Comparative Effectiveness Research Review Disposition of Comments Report

Research Review Title: An Evidence Review of Active Surveillance in Men with Localized Prostate Cancer

Draft review available for public comment from August 16, 2011 to September 13, 2011.

Research Review Citation: 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/Technology Assessment No. 204. (Prepared by Tufts Evidence-based Practice Center under Contract No. HHSA 290-2007-10055-I.) AHRQ Publication No. 12-E003-EF. Rockville, MD: Agency for Healthcare Research and Quality. December 2011. Available at: www.effectivehealthcare.ahrq.gov/reports/final.cfm.

Comments to Research Review

The Effective Health Care (EHC) Program encourages the public to participate in the development of its research projects. Each comparative effectiveness research review is posted to the EHC Program Web site in draft form for public comment for a 4-week period. Comments can be submitted via the EHC Program Web site, mail, or email. At the conclusion of the public comment period, authors use the commentators' submissions and comments to revise the draft comparative effectiveness research review.

Comments on draft reviews and the authors' responses to the comments are posted for public viewing on the EHC Program Web site approximately 3 months after the final research review is published. Comments are not edited for spelling, grammar, or other content errors. Each comment is listed with the name and affiliation of the commentator, if this information is provided. Commentators are not required to provide their names or affiliations in order to submit suggestions or comments.

The tables below include the responses by the authors of the review to each comment that was submitted for this draft review. The responses to comments in this disposition report are those of the authors, who are responsible for its contents, and do not necessarily represent the views of the Agency for Healthcare Research and Quality.

Source: http://www.effectivehealthcare.ahrq.gov/index.cfm/search-for-guides-reviews-and-reports/?productid=859&pageaction=displayproduct
Published Online: December 2011

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer 1	Executive Summary	On Page ES-10, the authors should clarify that cohorts have allowed Gleason 3+4 for AS, as detailed in the main sections.	We have specified that cohorts generally used Gleason score 6 or less (or no pattern 4 or 5). Three cohorts allowed Gleason pattern 4.
Peer Reviewer 5	Executive Summary	p.14. The flow chart should acknowledge that a third path exists after the diagnosis of prostate cancer: treatment without curative intent (ADT).	We have added "ADT alone (without curative intent)" as a treatment strategy.
Peer Reviewer 5	Executive Summary	p.17 line 36. The trend described here continued until 2005, as outlined in the body of the report (p. 51)	Thank you. We have added this information in the executive summary.
Peer Reviewer 5	Executive Summary	p. 19 (and p. 54-55). I have trouble with the way the information regarding trends in tumor grade (p. 19 lines 5-9) and histopathologic grading changes (lines 29-32) is presented. The correlation between the two should be more clearly drawn. I appreciate p. 56 line 13 – that the studies on this topic did not meet inclusion criteria – but to my mind the two sections should at least be next to each other so that the connection may be drawn more readily by a reader.	We did not change the order of the Key Questions. Instead we have provided a reference from the section on "trends in mortality" to the section on "trends in histopathologic grading" and added a note on the importance of the latter in interpreting the former.
Peer Reviewer 5	Executive Summary	p. 19 line 24. Number of cores: in my opinion, the more useful information to include is the fact that the mean number of cores increased, rather than the +0.41 cores/patient/year.	The "+0.41 cores/patient/year" is derived from linear regression of the number of cores obtained over diagnosis year. Thus, it represents the annual change in the mean number of cores. We have provided this information in the ES. We have also provided the actual mean number of cores observed in the first and last year of the study.
Peer Reviewer 5	Executive Summary	p. 19 line 48. The first sentence is unclear here and conflicts with the body of the report (p. 57) – should read "decreasing trends" and for clarity's sake defer mention of ADT until the body. In fact, the use of observational strategies (excluding ADT) has decreased over this period.	Thank you. We have corrected the inconsistency.
Peer Reviewer 8	Executive Summary	Pg 10. There is agreement although not universal that the GS should be 6 and excludes those with Gleason grade 4 or 5. The authors state there is little agreement which I would change to general agreement about the grade for inclusion with some exceptions.	We have specified that cohorts generally used Gleason score 6 or less (or no pattern 4 or 5). Three cohorts allowed Gleason pattern 4.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer 8	Executive Summary	On pg 15 the authors should make it clear that there are no prospective randomized studies comparing AS to initial treatment. There are reports of AS vs treatment for pts who would have met the criteria for AS but the pt or doctor chose treatment.	The study designs of the included studies are stated explicitly. We included all studies that met eligibility criteria.
NCI	Executive Summary	I'll look at their literature but I think they are setting up a false dichotomy. Watchful Waiting can occur with or without active surveillance. When there is active surveillance than the intent is treatment at a point where progression seems clear and cure remains possible. When there is not active surveillance the intent is palliation as needed because it is known that cure is not possible.	We are only providing an operational definitions of AS and WW so we can carry out the review. How these terms should be eventually defined is up to the Conference participants. We have further clarified our methods and logic for categorizing studies, and our lack of intent to formally define these strategies.
NCI	Executive Summary	This makes it sound like the future is hopeless. Another way to say this is that WW is used because there is minimal risk of progression so the treatment may be worse than the risk of death from the disease. Observation for symptoms may be an appropriate way to minimize morbidity. That is the question this review is undertaking.	We have improved our descriptions of the potential value of observational strategies in the Introduction.
NCI	Executive Summary	Without some clarification in the prior paragraph the "tradeoffs" in this sentence aren't obvious	We have clarified the paragraph in the introduction
NCI	Executive Summary - Analytic Framework	Where in this branching diagram is the arm which is watchful waiting? Aren't the considerations that go into WW with or without AS part of the discussion at this point.	The analytic framework is not designed to address every treatment option. We have added other observational strategies, including WW, to the cells with AS
NCI	Executive Summary - Results- Question 1	The paragraph would be strengthened if you could say the magnitude of the difference in incidence at its height, and when the last data was available. Eg. The overall age-adjusted incidence rose to y x the baseline and then fell through 2007 to z x the baseline.	We have provided this information both in the Executive Summary and the Results section of the report.
NCI	Executive Summary - Results- Q1	The absolute increase in cases is interesting but less interesting to me than the relative rate before and after PSA became available.	We have provided the baseline rates for this study in the Executive Summary.

Commentator & Affiliation	Section	Comment	Response
NCI	Executive Summary - Results- Q2 (1st paragraph)	Something here about summarizing the criteria that separates protocols with curative intent from those without curative intent would point to the important process clinicians face of deciding who is eligible for each approach, and for whom do we have data on the effects of each approach.	We have added explanation to explain how we operationalized the process of summarizing the various definitions.
NCI	Executive Summary - Results- Q2	Do you mean cohort here?	No. This was the wording of the original Key Question; it refers to the tissue cores obtained through biopsy.
NCI	Executive Summary - Results- Question 2	Might be the place to summarize that 3 used a Gleason score of 8. The fact that one used scores/region raises issues of whether this is standard. These characteristics are important though because they are critical to guiding providers to who might be eligible.	This section briefly compared protocol with curative intent with other observational strategies in the modern PSA era. We focused more on followup protocol. Besides, not all three cohorts used a Gleason score of 8 as part of their inclusion criteria were form in PSA era. Therefore we elected not to discuss this in the executive summary.
NCI	Executive Summary - Results- Question 4	It would be very helpful to spell out Radical Prostatectomy here. Yes, the summary was earlier but for people not steeped in the disease it is helpful to spell it out as a reminder of what RP stands for.	We have spelled out “Radical Prostatectomy”, as suggested.
NCI	Executive Summary - Results- Q4	Ditto. Spell out Radiation Therapy as a reminder.	We have spelled out “Radiation Therapy”, as suggested.

Commentator & Affiliation	Section	Comment	Response
Michael Krachon	Executive Summary-Background	Active surveillance (AS) and watchful waiting (WW) are two observational follow-up strategies that forgo immediate therapy in patients with prostate cancer. Active surveillance generally connotes the monitoring of a potentially curable prostate cancer and intervening with a curative-intent treatment at the earliest sign of worrisome progression. In contrast, watchful waiting generally connotes postponing therapeutic interventions until symptom development, with the primary objective being palliation of the symptoms rather than an attempt at a cure. Active surveillance often entails a multifactorial follow-up of patients—monitoring of PSA values, digital rectal examinations, prostate imaging, and periodic prostate biopsies—while watchful waiting is a relatively passive strategy—with interventions triggered by symptoms.	We have improved our explanations of AS and WW.
Gene Gardner	Executive Summary	STRUCTURED ABSTRACT I saw no relationship between the PURPOSE statement and the CONSLUSIONS offered. This was Very Disappointing and discouraged me from further reading of the report.	The relationship is that the current evidence only poorly addresses the key questions. Systematic review, as opposed to narrative review, describes the evidence as it is.
AUA	Executive Summary-Results- Question 2	In KQ2, page ES-10, under Eligibility criteria, the authors strive to make the point that the eligibility criteria are variable. I think this is an important finding, in the sense that the reader ought to come away with a sense of the extent of variability, however, the authors seem to over-emphasize this in the Executive Summary, out of proportion to the data presented. For example, in considering the Gleason score criteria, they state “there was little agreement across cohorts in the threshold used.” In reality, there is very little variability. Twelve of the 15 cohorts used definitions that are nearly identical (#1: <= 6; #2 <= 3+3 without Gleason 4 or 5 pattern). I suspect that those studies in group #1 had no more than a few patients with Gleason	We agree. We have rewritten the text to point out the commonly used thresholds and some exceptions.

Commentator & Affiliation	Section	Comment	Response
		4+2, 2+4, 5+1 or 1+5 disease, making these two definitions nearly uniform. Of the remaining 3, one also allowed Gleason 3+4 for patients over the age of 69 and the other two also allowed Gleason 7 patients. Thus, the description “little agreement” overstates the case. Similarly, regarding PSA criteria, the authors state in the Executive summary that studies “used a wide range of thresholds”, mostly ranging from 10 to 15. While there is some variability in individual criteria used, there is wide *agreement* that active surveillance be reserved for patients who are perceived to be low-risk for prostate cancer morbidity and mortality, either because of low-risk disease characteristics or because of advanced age and comorbidities. The differences between protocol eligibility criteria are merely variations on this theme. Of course it would be convenient to have greater agreement across protocols in order to evaluate outcomes across institutions, but I do not think it is fair to characterize the criteria as having “little agreement” or a “wide range of thresholds”. Because these protocols developed ‘organically’ at various institutions, there is no reason to expect that there would be universal agreement on exact eligibility criteria. Thus, data demonstrating variability in eligibility criteria should be presented neutrally, rather than being imbued with value judgments about its extent.	

