observational management strategies (without ADT) to active treatment, including RP, external beam RT (EBRT), or brachytherapy (BT), all with or without ADT. However, ADT monotherapy was not considered an active treatment. Outcomes of interest included: prostate-cancer mortality, all-cause mortality, morbidity of primary treatment, metastatic disease, quality of life, and costs.

All five EPC team members participated in screening and selecting studies. An iterative screening process was used for training and to ensure consistency in application of eligibility criteria. Abstracts were screened once. A very low threshold was used to mark a study as of possible interest.

<sup>b</sup>Available at <http://seer.cancer.gov/faststats/>; last accessed September 30, 2011.

During full-text screening, equivocal articles were screened by at least two team members.

## Data Extraction

We extracted bibliographic data, eligibility criteria, enrollment years, study duration, and sample size for all studies. For Key Question 1, we extracted data that allowed reconstruction of trends over time in incidence and mortality, as well as patient-, tumor-, and system-level characteristics of interest. We extracted data into tables of 5-year bins (e.g., 1980-84, 1985-89) from 1980 to 2010. We extracted reported statistical data regarding changes over time in factors of interest. For Key Question 2, we extracted data on patient- and tumor-level characteristics used as eligibility criteria, followup or monitoring parameters, and specific triggers for definitive treatment. We also extracted definitions of disease progression. For quantitative studies (multivariable models) related to Key Question 3, we extracted the definition of the observational strategy, factors of interest, and effect sizes. For qualitative studies (surveys) related to Key Question 3, we extracted the specific survey approach used, the definition of the observational strategy addressed, the qualitative summary of the key study findings, and information to assess the study validity (e.g., survey response rate, survey validation). For Key Question 4, we extracted details about the study population (including eligibility criteria and baseline characteristics), specific interventions compared, outcome definitions, study design, and effect sizes of outcomes of interest.

## Quality Assessment

We formally assessed methodological quality only for studies included for Key Question 4. Studies were graded using standard AHRQ EPC methodology with a three-level grading system (A, B, or C). For RCTs, we primarily considered the methods used for randomization, allocation concealment, and blinding, as well as the use of intention-to-treat analysis, the report of dropout rate, and the extent to which valid primary outcomes were described and clearly reported. Only RCTs and prospective comparative studies could receive an A grade. Retrospective studies could be graded either B or C. For all studies, we used the following in our assessment (as applicable): the report of eligibility criteria, the similarity of the comparative groups in terms of baseline characteristics and prognostic factors, the report of intention-to-treat analysis, important differential loss to

followup between the comparative groups or overall high loss to followup, and the validity and adequacy of the description of outcomes and results. Quality A studies are those judged to have the least likelihood of bias and are considered the most internally valid. Quality C studies have a substantial risk of bias and may not be valid. Quality assessment was performed by the team member responsible for primary data extraction. The quality grade was confirmed by at least one other team member.

## Data Synthesis

All included study data were tabulated into summary tables (provided in the report appendixes) that succinctly describe the important study characteristics and their findings. Time-trend data for Key Question 1 were graphed over the interval of interest (1980–2010). Although we considered generating forest plots for comparative effectiveness data for Key Question 4, the data were inadequate for forest plots to be informative (i.e., there were generally only one or two studies addressing a specific question).

## Grading the Body of Evidence

We graded the body of evidence only for the comparative effectiveness review portion of the systematic review (i.e., Key Question 4). We used standard AHRQ EPC methodology. We assessed the risk of bias of the studies based on their study design and methodological quality, the consistency of data across studies, the applicability of the studies to the U.S. population of men with localized prostate cancer, potential problems with measurement of outcomes in studies, and the precision and sparseness of data. The strength of evidence was rated on a four-level scale: High, Moderate, Low, and Insufficient. Ratings were assigned based on our level of confidence that the evidence reflected the true effect for the major comparisons of interest.

## Results

### Key Question 1

#### **How have the patient population and the natural history of prostate cancer diagnosed in the United States changed in the last 30 years?**

We identified 79 relevant primary observational studies and one systematic review. Of the primary observational studies, 51 analyzed the SEER database or a subset of its component

registries, 9 the linked SEER-Medicare database, 11 the Cancer of the Prostate Urologic Research Endeavor (CaPSURE) database, 5 the National Cancer Database (NCDB), and 3 examined other large U.S.-based databases. In addition, we queried the online SEER database and reviewed the latest version of the Cancer Statistics report prepared annually by the American Cancer Society,<sup>1</sup> and a recent SEER Survival Monograph.<sup>2</sup>

## **Trends in Prostate Cancer Incidence**

Prostate cancer incidence rates rose between 1975 and 1992 (from approximately 100 to more than 240 new cases per 100,000 men per year), and then fell until around 1995. After a period of nonsignificant increase from 1995 to 2000, rates declined again from 2000 to 2007 (to the current level of approximately 156 new cases per 100,000 men per year).<sup>3</sup> Overall, 33 studies provided information on trends of prostate cancer incidence stratified by factors relevant to Key Question 1.

### **Age**

Eleven studies (covering 1969–2005) reported prostate cancer incidence rates according to age group. Collectively, they indicated an increase within all age groups until 1992–93 and then a decline until 1995–99. One study reported the following: compared to the pre-PSA era (1986), the incidence rates in 2005 were 3.64 times higher for men aged 50–59 years, 1.91 times higher for men aged 60–69, and 1.09 times higher for men aged 70–79 years, but only 0.56 times as common in men 80 years or older.

### **Race/ethnicity**

According to the 17 studies (covering 1973–2005) reporting incidence rates stratified by race/ethnicity, all groups experienced increases in prostate cancer incidence since the mid-1980s. The incidence rate appears to have peaked in the early 1990s for all racial/ethnic groups.

