



Comparative Effectiveness Review Disposition of Comments Report

Research Review Title: *Management of Renal Masses and Localized Renal Cancer*

Draft review available for public comment from May 28, 2015, through June 25, 2015.

Research Review Citation: Pierorazio PM, Johnson MH, Patel HD, Sozio SM, Sharma R, Iyoha E, Bass EB, Allaf ME. Management of Renal Masses and Localized Renal Cancer. Comparative Effectiveness Review No. 167. (Prepared by the JHU Evidence-based Practice Center under Contract No. 290-2012-00007-I.) AHRQ Publication No. 16-EHC001-EF. Rockville, MD: Agency for Healthcare Research and Quality. February 2016. 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 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 E-mail. At the conclusion of the public comment period, authors use the commentators' submissions and comments to revise the draft 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.

Commentator & Affiliation	Section	Comment	Response
Public Reviewer #3 (Ithaar H. Derweesh, Moores UCSD Cancer Center)	Executive Summary	1. Executive Summary Introduction Section Page 15: (Background) a. the executive summary makes a statement about urothelial lesions; we need to state that we are focusing on cortical renal masses, and not urothelial lesions, whose overall prognosis and biology is significantly different.	All mentions of urothelial carcinoma in the ES have been removed.
		2. Executive Summary Introduction Section Page 15: (Diagnostic Evaluation) Would change from “sampling is offered” to “sampling may be offered” as this is not routinely carried out and omission of biopsy/FNA is not considered to be a breach of standard of care.	This has been changed. Thank you.
		3. Executive Summary Key Questions section (Page 16/Page 39): There are 5 questions addressed, one with one part and two questions with 2 parts. Questions 1 and 2 (2 parts) deal with efficacy and complications of renal mass biopsy, Question 3 is a comparative effectiveness question regarding different management modalities, whose second part deals with disparities. These are all indeed important questions, I do have one comment—given the emerging role of biopsy, we should give strong consideration for a comparative effectiveness analysis of biopsy vs. cross sectional imaging for diagnosis of kidney cancer—i.e., while it may seem intuitive that biopsy made add further information, how much are we gaining by proceeding down the biopsy route?	This is an excellent point. Renal mass biopsy studies do not provide granular imaging findings and are often a selected group. This would represent an additional key question that was beyond the scope of this report. A discussion of this limitation was added to the report (discussion, limitations of the evidence base, KQ2
		I would also suggest a focus on other controversial and important topics dealing concerning care of cT1/cT2 renal cancer patients which may impact clinical decision making—such as a) metabolic sequelae/cardiovascular sequelae of renal cancer management strategies—in my opinion we should consider this to be a separate set of key questions b) a more robust section on harm/complications (surgical and non-surgical) as part of the ‘key questions’, and c) A similar analysis on utility of serum/urine markers or different forms of imaging (though the second may overlap in terms of crossover with recently promulgated white paper by our radiological colleagues).	There is an emerging literature on this topic but unfortunately was beyond the scope of the key questions selected for this review. A discussion of this limitation was added to the report (discussion, limitations of the evidence base, KQ3a, Renal functional outcomes, health related quality of life, and perioperative outcomes and harms were reviewed as part of KQ3. Composite models to predict pathologic diagnosis were reviewed including imaging characteristics, demographic, clinical and other diagnostic tests (blood, urine, etc.) as part of KQ1. New emerging imaging techniques beyond standard (US, MRI, & CT) were not included. Assessment of these technologies not in widespread use was

Source: <https://www.effectivehealthcare.ahrq.gov/search-for-guides-reviews-and-reports/?pageaction=displayproduct&productid=2186>