Commentator & Affiliation	Section	Comment	Response
AUA	Executive Summery-Question 3	On page 61, it is noteworthy that the survey on New Zealand doctors, showing that only 3% would recommend WW if a patient's life expectancy was more than 10 years was a survey of *general practitioners.* Some have criticized the doctors who manage men with localized prostate cancer (urologists and radiation oncologists) for over-using their therapies (either for profit motive or because they ignore data that calls the therapies into question), and suggest that internists may be better suited to guide patients through prostate cancer screening and treatment decisions. This study demonstrates that internists lean toward recommending treatment in men with low-risk prostate cancer and a life-expectancy of 10 years or more. While it seems reasonable for internists and treating physicians to work together in guiding patients through prostate cancer management decisions, there is no reason to expect that internists have a better perspective for doing so. I think it is worth highlighting this concept in Executive Summary, to indicate that AS research and clinical decisions are perhaps best led by doctors who care for these patients, such as urologists.	This is an evidence review and we do not make practice recommendations. The opinion that you expressed is best discussed in an active surveillance consensus conference.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer 1	Introduction	Bio- and imagining markers of indolent versus aggressive disease. The current report does not address the substantial need for translational science to better decipher indolent from aggressive disease. Clinical staging in prostate cancer (based on DRE, Gleason score and PSA) is limited. Some patients classified as low risk by clinical staging will have cancers that behave aggressively. There is also great heterogeneity in the phenotype of high risk cancer. This topic would seem to fit naturally under research needs and is resonant with joint NIH/AHRQ focus on CER and personalized medicine (e.g. identification of subgroups most likely to benefit from various management strategies).	Thank you. This excellent point has been added to the future research section (KQ 5) for KQ 4.
Peer Reviewer 2	Introduction	please see general comments	Please see below for responses to general comments.
Peer Reviewer 3	Introduction	Page 1, para 3: Consider pointing out that there's a continuum of aggressiveness of follow-up between AS and WW. For example, in PIVOT, though the conservative arm was primarily a WW strategy, periodic bone scans were done in part to detect metastases before they became symptomatic.	Thank you. This concept has been added to this paragraph
Peer Reviewer 3	Introduction	Page 1, para 4: Clarify that an AS or WW strategy can be somewhat less effective in terms of preventing cancer mortality (a benefit delayed far into the future), as long as it offers substantial benefit in terms of delaying or avoiding treatment side effects (with the benefit often beginning to accrue immediately).	The text has been rephrased on the bottom of page 1.
Peer Reviewer 4	Introduction	well done and relevant in all respects very comprehensive	Thank you. No further response necessary.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer 5	Introduction	The introduction captures well the issues surrounding the overdiagnosis and – treatment of low risk prostate cancer in the PSA era. Please see attached file for specific comments.	Thank you. Please see below for responses to specific comments.
Peer Reviewer 6	Introduction	1:26. I don't think I'd agree that WW and AS are used "interchangeably" in the literature. The two definitions may be conflated by necessity if it cannot be determined which strategy was followed (e.g., in an analysis of administrative data), but especially with respect to intent, no investigators with substantial AS cohorts would agree that WW is an equivalent term.	Thank you. This section has been rewritten taking into account your comments.
Peer Reviewer 6	Introduction	1:56. The authors reasonably contrast AS to "curative intent" RP or RT. But at least some mention should be made of highly prevalent use of androgen deprivation as monotherapy for men with localized disease. This is non-curative in most cases and not endorsed by most guidelines, but is commonly used. It's an important point because multiple high-profile papers based on SEER or other registries which cannot identify ADT use reliably have conflated WW, AS, _and_ primary ADT as "conservative management."	Thank you. This has been added to Introduction.
Peer Reviewer 6	Introduction	2:24. I think "risk perception" is an important line item under physician factors as well as patient factors, and in the case of MDs would apply not only to patient's risk of tumor progression, but also to fear of medicolegal exposure in the event that a patient on AS did progress to incurable disease.	Unfortunately, we cannot change the Key Questions.
Peer Reviewer 6	Introduction	2:31. I'm surprised there's not an incentive line specifically referring to MD payments for radiation vs. surgery vs. ADT vs. AS… 2:32. Should add to the list of tumor characteristics (and to methods / results): disease risk, as assessed by some multivariable instrument (D'Amico, nomogram, CAPRA score, etc). This is arguably much more important than trends in any individual characteristic.	Unfortunately, we cannot change the Key Questions.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer 6	Introduction	2:40. At some point should address the issue of nomenclature: whether some minimal-risk prostate cancers should be re-named (e.g., “IDLE” tumors, as per Esserman/Thompson JAMA 2009)	We believe this is beyond the scope of this report. We did review the mentioned article to check if it met eligibility criteria.
Peer Reviewer 6	Introduction	5:34. I couldn’t find this list… would be interesting to review 6:43.	This list was redacted from the draft version.
Peer Reviewer 6	Introduction	6:43. “Morbidity” should include morbidity (not just mortality) of progressive cancer if the decision is not to treat. Next line: ”disease” should be singular not plural.	We have clarified the list of outcomes.
Peer Reviewer 1	Methods	Management of prostate cancer in the older elderly. The current report does not address management of prostate cancer among those > 75 years. While few studies have reported on this population, demographic trends point to substantial growth of this (often quite healthy) population of men over the next 20 years. How should men on observation (either AS or WW) who transition into this age bracket be managed? There is a very brief reference to this issue in the research section, but it may be worthwhile to expand this topic.	Studies of the older elderly (eg, >75) are included. We described age-related results as they were analyzed and reported in the original studies.
Peer Reviewer 2	Methods	please see general comments	Please see below for responses to general comments.
Peer Reviewer 3	Methods	This section is clear, and the decisions made regarding the scopes of the various literature reviews thoughtful. Overall, outstanding methods. No specific comments.	Thank you. No further response necessary.
Peer Reviewer 4	Methods	well done and relevant in all respects very comprehensive	Thank you. No further response necessary.
Peer Reviewer 5	Methods	The methods are appropriate to the question raised. Please see attached file for specific comments.	Thank you. No further response necessary.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer 6	Methods	7:49. If systematic review ref lists were fair game for identifying studies, perhaps narrative review ref lists should also have been considered—I believe there are more recent narrative reviews published (one very recent on in JCO 2011 comes to mind).	We have followed standard systematic review protocols, which generally include screening reference lists from previous systematic reviews. We did not download and screen through reference lists of the thousands of review articles on prostate cancer.
Peer Reviewer 6	Methods	10:21. I would have argued for including studies which compared expectant management to ADT monotherapy. As per my comment above, ADT is commonly used, and some high - quality studies have compared ADT to expectant management which give good insights into outcomes in the absence of treatment. Furthermore, the adverse impacts of ADT monotherapy may be at least as significant (or even more so in some cases) than those of "curative" local treatments.	We followed the guidance of the Technical Expert Panel for this report and excluded ADT monotherapy for this analysis. The main focus of this report is on AS, not all forms of noncurative treatment.
Peer Reviewer 6	Methods	10:37. Just a general caveat that CaPSURE may be nationally representative, but it is not a statistically valid population sample, unlike PCOS, SEER, etc.	We agree that CaPSURE does not use a probability-sampling scheme. We also note that, within each region covered, SEER attempts to capture all cancer cases and SEER regions are not randomly selected (instead they target specific population groups). Thus, SEER studies also do not represent a probability sample of the total US population. For this reason, we have used the term "databases sourced from the U.S. population" to indicate that the databases draw cases from large underlying populations, but not necessarily using a formal sampling approach. We have added a note relevant to CaPSURE in the opening section of KQ1.
Peer Reviewer 6	Methods	10:39. I might have added VA studies to the list of data sources explicitly searched for evidence on trends and outcomes.	VA studies met criteria. We have added the VA as another example.
Peer Reviewer 6	Methods	10:39. Would also explicitly refer to the NIDDK/UDA compendium and associated published studies which cross multiple large data sources.	This section describes the primary studies that were included, in contrast to the mentioned compendium. We reviewed and screened out the compendium.

Commentator & Affiliation	Section	Comment	Response
Donald Fuller, MD (CyberKnife Coalition)	Methods	Lack of class one data still hamper optimal decision making surrounding the management of prostate cancer. Randomized controlled studies (RCT's) are the method of choice to determine the appropriate care for patients suffering from prostate cancer but executing RCT's presents many problems, including a significant incidence of patient refusal to be randomized to a specific treatment versus a WW or AS study arm. Single arm studies and observational approaches such as registries have been considered viable alternatives.	We appreciate your suggestions. However, in conformity with the Key Question, we systematically reviewed only comparative studies that directly compare AS (or other observational management strategies) to immediate treatment strategies. This review does not address evaluations of long-term clinical outcomes reported in noncomparative cohort studies, which are beyond the scope of this comparative Key Question.
Donald Fuller, MD (CyberKnife Coalition)	Methods	We are aware of two large prospective single arm clinical trials evaluating two different dose volume schedules for the newer method of radiation therapy, CyberKnife SBRT: CyberKnife Radiosurgery for Organ-Confined Prostate Cancer: Homogenous Dose Distribution: NCT00643994; and CyberKnife Radiosurgery For Low & Intermediate Risk Prostate Cancer: Emulating HDR Brachytherapy Dosimetry NCT00643617, each of which has now reached its accrual goal. There is also a national multi-institutional registry, named Multi-Institutional Registry for Prostate Cancer Radiosurgery (RPCR), NCT01226004, gathering important data needed to determine which patients benefit the most from SBRT therapy.	We did not look at intra-modality comparison in this report. Intra-radiation modality comparison was addressed in our previous report (https://www.cms.gov/coveragegeninfo/downloads/id69ta.pdf)

Commentator & Affiliation	Section	Comment	Response
Donald Fuller, MD (CyberKnife Coalition)	Methods	For the purposes of this submission, it is important to note that our clinical experience with CyberKnife SBRT reveals that this method has significantly lower acute toxicity than prostate resection or other invasive methods and appears comparable with IMRT, with a more efficient delivery schedule. As such, we feel that CyberKnife SBRT should be taken into consideration in this report, when evaluating the body of data available today, as AHRQ has done in its most work in this area1. Most importantly the 5 year disease-free survival outcomes appear comparable and the patients' benefit from a much more convenient and less stressful treatment regime with 4-5 CyberKnife SBRT treatment visits versus 40-45 treatments with typical IMRT or proton beam therapy regimens. In a recent patient survey2 it was found that 99% respondents said they would choose CyberKnife SBRT again.	We appreciate your suggestions. However, in conformity with the Key Question, we systematically reviewed only comparative studies that directly compare AS (or other observational management strategies) to immediate treatment strategies. This review does not address evaluations of long-term clinical outcomes reported in noncomparative cohort studies, which are beyond the scope of this comparative Key Question.
Donald Fuller, MD (CyberKnife Coalition)	Methods	The CKC membership is curious as to why some treatments being offered in clinical practice such as the CyberKnife SBRT were not mentioned in the discussion about radiation therapies. Looking forward, we hope that you will recognize hypofractionated CyberKnife SBRT as a viable treatment option for patients today.	The focus of this review is on observational management strategies. Different forms of prostatectomy (e.g., robotic), brachytherapy, or external radiation therapy were not analyzed unless when compared to observational management strategies.