### **Tumor stage**

Fifteen studies (covering 1969–2005) reported incidence data, stratified by tumor stage. These studies consistently demonstrated that early-stage (localized and regional) prostate cancer cases were responsible for the observed increase in prostate cancer incidence from the mid-1980s up to the mid-1990s. Studies also consistently indicated a decrease in incidence rates for all disease stages from the mid-1990s to 2000. A single study investigated trends in incidence stratified by T stage, and demonstrated that

compared to 1988–89 the incidence rate in 2004–05 reflected an increase of 76 cases per 100,000 person-years for T1 tumors (from a baseline rate of 42.3 cases per 100,000 person-years) and 11.2 cases per 100,000 person-years for T2 tumors (from a baseline of 95 cases per 100,000 person-years). In contrast, the incidence of T3 or T4 tumors (combined) decreased by 47.1 cases per 100,000 person-years (from a baseline of 55.5 cases per 100,000 person-years) over the same time period.

### **Tumor grade**

Five studies (covering 1973–2005) reported prostate cancer incidence rates stratified by tumor grade (level of tumor differentiation or Gleason score). In studies using SEER data, the increase in prostate cancer incidence observed from the mid-1980s to early 1990s was mainly due to an increase in the incidence rate of moderately differentiated tumors (corresponding to Gleason score 5–6). A single study using SEER data after 2000 reported a continued increase in incidence rate of moderately differentiated tumors from 1988 to 2005 and a concomitant decrease in the incidence rate of well differentiated tumors (corresponding to Gleason score 2–4).

## **Trends in Prostate Cancer Mortality and Survival Rates**

For the overall U.S. population, the NCI's Cancer Trends Progress Report (2009/10) indicates that, after increasing from 1975 to 1991, prostate cancer death rates fell from 1994 to 2007. The baseline prostate cancer mortality in 1975 was 31 deaths per 100,000 men per year and has declined to the current level of approximately 24 deaths per 100,000 men per year. Overall, 17 studies provided information on trends of prostate cancer mortality or survival after diagnosis.<sup>3</sup>

### **Age**

Nine studies (covering 1969–2003) reported prostate cancer mortality and survival rates stratified by age group. Collectively, they demonstrated decreases in the mortality rate for all age groups between the early 1990s and 1999. One study of prostate cancer-specific survival indicated that over time (1988–95) the proportion of patients diagnosed with prostate cancer who died of their cancer had decreased (i.e., patients with prostate cancer have increasingly died of other causes) across all the age groups considered (>50 years old).

### **Comorbidity (other primary cancers)**

One study demonstrated that throughout 1988 to 1995 prostate cancer patients with other primary tumors were consistently less likely to die of prostate cancer compared to patients with no other primary tumors.

### **Race/ethnicity**

According to 15 studies (covering 1969–2000) reporting prostate cancer mortality and survival rates stratified by race/ethnicity, mortality rates among blacks were consistently higher compared to those of non-Hispanic whites. Similarly, black patients with prostate cancer were at higher risk for death due to prostate cancer compared to non-Hispanic whites, although the difference between these two groups appeared to decrease over time.

### **Tumor stage**

Six studies (covering 1969–2001) reported information on prostate cancer mortality stratified by tumor stage. Data indicated that over time the proportion of deaths due to prostate cancer among patients diagnosed with the disease had decreased, particularly for patients with early-stage (localized or regional) disease at diagnosis.

### **Tumor grade**

Three studies reported information on prostate cancer mortality stratified by tumor grade (of patients diagnosed in 1973–97). One study demonstrated that the probability of dying from prostate cancer among patients diagnosed with prostate cancer decreased during the study period (1988–95). Although the decrease was observed for all cancer grades, it was more pronounced among patients with well- and moderately-differentiated tumors. The second study demonstrated that, compared to patients with well differentiated tumors, patients with moderately and poorly differentiated tumors had a higher probability of prostate cancer death (more than 2-fold and more than 4-fold higher, respectively). These differences were relatively constant over the time period covered by the study (1988–95). The third study demonstrated that patients with low grade tumors have the highest relative survival compared to those with higher grade disease; improvements in relative survival over time were observed for all tumor grades.

## **Patient, Tumor and System-Level Characteristics at Diagnosis**

We identified 52 observational studies reporting on patient characteristics at presentation.

### **Age**

Twenty-one studies (covering 1973–2005) reported information on patients' age at presentation. Among seven studies evaluating average age at diagnosis of prostate cancer, five found reductions in the average age of patients whereas two studies did not report any changes during their respective time periods. Fifteen studies that evaluated the distribution of patients' ages into discrete categories generally supported a trend toward younger age at diagnosis (the effect was significant in four of the six studies reporting statistical tests).

### **Comorbidity**

Two studies (covering 1997–2003) reported information on comorbidities at diagnosis. The CaPSURE analysis found no statistically significant difference in the distribution of patients with no, one or two, or three or more comorbidities, when comparing 1997–99 versus 2000–03. The POCS analysis found that the proportion of patients with no comorbidity increased from 78.3 percent in 1998 to 87.4 percent in 2002.

### **Race/ethnicity**

Eighteen studies (covering 1973–2003) reported information on race/ethnicity. No consistent pattern in the racial or ethnic distribution of cases over time was found: some studies indicated that the number of whites increased over time, others that it remained stable, and others that it decreased. Studies using the same database often provided discrepant results even for overlapping time periods; these findings may be a reflection of the different inclusion criteria used in each study.

### **Tumor stage**

Twenty-two studies (covering 1973–2007) reported information on trends in the distribution of prostate cancer stage at diagnosis. Studies reporting on cancer stage consistently demonstrated decreases in the proportion of patients presenting distant disease and concomitant increases in the proportion of patients with localized or regional disease, over their respective time periods. Studies consistently demonstrated reductions in the proportion of

patients presenting with higher T stages. The two studies reporting on T1/T2 tumors both demonstrated a decrease of T1a/T1b tumors and T2a tumors and an increase in T1c tumors.

### **Tumor volume**

We did not find data on changes in tumor volume.