Published Online: February 24, 2016

		<p>4. Executive Summary Methods Section, Page 17/39 It is evident by the reading the subsequent results and Discussions sections of the executive summary, that studies specifically looking at morphometric systems (RENAL score, PADUA score, C-index, etc.) which a number of studies have shown to correlate with risk of complications, surgeon choice, as well as of malignancy and prognosis were not included in the analysis. I see one such study referenced in Key Question 1 (Mullins et al, 2012)—there is a more robust literature using these systems? Why were other papers excluded?</p> <p>5. Executive Summary Key Question 2a, Page 19 I have a concern about the methodology and definitions for false negative/nondiagnostic biopsy. I disagree with the statement “However, benign or non-diagnostic biopsies do not necessarily proceed to surgical extirpation, limiting the analysis and making the exact false negative rate difficult to ascertain.” A non-diagnostic rate should thus not be included in the false negative calculation. This is an important point for the panel to clarify for the calculation as well as from a clinical principle/management standpoint—that is, in patients with nondiagnostic biopsy, strong consideration should be given to rebiopsy or definitive treatment</p> <p>6. Executive Summary Key Question 3a/3b, Page 20/21</p> <p>a. Oncological Outcomes: The writing here is a bit too generalistic. While the gestalt is captured, it is most accurate for cT1a tumors. Given the very different prognosis of cT2 lesions as well as larger cT1b lesions, this section should be broken down by clinical T/AJCC stage, in my opinion.</p> <p>b. Renal Functional Outcomes: Should also be broadened to include cardiovascular and metabolic endpoints. Can be part of a separate key question or just part of this analysis.</p>	<p>beyond the scope of this review. This is discussed in Research Gaps, KQ1.</p> <p>All studies meeting appropriate criteria were included. For KQ1, studies needed to include predictors of malignancy from a cohort of patients with both benign and malignant disease (many studies included only tumors confirmed to be cancer). Tumors must also have been clinically localized. For KQ3 evaluating complications, only comparative studies were included where treatment arms could be compared (i.e. not a combined cohort).</p> <p>It is a fact that the biopsy literature suffers from a verification bias. For accuracy measures, only cases with surgical pathology were used. Non-diagnostic biopsies were NOT included as “negative” in these calculations. We analyzed and report on what is known regarding re-biopsy for a non-diagnostic biopsy. False negative rate in the report is calculated using standard definitions (defined in Methods under Data Synthesis).</p> <p>a. Most studies including larger tumors are often heterogeneous and do not allow separation from cT1a tumors in the cohort. When possible, oncologic and other outcomes are separated by T stage (e.g. Table 23 for cancer-specific survival). Otherwise, individual studies dealing with T1b and T2 tumors specifically are commented on for radical and partial nephrectomy – where the data is strongest. The ES now reports stage-specific outcomes.</p> <p>b. Cardiovascular and metabolic endpoints specifically were not part of the KQ in this report. There is an emerging literature regarding the cardiovascular sequelae of renal interventions,</p>
--	--	--	---

Source: <https://www.effectivehealthcare.ahrq.gov/search-for-guides-reviews-and-reports/?pageaction=displayproduct&productid=2186>

Published Online: February 24, 2016

		<p>7. Executive Summary Key Question 3, Page 22 Health Related Quality of life, I can understand including depression and anxiety in this domain, but how valid is radical nephrectomy providing greater quality of life for cancer control? This would seem to contradict the findings of the section on oncological outcomes.</p> <p>8. Executive Summary Key Question 3b, Page 23/Perioperative Outcomes Page 140 I have a significant issue, from an ethical and medico-legal standpoint, from classifying 'conversion to open surgery' as a harm. If the emphasis is on preventing harm, then conversion to open surgery for the sake of safety from an ethical standpoint should not be considered a harm. Furthermore, this would expose urologists, financially and legally, and paradoxically decrease adoption of minimally invasive surgical technique.</p> <p>9. Executive Summary, Discussion Section, Page 28 (Discussion regarding role of biopsy) The sentence "For example, renal mass biopsy could be performed prior to thermal ablation when its results could help determine appropriate follow-up and treatment efficacy," is not in keeping with the initial clinical T1 guidelines (Campbell et al., J Urol, 2009) or the follow up guidelines (Donat et al., J Urol, 2009) and should be strengthened to reflect the recommendation for biopsy, which should read as follows: "For example, renal mass biopsy should be performed prior to thermal ablation when its results could help determine appropriate follow-up and treatment efficacy."</p> <p>10. Executive Summary, Discussion Section, Page 29 (Discussion regarding role of ablation) I take issue with the statement "The higher local recurrence rate does not necessarily indicate a biological failure of treatment, but may be indicative of a higher rate of multiple treatments being required to achieve a cancer-free state and uncertain</p>	<p>however the data is inconsistent and beyond the scope of this review.</p> <p>The strength of evidence for the four HRQOL studies included in this analysis was insufficient/low. The review simply reports the data in these studies. Perception of patients may dictate HRQL measures and increased awareness of the oncological outcomes may influence patient perception of the treatment they received. Language has been added to the limitations section that may assist in addressing the contradiction perceived by the reviewer. (Discussion, limitations of the evidence base, KQ3a,</p> <p>We have made it explicit in the revised report that this is not a harm but rather a metric that is important for patients and surgeons to understand when considering either minimally-invasive and/or nephron-sparing approaches. Patients opting for a certain procedure want to know the frequency of a conversion. We have added language to the appropriate section to clarify this point.</p> <p>The discussion section is discussing hypothetical settings where renal mass biopsy may be helpful. The AUA and EAU guidelines regarding biopsy are justified by the data in this systematic review. We have adjusted the language to address this issue. This section of the review is not meant to endorse a guideline but rather discuss the data surrounding renal mass biopsy in different settings.</p> <p>We removed this sentence. We clearly state the worse local recurrence free rate associated with thermal ablation.</p>
--	--	--	---