Commentator & Affiliation	Section	Comment	Response
Donald Fuller, MD (CyberKnife Coalition)	Methods	Last but not least, we wanted to make you aware of a cost-effective analysis published in *Value in Health* (volume 14, #3, 2011) comparing CyberKnife to IMRT, Proton Therapy, and surgery. The study found CyberKnife SBRT to be less expensive than IMRT and Proton Therapy. CyberKnife provided the highest Quality of Life Years per dollar to patients, higher than surgery, IMRT or Proton Therapy. On a Medicare allowable cost comparison basis, a full course of CyberKnife SBRT costs considerably less per case to deliver than a full course of IMRT or proton beam radiotherapy and "CK was found to be cost-effective versus surgery."	Thank you for this suggestion. The study is presented in abstract form only; as such it does not meet our inclusion criteria. Even if it had been published, we would have included this study only if it reported information on a treatment arm managed with AS (or any other observational management strategy).
NCI	Methods	This graphic doesn't seem to tell the story. In fact there is considerable concern about how to sort into Active surveillance when the expectation is that cure is still possible if progression appears. On the other hand if there is no hope then surveillance occurs for chances to provide palliative care. The terminology I'm used to is watchful waiting with and without active surveillance. It would clarify the review's scope if the arm for watchful waiting without active surveillance were also be included in the graphic and then it was clarified where that literature was used or not.	This is just one of several definitions of AS vs. WW we have come across. The primary focus of the Key Questions was on AS. We have added WW to the analytic framework, without an attempt here to differentiate them.
Peer Reviewer 1	Results	Concordance and discordance in national guidelines. The current report does not review guidelines from national societies (NCCN, AUA, US Preventative Task Force, and others). It may be helpful to better understand how national guidelines advise on AS and WW and whether these guidelines are consistent with the evidence presented in the report.	A review of clinical practice guidelines would be outside the scope of this report.
Peer Reviewer 2	Results	please see general comments	No response necessary.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer 3	Results	Though the data in this section are necessarily voluminous considering the scope of the topic, the summaries of findings are clear and fairly easy to follow. For example, the reviewer particularly likes Figure 1.1, as the time interval for a study is clearly important, given the time trends in incidence, mortality, and treatment observed overall.	Thank you.
Peer Reviewer 3	Results	Page 22, para 3: In the discussion stage trends, it's important to remind readers that these datasets only provide stage at diagnosis. Men who initially present with localized disease are not re-reported if they later develop advanced stage disease. Therefore, given the lead time of PSA screening, the fall in advanced-stage disease is in part a reporting artifact, and overestimates the "true" fall in the incidence of advanced-stage disease.	We have emphasized that stratification was based on disease stage at diagnosis. We agree that a major determinant of changes in disease stage at diagnosis is wide application of PSA screening. However, we have not adopted the distinction between "artifactual" and "true" changes in incidence, suggested here. Because incidence is a descriptive measure of the observed population-level density (over person-time) of events, it can only be based on the observed distribution of events (stratified by disease stage). We agree that partitioning the observed changes in incidence into components attributable and non-attributable to PSA screening is of interest and can be accomplished using mathematical models of prostate cancer diagnosis (e.g. cohort simulation). However, such research was considered to be outside the scope of the present report.
Peer Reviewer 3	Results	Page 27, line 25:...symptoms or signs (such as a suspicious DRE) Page 27, para 3: Nothing else, particularly more recent than 1991, on the population-based incidence of biopsies over time? That's critical to interpreting incidence trends. The reviewer suspects PSA thresholds to recommend biopsies have been dropping over the "PSA era".	We have adopted the suggested wording. We did not review thresholds or protocols for performing biopsies. Our updated searches have identified one additional study reporting trends in the age-and race-adjusted rate of prostate biopsies using SEER-Medicare data. We have presented findings from this study in the Results section and the Executive Summary.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer 3	Results	Page 29, para 2: The reviewer would favor some early discussion of the possible explanations for the findings in the reviews. For example, decreasing mortality among diagnosed patients may represent the effects of lead- length-, and particularly overdiagnosis bias.	In the "Summary/Conclusions" section we have added a note indicating that there are multiple potential explanations for the observed population trends, including those related to mortality.
Peer Reviewer 3	Results	Page 31, Table 2.1: The reviewer would favor also providing a time frame for the AS cohorts, similar to Figure 1.1.	We have revised Table 2.1 to include enrolment start date.
Peer Reviewer 3	Results	Page 44, Figure 2.5. Same comment.	We have revised Table 2.5 to include enrolment start date. This table has been renumbered to Table 2.7.
Peer Reviewer 3	Results	Page 61. In the Section on physician factors, the reviewer would favor giving the year of key studies cited in the text, as AS has been a bit of a "moving target" that has seemed to gain some traction (at least in academic centers) over time. For example, what was the date of the study referred to in the 1st para on page 63?	The author and year of publication has been added to this section.
Peer Reviewer 3	Results	Page 73, para 2: At some point, the authors should reflect on the potential for residual confounding by comorbidity in observational studies using overall mortality as an outcome. For example, in reference #197, the decrease in overall mortality was largely attributable to a decrease in non-prostate cancer mortality, which is implausible for RP. Getting RP versus WW in the first place is probably a strong surrogate for comorbidity, better than some indices of comorbidity.	We have added a section about potential confounding in retrospective studies in the Discussion.
Peer Reviewer 3	Results	Page 73, para 3: Typo? Were urethral dilations really more common after WW?	We have corrected this sentence. Urethra dilation procedures were less common in the WW group.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer 3	Results	Page 75, para 2: Last sentence. The CE ratios were sensitive to the assumption that all men who "crossed over' top active treatment from AS would get relatively expensive IMRT.	We agree that the results of economic modeling may be affected by the underlying model assumptions. We have provided additional details regarding the modeling strategies used in the ICER economic evaluations in the "Costs" section of KQ4. Please note that discussion of cost-effectiveness ratios was out of the scope of the current report (we did not consider the cost-effectiveness results presented in the ICER reports).
Peer Reviewer 3	Results	Page 76, para 2: Excellent point, as above. Might comment on the apparent reduction in non-prostate cancer mortality associated with RP in some of these observational studies.	Thank you. We have added a comment discussing potential confounding in observational studies of treatment effectiveness. We have also included a primary research study evaluating non-prostate cancer mortality causes that has been cited to highlight this point.
Peer Reviewer 4	Results	well done and relevant in all respects very comprehensive	Thank you. No further response necessary.
Peer Reviewer 5	Results	p. 35. Line 50. Were the bibliographies of the systematic reviews performed by ICER regarding effectiveness reviewed? It is unclear what "not considered" means here.	We did not use the bibliographies of ICER reports. We have clarified that we did not use the efficacy or cost-effectiveness sections of the ICER reports.
Peer Reviewer 5	Results	p. 52 line 33/49. I would include the following study comparing outcomes for conservatively-managed men with clinically localized disease in SEER-Medicare in the pre- and post-PSA era: Lu-Yao GL, Albertsen PC, Moore DF, et al. Outcomes of localized prostate cancer following conservative management. Jama. Sep 16 2009;302(11):1202-1209.	This study did not meet eligibility. It was not a comparative study.
Peer Reviewer 5	Results	p. 57, line 32. I would include mention of the Will Rogers phenomenon here.	We have added a sentence discussing the impact of PSA screening and histopathologic grading changes on tumor grade at diagnosis and tumor-grade-adjusted mortality statistics.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer 5	Results	General comment: I feel this section would be more useful to a panel considering recommendations for AS if evidence comparing outcomes related to eligibility/monitoring schedules were included. While such data is sparse, examples of major issues with some data available include 1) whether to include patients with intermediate risk disease, for example (addressed specifically in Cooperberg MR, Cowan JE, Hilton JF, et al. Outcomes of active surveillance for men with intermediate risk prostate cancer. J Clin Oncol. Jan 10 2011;29(2):228-234.) and 2) the use of PSA kinetics and other factors as triggers for repeat biopsy, for example (addressed in Ross AE, Loeb S, Landis P, et al. Prostate-specific antigen kinetics during follow-up are an unreliable trigger for intervention in a prostate cancer surveillance program. J Clin Oncol. Jun 10 2010;28(17):2810-2816. 3) San Francisco IF, Werner L, Regan MM, Garnick MB, Bubley G, DeWolf WC. Risk stratification and validation of prostate specific antigen density as independent predictor of progression in men with low risk prostate cancer during active surveillance. The Journal of urology. Feb 2011;185(2):471-476. 4) Ng MK, Van As N, Thomas K, et al. Prostate-specific antigen (PSA) kinetics in untreated, localized prostate cancer: PSA velocity vs PSA doubling time. BJU international. Apr 2009;103(7):872-876.	Thank you for these suggestions. As noted elsewhere in this document, we only considered comparative studies of AS (or other observational management strategies) versus immediate active treatment for KQ4. Regarding the specific study suggestions: 1) The main analysis of this study is not a comparison of AS vs. active treatment (thus, this study is not eligible for KQ4). The subanalysis comparing patients who received surgery after having been assigned to AS versus patients who were assigned to surgery, is not within the scope of the report. 2) This study has been included in our review of definitions of AS strategies (however, it may overlap with other publications from the same research team). It is not a comparative study of AS versus active treatment; as such, it is not eligible for KQ4. 3) This study has been included in our review of definitions of AS strategies (however, it may overlap with other publications from the same research team). It is not a comparative study of AS versus active treatment; as such, it is not eligible for KQ4. 4) This study has already been considered for KQ2 (however it may overlap with other publications from the same research team). It is not a comparative study of AS versus active treatment; as such, it is not eligible for KQ4.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer 5	Results	p.58. While I appreciate the difficulty in distinguishing AS from WW, the introduction is wordy and confusion. I believe the difference can be summarized by intent: AS is curative in intent, and WW is palliative. AS is appropriate in men with disease believed to be indolent, men who may not require therapy. Because prediction tools are not perfect, 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, watchful waiting is generally considered, as it offers palliative therapy for symptomatic disease. Line 26. Studies of AS may be identified by their predefined protocol of surveillance, typically more intensive than that for WW, and in most cases the predefined definition of "progression".	This section has been rewritten adopting your suggestions. Thank you.
		p. 59. Table 2.1. The Canary PASS study merits inclusion, I believe. Newcomb LF, Brooks JD, Carroll PR, et al. Canary Prostate Active Surveillance Study: design of a multiinstitutional active surveillance cohort and biorepository. Urology. Feb 2010;75(2):407-413.	This study has been included in our review of definitions of AS strategies.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer 5	Results	p. 61 line 41 (and page 20). Imaging: Regarding TRUS-guided biopsy "to confirm the diagnosis of prostate cancer", it is important to distinguish between two types of biopsies performed: those performed to confirm the diagnosis/Gleason score prior to enrollment in an AS program, and those performed as part of the surveillance protocol once a patient has been followed for a period of time. Omitting "to confirm…" from this sentence would remove this ambiguity. It might be worth including a section describing biopsy requirements (path review at central location, repeat at surveying institution, etc) in more detail than on page 60 section iii.	The first mention of TRUS is in a paragraph clearly related to study eligibility, and thus cancer diagnosis. The second mention, under imaging, has been removed since the TRUS-guided biopsy is not an intervention used for "imaging". TRUS has also been removed from the table and text under imaging. We did not change the key questions; therefore we did not add a biopsy section.
Peer Reviewer 5	Results	p.61 section vii. Behavioral indicators. In contrast to the rest of this section, the data included here is descriptive of behavioral factors – why people chose AS – but does not relate to eligibility.	"No behavioral indicator was used explicitly as a criterion for AS program enrollment, beyond the requirement to participate in the study." We agree that the rest of the paragraph is unnecessary here and is repetitive with KQ3. It has been deleted.
Peer Reviewer 5	Results	p.91-2. Adherence section. I would recommend stating more explicitly which clinical features present at diagnosis vs. which features identified during follow up were predictive of seeking treatment, particularly with regard to PSA kinetics. In addition, it should be stated more explicitly whether PSA kinetics were a trigger in the studies in which it predicted treatment.	The multivariable analyses did not explicitly distinguish variables at diagnosis versus those at followup (although some of them were self-evident, like demographic data). PSA kinetics were assessed during followup; this has been clarified in the paragraph.
Peer Reviewer 5	Results	p. 92 line 3. Histopathology. Should clarify that this section describes tumor differentiation.	We have added text to clarify.
Peer Reviewer 5	Results	KQ4. General comments: 1. The majority of evidence regarding active surveillance comes from the small observational studies described in the section addressing KQ2, and it will be many years before RCT data comparing AS specifically to initial treatment is available. I feel it is important to include a	Studies were summarized first by comparisons and then by outcomes. We have added outcome subheadings under each comparison.