### **Tumor grade**

Sixteen studies (covering 1973–2006) reported information on trends in the distribution of tumor stage at diagnosis. These data consistently demonstrated reductions in the proportion of patients diagnosed with well- or poorly-differentiated tumors (including undifferentiated tumors) with concomitant increases in the proportion of patients with moderately-differentiated disease.

### **Prostate specific antigen**

Eight studies (covering 1989–2007) reported PSA data. These studies indicated that PSA values at diagnosis have decreased over time (i.e., that a larger number of patients are currently diagnosed with PSA concentrations below 10 ng/mL).

### **Biopsy frequency**

Four studies (covering 1982–2001) reported information on trends in the performance of prostate biopsies. A study based on the SEER-Detroit database reported that the proportion of prostate cancer patients diagnosed through biopsy (compared to those diagnosed through other procedures, such as transurethral resection of the prostate) increased over time. A similar trend was observed in a study based on the SEER-New Mexico registry. A SEER-Medicare study demonstrated an increase in the age-adjusted rate of biopsy procedures (from 685 to 2,600 per 100,000 men) between 1986 and 1991. An update from the same database reported that there was no statistically significant change in the population biopsy rate between 1993 and 2001.

### **Number of cores**

One study examined trends in the number of biopsy cores obtained during diagnostic workup, and found that between 1997 and 2002 the average number of cores obtained per patients had increased by 0.41 cores annually (from a mean of 7.5 to a mean of 9.8 cores per patient).

### **Histopathologic grading changes**

One study reported the results of regrading in 2002–04 pathology slides from patients diagnosed in 1990–92. The regrading resulted in the assignment of significantly higher Gleason scores compared to the original readings (mean score increase from 5.95 to 6.8).

### **Differences in geographical access and other system-level factors**

Four studies (covering 1986–2003) reported information on changes in the distribution of patients by system-level factors. Among three studies on trends in the distribution of patients' insurance status at diagnosis, the two CaPSURE analyses demonstrated a decrease in the proportion of patients with Medicare coverage at the time of diagnosis over the time periods covered (1997–2003 and 1989–2001). The POCS analysis did not demonstrate a change in the distribution of insurance status over time (1998–2002). An analysis of POCS comparing 1998 to 2002 reported an increase over time in the number of patients residing in areas of higher median income. Patterns in the distribution of income are difficult to interpret because sampling strategies changed and different regions were included at the different time points. An analysis of NCDB found little evidence of change in the distribution of patients by hospital caseload over time (1986–87 and 1992).

### **Trends in Treatment Patterns**

Among the 21 studies (covering 1973–2008) from which data could be gleaned regarding treatment patterns over time, most demonstrated decreasing trends in the proportion of patients being managed with observational strategies of no active treatment (AS, WW or expectant management), with or without androgen deprivation therapy (ADT). In all seven studies providing data since 2000, the proportion of patients receiving AS or WW was less than 10 percent; this also held true for the subgroups of patients with “low-risk disease” investigated in two studies.

### **Key Question 2**

#### **How are active surveillance and other observational management strategies defined?**

The terms AS and WW (as well as others) have been used by investigators to denote strategies both with and without curative intents. There is a broad spectrum of approaches for observational strategies described in research publications.

For the purpose of operationalizing the process of summarizing the various definitions, we divided protocols into those clearly described as having curative intent and those in which their aims were either unclear or primarily palliative, regardless of how these regimens were labeled. This categorization was applied for practical reasons, not to suggest what the definitions or protocols for AS, WW, or any other observational strategy should be.

### **Strategies With Curative Intent**

Sixteen unique cohorts reported criteria and protocols for AS (i.e., studies that met our criteria of monitoring triggers for curative treatment of prostate cancer other than symptom progression). In all cohorts, AS was offered to men with low-risk or clinically localized prostate cancer.

#### **Eligibility Criteria**

Other than restriction to men with clinically localized prostate cancer (T1 or T2), eligibility criteria for AS varied across the cohorts. The most commonly used eligibility parameters were Gleason score (12 cohorts), PSA (10 cohorts), and number of biopsy cores positive for cancer (8 cohorts).

#### **Age**

Only three studies used age as an eligibility criterion, restricting subjects to men under age 70 or 80 years.

#### **Gleason score**

Twelve cohorts based eligibility for AS on the Gleason score. Generally, cohorts used Gleason score of 6 or less (no pattern 4 or 5). Three cohorts allowed Gleason pattern 4.

#### **Number of cores positive for cancer**

Eight cohorts used a maximal number of biopsy cores positive for cancer as part of the eligibility criteria for AS. Five cohorts allowed two or fewer cancer-positive cores; three cohorts allowed three or fewer. Some cohorts used sextant, some octant, and some extended (>10 cores) biopsies.

#### **Percentage cancer involvement in each core**

Five cohorts used “low-volume disease” as part of the patient eligibility criteria for AS. In three cohorts, the definition of “low-volume disease” was involvement of less than half of any individual core with cancer. In the other two cohorts, the criterion percent of biopsy cores with cancer involvement was described variably as less than half of two

biopsy cores, less than 20 percent in one or two biopsy cores, and less than 33 percent of biopsy cores.

#### **Prostate-specific antigen**

Ten of 16 cohorts used PSA as part of the eligibility criteria for AS. Three PSA thresholds were used:  $\leq 10$  (7 cohorts),  $\leq 15$  (3 cohorts), and  $\leq 20$  (2 cohorts) ng/mL. Two cohorts used PSA density (PSA per volume of prostate tissue) thresholds.

#### **Imaging**

One cohort required a chest radiograph as part of patient eligibility criteria, and another cohort noted that magnetic resonance imaging was selectively used at diagnosis.

#### **Behavioral indicators**

No behavioral indicator was used explicitly as a criterion for AS enrollment.