Source: <https://www.effectivehealthcare.ahrq.gov/search-for-guides-reviews-and-reports/?pageaction=displayproduct&productid=2186>

Published Online: February 24, 2016

		<p>outcomes related to the radiographic appearance of ablated tissue. Regardless, patients should be counseled that an equivalent local control rate with thermal ablation may require more than one treatment.”</p> <p>By definition of more than one treatment is required, IT DOES INDICATE BIOLOGICAL FAILURE of the treatment. Furthermore, this contradicts the prior endorsement of the Definition of primary treatment failure (promulgated by the working group on image guided therapy and published in Radiology in 2005) which was endorsed by the initial clinical T1 guidelines (Campbell et al., J Urol, 2009) or the follow up guidelines (Donat et al., J Urol, 2009)</p> <p>11. Executive Summary, Discussion Section, Page 29-30 (Executive Summary Discussion regarding renal functional outcomes)/Page 147 of the draft/Page 190, 194 of the Discussion</p> <p>I disagree with the conclusion expressed in on Executive Summary Page 29/30 (and page 147 of the draft), the following statement is problematic: “While a greater decrease in serum creatinine may be seen with radical nephrectomy, the incidence of progressive chronic kidney disease and end-stage renal disease was low across all management strategies. Additionally, the data is limited by the wide range of reported outcomes and the time interval at which they are reported. Therefore, the clinically significant implications of nephron-sparing management strategies are not well elucidated by the current data. Future studies of renal functional outcomes should report baseline renal function (estimated glomerular filtration rate), short-term (1-6 month) and long-term (1 year and longer) outcomes in an attempt to better compare management strategies. Based largely on results from partial nephrectomy series, this review suggests that patients with excellent renal function and those with the poorest renal function (chronic kidney disease stage 3 or greater) may not benefit from nephron-sparing surgery – while those at intermediate risk (chronic kidney disease stage 2) have the most to gain from nephron-sparing approaches. This may be due to the significant renal reserve at normal levels of renal function and low renal reserve at very low levels of renal function.”</p> <p>This statement is problematic and incorrect for two main reasons—a) an incorrect assumption about the nature of renal function and potential change over time, b) a problematic scope of the analysis, which limits its applicability to most patients with cortical renal tumors.</p> <p>a. The statement makes recommendations based on the incorrect assumption that renal function is static, and that renal functional</p>	<p>a. We agree with the reviewer's comment, in that this subgroup analysis is limited by power and by</p>
--	--	---	---

Source: <https://www.effectivehealthcare.ahrq.gov/search-for-guides-reviews-and-reports/?pageaction=displayproduct&productid=2186>

Published Online: February 24, 2016

	<p>outcomes are stable and immutable.</p> <p>i. Even a patient without CKD-Stage I or II, if young enough will experience renal functional decline over time. The fact that surgical reduction of nephrons then sets a new baseline GFR, and where that GFR is, in relation to change, and what we know about the rate of change in GFR over time is still important. For example a new baseline GFR of 50 for a 75 year old male may not be an issue, 57 year old male (median age of diagnosis for RCC) who has a 20 year life expectancy otherwise may certainly be.</p> <p>ii. Even in the setting of surgically induced nephron loss, a proportion of patients may nonetheless experience inexorable decline of their renal function. A subset of these patients may indeed have normal GFR and may have medical drivers towards CKD, but such a population nonetheless exists.</p> <p>b. The scope of the analysis is problematic, and has limited applicability to clinical decision making and the vast majority of patients presenting with renal cortical neoplasms</p> <p>i. There are unreferenced studies with far larger N values and rates of CKD that were not referenced and used in the meta analysis. Why?</p> <p>ii. The methodology should be broadened to include rates of CV/metabolic endpoints. As it stands the question being asked frankly has only limited clinical utility. While Progression to ESRD may seem to be a solid and valid endpoint, however, renal insufficiency is associated with a variety of sequelae that can impact life expectancy and quality of life. While the strength of the conclusions to be derived from retrospective analyses (large single/multicenter and national database studies such as SEER-Medicare) will be criticized; data nonetheless</p>	<p>limited data on long-term renal functional outcomes (beyond 1 year). We have changed the text to the following statement, "Our synthesis of studies suggests that patients at the lowest (preoperative eGFR <45 ml/min/1.73 m²) and highest levels (preoperative eGFR >90 ml/min/1.73 m²) of kidney function may not experience renal functional benefits from nephron sparing procedures compared with radical nephrectomy. However, this is likely due to decreased numbers of studies reporting these subgroups and outcomes, and the few studies reporting follow-up beyond one year. Further research should strive to identify the patients most likely to benefit from nephron-sparing approaches from a renal functional standpoint, and in particular long-term development of chronic kidney disease (CKD) and/or end stage renal disease (ESRD)."</p> <p>b. We thank the reviewer for the comment. While a synthesis of heterogeneous data is problematic, we do feel our synthesis of these comparative studies provides the best evidence in reporting kidney function outcomes to providers and patients. In addition, part of the process is to identify research gaps to guide future research and improve patient care.</p> <p>i. We have included comparative studies of localized renal masses, and thus excluded some of the large uncontrolled single-arm studies or studies without adequate descriptions of tumor localization. While these studies do give valuable information on trajectories of kidney function, they do not allow for adequate comparisons.</p> <p>ii. The impact of therapy on CV/metabolic endpoints is beyond the scope of the KQs outlined for this review. Please see similar comment above for a more complete response. In our discussion (research gaps), we added information regarding these potential sequelae as important points to consider when evaluating these</p>
--	---	--