Commentator & Affiliation	Section	Comment	Response
		discussion of the disease-specific outcomes of these studies, acknowledging that it will give only a general sense of the safety of this approach, that follow up is relatively short for most, and that the key questions called for a review of comparative studies. The inclusion of watchful waiting (or indeterminate observation) in this discussion is most useful in that it can give a sense of the "worst case scenario" in active surveillance, as outcomes would be expected to be better on AS than WW/other observation. Therefore a discussion of these outcomes would be useful as well. 2. Quality of life is included as a subset of comparative studies discussion but not as an independent section. Given that a primary purpose of observation is to improve quality of life, I feel further description of the literature on this topic is warranted. Studies to consider for inclusion are: Burnet KL, Parker C, Dearnaley D, Brewin CR, Watson M. Does active surveillance for men with localized prostate cancer carry psychological morbidity? BJU international. Sep 2007;100(3):540-543. van den Bergh RC, Essink-Bot ML, Roobol MJ, et al. Anxiety and distress during active surveillance for early prostate cancer. Cancer. Sep 1 2009;115(17):3868-3878. Litwin MS, Lubeck DP, Spitalny GM, Henning JM, Carroll PR. Mental health in men treated for early stage prostate carcinoma: a posttreatment, longitudinal quality of life analysis from the Cancer of the Prostate Strategic Urologic	2. Thank you for these suggestions. We have revised the Results section to have a separate sub-section on quality of life. We have reviewed the suggested studies and made the following determinations: This was a cross-sectional study. As such, it was not considered eligible for KQ4 based on predefined criteria. We have included this study for KQ3 of the report. This study has already been included for KQ4 of the report. This study was not eligible for KQ2 because it did not provide required details regarding the definition of the observational management strategy employed in the study (WW). The

Commentator & Affiliation	Section	Comment	Response
		Research Endeavor. Cancer. Jul 1 2002;95(1):54-60. Arredondo SA, Downs TM, Lubeck DP, et al. Watchful waiting and health related quality of life for patients with localized prostate cancer: data from CaPSURE. The Journal of urology. May 2008;179(5 Suppl):S14-18. …among others	study also did not meet the eligibility criteria for KQ3 (no multivariable model) or KQ4 (single arm, non-comparative study). Many more studies similar to those suggested were identified by our searches. We based our decisions of inclusion or exclusion on predefined criteria.
Peer Reviewer 5	Results	p.99, lines 21 plus. I recommend dividing the discussion of outcomes into disease-specific findings, adverse effects of treatment, and quality of life measures.	Subheadings have been added.
Peer Reviewer 5	Results	p.100 line 26: PIVOT trial. I agree with not including the PIVOT trial in a table, but I think some discussion of preliminary results is important in a discussion of observation vs. RP.	We only included peer-reviewed results in this report. A link has been added for those readers interested in the preliminary information presented at the 2011 AUA meeting.
Peer Reviewer 5	Results	p.100 lines 29 plus. I would include discussion of SPCG-4 subgroup analysis results for men with low-risk disease, the cohort most likely to be followed on AS.	We have added this information in the report.
Peer Reviewer 6	Results	20:3. Where is Siegel / CA Cancer J Clin 2011?	Note that the Siegel et al publication had not been published when we conducted our search (it was however cited as background information). We have now updated our searches of the published literature. We have cited the study as informative on overall trends in the U.S. population. However, the study reports no analyses stratified by factors of interest to KQ1 (e.g., for some outcomes it does not report trends; when trends are reported they are not stratified by factors relevant to Key Question 1; or presented evidence comes from other primary sources, external to the specific paper).

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer 6	Results	23:4. I'm sure the authors know this, but the notion that Gleason score 5-7 corresponds to some uniform group of "moderately differentiated" tumors is outdates and clinically of very low utility. The drop in Gleason 2-4 tumors reflects a change in grading practices, not in biology. SEER (till recently) does not allow the more important distinction to bet determined between trends in GS 2-6 vs. 7 (and within 7, 3+4 vs 4+3).	We agree that the observed patterns are the result of changes in histopathological grading practices as well as the way SEER classifies and reports information. We have added text in the results section to indicate the latter point. A separate subsection of the Report deals with changes in histopathologic grading practice.
Peer Reviewer 6	Results	23:18. The actual decline in prostate cancer mortality rates should be noted here, not just that there was a decline (I believe it's roughly 40% since the early 1990s). This is an important point because while there is unquestionably a major problem with overdiagnosis and overtreatment of low-risk disease, the debates on screening and treatment often gloss this fact of a major improvement in mortality which are difficult to attribute to factors other than screening and/or improved treatments for higher-risk disease.	We have reported estimates of the change in population mortality rates.
Peer Reviewer 6	Results	26:22. This section on tumor characteristics is probably unnecessarily long—and at the same time incomplete in terms of the literature review. Stage / risk migration is pretty well documented and understood, and there's little controversy about it, likewise Gleason migration etc. The review here is not comprehensive—and shouldn't be. A few paragraphs reviewing key findings should be sufficient.	We believe the section length is appropriate. We did not determine length of write up based on level of controversy. We only attempted to examine factors relevant to the Key Questions, and we believe to have done so systematically.
Peer Reviewer 6	Results	26:47. "Tumor volume" is probably the wrong search term. Clinical stage is a (poor) proxy for tumor volume, as is the percent of positive biopsy cores or some other measure of extent to biopsy tissue involvement. I know at least one of the CaPSURE studies has reported on trends in % of cores involved (J Urol 2007), and I believe other investigators have as well.	Our original search should have captured all relevant studies, but since no studies were found, we performed an additional search with the term used in the key question. We did not include % of cores in tumor volume.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer 6	Results	27:43. I don't think this is really the only study to report trends in biopsy templates… 28:12. There are other important refs re: Gleason grade migration (e.g., Smith in Cancer 2002). Also, I think this is better thought of as a "Lake Wobegon" phenomenon (i.e., all the children are above-average"). Will Rogers is more an issue when it comes to questions like extent of lymphadenectomy and quality of imaging. Grade migration ('grade inflation') is more of a pure artifact of changes in definitions.	This is the only study we found that fulfills the eligibility criteria. Smith in Cancer 2002 is a single center study and thus ineligible. We believe we used the term Will Rogers phenomenon correctly. The reclassification did not put everyone above average but simply changed the scale.
Peer Reviewer 6	Results	28:27. May want to add here papers on regionalization of care (e.g, J Urol 2009 from NIS, among others) 28:20. I would be very careful in using CaPSURE to draw conclusion about time trends in settings of care. There are not enough sites and they are not randomly chosen. I'm not sure they reflect changes in academic vs. community care, for example. NCDB, on the other hand, could do this.	Thank you. Regionalization of care (PMID 17870128) with respect to inpatient care is outside the scope of this review. We agree and have provided a note within the paragraph that the changes in the distributions may be difficult to interpret.
Peer Reviewer 6	Results	29:20. I think the summary statement reinforces my feeling that this whole section (KQ1) doesn't really generate any new knowledge or understanding. It's good background for considering AS issues, but frankly not as systematic as some prior efforts—and really shouldn't have to re-invent that wheel.	Thank you for your comment. We have addressed the questions of interest to the committee.
Peer Reviewer 6	Results	30:21. Again, I would strongly disagree that AS and WW are/can be/should be used interchangeably. They may be comingled by necessity depending on the data sources examined, but WW is frankly a fading concept and AS is by far the more important from both research and public health standpoints. This important in terms of framing the research strategy for KQ2: the question is defining protocols / strategies for AS, period, not for WW or other constructions of "conservative management".	This section has been rewritten to address your concern.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer 6	Results	32:4. For some of the cohorts, it's important to differentiate between ideal inclusion criteria for the AS protocol and de facto characteristics of the patients enrolled. Both the Toronto and UCSF cohorts, for example, include and report on some men w/ GS 3+4, as per ref 106.	We have added a new Table 2.2. to provide these details more clearly.
Peer Reviewer 6	Results	33:3. PSA density (PSAD) is also included in some (e.g., Hopkins).	Yes, this information was included in the original table 2.2. We have added a new Table 2.2. to provide this information more clearly.
Peer Reviewer 6	Results	33:49. I'm not sure why behavior is even listed here. A cohort describing pts' choices etc. is not the same as restricting eligibility based on, for example, baseline levels of anxiety, which of course none of the cohorts currently does. This heading belongs under descriptive characteristics but not under eligibility criteria.	We were unable to change the key questions and have organized the report as per the structure of these questions.
Peer Reviewer 6	Results	34: Same comments apply to the behavioral column in Table 2.2: these are descriptive but were never intended by any of the investigators to be considered inclusion criteria per se.	See comment above
Peer Reviewer 6	Results	37: Similar comment again for table 2.3.	See comment above
Peer Reviewer 6	Results	41: I would argue that table 2.4 (monitoring schedule) should come before 2.3 (triggers for intervention). Also, p41 really should refer to table 2.3, not 2.4. This may be semantic, but the followup protocol per se just specifies frequency of biopsy, not of Gleason, number of cores, etc. (subheadings I, ii, and iii) Those are all part of doing a biopsy, and may define progression but are not separate components of the followup protocol.	We have rearranged the tables according to your suggestion.
Peer Reviewer 6	Results	42: Switching row headings from cohort names to "cohort 1" etc. makes for a more compact table but is a bit confusing, esp as they have not been numbered previously.	We have changed "cohort 1 , 2 , 3, etc." to the cohort names in the table.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer 6	Results	43:30. I'm not sure how much the bar charts add beyond the text, esp fig 2.3, unless the point is just to emphasize inconsistency/heterogeneity across the cohorts—but even still, if the values are all either 1 or 2, a bar chart doesn't the best choice of presentation format.	We have replaced the bar charts with tables.
Peer Reviewer 6	Results	44:4. Performing a repeat TRUS-guided biopsy does NOT indicate use of imaging to guide management. In most cohorts it's the biopsy results, not the TRUS, that are key (per criteria I, ii, and iii above). I believe UCSF is the only exception, in which TRUS findings per se (i.e., growth of a TRUS-visible lesion) are considered explicitly in the protocol.	See response to Peer Reviewer 5 Results Page 61 line 41.
Peer Reviewer 6	Results	44:12. Again, de facto treatment due to patient anxiety is a characteristic of all surveillance cohorts, but none of them explicitly encourage treatment based on anxiety alone. The U Conn cohort is not unique in this respect. (A CaPSURE paper actually looked at treatment driven by anxiety explicitly, see Latini et al in J Urol 2007).	We have clarified that no study used formal assessment of any behavioral indicators.
Peer Reviewer 6	Results	44:17. I'm not convinced re: the relevance of reviewing these studies, most of which are pre-PSA era and not reflective of contemporary practice. For example, the age criteria for SPCG-4 and PIVOT (45:15) had nothing to do with WW per se, but rather reflected the fact that patients were being randomized vs. active treatment. It makes little sense to argue that older men should be less eligible for WW.	We are not implying what the eligibility criteria should be for WW. As requested, we are simply reporting what the eligibility criteria were for cohorts or trials that used an observational strategy.
Peer Reviewer 6	Results	55:5. If you do want to include studies which mix AS and WW, you could also add CaPSURE here as some useful AS/WW studies have been conducted based on this cohort.	CaPSURE (and other database registries) did not meet eligibility criteria for KQ 2 because there were no predefined followup criteria.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer 6	Results	61:16. A much better reference looking at impact of both primary care doc and specialist visits on tx decisions is Jang et al. Arch Intern Med 2010.	This study did not meet our eligibility requirement of a multivariable analysis. It has been added as a footnote in the text.
Peer Reviewer 6	Results	62:48. Again, the Jang ref is useful re: rates of AS/WW after consult w/ just a urologist vs. a urologist + a radiation and/or medical oncologist.	This study did not meet our eligibility requirement of a multivariable analysis. It has been added as a footnote in the text.
Peer Reviewer 6	Results	63:47, Again, would also look at Latini J Urol 2007 which is a good CaPSURE report on factors leading to discontinuation of AS.	This study is already in the results section for this key question.
Peer Reviewer 6	Results	67:17. I'm not sure this is true: at least one recent study looked at the economic impact of various tx options for the providers, and found that AS actually winds up favorable in the long run given the fees for biopsies and eventual treatment in many cases.	This is an interesting question, but not relevant to the key questions.
Peer Reviewer 6	Results	67:22. I think of insurance type as a patient characteristic along w/ other SES variables, not as an incentive per se.	This was categorized by the committee that wrote the key question, and is unalterable.
Peer Reviewer 6	Results	67:51. It's not population-based, but one CaPSURE study (Cooperberg et al in JCO 2010) did report variation in WW/AS use across the CaPSURE sites. The Jang et al study mentioned above (Arch Intern Med) also considers SEER sites.	Cooperberg 2010 was included. The Jang study has been added as a footnote.
Peer Reviewer 6	Results	71:11. This is not true – some of the AS cohorts have reported at least short-term outcomes for men who were treated immediately vs. those who were treated after a period of surveillance (see the UCSF study in JCO 2011).	This study did not directly compare all eligible patients who received AS to all eligible patients who had immediate treatment. Only those who had AS and then surgery were included. Thus it did not meet eligibility criteria.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer 6	Results	71:40. PIVOT has not been published but it has been presented, and seems an obvious key study for inclusion here. I would expect Dr. Wilt would be willing to share at least the publicly presented data, esp as a PIVOT paper will presumably be forthcoming soon… 72:26. This is a clear example of the dangers of conflating "observational management" strategies. The studies reviewed used WW in the observation arms, and the outcomes clearly are not reflective of those expected in a contemporary AS practice. I think this distinction would be known to most readers, but should definitely be more explicit.	We are only including peer-reviewed published studies in this report. However, in the discussion, we now note that PIVOT is forthcoming
Peer Reviewer 6	Results	75:5. Ref 200 is out of date, and is superseded by Wilson et al Cancer 2007 76:3. Obviously head-to-head trials of surveillance vs. active treatments are lacking and are needed. But why review old trials comparing WW to prostatectomy but completely omit contemporary AS cohort studies? Several of the major AS institutions have reported outcomes out to 3-5 yrs median f/u, and while these are generally short-term and reflect only progression using definitions with limitations discussed earlier in the review, they amount to over 3000 patients, mostly with high-quality data collected prospectively. Of course these analysis are not without limitation, but I don't understand the rationale for omitting them entirely in a discussion of the clinical outcomes of contemporary surveillance.	Reference 200 (Penson) meets criteria, even if the study is old. But Wilson 2007 is not multivariable adjusted and therefore did not meet eligibility criteria. The AS cohort studies are not comparative and therefore also do not meet criteria.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer 6	Results	93:17. I know I'm repeating myself, but I want to stress once again in this paragraph the reviewer here is conflating two different problems. One is that prospectively-accrued contemporary AS cohorts use varying eligibility criteria, and two is that retrospective studies of older patients and those based on administrative data cannot distinguish WW from AS. That does not mean that the two cannot be distinguished in the literature. The goal going forward is to move toward consensus across prospective cohorts, and this issue is not impacted in any way be the second problem.	Thank you. This paragraph has been modified taking into account your concerns.
Peer Reviewer 6	Results	94:40. Again, don't forget complications of progressive cancer (or BPH) in evaluating the relative merits of AS and treatment.	"Complications from deferred treatment" is mentioned in the text.
Peer Reviewer 6	Results	95:6 I think the major cohort reports are already meeting most of these metrics, no?	As we discuss, many published cohorts do not clearly define AS (or WW) and are often vague about what the goal of the deferred active treatment is (curative or palliative). Several do not explicitly define progression or how protocols changed over time. Few reported why and how often patients and clinicians chose to not follow protocols.
Peer Reviewer 6	Results	95:15 These issues need to be considered in the context of overall adequacy of discussion with patients regarding management options—and documentation of those discussions. These are candidate quality of care metrics (see papers from Spencer / Litwin / Miller) and in general these discussions are not well documented.	Thank you. Your suggestions have been added to this section.
Peer Reviewer 6	Results	95:37 I might argue based on the findings of this review that database analyses in contexts like SEER are really not going to answer the key questions for AS, and that more attention (including more funding) should be directed to high-quality prospective clinical cohorts, including establishing/supporting cohorts outside of the traditional tertiary centers.	We agree and have amended the wording, in part to include prospective observational studies.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer 6	Results	96:27. It might be worth a few extra lines on the ProtecT study—especially considerations of how it succeeded accrual where US efforts have failed, also including what is expected to be learned and what its limitations are.	Done.
Peer Reviewer 6	Results	96:40. I believe a nod to the Canary / PASS investigators is warranted (see Newcomb et al Urology 2010); though the primary goal of this collaboration is biomarker research, the effort has succeeded in synchronizing definitions and protocols across 9 major centers.	This study is now described in Key Question 2.
Peer Reviewer 6	Results	96:57. It's also the case that novel biomarkers and/or imaging modalities are needed to better risk-stratify patients—to identify those who appear to be good AS candidates but may actually be at risk for progression, those who will likely do well on AS, and those who have such indolent disease characteristics that they would be good candidates for more of a WW approach (i.e., those with lesions, as mentioned above, which probably should not even be called cancer). A key question germane to ongoing biomarker studies is how the results of a new test will be used in practice. Say a new test increases accuracy of prediction from 80% to 90%--will that drive more clinicians and/or patients to accept and stay on AS?	See response to Peer Reviewer 1, Introduction, about bio- and imaging markers.
NCI	Results- Question 2	As noted in the exec sum another common terminology is "watchful waiting" for any approach that does not include active treatment. Then within WW there is active or passive surveillance. The choice depends upon the expectations for the natural history (potential for cure, no potential for cure). This whole sorting issue is key to the discussion and the incorporation of the decision process into care. Docs have to do the sorting and then decide which data applies to whom.	Thank you. We agree this is a key issue for the conference attendees to discuss.