#### **Followup Protocols**

All 16 cohorts included regular PSA testing in the followup protocol. Twelve cohorts included regular digital rectal examination (DRE). Fourteen cohorts performed routine rebiopsy. The testing frequency of PSA, DRE, and rebiopsy varied across the cohorts. One cohort also incorporated a regular bone scan schedule. Criteria for recommending curative treatments varied across the cohorts. The recommended treatments were not standardized and were left at the discretion of treating physicians and patients in many of the cohorts.

#### **Gleason score**

Twelve cohorts described using the Gleason score as part of their monitoring criteria for disease progression. Generally, disease progression was defined as a Gleason score or pattern greater than those used in the eligibility criteria for AS.

#### **Number of cores positive for cancer**

Eight cohorts included the minimum number of biopsy cores positive for cancer as part of their monitoring criteria for disease progression. Two criteria were used: three or more and greater than four positive biopsy cores (6 and 3 cohorts, respectively). Rebiopsy frequencies varied across the cohorts.

#### **Percentage cancer involvement in each core**

Six cohorts used more than 50 percent cancer involvement in each biopsy core as part of monitoring criteria for disease

progression. Two other cohorts considered an increase in tumor volume as part of the monitoring criteria for disease progression, but specific percentage cancer involvement was not reported.

### **Prostate-specific antigen**

All 16 cohorts included regular PSA testing in the followup protocol. Six cohorts considered rising PSA and/or PSA kinetics as part of triggers for treatment but did not specify the detailed criteria. Nine cohorts used a variety of PSA triggers for treatment.

### **Imaging**

One cohort performed an annual bone scan for the first 2 years and biennially thereafter. Another cohort reported that magnetic resonance imaging of the prostate was selectively performed every 1 to 3 years during followup.

### **Behavioral indicators**

No study used a formal assessment of any behavioral indicator for triggering active treatment as part of their followup protocol, but one cohort reported that some patients requested treatment due to anxiety related to increasing PSA concentration.

### **Observational Management Strategies With Palliative Intent**

Thirteen cohorts reported followup protocols for patients who initially received no treatment and who were subsequently treated only for symptomatic progression.

### **Eligibility Criteria**

The six cohorts that enrolled patients in the pre-PSA screening era primarily based enrollment on clinical staging alone. In the PSA era, the seven cohorts mostly enrolled patients with stage T1 or T2 cancer or without evidence of nodes or metastases. The commonly used patient eligibility criteria were PSA (5 cohorts), age (4 cohorts), Gleason score (4 cohorts), and normal bone scan findings (4 cohorts).

### **Age**

Four cohorts included age as part of their eligibility criteria. The different thresholds used were less than 75 years (2 cohorts), less than 85 years, and between 50 and 75 years.

### **Gleason score**

Four cohorts used Gleason score thresholds. Three used a threshold of less than 8. One required that less than 25 percent of the tumor was Gleason grade 4 and less than 5 percent grade 5.

### **Number of cores positive for cancer**

No cohort used this factor.

### **Percentage cancer involvement in each core**

No cohort used this factor.

### **Prostate specific antigen**

Five cohorts used PSA as part of their eligibility criteria, with thresholds of less than 50 ng/mL (4 cohorts) and less than or equal to 15 ng/mL (1 cohort).

### **Imaging**

Four cohorts required normal bone scan findings. One of these cohorts also required normal chest radiograph findings.

### **Behavioral indications**

No cohort used this factor.

### **Followup Protocols**

Five of the six cohorts in the pre-PSA screening era included regular prostate acid phosphatase (PAP) testing and bone scan in the followup protocol. The sixth cohort reported regular PSA and DRE in the followup protocol for patients who received no treatment after the introduction of PSA in 1990. All seven cohorts in the PSA screening era included regular PSA testing. Compared with AS cohorts (see previous section), rebiopsy was not commonly included in the followup protocol among WW cohorts.

### **Gleason score**

No cohort used this factor.

### **Number of cores positive for cancer**

No cohort used this factor.

### **Percentage cancer involvement in each core**

No cohort used this factor.

### **Prostate-specific antigen**

Three cohorts formed in the pre-PSA screening era reported that PSA testing became part of followup protocol after PSA became available. All six cohorts in the PSA screening era included regular PSA testing as part of followup protocol. However, rising PSA concentration alone was not used as a trigger for treatment in five cohorts. The sixth cohort reported that “hormonal manipulation was demanded by the protocol when the PSA rose to 50 ng/mL.”

### **Imaging**

Five cohorts in the pre-PSA screening era included regular bone scan in the followup protocol. One cohort also included regular chest and skeletal radiographs in the followup protocol. Another cohort reported that computed tomography of the pelvis was conducted infrequently. Three cohorts in the PSA screening era included regular bone scans and chest radiographs in the followup protocol. Another cohort reported that all patients underwent “multiple bone scans” during followup.

### **Behavioral indications**

No cohort explicitly used this factor.

Implicit in the Key Question is a comparison between AS and other observational strategies in the modern PSA era. Thus, we compared the 16 unique cohorts reporting formal protocols to monitor triggers for curative treatment with the 7 unique cohorts of other observational strategies with primarily palliative intent in the PSA screening era. Enrollment into AS protocols more commonly used Gleason score as a threshold than other observational strategies. They also used the number and percentage of cores positive for cancer as a threshold, while none of the other strategies used these factors. Both sets of strategies generally used some sort of PSA criteria, but the thresholds in AS were generally lower (10–15 ng/mL) than the other observational strategies (15 or 50 ng/mL). AS protocols had more clearly defined followup processes than other observational management strategies, with explicit indications for curative treatment including increase in Gleason scores, number and percentage of positive cores (on rebiopsy), and PSA velocity. AS protocols generally did not include imaging as part of their followup processes. In contrast, other observational strategies typically included imaging in their followup, specifically bone scan and chest radiography. They also generally did not employ rebiopsy but did use

PSA in their followup. Comparison of the followup frequencies between AS and other observational strategies showed that PSA testing and DRE were common in both strategies, but somewhat more frequent with AS protocols, at least within the first year of followup.