Source: <https://www.effectivehealthcare.ahrq.gov/search-for-guides-reviews-and-reports/?pageaction=displayproduct&productid=2186>

Published Online: February 24, 2016

		<p>suggest an impact long-term from renal surgery, on cardiovascular and metabolic sequelae, with a potential increased risk from radical nephrectomy. Even if these studies are to be discounted as resulting from selection bias, they should nonetheless be weighted in any conclusion regarding recommendation of renal surgery, and in the least, these data suggest that surgical nephron loss, by whatever ultimate procedure, may indeed lead to further 'downstream' sequelae. Using only ESRD as the ultimate endpoint is very problematic and primitive, and lays into question the validity (as in the actual clinical utility and applicability) of the conclusion. Even progression on various stages of CKD is, in and of itself, problematic without context. For this reason, CV and metabolic endpoints should be considered in the analysis. Indeed, while the conclusion which reached may reflect the poor quality of the data (all except one RCT for example), it would nonetheless shed light on the current state of knowledge in this field and may help shape practice based recommendations and future investigational questions.</p> <p>12. Executive Summary, Page 31: The following statement is incorrect and should be removed “As nephron-sparing approaches are only indicated for clinically localized tumors, these studies were included regardless of the reporting of clinical stage.”</p> <p>Research Gaps, Page 32, Key Question 1: The report makes mention of emerging biomarkers and states that they should be incorporated into composite models and validated prospectively. While that may be correct, the analysis could also touch upon markers such as CRP, platelet count, CAIX and other putative markers. As mentioned above, consideration should be made to analyze these.</p>	<p>treatment options.</p> <p>We changed “only” to “mostly” in this sentence. We found multiple studies comparing partial nephrectomy and radical nephrectomy. The partial nephrectomy cases can generally be assumed to involve clinically localized disease at presentation, as this is the core indication for this approach. For radical nephrectomy however we found that non-localized tumors are often included. Methodologically we required that radical nephrectomy studies explicitly state or show that the cases included were clinically localized and within the scope of this review. We acknowledge an emerging literature on partial nephrectomy for renal vein involvement or for metastatic disease but it remains that the overwhelming majority of studies using partial nephrectomy patients have patients with clinically localized disease. This is not the case for the radical nephrectomy literature. This is clarified in Methods, Data Synthesis section</p> <p>These were added to the ES research gaps.</p>
--	--	---	--

Source: <https://www.effectivehealthcare.ahrq.gov/search-for-guides-reviews-and-reports/?pageaction=displayproduct&productid=2186>