Commentator & Affiliation	Section	Comment	Response
Michael Krachon	Results- Question 4	There are no suitable studies in the clinical literature that show active surveillance is an appropriate or preferable clinical option in comparison to brachytherapy or other immediate definitive treatments for localized prostate cancer. There are two high-quality evaluations described in the draft report demonstrating that immediate treatment with brachytherapy is less costly than active surveillance. **CAB Comments:** Providers and policy makers should promote brachytherapy (BT) for curative treatment of localized prostate cancer due to lower mortality rates, similar treatment complications and potentially decreased health care expenditures compared to observational management strategies. In addition to the literature referenced in this report, there are numerous peer-reviewed studies demonstrating outstanding patient outcomes and very low complication rates when brachytherapy is used for treatment of localized prostate cancer.	Thank you for these suggestions.
Michael Krachon	Results	With respect to cost effectiveness, we strongly support the inclusion within the draft report of information from the evaluations performed by the Institute for Clinical and Economic Reviews (ICER). The ICER evaluations demonstrate that the total cost of brachytherapy treatment for localized prostate cancer is less than the costs of active surveillance, intensity-modulated radiation therapy and proton beam radiation therapy. ICER is well-established, and the methodology used in these two evaluations is sound.	Thank you for the suggestion. We have provided additional details regarding the ICER methodology in the Results section of the report.
Michael Krachon	Results	We urge AHRQ to place greater emphasis on ICER's findings, which directly involve evaluations of the costs associated with active surveillance (the subject of the key question underlying this section of the	We believe that the Methods and Results of the ICER economic evaluations are covered adequately by the report. We have pointed out in the Results section that the ICER results are based on modeling and as such are sensitive to the model assumptions, parameterization and data sources.

Commentator & Affiliation	Section	Comment	Response
		report). In its current form, the draft report runs a significant risk of misleading readers with the lengthy discussion of the costs of other observational management strategies (such as watchful waiting) that are not referenced in the key question. In a number of instances throughout the draft report, the benefits of brachytherapy in treating localized prostate cancer are referenced, including the following: • AHRQ reported that studies generally reported that men treated with radiation therapy (includes brachytherapy) had lower all-cause mortality rates than men on watchful waiting (Stattin 2010, Wong 2006). • AHRQ reported that study of active treatments (radical prostatectomy, radiation therapy, brachytherapy considered together) resulted in lower all-cause and prostate cancer-specific mortality rates compared to watchful waiting (Wong 2006). • Mortality associated with brachytherapy is less than watchful waiting. o One retrospective cohort study found significantly better disease-specific survival in men treated with brachytherapy (Zhou 2009). • Complications associated with brachytherapy are no different than watchful waiting or no treatment. o One prospective cohort study reported no difference in sexual function between brachytherapy and no treatment (Choo 2010). o Three studies reported gastrointestinal or genitourinary toxicity outcomes and found no difference between brachytherapy and no treatment (Elliott 2007). o One report analyzed the incidence	The cited studies have already been included in the AHRQ reports we reviewed or the current report. Thus, no further report is necessary.

Commentator & Affiliation	Section	Comment	Response
		of treatment for urethral stricture captured in the CaPSURE registry and did not find a significant difference between patients on watchful waiting and patients treated with brachytherapy over a median follow-up of 2.7 years (Elliott 2007).	
NCI- Dan Ollendorf	Results	(1) The authors did not sufficiently emphasize the impact of age and/or life expectancy on outcomes in comparisons of observation strategies to immediate treatment. In particular, the one available RCT (the SPCG-4 study) of RP vs. WW shows reduced mortality and incidence of metastases in patients randomized to RP, but this effect was observed ONLY in patients <65 years of age. No significant differences in these outcomes were observed in older patients, even among those with low-risk disease. As this is a critical issue for determining the appropriate candidates for observation, and given that age subgroups were pre-specified in SPCG-4, these findings should be duly noted in the Results section. (2) It appears that the discussion of harms is limited to those reported for both observation and active comparator groups. This is incomplete, as many studies provide details only on procedure-related complications. Again using SPCG-4 as an example, patients considering observation vs. surgery will want to know that the cumulative incidence of impotence at 1 year post-RP was 58%, yet this was not mentioned in the report. While these data are not comparative in nature, they are important to describe, as the principal goal of observation strategies is to delay treatment and associated side effects.	We had provided the subgroup estimates of the hazard ratio as well as a detailed discussion of the interaction analyses performed. 2. We have provided additional information on harms from randomized trials. Unfortunately, observational studies rarely reported information on treatment related harms.

Commentator & Affiliation	Section	Comment	Response
AUA	Results- Question 2	In KQ1, page 22, under Tumor Characteristics, Stage, the authors attempt to make the point that most of the rise in prostate cancer incidence after the introduction of PSA-based screening is among low-stage tumors. This is of course correct. However, in making this point, they conflate localized and regional disease (which includes node-positive and extra-prostatic disease) into "early-stage" disease, which is very different from the focus population of the document ("Localized prostate cancer"). This is probably because the SEER registry inappropriately groups stage this way, but it still does not inform the present analysis. The "single study" mentioned in that paragraph (reference #42) is more relevant in showing the stage migration, showing a rise in localized cancers in lieu of locally advanced cancers.	The reviewer is correct; the aggregate mention to localized and regional disease reflects limitations of the underlying data. We have provided additional information on this point in the Results section.