### **Key Question 3**

#### **What factors affect the offer of, acceptance of, and adherence to active surveillance?**

We included three types of studies to address this Key Question. We included multivariable database analyses of predictors for the offer or acceptance of or adherence to AS (or WW). We included survey or questionnaire studies addressing the same issues. We also searched for experimental studies evaluating the effect of tools, such as decision aids, on the offer or acceptance of or adherence to AS (however, no such studies were found). Of note, the outcomes of many of the examined studies were either treatment with an observational strategy or interruption (cessation) of the observational strategy. Studies generally did not directly analyze the offer or acceptance of or adherence to AS.

#### **Primary care**

One survey of New Zealand general practitioners found that 45 percent would offer observational management if the patient’s life expectancy was <10 years, but only 3 percent would offer observational management to patients with a longer life expectancy. Five surveys of patients reported that their physician’s treatment recommendation was the most influential factor in deciding on their treatment. In one survey, 81 percent of men on observational management who ultimately received active treatment believed that the treatment was favored by their physicians; in contrast, only 24 percent of the physicians’ notes documented that the physician recommended treatment.

#### **Diagnosing physician**

One survey of patients on observational management strategies reported that observational management strategies were offered by 36 percent of the physicians who had made the initial diagnosis.

#### **Consultant—second opinion**

One survey was of men diagnosed with early-stage cancer. They had not yet decided on treatment and were recommended by their urologists to seek a second opinion.

None of the men followed through with the recommendation to seek a second opinion, but the offer reinforced their trust and confidence in their urologists. A survey of Australian men who had a urological consultation reported that 71 percent of the urologists discussed observational management strategies, compared with 92 percent who discussed RP and 87 percent RT. One survey of urologists regarding men with localized cancer and few comorbidities found that 4 percent preferred observational management strategies; two-thirds preferred RP. The same study reported that 20 percent of patients thought that treatment options were not discussed, while only 1 percent of the urologists thought so. In a survey of men and their urologists, the urologists, in an initial consultation setting, recommended observational management strategies to 25 percent of men and offered 0.5 more treatment options than the urologists in a second opinion visit setting, who recommended observational management strategies to only 16 percent of men.

### **Clinical factors**

One survey of urologists and radiation oncologists reported that about 10 to 20 percent would recommend observational management strategies for a 65 year old man with a low PSA, a Gleason score of 4 or 5, in good health, with negative DRE, and no evidence of nonlocalized disease. Almost none would recommend observational management strategies for those with higher PSA or Gleason scores. The responses of urologists and radiation oncologists did not differ significantly. Numerous multivariable analyses found that receipt of observational management strategies was predicted by older age, an increased number of comorbidities, lower Gleason score, well-differentiated tumor, lower stage disease, lower PSA, or low-risk on the D'Amico scale. Multivariable analyses also found that interruption of observational management strategies was predicted by higher stage disease, higher PSA at diagnosis, decreased free-to-total PSA ratio, or more rapid PSA increase, but not comorbidities or Gleason score; two of four studies found an association with younger age and one of three with higher D'Amico risk score.

### **Family involvement**

In two surveys, advice from family and friends was the most influential factor in deciding treatment in 19 and 9 percent of men surveyed. In a focus group, half the men reported relying on influential others to make a treatment decision

(either for or against observational management strategies). In an open-ended interview of men with localized disease, 4 percent reported that family opinions were a reason for not choosing observational management strategies.

### **Personal preferences**

One analysis compared men who refused randomization, but instead selected AS (i.e., men who did not participate in the trial), to men in the trial who were randomized to AS. It found that lower baseline anxiety was associated with the decision to choose AS (and not be randomized). Four surveys found that concern for treatment side effects (primarily impotence and incontinence) were reasons that men chose observational management strategies. Three multivariable analyses found predictors of choosing observational management strategies included the desire to avoid side effects or having current bowel problems, urinary dysfunction, or other urinary conditions; sexual dysfunction was predictive of choosing RT over observational management strategies. One multivariable analysis also reported that increased anxiety was associated with an increased probability of interruption of observational management strategies.

### **Risk perceptions**

One set of interviews in men with low-risk prostate cancer reported that physician description of prostate cancer affected treatment choice. One survey of men with early stage prostate cancer reported that men who chose RP over RT or observational management strategies perceived prostate cancer as a significantly more serious disease. Another survey of men with localized prostate cancer reported that fear of consequences was the most common reason for not selecting observational management strategies.

### **Family history**

Two multivariable analyses reported that family history was not a significant factor in predicting interruption of observational management strategies.

### **Social support**

Four multivariable analyses reported that not being married or in a permanent relationship were associated with an increased probability of receiving observational management strategies. One survey of couples in which the men were diagnosed with early-stage cancer but had not yet

decided on treatment concluded that couples ruled out options based on both formal (provided by the physicians) and informal (provided by family and friends) information, and that they also “considered both their own individual histories and concerns and their shared life experiences.” One multivariable analysis reported that marital status was not associated with time to interruption of observational management strategies.

#### **Insurance type**

Two multivariable analyses reported that having Medicare insurance increased the probability of receiving WW/AS compared with private or Veterans Administration insurance. One analysis reported that having preferred provider organization or health maintenance organization coverage decreased the probability of receiving observational management strategies versus RP. It also reported that Medicare supplemented with fee-for-service, health maintenance organization, or preferred provider organization coverage decreased the probability of receiving observational management strategies versus RP. One multivariable analysis reported that insurance status was not a significant factor in predicting interruption of observational management strategies.

#### **Availability of technology**

No study addressed this factor.

#### **Small area variation**

No study addressed this factor.

#### **Regional variation**

One multivariable analysis comparing the registries in the National Cancer Institute’s Patterns of Care study claimed that men who resided in New Jersey had an increased probability of receiving observational management strategies compared with men in California (excluding three major cities). Comparisons among other registries were nonsignificant. Another multivariable analysis reported that men in Northeast had a decreased probability of selecting observational management strategies (versus active treatments) compared with men in California (excluding three major cities).