Published Online: February 24, 2016

<p>Public reviewer # 4 (Jonathan Himmelfarb, American Society of Nephrology)</p>		<p>ASN agrees with the findings and statement that overall survival and renal outcomes are favorable for nephron sparing surgery as addressed in Question 3a. However, the Executive Summary and Key Question 3b (pages 146-147) discourages nephron sparing surgery for those with “excellent renal function and those with the poorest renal function (Stage 3 or greater).” The society disagrees with this statement since the recommendations are largely based on a few papers and without substantial evidence to support the assertions.</p> <p>The largest study of the four citations was the retrospective evaluation of the EORTC study. The reviewers interpreted that those with “lower” glomerular filtration rates (GFRs)—Stage 3 and above—did not benefit from nephron sparing surgery (NSS) as this group did not have significantly better survival in comparison to the radical nephrectomy (RN) group. However, the majority of the study population (497 patients) had baseline creatinine <1.25 x upper limit of normal, and only 34 total patients had baseline creatinine >1.25 x upper limit of normal. Thus, this study was underpowered to allow a true assessment of treatment effect.</p> <p>The Woldu paper also had too few patients (262 of 1306) in the Stage 3 group, with only 74 Stage 3B patients (GFR 30-44). Takagi showed benefit of NSS for those with baseline GFR 45-59 (Stage 3A) but not Stage 3B. This was also a small retrospective study with a total of 118 patients, and only 27 patients in the Stage 3B group. Clearly, the numbers are very small in the advanced CKD categories. In fact, there is a trend for better kidney function in the eGFR <30 group treated with NSS in the EORTC study and the freedom from new onset lower GFR in the eGFR 30-44 group in the Takagi paper.</p> <p>In summary, the Executive Summary conclusions (page ES-15) seem to be overstated given the limited data. Until further studies are performed among those with more advanced CKD (Stage 3B and above), it is premature to have strong conclusions about groups who should have nephron sparing surgery or radical nephrectomy.</p>	<p>We agree with the reviewer’s comment, in that this subgroup analysis is limited by power and by limited data on long-term renal functional outcomes (beyond 1 year). We have changed to the following statement, “Our synthesis of studies suggests that patients at the lowest (preoperative eGFR <45 ml/min/1.73 m²) and highest (preoperative eGFR >90 ml/min/1.73 m²) levels of kidney function may not experience renal functional benefits from nephron sparing procedures compared with radical nephrectomy. However, this is likely due to decreased numbers of studies reporting these subgroups and outcomes, and the few studies reporting follow-up beyond one year. Further research should strive to identify the patients most likely to benefit from nephron-sparing approaches from a renal functional standpoint, and in particular long-term development of chronic kidney disease (CKD) and/or end stage renal disease (ESRD).”</p> <p>We have also updated the executive summary to reflect more of this uncertainty “The incidence of end stage renal disease was low in all interventions; however, most studies have limitations with few patients and events; and short-term follow-up”</p>
---	--	--	---

Peer reviewer #1	Executive Summary	<p>Executive summary, background: The incidence has increased "dramatically"; I'd prefer to see numbers about relative and absolute increase. Would also reconsider choice of wording.</p> <p>> 2. I would consider making a distinction between active surveillance with curative intent versus watchful waiting with no such intent. In the case of the latter, such as a patient with limited life expectancy due to coexisting medical comorbidities no treatment (or follow-up for that matter) may be indicated.</p> <p>> 3. Therapeutic interventions and outcomes: Please be precise when you use the terms "standard", "recommendation" and "option" and are referring to the strength of recommendations used by the AUA for the relevant guideline you are citing. A "standard" recommendation should not be equated with the standard of care. It may be helpful to put these terms into quotation marks. I would also consider explaining what they meant for this guideline, especially since a great guideline methodology has been evolving over the years.</p>	<p>The increase in RCC incidence is a well-documented phenomenon, demonstrating an increase in incidence of 2-3% per year since the 1970's. One of the seminal references documenting this observation has been added.</p> <p>The difference between active surveillance (AS) and watchful waiting (WW) has been clarified by the following language: It is important to note a difference between active surveillance with curative intent versus watchful waiting. The latter constitutes a strategy where treatment is never entertained and surveillance imaging is infrequent or does not occur at all. Studies of watchful waiting are not examined in this report. (Introduction, Therapeutic Interventions and Outcomes, paragraph 1) (Methods, Therapeutic Interventions and Outcomes). This report does not include WW without therapeutic intent.</p> <p>More precise wording was used and the terms were put in quotations so as not to equate them with "standard of care."</p>
TEP #2	Introduction	well laid out problem	Thank you.
TEP #3	Introduction	This section is well written. The definitions of fine needle aspiration and core biopsy should be stated in this section.	In the introduction section we state "Percutaneous renal mass sampling can be done by fine needle aspiration with a reading of the sample by a cytopathologist or via core biopsy with a reading by a surgical pathologist." This is to define these terms early.
TEP #4	Introduction	Clear and concise; no changes	Thank you