Commentator & Affiliation	Section	Comment	Response
AUA	Results- Question 2 Both in Executive Summery and the actual results	In KQ2, page ES-10 to ES-11 and page 41, under Followup protocols, I would mention again that the tone here is critical of the finding of non-uniformity in followup protocols, whereas there ought not to be such an expectation. For example, the authors state, "All 15 cohorts included regular PSA testing in the followup protocol but there were no uniform followup frequencies." This is not informative with respect to the range of frequencies – it merely tells the reader that the authors are dissatisfied by the amount of variability. Why not simply state the follow up protocols (e.g., 3 studies recommended annual PSA, whereas 12 used more frequent testing) in a neutral fashion? There is no "right" number of PSAs per unit time, so why present this in a negative light? Similarly, the statement "The recommended treatments were also not standardized and were determined by the physicians in many of the cohorts" is filled with negative connotations about unnecessary variability and paternalistic medicine. The reality is that a number of treatment choices exist (i.e., there is no "standard treatment") and most often the patient and doctor come to a treatment decision that is consistent with the patient's wishes and the doctor's risk assessment. On the other hand, demonstrating the variability in protocols is important. For example, it is worth acknowledging that some doctors are ordering expensive imaging studies repeatedly on men with very low risk prostate cancer, whereas other doctors, presumably with similar outcomes, find that it is unnecessary. Again, I think the aim here should be to inform the reader of the range of followup protocols, their potential risks, benefits, and costs in a factual, rather than judgemental way.	We agree. We have rewritten the text to describe the most commonly used criteria and mentioned some exception. We have rephrased the sentences to remove any "judgemental" interpretations.

Commentator & Affiliation	Section	Comment	Response
AUA	Results	In KQ2, under Observational management strategies with palliative intent, the descriptions of Eligibility criteria and Followup protocols achieve the neutrality to which I allude above.	Thank you. No further response necessary.
AUA	Results- Question 3	On page ES-13 and page 64, under clinical factors, the results of reference #148 appear to be mis-stated. It states that "interruption of observational management strategies was predicted by… increased free-to-total PSA ratio…" In fact, risk of progression is inversely proportional to free-to-total PSA, as is likelihood of coming off surveillance: median % F/T PSA was 19% in those who did not get radical treatment and 10% in those who did get radical treatment.	Thank you. This has been corrected.
AUA	Results- Question 5	In KQ5, I agree with the call for retaining clinical staging in SEER and perhaps adding to the sites in order to capture more minority patients.	Our role is to describe how future studies could address the key and related questions. We do not think that this report is the place to offer suggestions for how SEER or other registries should be changed.
AUA	Results- Question 5	In KQ 5, I would reiterate that there is actually a surprising amount of agreement among clinicians and researchers with regard to eligibility, criteria for intervention and management once they have shown signs of progression, considering the fact that these protocols developed at different sites and at different times and by different clinicians/investigators. They are not easy to study in aggregate since they differ in small, but important ways. Clinicians and researchers in the field of AS should be lauded for their efforts to advance this field and enhance the safety of observation for low-risk patients, rather than being criticized for variability between individual studies.	We in no way meant to criticize clinicians and researchers in the field. By its nature, suggestions for future research must highlight the gaps and deficiencies in the current evidence (and thus how studies have been conducted). We think it is more important to emphasize the lack of consistency among clinicians and researchers in definitions of AS and WW, rather than encourage continued lack of clarity or definitions.
Charles M. Washington- The Proton Therapy Consortium	Results- Question 4	With respect to the Draft Report, Key Question 4 asks: "What are the comparative short- and long-term outcomes of active surveillance versus immediate treatment with curative intent for localized prostate cancer?" The Draft Report focuses on a number of outcomes, including costs. With respect to	It should be noted that we included the cost results, not the cost-effectiveness analysis results. We also noted that the data are based on models.

Commentator & Affiliation	Section	Comment	Response
		costs the Executive Summary of the Draft report provides that: “part of draft report ” In citing two economic evaluations prepared by the Institute for Clinical and Economic Research ("ICER”) of the Massachusetts General Hospital, the Draft Report notes that: “part of draft report” We are very concerned that the cost comparisons included in the Draft Report will mislead patients, physician, and other stakeholders. While primary studies comparing the cost of active surveillance with active treatment strategies do not exist, the cost comparisons and underlying ICER economic evaluations inadequately addressed the multiple facets involved in the "cost"' of this disease, including the relevant complications and side effects associated with each option. The complications and side effects include gastrointestinal and genitourinary issues, erectile dysfunction, secondary tumors, patient anxiety and other clinical and quality of life issues. These indirect costs to the patients, patient's family, payor, employer and others are all well beyond the direct treatment costs relied on in the Draft Report and ICER economic evaluations. In fact the ICER economic evaluation provided that their estimates of side effects and complications of treatment were based on *'low-quality evidence from individual case series… we are aware that the side-effect estimates produced by the model may appear to be lower than rates from patient reported instruments or even from clinical reports."*[9] In reality, any economic studies are premature and should be labeled as simple cost rather than cost effectiveness analyses, for their conclusions arc based on unwarranted efficacy assumptions. The conclusions of the Draft	

Commentator & Affiliation	Section	Comment	Response
		Report are highly likely to mislead patient, physicians and other stakeholders because it only includes a comparison of direct treatment costs, not the true cost. Moreover, emerging evidence suggests that proton therapy may be superior to other options in reducing the indirect cost associated with the treatment of prostate cancer.	
Charles M. Washington- The Proton Therapy Consortium	Results- Question 4	The direct treatment costs cited in the Draft Report and economic evaluations arc not accurate. Actual direct treatment costs vary significantly by treatment plan, level of service, site of service, geographic location and payor. For example, 2011 Medicare reimbursement for one treatment session of proton therapy varies from S536. 19 to $ 1,349.61, depending on the level and site of service and geographic locale.[1O] The Medicare Hospital Outpatient Prospective Payment System ("HOPPS") proton therapy rates referenced in the Draft Report, and by ICER, only apply to hospital outpatient departments. As such, neither the Draft Report nor ICER considered the rates for free-standing proton therapy centers, which are established by the Medicare Administrative Contractors and not through national policyrnaking rules. The Draft Report cited only one figure (S53,828) for a presumed proton therapy treatment plan, level of service, site of service and payor which is significantly over-simplified and may not reflect current practice patterns and actual direct treatment costs.	Please see our previous response.
Peer Reviewer 1	Discussion (or a different section?)	Several studies have been published in 2011 that would be included in this report had the review been updated through July, 2011. Some of the studies (for example, the 2011 update of the Swedish RCT) have been included but others have not. Can the review be updated to reflect the more current literature?	Update has been done.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer 2	Discussion/Conclusion	p. Regarding the statement "Because of the different usages 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 which patients had true AS or WW...": the authors delineated in the Methods section the criteria used to distinguish AS from WW. While we agree that there are no standard definitions to look to currently, and that AS and WW exist on a continuum, the presentation of this construct as an issue with current research can be made separately from the treatment of AS and WW as discrete approaches for the purposes of answering the Key Questions (as the authors did). (2) As in the Results section, more mention should be made of the differing results by age in the SPCG-4 comparison of RP and WW. Given these findings, is it not feasible for the authors to conclude that, despite the lack of RCT data specifically on AS, its performance is likely to be at least as good as WW in older patients?	1. This section has been rewritten. 2. We have provided additional details on the interaction analyses reported by the SPCG trial. We agree that in terms of efficacy WW can be viewed as a low bound of the expected efficacy of AS.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer 3	Discussion/Conclusion	The authors do a good job of summarizing the results of a sparse and confusing body of largely nonexperimental evidence. The reviewer believes readers of the executive summary should be cautioned again about the hazards of interpreting therapeutic effects from observational studies (particularly ES-15, paras 4-6). It is unfortunate about the timing of publication of the PIVOT trial, which provides level I evidence about RP vs WW among men with most PSA-detected cancers (though indeed more WW that AS) and bears most closely on Key Question 4. Hopefully, the PIVOT main paper will be available by the time of the planned meeting to supplement the information in this report. The reviewer finds the future research section clear and straightforward. On pages 96-97, a stronger statement on the need for an RCT of AS vs RP vs XRT needs to be considered, along with a description of how the ongoing ProtecT trial in the UK measures up to the recommendations. These trials take a long time (PIVOT has been ongoing for 15 years or so), and the sooner started, the better.	A cautionary statement has been added per your suggestion. The need for an RCT between AS and immediate treatments has also been added. Details of ProtecT on recruitment has also been provided.
Peer Reviewer 4	Discussion/Conclusion	well done and relevant in all respects very comprehensive	Thank you. No further response necessary.
Peer Reviewer 5	Discussion/Conclusion	The future research section is translatable into new research. In the discussion of KQ2, line 47, the discussion states that the question of the optimal tests for monitoring on AS was not raised by KQ2. I would argue that a discussion of the little that is known about this area does fall under this question and should be included. I agree prospective studies are needed.	We respectfully disagree. To assess the optimal test implies comparing outcomes when different tests are used, which is not asked by the key question.
Peer Reviewer 5	Discussion/Conclusion	p.101 discussion of observation vs. RP (and RT later). I would include discussion of results described for low risk patients when available as well as overall localized disease results.	We have provided information on low-risk subgroups when available.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer 5	Discussion/Conclusion	p.103. line 5. Two additional cost analyses to consider are: Corcoran AT, Peele PB, Benoit RM. Cost comparison between watchful waiting with active surveillance and active treatment of clinically localized prostate cancer. Urology. Sep 2010;76(3):703-707. Wilson LS, Tesoro R, Elkin EP, et al. Cumulative cost pattern comparison of prostate cancer treatments. Cancer. Feb 1 2007;109(3):518-527.	Thank you for these suggestions. These papers had been identified by our searches and we made the following determinations: This study reports cost estimates based on a simulation model. We excluded such studies (with the exception of the ICER report). Thank you. We have included this study.
Peer Reviewer 5	Discussion/Conclusion	p. 103 line 22. I do not feel that inclusion of the Swedish data is informative; it can be omitted in my opinion.	Although Swedish costs are not applicable to the US setting, the relative ranking of treatment costs could very well be similar across countries. Because our inclusion criteria did not limit the studies considered for KQ4 to those conducted in the US, we have not excluded this study. We have added a comment to indicate that Swedish costs are unlikely to be applicable to the US.
Peer Reviewer 5	Discussion/Conclusion	p.103 line 32 plus. The ICER reports used a lifetime horizon, not 15 year.	Thank you. We have clarified that the cost model had a lifetime horizon.
Peer Reviewer 5	Discussion/Conclusion	p. 124 KQ4. Mention of the recent closure of START due to poor accrual deserves mention in the first paragraph.	This has been added.
Peer Reviewer 6	Discussion/Conclusion	The discussion essentially provides a factual summary of the major findings and limitations. There are few 'action items', and the future research section is quite sparse. Some possible suggestions are noted above (e.g., re: biomarkers, better prospective cohorts, etc). I might have expected a section on policy implications / recommendations based on the current state of the admittedly limited evidence.	It is correct that the discussion sticks to a factual summary of the findings and limitations. The EPC is not in a position to make policy or clinical recommendations.