#### **Urban versus rural**

One survey of men with prostate cancer in North Carolina reported that there was no significant difference between urban and rural residents in North Carolina as to whether the option of observational management strategies was

discussed with their physicians. One multivariable analysis reported that men who resided in urban areas (vs. rural areas) had a decreased probability of receiving observational management strategies versus RP or RT. The survey in North Carolina reported that there was a difference in whether physician recommendation was the most influential factor in the treatment decision between urban and rural residents (62 percent vs. 44 percent, respectively).

#### **Academic centers versus private practice**

One multivariable analysis reported that treatment facility status (academic vs. community practice) was not a significant factor in predicting receiving observational management strategies versus active treatment.

#### **Risk assessment, predictive models**

No study addressed this factor.

#### **Decision-making tools and aids specifically for AS**

No study addressed this factor.

### **Key Question 4**

#### **What are the comparative short- and long-term outcomes of active surveillance versus immediate treatment with curative intent for localized prostate cancer?**

In order to understand the effectiveness of AS relative to active treatment options, studies of AS need a control group for comparison. As such, we did not include single-arm AS cohort studies, which cannot address comparative effectiveness questions. However, no study reported clinical outcomes specifically for AS management strategies with deferred treatment with curative intent versus immediate definitive treatment. Therefore, there is insufficient evidence to evaluate the comparative effectiveness of AS management with curative intent versus immediate definitive treatment in men with localized prostate cancer.

Faced with a lack of studies comparing AS to immediate active treatment, we elected to evaluate studies that compared other observational management strategies (largely resembling WW) with immediate treatment. In addition to previously published systematic reviews and evidence reports, our searches identified two updates from multicenter RCTs (four publications, three on clinical/quality of life outcomes and one on costs) and 16 cohort studies (3 prospective and 13 retrospective). Notably, the majority of evidence for this Key Question came from

retrospective analyses of observational studies. Confounding by indication is likely in these studies, due to the differences in patient characteristics and risk profile between patients treated with observational strategies and those who received active treatment.

### **Observational Management Strategies Versus Radical Prostatectomy**

Studies generally reported that men treated with RP had lower all-cause or prostate cancer-specific mortality rates than men on WW. The development of metastatic disease was assessed by a single study that found a significant benefit for RP compared to WW. Morbidity of primary treatment was reported by two studies that suggested an increased risk for urethral stricture (and procedures to treat it) were less likely among patients on observational management. Quality of life (QoL) was reported in three studies; the results varied across different domains of QoL measure.

### **Observational Management Strategies Versus Radiation Therapy**

Studies generally reported that men treated with RT had lower all-cause mortality rates than men on WW. One study reported prostate cancer-specific mortality information and did not find a statistically significant difference between RT and observational management. No study reported on treatment comparisons for the development of metastatic disease. One study did not find a significant difference in morbidity between observational management and BT or EBRT. QoL measures and satisfaction with treatment were reported in four studies; the results varied across different domains of QoL measure.

### **Observational Management Strategies Versus Combined Active Treatments or Combined Radiation Treatment Modalities**

One study reported that active treatments (RP, RT, and BT considered together) resulted in lower all-cause and prostate-cancer-specific mortality rates compared to WW. Morbidity of primary treatment was reported by only one study, which found that a group of patients receiving EBRT and BT (combination therapy) had a higher rate of receiving treatments for urethral stricture compared to a group managed using observational management strategies.

### **Costs**

Short- and long-term costs appear to be higher for active treatment strategies (RP or RT) compared to WW; however, evidence originated from small studies (or studies where the subgroup of patients receiving observational management was small) using heterogeneous measurement methods. We did not identify any primary study comparing actual costs of AS versus active treatment strategies; economic modeling using U.S. prices suggested that within 10 to 15 years of diagnosis AS may be less costly compared to active treatments; a study using a lifetime horizon indicated that AS may be associated with higher costs compared to RP and BT, but lower costs compared to intensity modulated RT (IMRT) and proton beam RT. We note that model based costs are sensitive to the model assumptions and choice of inputs.

### **Key Question 5**

**What are the research needs regarding active surveillance (or watchful waiting) in localized prostate cancer?**

The evidence directly addressing the four principal Key Questions is largely incomplete. There is not yet consistency among clinicians or researchers as to the definitions of AS or WW, the standard protocols for the interventions, or how to manage patients whose cancers show signs of progression. There are also many gaps in the evidence regarding the numerous specific factors and subgroups of interest to the conference.

This review implicitly assumes that it is possible to identify men who are at sufficiently low risk of progression of their prostate cancer that AS can be a safe and appropriate option for them. However, additional basic and clinical research is needed to more accurately classify or predict those men whose diseases are indeed at a low risk of progression. These are the men that presumably would be most appropriate to consider offering AS.

### **Key Question 1—Patient Population and Natural History Changes in the Last 30 Years**

Better understanding of time trends can be gained by improving the data collected and expanding the scope of major U.S. databases. In particular, stage and grade information are often incomplete, requiring researchers to create broad categories that place major limitations on

analyses. The SEER database often appears inadequate for analyses on races other than blacks and whites; this may require adding new registries to SEER that better represent other races.

A misclassification bias is likely in the analyses of SEER using the “best available information” on staging information, because the “best available staging information” depends on the treatment the patients receive. Patients having surgery are staged more accurately than those with clinical or imaging staging alone. This bias could be reduced if the SEER database maintained the staging information available prior to surgery.

### **Key Question 2—Definition of Active Surveillance**

Little new research is needed to address how active surveillance has been defined by researchers. However, interpretation of future studies would be best served if there were a standard, agreed-upon definition of AS that clearly distinguished it from WW and other forms of withheld or noncurative treatments. A consensus conference may be the most appropriate forum to define AS. Features of the definition will need to include (1) the goal or intent of the intervention; (2) the “eligibility criteria,” a determination of which patients should be offered AS based on disease and patient characteristics; (3) the “followup protocol,” the minimum set of tests that should be followed and their timing; and (4) the criteria or triggers for stopping AS to seek definitive treatments.