TEP #5	Introduction	Appropriate. It explains the need of an evidence review to be performed now.	Thank you
Public Reviewer #2 (Nenellia Bronson, American Urological Association Education and Research Inc)	Introduction	<p>1. The abstract does not adequately reflect the executive summary and/or overstates the findings. The abstract never defines localized RCC target population. Is not defined until ES17 making interpretation of abstract unnecessarily obscure.</p> <p>Abstract and ES 8 I believe the wording thermal ablation offered the most favorable perioperative outcomes is misleading and too strong. Thermal ablations are not operations they are percutaneous procedures so comparing a procedure to an operation is apples and oranges</p> <p>Ablation does not consistently demonstrate the most favorable outcomes, which implies multiple metrics need to specify that ablation had the shortest LOS lowest EBL and fewest conversions which is a metric that can't be measured against open cases.</p> <p>Abstract you need to be very careful when comparing AKI rates among modalities. It is expected that a more precipitous drop in eGFR and bump in Scr will occur after RNx. I did not find a definition in this document of AKI and I am afraid they are using a change in GFR as AKI which then disproportionately affects RNx data interpretation.</p> <p>d. Abstract - I disagree with the conclusion percutaneous renal mass sampling with core biopsy is a safe and sensitive procedure but has a high nondiagnostic rate. A 16 nondiagnostic rate is not high and moreover reflects the results of a single attempted core bx. the rate is far lower on re-biopsy and with improved hospital processes at most major centers i.e. cytopathologist or tech in radiology at time of bx. nonetheless a 16 nondiagnostic rate is rather low.</p>	<p>The first sentence of the Structured Abstract clearly defines the study population as patients with a renal mass suspicious for RCC.</p> <p>Perioperative outcomes are the consistent terminology used throughout the report. The ablation literature includes both percutaneous and laparoscopic approaches. The argument that ablation constitutes a "procedure" and not an "operation" is largely semantic and, for consistency, we maintain the title "perioperative outcomes" to define this measure for any intervention throughout the report, including percutaneous ablation.</p> <p>We respectfully disagree given that ablation can be performed laparoscopically and percutaneously, but also because percutaneous procedures could have a rate of conversion and blood loss. We acknowledge the limitations in these studies and data, however, we included existing comparative studies accepted by peer-reviewed journals that reported on these metrics.</p> <p>AKI is evaluated in both the renal functional outcomes section and the harms section of KQ3a. In the harms section, AKI is tabulated as reported by study authors. In the renal functional outcomes section, AKI is defined by strict GFR definitions. We have clarified this point.</p> <p>We changed the language to state: "but is associated with a significant non-diagnostic rate." While re-biopsy increases the diagnostic yield, they are only selectively performed and thus some patients having a non-diagnostic biopsy patients live with uncertainty or proceed to treatment. Language was added to the Research Gaps section to address this issue.</p>

Source: <https://www.effectivehealthcare.ahrq.gov/search-for-guides-reviews-and-reports/?pageaction=displayproduct&productid=2186>

Published Online: February 24, 2016

		<p>Abstract is inconsistent with harms. In the results it states when evaluating postoperative harms all interventional strategies were approximately equivalent while in the conclusions it states thermal ablation has the highest local recurrence rate but favorable perioperative outcomes and harms. All strategies cannot have equivalent harms but ablation has favorable ones.</p> <p>2. In the executive summary. ES 1 states active surveillance has emerged as an option for patients with small renal masses, a low likelihood of aggressive malignancy, and/or a limited life expectancy. it is also an option for patients with significant operative risks not just limited life expectancy</p> <p>b.ES5 states ...and negative predictive value of 60.0 percent but 16 percent of biopsies were nondiagnostic. The majority of nondiagnostic biopsies were found to correspond with malignant surgical pathology 88.9. The negative predictive value of biopsy was estimated to be 60 percent. However benign or nondiagnostic biopsies do not necessarily proceed to surgical extirpation limiting the analysis and making the exact false negative rate difficult to ascertain. Therefore, the strength of evidence for diagnostic accuracy of renal mass sampling was graded as moderate. The reviewers need to distinguish nondiagnostic biopsies missed, the tumor got only fat or normal kidney, from negative biopsies. This is an acknowledged problem with the literature and adversely affects the NPV. One way to evaluate this is to look at the rebx data for nondiagnostic biopsies. Admittedly, limited but the way this is currently written underrepresents the value of the renal mass biopsy. The authors are willing to concede the use of reablation ES7 improves outcomes and should so the same for rebx.</p> <p>ii. The data on the strength of biopsy are best considered by 3 distinct categorical variables for those that go on to be removed cancer y/n, histology right/wrong, grade up/down.</p> <p>iii. The data note a difference in undiagnosed and treated SRMs. this should be mentioned in with the bx data. What of those lesions that are treated by ablation NSSRNAS never obtain a histologic dx a downside to ablation and ASc.</p>	<p>The confusion here may stem from the difference between “perioperative outcomes” and “harms.” These are defined in the PICOTS (Table 1) and clarified in the KQ3a sections. The abstract was changed to indicate that TA has more favorable perioperative outcomes. The abstract and text clearly demonstrate the differences in harms among treatment strategies.</p> <p>We added “procedure limiting comorbidity” to the list.</p> <p>These are excellent points. Non-diagnostic biopsies were NOT used to calculate the NPV and other parameters. Only negative biopsies with a known benign entity was used. Fat and normal kidney when delineated by the authors was categorized as non-diagnostic and NOT used in the calculations. We have calculated accuracy following a second biopsy (79.8%), but caution that authors only biopsy a selective group the second time around. Language was added to the discussion (KQ2) to the point of repeat biopsy.</p> <p>We agree. All three metrics are addressed in the results section of the report, KQ2: histology (cancer vs benign, clear cell vs other, and grade concordance).</p> <p>Thank you. The following statement was added to ES: “Although histologic confirmation was not required for thermal ablation studies, most institutional studies only included biopsy</p>
--	--	--	--