Commentator & Affiliation	Section	Comment	Response
Donald Fuller, MD (CyberKnife Coalition)	Discussion/ Conclusion	We ask you to consider including a discussion of hypofractionated CyberKnife SBRT treatment, including the latest cost-effectiveness research published in Value in Health, and adding disclaimers regarding the (in)conclusiveness of the findings, given the ongoing lack of comparable long-term data and the lack of definitions and consistent clinical application of AS and WW.	The focus of this review is on observational strategies. Different forms of prostatectomy (e.g., robotic), brachytherapy, or external radiation therapy were not analyzed. To report specifically on CyberKnife is outside the scope.
Donald Fuller, MD (CyberKnife Coalition)	Discussion/ Conclusion	In summary, taking on the position of trying to define the two approaches known as WW and AS begs the question: Can we or should we define treatment versus observational methodologies before we can accurately diagnose which patients truly present a dangerous prostate cancer?	This is a good question for the conference attendees to discuss.
NCI	Summery/ Conclusion- Question 2	"Thus, we compared the 15 unique cohorts reporting formal protocols to monitor triggers for curative treatment with the seven unique cohorts of other observational strategies with primarily palliative intent in the PSA screening era." This is important and didn't come out in Figure 1 very clearly.	The Analytic Framework was designed specifically to review the key questions. It was not meant to summarize all analyses in the report.
NCI	Summery/ Conclusion- Question 4	"Therefore, there is insufficient evidence for the comparative short- and long-term outcomes of AS versus immediate definitive treatment for localized prostate cancer. We identified an updated analysis from a multicenter RCT and 10 multicenter cohort" this is very important.	Thank you. No further response necessary.

Commentator & Affiliation	Section	Comment	Response
Michael Krachon	Discussion/ Conclusion	AHRQ states that they have insufficient evidence to evaluate comparative effectiveness of active surveillance management with curative intent versus immediate definitive treatment in men with localized prostate cancer. Due to the lack of studies comparing active surveillance to immediate treatment, AHRQ evaluated studies that compared observational management strategies (largely resembling watchful waiting) with immediate treatment. **CAB Comments:** We strongly agree with the draft report's conclusion that there are no suitable studies to compare active surveillance against brachytherapy and other immediate definitive treatment for localized prostate cancer. We urge the authors to place greater emphasis on this finding in the report since this finding is the answer to the key question that underlies this entire section of the report. The significant number of pages in the report devoted to the data on other observational management strategies (such as watchful waiting) is not directly responsive to the key question and is likely to be misleading to many readers	We believe we have conveyed our message clearly in the abstract, executive summary, results section, and discussion. There is little more that can be said about the lack of evidence regarding AS vs immediate treatment. We think the ancillary evidence about other observational strategies is appropriate.

Commentator & Affiliation	Section	Comment	Response
Michael Krachon	Discussion/ Conclusion	A significant number of active surveillance patients eventually either opt for curative treatment for localized prostate cancer or require definitive treatment due to disease progression. **CAB Comments:** Given that approximately one-third of active surveillance patients transition to curative treatment[1,2] within a few years of initial diagnosis, providers and policy makers should not support active surveillance as an appropriate option for men with localized prostate cancer due to increased overall expenditures to the health care system for both active surveillance and the cost of a curative therapy at a later date.	Thank you for your comment. This report does not make policy or clinical recommendations.
Michael Krachon	Discussion/ Conclusion	AHRQ states that there needs to be a study to evaluate whether men offered active surveillance will accept this strategy and adhere to it. If men feel a strong need to do something to definitively treat the cancer, and if active surveillance is rarely chosen or not adhered to, then the impact of offering this strategy will be small.	Our future research recommendation was based on the assumption that the key question is a question of interest that should be answered by future research.

[1] 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.* 2011 August 25 [Epub ahead of print].

[2] Klotz L. Active surveillance for prostate cancer: trials and tribulations. *World J Urol.* 2008; 26:437.

Commentator & Affiliation	Section	Comment	Response
NCI- Dan Ollendorf	Discussion/ Conclusion	(1) Regarding the statement "Because of the different usages 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 which patients had true AS or WW...": the authors delineated in the Methods section the criteria used to distinguish AS from WW. While we agree that there are no standard definitions to look to currently, and that AS and WW exist on a continuum, the presentation of this construct as an issue with current research can be made separately from the treatment of AS and WW as discrete approaches for the purposes of answering the Key Questions (as the authors did). (2) As in the Results section, more mention should be made of the differing results by age in the SPCG-4 comparison of RP and WW. Given these findings, is it not feasible for the authors to conclude that, despite the lack of RCT data specifically on AS, its performance is likely to be at least as good as WW in older patients?	(1) We have added explanation to explain how we operationalized the process of summarizing the various definitions (2) Discussion on differing results by age in the SPCG-4 comparison of RP and WW has been added.
AUA	Discussion/ Conclusion	Consider the following reference in discussion of the role of comorbidity in use of AS: Daskivich TJ, Chamie K, Kwan L, Labo J, Palvolgyi R, Dash A, Greenfield S, Litwin MS. Overtreatment of men with low-risk prostate cancer and significant comorbidity. Cancer. 2011 May 15;117(10):2058-66. doi: 10.1002/cncr.25751. Epub 2010 Nov 29.	Thank you for this suggestion. This manuscript considers patients receiving ADT together with those managed with WW. The results of the study do not differentiate between these two treatments in analyses of factors predicting treatment or in analyses of the effect of treatment on outcomes. As such the study is not eligible for any of the report's Key Questions.

Commentator & Affiliation	Section	Comment	Response
AUA	Discussion/ Conclusion	In discussing the cost studies, it is worth noting that, since the natural history for low-risk prostate cancer is so long, studies focusing on long-term costs (>5 years) are more relevant than those focusing on short-term costs (1-5 years). It also may be worth mentioning that the cost differences will most likely prove sensitive to the surveillance regimen on AS and proportion who move from AS to delayed treatment.	We have added additional information on primary cost studies and model-based cost estimates in the results section.
Peer Reviewer 1	General	This review of the state of the evidence for active surveillance is appropriately thorough and balanced. I am in agreement with questions posed, the search strategy outlined, the presentation of results (both in narrative and graphical form) and the discussion. I believe this review offers a thoughtful and even keeled appraisal of the evidence for active surveillance, and the evidence review panel should be congratulated on a substantial task that will set the foundation for discussion at the state-of-the-science meeting in December.	Thank you. No further response necessary.
Peer Reviewer 2	General	I believe this report is misleading and does not represent a suitable document to support the NIH consensus meeting. As written, the report quickly notes that there are no RCTs of active surveillance (AS) vs. immediate treatment, and then goes on to summarize earlier data on watchful waiting without any discussion of the data on patient outcomes from observational cohorts of AS. Since observational data were within the scope of this report, I'm not clear why the executive summary includes not one single sentence related to clinical outcomes of AS. This omission will leave all readers without knowledge of the data on AS that currently guides its use in practice. Moreover, by summarizing the WW literature with so little discussion of how it relates to the use of AS will be misleading. In particular, the distinction of evidence and applicability for	We understand your concern. However, in conformity with the Key Question, we systematically reviewed only comparative studies that directly compare AS (or other observational management strategies) to immediate treatment strategies. This review does not address evaluations of long-term clinical outcomes reported in noncomparative cohort studies, which are beyond the scope of this comparative Key Question.

Commentator & Affiliation	Section	Comment	Response
		patients at different ages/life expectancies is entirely missing from this report. The one available RCT (the SPCG-4 study) of RP vs. WW shows reduced mortality and incidence of metastases in patients randomized to RP, but this effect was observed ONLY in patients <65 years of age. No significant differences in these outcomes were observed in older patients, even among those with low-risk disease. As this is a critical issue for determining the appropriate candidates for observation, and given that age subgroups were pre-specified in SPCG-4, these findings should be duly noted in the Results section. Another glaring problem are the data presented to address Key Question 4. The clinical outcomes of AS that are important to consider include not only its impact on mortality and disease-specific mortality, but the rates of men who progress while on AS, the length of "treatment free" time, cumulative urinary symptoms, and quality of life. There are data on these outcomes in the observational literature and to ignore them here seems not in keeping with the stated scope of the review. It also appears that the discussion of harms is limited to those reported for both observation and active comparator groups. This is incomplete, as many studies provide details only on procedure-related complications. Again using SPCG-4 as an example, patients considering observation vs. surgery will want to know that the cumulative incidence of impotence at 1 year post-RP was 58%, yet this was not mentioned in the report. While these data are not comparative in nature, they are important to describe, as the principal goal of observation strategies is to delay treatment and associated side effects.	

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer 3	General	Overall, I find the report comprehensive and accurate. The key questions are clear and the analytic framework makes sense.	Thank you. No further response necessary.
Peer Reviewer 4	General	well done and relevant in all respects very comprehensive	Thank you. No further response necessary.
Peer Reviewer 5	General	The report addresses a very important clinical question with far-reaching implications both for patients and for the health care system. The target population is clearly stated. The key questions are appropriate, although the limitation of key question 4 to comparative studies severely inhibits the report's ability to describe the current state of knowledge about the clinical outcomes of AS.	Thank you. No further response necessary. We understand your concern. However, in conformity with the Key Question, we systematically reviewed only comparative studies that directly compare AS (or other observational management strategies) to immediate treatment strategies. This review does not address evaluations of long-term clinical outcomes reported in noncomparative cohort studies, which are beyond the scope of this comparative Key Question.
Peer Reviewer 5	General	The authors are to be commended for a thorough and well-written summary of much of the available evidence regarding active surveillance. I have interspersed general comments below, but my chief reservations have to do with what I feel are important omissions, though I recognize the scope of the report is limited by the key questions. In particular, the omission of data regarding comparisons between AS monitoring protocols and of outcomes of observational studies describing experience with active surveillance. That being said, the report does an excellent job of answering the questions asked given the dearth of data available.	We understand your concern. However, in conformity with the Key Question, we systematically reviewed only comparative studies that directly compare AS (or other observational management strategies) to immediate treatment strategies. This review does not address evaluations of long-term clinical outcomes reported in noncomparative cohort studies, which are beyond the scope of this comparative Key Question.
Peer Reviewer 5	General	There are numerous typographical and grammatical errors I have not corrected but should be addressed.	We have reviewed again and have corrected the typos we found.
Peer Reviewer 6	General	I think I cover these questions in the details below (all given as page:line number)	Thank you. No further response necessary.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer 8	General	This is a wonderful review of the entire subject of active surveillance (AS) for the initial approach to men with low risk prostate cancer (PC). The amount of work in researching the subject is impressive and hopefully this document will be available to health care workers as well as men with PC. As I read the document I do have a few salient comments.	Thank you. Please see above for responses to specific comments. The full text of the report will be freely available at http://www.effectivehealthcare.ahrq.gov/index.cfm
Peer Reviewer 8	General	WW does not necessarily indicate waiting until symptoms occur. For example a man with PC who does not initially have treatment may wait several years on AS and when he reaches 80 yrs old his PSA begins rising much faster or his prostate grows and is firmer. He is felt not to be a candidate for radiation or surgery and thus not a curative treatment but treatment by androgen deprivation might be used. Thus one often treats before any symptoms occur. The usual definition of WW as opposed to AS is that the patient will not undergo a curative approach but rather because of his age or co morbid problems a palliative treatment.	We agree that there are many variations of WW or AS. The purpose of this conference is to clarify those variations.
Donald Fuller, MD (CyberKnife Coalition)	General	There are several important points we hope that you consider as you move forward. The first is to clearly define your position regarding Watchful Waiting (WW) and Active Surveillance (AS.) There are different opinions about the definitions of WW and AS and no consensus exists at this time. Many organizations and groups use these terms loosely and interchange them frequently, which creates inconsistencies and confusion. The lack of definitions for AS and WW threaten to weaken the validity and acceptance of the findings and conclusions.	We do not have “a position” regarding the comparative effectiveness WW and AS, or about their definitions. We have aimed to summarize the evidence in a neutral manner. We hope the upcoming conference will start the process to better define these terms.