Assuming that AS is an intervention plan that many patients may select (if offered) in order to avoid the side effects from immediate invasive treatment for a potentially nonlethal disease, it would be desirable to determine the best AS protocol that would minimize prostate-cancer specific morbidity and mortality, and that patients and caregivers would adhere to. This best AS protocol should be investigated by randomized or other prospective comparative studies that directly compare different protocols. Examples of comparisons for future trials could include use of different combinations of followup testing, different timing for the tests, and different definitions of progression that would determine when curative treatment is offered. The outcomes of greatest clinical importance are those that are most pertinent to patients’ health, well-being, and longevity. Examples include all-cause mortality, prostate-cancer-specific mortality, symptomatic disease, urological and other

complications (from testing, treatment, or deferring treatment), quality of life, anxiety, and family dynamics. Also of interest would be overall costs, use of resources, and numbers of negative invasive tests (i.e., biopsies showing no progression, thus arguing they were unnecessary).

At a minimum, future study reports should be very explicit and clear as to their definitions of AS (or WW), the goals of the intervention, the exact protocols, the exact definitions of progression, how and when protocols or standards changed during their study (and why), and how often and why patients and clinicians chose not to follow the protocols.

### **Key Question 3—Factors That Affect Offer of, Acceptance of, and Adherence to Active Surveillance**

Current databases tend to have data only about what treatment patients received and when. Therefore, whether different treatment options were offered to them, whether they accepted those options, and whether they adhered to their initial choices could only be inferred. Even the best analysis of predictors of initial treatment cannot adequately address this Key Question. Thus, full statistical analyses of predictors will require the prospective collection of data specifically about what interventions were offered to each patient, which treatments the patients accepted, and when they chose to receive curative treatment despite lack of evidence of progression. These datasets will need to be sufficiently large to allow for testing of multiple predictor variables. In addition, future studies should only perform complete analyses of all treatment options without arbitrarily grouping treatments or selectively excluding treatments. This will minimize bias and increase clarity about what is being tested.

Future database analyses and prospective observational studies should focus on those predictors that are amenable to change or that can be acted upon. Researchers should avoid interpreting analyses to suggest that men with certain demographic (or other nonmodifiable) features are more likely to accept treatment and thus other men should not be offered treatment.

Further surveys of patients, their families, and their clinicians are warranted. To improve reliability, these should be adequately powered to ensure that sufficient numbers of men were treated with different interventions and to allow full analyses of the tested predictors. Studies should use

established methods including standardized qualitative research designs and, ideally, validated questionnaires to elicit preferences. Studies of this sort also need to consider the overall adequacy of discussion with patients regarding management options—and documentation of those discussions. Adequate documentation of these discussions will surely improve the veracity of some of these survey data.

Future Key Questions of interest could include comparisons of interventions that improve the likelihood that eligible men are offered AS, that improve acceptance of AS, and that improve adherence with AS, so long as it remains the most appropriate treatment. Arguably, it is more important to first establish how to successfully get men offered, accepting, and adhering to AS, before determining which men are at greatest risk of failing to receive AS.

#### **Key Question 4—Active Surveillance Versus Immediate Curative Treatment**

A randomized controlled trial with long followup would provide the best evidence to adequately assess the differential effects between AS and immediate curative treatment. The least biased, most reliable study design comparing two interventions is the randomized controlled trial that adheres to modern methodological standards. Outcome assessors—particularly those who conduct psychometric testing—should be blinded. The primary outcomes of interest should be patient-centered outcomes, including mortality and disease-free survival, psychometric measurements, adverse events, resource utilization, and costs. Trials need to be of sufficiently long duration to collect data on the clinically relevant outcomes. However, limited resources and possible difficulties recruiting patients may preclude such a study.

In lieu of randomized trials, adequate findings may possibly be extracted from long-term databases with prospectively collected data. However, these studies, too, should use AS protocols that are defined a priori and undergo minimal change over time or between centers. The determination of which patients are potentially eligible for AS should also be made a priori. These studies will need to use multivariable analyses, propensity scores, or other validated methods to adjust for the broad range of factors that affect the decision to use AS. We do not believe that retrospective studies are capable of providing adequate data for unbiased analyses.

Subgroup analyses of both randomized trials and prospective observational comparative studies should be conducted to look for particular sets of men who may benefit most (or least) from one approach or the other. Preferably, these subgroups should be considered a priori. The factors listed in Key Questions 1 and 2 form a good starting point to consider which subgroups may be of interest. In addition, future studies that could uncover better bio- and imaging markers of indolent versus aggressive disease, and thus could better stage patients as having either low or high risk disease, are necessary to better inform which patients are most likely to benefit from observational versus active treatment.

#### **Discussion**

Prostate cancer epidemiology is affected by population-level trends, such as the aging of the U.S. population, but also by changes in the application of screening and diagnostic technologies among the population at risk. Keeping these caveats in mind, studies indicate that men in all racial/ethnic groups experienced increases in prostate cancer incidence since the mid-1980s, with rates peaking in the early 1990s. For all groups, incidence rates declined between the early-1990s and 1999. Studies have consistently demonstrated that early-stage (localized and regional) prostate cancer cases were responsible for the observed increase in prostate cancer incidence from the mid-1980s up to the mid-1990s. Studies also demonstrated decreases in the prostate cancer-specific mortality rate for all age groups between the early-1990s and 1999. Mean age of diagnosis has also decreased over time for both blacks and whites. Another consistent trend in SEER data has been the decrease in low-grade (Gleason score 2–4) and high grade ( $\geq 7$ ) tumors, and a concomitant increase in intermediate grade tumors (Gleason 5–6). It has been hypothesized that this effect is due to changes in histopathological grading guidelines, a preference towards avoiding assigning Gleason 2–4 scores based on prostate cancer biopsy samples, and PSA test's ability to detect moderately differentiated tumors with higher accuracy (compared to poorly-differentiated tumors). Most studies demonstrated decreasing trends in the proportion of patients being managed with strategies other than RP or RT throughout their respective time periods. Studies explicitly reporting on AS/WW-type strategies indicated decreases in the proportion of patients receiving such treatments over time; this was true even for subgroups of men with “low-risk disease.”