Source: <https://www.effectivehealthcare.ahrq.gov/search-for-guides-reviews-and-reports/?pageaction=displayproduct&productid=2186>

Published Online: February 24, 2016

		<p>ES6 The strength of evidence was moderate for the finding of equivalent cancer-specific survival for thermal ablation versus radical nephrectomy and low for thermal ablation versus partial nephrectomy. Metastasis free survival did not differ between any treatment modalities with low strength of evidence on pairwise comparisons except for partial nephrectomy vs. thermal ablation. where there was moderate strength of evidence for equivalent metastasis free survival it is unclear how the strength of evidence differs comparing thermal ablation to radical v partial NTX as there are no studies that make said comparisons.</p> <p>d.ES 16 states Based largely on results from partial nephrectomy series this review suggests that patients with excellent renal function and those with the poorest renal function chronic kidney disease stage 3 or greater may not benefit from nephron sparing surgery while those at intermediate risk chronic kidney disease stage 2 have the most to gain from nephrons paring approaches. This may be due to the significant renal reserve at normal levels of renal function and low renal reserve at very low levels of renal function. Further research should strive to identify the patients most likely to benefit from nephron sparing approaches from a renal functional standpoint. The data are not strong enough to determine at what eGFR CKD status NSS is best suited. The immediate goal of NSS is to maintain enough nephrons to maintain renal function and avoid HD which means NSS is appropriate for many pts with CKD3 and to avoid more rapid future decline in those with adequate renal function in those with CKD 4 and medical renal disease. The statement as written is not an accurate reflection of the published literature or current clinical management.</p> <p>Data about urine leaks and other urologic complications such as abscess appears to be missing and clearly is more common after PNx than RNx. Studies from MSKCC show more such complications after PNx than RN. Also the randomized trial showed this, more urine leaks after PN, more patients needing repeat surgery etc. Expert opinion will clearly support this.</p>	<p>confirmed tumors and the rate of histologic confirmation was lower for thermal ablation patients in SEER studies.”</p> <p>Strength of evidence was rated as moderate for cancer-specific survival for thermal ablation vs. radical nephrectomy due to consistent and precise study findings with and medium study limitations (mostly selection bias, TABLE 27). Thermal ablation vs. partial nephrectomy had low SOE due to inconsistent findings and high study limitations of studies. Two included studies do include all three treatments.</p> <p>We agree with the reviewer’ comment, in that this subgroup analysis is both limited by power and limited by long term renal functional outcomes (beyond 1 year). We have changed to the following statement, “Our synthesis of studies suggests that patients at the lowest (preoperative eGFR <45 ml/min/1.73 m2) and highest levels (preoperative eGFR >90 ml/min/1.73 m2) of kidney function may not experience renal functional benefits from nephron sparing procedures compared with radical nephrectomy. However, this is likely due to decreased numbers of studies reporting these subgroups and outcomes, and the few studies reporting follow-up beyond one year. Further research should strive to identify the patients most likely to benefit from nephron-sparing approaches from a renal functional standpoint, and in particular long-term development of chronic kidney disease (CKD) and/or end stage renal disease (ESRD).”</p> <p>Thank you. It is highlighted in the text more explicitly. Table 43 lists the partial nephrectomy vs. radical nephrectomy harms and bold faced items are deemed to be significant. This includes ureteral injury, urine leak, and abscess. .</p>
--	--	---	--

Source: <https://www.effectivehealthcare.ahrq.gov/search-for-guides-reviews-and-reports/?pageaction=displayproduct&productid=2186>