Commentator & Affiliation	Section	Comment	Response
AUA	General	In general, this document was very lengthy and repetitive; it was not an easy read. Overall though, we were very impressed with the reviewers' questions and their methodological approach for answering them. We believe they did find all of the important studies (to our knowledge) on this subject. We agree with the research questions which they suggest and with the basic concepts, like the need to standardize nomenclature and approach.	Thank you. No further response necessary.
AUA	General	One concern in this document is that we felt that when the authors described the changes in prostate cancer epidemiology over the past 30 years they never explicitly stated that the decrease in prostate cancer specific mortality in the nation could be attributable to the increasing use of aggressive, curative therapy for the disease. (It is not the only potential reason, but it is certainly one of them.) We believe that most urologists believe this to be the case, and, even though there are few specific well-designed trials addressing mortality improvements after therapy, this is the reason why active surveillance is not offered more.	In the main body of the Results section we have refrained from discussing potential explanations of the observed trends. We agree that increasing availability and use of an effective treatment is a tenable explanation for the decline in population mortality. In the "Summary/Conclusion" section of KQ1 we have commented that there are several potential explanations for the observed secular trends. We have used population mortality rates as an example of this complexity.
AUA	General	Another corollary to this might be a suggestion to explore the reasons why physicians recommend or do not recommend active surveillance (perhaps using qualitative methods) rather than only examine the physician-level correlates with this decision.	Thank you. This has been presented in the report.

Commentator & Affiliation	Section	Comment	Response
AUA	General	I completely agree that future research would be facilitated by more uniform definitions of AS vs. WW, eligibility criteria, followup protocol and triggers for stopping AS, and that clinical trials are an appropriate methodology. However, keep in mind that this is a variable disease that manifests in individuals with a variety of medical conditions, ages, values and beliefs, etc. As clinicians, we are trained to use judgment in areas of uncertainty. So we can expect to see quite a bit of variability outside of clinical trials. It is reasonable to compare protocols by starting with a reasonable option (by consensus) and comparing it to more intense and less intense options. It is unlikely that we will refine the protocols in perfect detail, for example, to distinguish between the comparative effectiveness and cost effectiveness of 2 vs. 3 PSA's per year. On the other hand, there are some clinical activities that seem to be unusually costly and probably do not add much (like annual imaging) which could be studied and hopefully eliminated from protocols.	Thank you. No response needed.
AUA	General	Some mention should be made of the challenges clinicians face in distinguishing inconsequential low-risk prostate cancer from low-risk prostate cancer that is capable of progressing to symptomatic or life-threatening disease. We have some tools for predicting this, but they are weak. Some future research may focus on improvement in predictive models of disease progression in low-risk prostate cancer and biomarkers to distinguish between indolent and non-indolent disease	We have added in a call for future research on bio- and other markers to better categorize disease.

Commentator & Affiliation	Section	Comment	Response
AUA	General	Overall, this is a very comprehensive and factually accurate review of the evidence on active surveillance and its many limitations. It is designed to provide guidance for an upcoming consensus conference on active surveillance and very appropriately avoids any temptation of being overly prescriptive or taking sides in any way. Both opponents and advocates will see their side represented. Its main value will lie in serving as a blue-print for future research in this arena. The central expertise for this condition squarely rests in the urological community, and the AUA appreciates the opportunity to provide feedback and would welcome the prospect of further involvement on this topic.	Thank you. No further response necessary.

Commentator & Affiliation	Section	Comment	Response
Charles M. Washington-The Proton Therapy Consortium	General	The Proton Therapy Consortium (the "Consortium") applauds the Agency for Healthcare Research and Quality's ("AHRQ's") efforts to improve the quality of health care in the United States. We welcome the opportunity to serve as a resource to AHRQ and the evidence-based practice center' (" EPC") on the Draft Evidence Report: An Evidence Review of Active Surveillance in Men with Localized Prostate Cancer (the "Draft Report"). We believe that several key considerations were not appropriately addressed in the Draft Report which, if not correctly addressed in the final report, will mislead patients, physicians and other stakeholders. The Consortium is, a nonprofit corporation, whose mission is to ensure that patients have availability and access to proton therapy. Our members are world-renowned cancer centers that provide life saving treatments to patient, including the Lorna Linda University Medical Center, the University of Texas M.D. Anderson Cancer Center, the Indiana University Health Proton Therapy Center, the ProCure Proton Therapy Centers and the University of Florida Proton Therapy Institute. Our centers believe in a collaborative approach of care, where physicians, nurses, radiation therapists, and other health professionals promote improving patient choice, facilitating the appropriate use of proton therapy, and encouraging cooperative research through collaboration with AHRQ and other government agencies, payors, purchasers and stakeholders.	Thank you for your suggestions. Please see below for a reply to the specific comment.

Commentator & Affiliation	Section	Comment	Response
Charles M. Washington-The Proton Therapy Consortium	General	Patient, and physician have a number of options when approaching prostate cancer and we believe that all options, including active treatment options, should be available through an informed decision making process. Proton therapy delivers a focused dose of radiation energy to the targeted area while surrounding normal tissue receives minimal radiation, releasing its highest percentage of energy at the end of its path (i.e., Bragg peak), depositing virtually all of the dosage at the targeted tissue. In contrast conventional external beam radiation therapy delivers radiation to all involved tissue, diseased and normal, with targeted tissue receiving *60-70%* of the intended dose. While radiation therapy has a number of benefits, the increased precision of proton therapy is used to reduce unwanted side effects by limiting the dose to normal tissue, resulting in fewer complication, and side effects.[1234] Studies have shown a reduction in long-term rectal and genito-urinary damage when treating with proton therapy compared to photon therapy, which provides considerable outcome benefits to patients[5] -Table are included after this section in the comments-	Thank you. Descriptions of the theoretical benefits and harms of the different active treatments for prostate cancer was not a goal for this report.

Commentator & Affiliation	Section	Comment	Response
Charles M. Washington-The Proton Therapy Consortium	General	There are several reasons why there have not been randomized controlled trials ("RCTs") comparing conventional radiation to proton therapy. Given the demonstrated facts that dose distributions of proton therapy are superior to x-rays (photons), that proton therapy delivers two to three times less energy to normal, healthy tissue outside the prostate, that tissue response per unit dose between protons and x-rays is virtually identical, and that radiation damages normal tissues and healthy organs, there are ethical questions about whether RCT comparing proton therapy to photon therapy in treatment of prostate cancer should be pursued.[11] Ethical concerns arise from the fact that the major clinical difference between photon irradiation (IMRT) and proton therapy lies in the volume of normal tissue exposed to radiation. The main point of a comparative trial would be to determine whether (if one assumes the same total dose delivered to the target volume) the difference in volume integral dose results in detectable clinical differences----presumably in side effects and second malignancies-over time. A RCT comparing photons to protons would require researchers to expose patients in the photon therapy group to normal tissue radiation. In view of the evidence that all radiation is harmful how could one ethically design a study wherein half of the participants would be receiving two to three times more radiation to normal tissue with no expected clinical benefit?[12]	Thank you for these comments. We did not evaluate (or search for) comparative studies of conventional radiation versus proton therapy. Such studies would be outside the scope of the current report (which required at least one of the comparators to be "observational management strategies").

Commentator & Affiliation	Section	Comment	Response
Charles M. Washington-The Proton Therapy Consortium	General	Based on the lack of consensus of what should be included in a cost analysis of prostate cancer treatments, and the misconceptions regarding the costs of proton therapy, we respectfully suggest that the final report accurately conclude that there are no primary studies comparing the cost of active surveillance, with active treatment strategies, including proton therapy. Any cost comparisons and economic evaluations should either (i) not be included in the Final Report, or (ii) be appropriately qualified to address the important considerations addressed above. Otherwise, we fear that patients, physicians and other stakeholders could inappropriately rely on the oversimplified cost comparisons which could adversely impact the quality of healthcare.	Our conclusion was that evidence is insufficient for the comparison of AS to active treatments for all outcomes considered. We believe this description captures the uncertainty alluded to by the reviewer. We cannot remove the discussion of costs, as the relevant studies fulfilled predefined inclusion criteria. We have highlighted the inherent limitations of model-based cost estimates.
Charles M. Washington-The Proton Therapy Consortium	General	Ultimately, all facets of a disease, direct and indirect, should be considered when patients and their physicians mutually decide upon an appropriate course of treatment when presented with a disease such as prostate cancer. The Consortium is dedicated to advancing the science and practice of proton therapy to enhance clinical outcomes and patient quality of life through an evidence-based approach. We believe that the AHRQ should embrace the same and not rely on misleading and inaccurate "costs" when the Final Report is issued. The Consortium appreciates the opportunity to serve as a resource to AHRQ and the EPCs on the Draft Report and we welcome the opportunity to continue to serve as a resource in the future. Should you have any questions, please contact Jason Caron at 202-861-4190 or jcaron@ebgtaw.com.	Thank you for this suggestion. Please see above for our replies to specific comments regarding costs. Thank you. No further response needed.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer 1	Clarity and Usability	The current report does not appear to clearly articulate the limitations of its review and the potential biases of the search strategies employed. Can the evidence review panel comment in the report on what the limitations are?	We have added a limitations section in our Discussion.
Peer Reviewer 2	Clarity and Usability	please see general comments	Please see above for responses to general comments.
Peer Reviewer 3	Clarity and Usability	Well, as the underlying evidence is unclear, the reviewer believes the authors have done as good a job as possible summarizing it! The bottom line is though AS sounds promising, we don't know how it compares to traditional attempted curative treatment strategies in terms of mortality and side effects. However, the same can also be said for the attempted curative treatments themselves for men with prostate cancer diagnosed in the PSA era. PIVOT should shed considerable light here, and these data can help in designing the next trial, which will need to be long, large (particularly for meaningful subgroup analyses), and expensive, the reviewer fears. If a trial to compare AS vs. RP for lower-risk men with PSA-detected cancers is too large to be practical, then perhaps we already know that both are reasonable management strategies.	Thank you. We believe that our grading of the evidence on the comparative effectiveness of AS versus active treatment as “Insufficient” captures this uncertainty. PIVOT uses a WW strategy in its observational arm. As such it may not be informative regarding the comparison of AS vs. active treatment. The results of ongoing trials comparing AS with active treatment are likely to provide valuable information on the relative benefits and harms of the two treatments. There remains equipoise. ProtecT is ongoing.
Peer Reviewer 4	Clarity and Usability	well done and relevant in all respects very comprehensive	Thank you. No further response necessary.
Peer Reviewer 5	Clarity and Usability	The report is well structured and generally well organized. Please see attached file for specifics.	Thank you. No further response necessary. Please see above for responses to specific comments.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer 6	Clarity and Usability	Overall this is a comprehensive review, clearly reflecting a great deal of work. At the end of the day, though, aside from a more thorough literature review, I'm not entirely clear what has been gained beyond existing published reviews on this topic, and I think more time on synthesis / conclusion / future directions / etc. would be well spent. Conversely, as I mentioned earlier, I think the whole section on epidemiology and trends could be substantially shortened. I've made most other specific comments earlier. The organization of results for KQ3 — separating offer/acceptance/adherence to AS for each factor seems forced and doesn't really reflect the literature well given how many holes there are. As I mentioned re: the discussion section, the conclusions are a bit sparse, and definitely could include more potential recommendations for policy/practice.	Thank you. No further response necessary.
Peer Reviewer 8	Clarity and Usability	The major part of the document is the background data which supports the conclusions. Although this part of the document will not be helpful for patients it is a tremendous resource for physicians active in this field. The reference list is clearly over the top and extremely comprehensive.	Thank you. No further response necessary.
Peer Reviewer 5	Tables	Tables 2.2-.4, 2.6-.8. The authors are to be commended for clear, user-friendly tables.	Thank you. No further response necessary.
AUA	Tables	There is a typo on the last line of table 2.2 on page 34. UCSF, US Gleason score criteria is "Gleason sum </=6 with no pattern 4 or 5" according to the publication (reference #101).	Corrected.