There is not yet a consensus among clinicians or researchers as to the definitions of AS or of WW, the standard protocols for the interventions, or how to manage patients whose cancers show signs of progression. This is evidenced by the 16 unique cohorts formed in the PSA screening era that used different formal protocols to monitor triggers for curative treatment of prostate cancer. In all these cohorts, AS was offered to men with low-risk or clinically localized prostate cancer although no uniform criteria were used to identify these men, with the exception that no cohorts enrolled patients with tumors of a clinical stage greater than T2. They employed different combinations of periodic DRE, PSA testing, rebiopsy, and/or imaging findings to determine different thresholds used for seeking definitive treatments. The AS followup protocols also varied across these cohorts.

Owing to the variation in usage of the terms AS and WW, and their intended and often mixed treatment objectives (both curative and palliative), it is often difficult when reviewing the studies to know whether patients received true AS or WW, or were simply not treated (for a variety of reasons) or experienced delays in their treatment (and thus initially had no treatment).

Only two studies specifically examined factors related to men who were enrolled in an active monitoring protocol with triggers for curative treatments. The first found that the free to total PSA ratio and T stage were independent predictors of time to radical treatments in patients on the protocol, while initial PSA, PSA density, Gleason score, number of positive cores, and prostate volume were not independent predictors. The second study found that men with decreased baseline anxiety and higher socioeconomic status were associated with decreased probability of willingness to consent to randomization for AS versus definitive treatment (i.e., these men proactively selected AS). The rest of the heterogeneous studies reported on men who did not receive treatments or initial treatments. Therefore, whether they were on AS or WW could not be readily discerned. The following patient and clinical variables are potentially important in increasing the probability that a patient receives an observational management strategy: older age, presence of comorbidities, lower Gleason score, lower tumor stage, lower diagnostic PSA, lower risk groups, or decreased baseline anxiety. The following patient and clinical variables are potentially important in increasing the probability that a patient

interrupts an observational management strategy to seek definitive treatments: younger age, higher tumor stage, higher diagnostic PSA, higher PSA velocity, higher risk groups, or increased anxiety.

As most of these tentative conclusions are drawn from multivariable analyses of large databases that did not specifically address the factors that affect the offer or acceptance of or adherence to AS, whether different treatment options were offered to the patients, whether they accepted those options, and whether they adhered to their initial choices could only be inferred from whether they received the treatments or not. In addition, retrospective studies could not provide adequate data for unbiased analyses, because patient characteristics are strongly associated with initial treatment choice.

No trial provided results from comparisons of AS with RP or RT in men with localized diseases. One trial reported that men who underwent RP had lower mortality than men on WW; one trial reported that there was no difference in mortality comparing men having undergone RP with men in WW. Retrospective studies suggest that men on conservative management had a higher prostate-cancer-specific mortality than men treated with RP. Men who had RPs had more urinary complications than men on WW. Retrospective studies also reported that men treated with RT had lower mortality than men on WW. They also reported higher rates of urinary strictures in men treated with RT compared with men on WW. Definitive conclusions for men with low-risk disease on AS or WW versus RP or RT will have to await results from two ongoing trials: Prostate cancer Intervention Versus Observation Trial (PIVOT: Observation vs. RP) and Prostate Testing for Cancer and Treatment trial (ProtecT: AS vs. RP or RT).

Although cost calculations using retrospective primary data were performed using different methods and followup durations in each study, it appears that generally WW is associated with lower treatment costs compared with active treatment. On the other hand, model-based cost analyses of AS compared to active treatments suggested that AS costs accumulate over time. In these models, at 10 and 15 years of followup, AS appeared to be less expensive than active treatments. However; based on a model with a lifetime horizon, the costs of AS may exceed those of RP and BT with long term followup, and may be lower than those of IMRT or proton beam RT.

## References

1. Siegel R, Ward E, Brawley O, Jemal A. Cancer statistics, 2011: The impact of eliminating socioeconomic and racial disparities on premature cancer deaths. *CA Cancer J Clin.* 2011;61:212-236.
2. Ries, LAG, Young JL, Keel, GE, et al. SEER Survival Monograph: Cancer Survival Among Adults: U.S. SEER Program, 1988-2001, Patient and Tumor Characteristics. Bethesda, MD: National Cancer Institute; 2007.
3. National Cancer Institute. Cancer Trends Progress Report - 2009/2010 Update. Bethesda, MD: National Institutes of Health; 2010.

## Full Report

This executive summary is part of the following document: Ip S, Dahabreh IJ, Chung M, Yu WW, Balk EM, Iovin RC, Mathew P, Luongo T, Dvorak T, Lau J. An Evidence Review of Active Surveillance in Men With Localized Prostate Cancer. Evidence Report No. 204. (Prepared by Tufts Evidence-based Practice Center under Contract No. HHS 290-2007-10055-I.) AHRQ Publication No. 12-E003-EF. Rockville, MD: Agency for Healthcare Research and Quality. December 2011.  
[www.effectivehealthcare.ahrq.gov/reports/final.cfm](http://www.effectivehealthcare.ahrq.gov/reports/final.cfm).

## For More Copies

For more copies of An Evidence Review of Active Surveillance in Men With Localized Prostate Cancer: Evidence Report Executive Summary No. 204 (AHRQ Pub. No. 12-E003-1), please call the AHRQ Clearinghouse at 1-800-358-9295 or email [ahrqpubs@ahrq.gov](mailto:ahrqpubs@ahrq.gov).





AHRQ Pub. No. 12-E003-1  
December 2011