Published Online: February 24, 2016

<p>Public Reviewer #3 (Ithaar H. Derweesh, Moores UCSD Cancer Center)</p>	<p>Introduction</p>	<p>Introduction Section, Page 36 In discussing the prognosis of Stage I and Stage II RCC, a greater distinction needs to be drawn (for example the text states “5 year survival rates better than 85%”— that may be so for both, but for Stage I, these rates are usually better than 90%, whereas for Stage II, these are in the upper 70s-low 80s”).</p> <p>Therapeutic Interventions and Outcomes, Page 37 The statement “Controversies exist regarding the ideal management for renal masses of different stages. For example, partial nephrectomy has emerged as the recommended treatment for T1 renal masses, yet the single randomized, prospective study demonstrated improved overall survival with radical nephrectomy.²⁶” Should be amended. When analysis was confined to Renal Cell Carcinoma patients, EORTC 30904 did not show OS advantage for RN. This should be added to the above mentioned statement.</p> <p>Current Guidelines and Shortcomings, Page 38 The statement “In 2009, the AUA published the guideline used most widely by the United States urological community. This guideline was largely based on expert opinions and the best studies available at the time, which were observational and retrospective in design.²³” Should be amended to reflect that fact that there was a meta-analysis of the literature in the guidelines. Would suggest the following: “In 2009, the AUA published the guideline used most widely by the United States Urological community. This guideline was based on expert opinions and a meta-analysis of the best studies available at the time, which were observational and retrospective in design.²³”</p> <p>Current Guidelines and Shortcomings, Page 38 The statement “Patients with surgical chronic kidney disease (defined as chronic kidney disease as a result of surgical nephron loss or injury) are different, with stable long-term renal function and improved overall survival, from patients with chronic kidney disease resulting from medical renal disease.” Should be amended to reflect the fact that even with CKD-S a subset of patients may experience decline. The tone of this sentence may need to be softened. Would suggest the following: “Patients with surgical chronic kidney disease (defined as chronic kidney disease as a result of surgical nephron loss or injury) are different, with more stable long-term renal function and improved overall survival for most, compared to patients with chronic kidney disease resulting from medical renal disease.”</p>	<p>Thank you. The introduction has been amended.</p> <p>Thank you for this excellent point. We amended the sentence to reflect this important conclusion.</p> <p>Thank you. This is now changed in the text.</p> <p>Thank you. The reviewer's wording is more precise and this change has been made.</p>
---	---------------------	--	--

Peer Reviewer #2	Introduction	The introduction is clear and well written.	Thank you
Peer Reviewer #3	Introduction	Well done, clear no concerns.	Thank you
TEP #2	Methods	appropriate inclusion/exclusion	Thank you
TEP #3	Methods	The inclusion and exclusion criteria are justifiable. The search strategies are explicitly stated and they are logical. The definitions or diagnostic criteria for the outcome measures are appropriate. The statistical methods used are appropriate.	Thank you
TEP #4	Methods	Well-reasoned; no changes.	Thank you
TEP #5	Methods	The inclusion and exclusion criteria are justifiable. The search strategies are stated explicitly. The definitions and diagnostic criteria for the outcome measures are appropriate. The statistical methods used are appropriate.	Thank you

Peer Reviewer #1	Methods	<p>1. I perceive this review took place an undue amount of emphasis on positive and negative predictive values as a measure of diagnostic accuracy. The issue with this is that these measures are highly prevalent dependent. Instead, I would like to see the reporting of positive and negative likelihood ratios, which would also allow the incorporation of non-dichotomous outcomes.</p> <p>> 2. I did not see the rationale for separate analyses for SEER and non-SEER based data reported in the methods, which does not strike me as intuitively indicated. Meanwhile, I would've expected a separate analysis for the single randomized controlled trial versus all the other comparative observational studies. Also, the decision to pool across the studies should be explicitly justified.</p> <p>> 3. This review includes studies of prognosis, clinical prediction rules, and studies of diagnostic accuracy as well as studies of therapy. I do not see a lot of information about how studies of prognosis and clinical prediction rules were rated for the quality of evidence that they provided. I suspect that none of the studies that sought to predict outcomes were prospective in nature let alone prospectively validated in an independent sample. Is maybe an issue that could also be featured in the discussion?</p>	<p>1. The focus is on standard diagnostic accuracy parameters while working within the confines of the literature. Since many studies have 100% reported sensitivity or specificity (often affected by verification bias due to missing surgical pathology after negative biopsy result), LRs cannot be calculated by study (leads to division by 0 situations). Based on the suggestion, we have calculated overall LR+ and LR- and included them in KQ2. However, we also provided PPV and NPV since these measures may be more clinically relevant given that we believe the studies do include a representative prevalence of small renal masses.</p> <p>2. The rationale for separating SEER studies has been added to KQ3: "A further limitation of the SEER database is a potential selection bias for patients selected to undergo partial nephrectomy as compared to radical nephrectomy, which has been repeatedly suggested in the literature as well as analyzed in a well-designed study [Shuch et al 2013, Cancer 10.1002/cncr.28141]. Therefore, when possible, results are stratified by SEER and non-SEER studies to account for this effect."</p> <p>3. Studies focused on prognosis (and predictors of prognosis) are not a focus of the review. Prognosis is only evaluated where it relates directly to comparisons of treatment efficacy, for which risk of bias and strength of evidence are provided with detailed justification. For diagnostic studies, risk of bias and strength of evidence are also evaluated in a standardized format using the QUADAS-2 tool and standard EPC domains (Tables 9 and 18).</p>
Peer Reviewer #2	Methods	The methods are well done; the search strategies are explicitly stated and logical. Statistics appear to be appropriate.	Thank you.

Peer Reviewer #3	Methods	See general comments. While the criteria for inclusion and exclusion are fairly well stated and logical it appears that there were many articles excluded at the "article review level" inappropriately. If they were excluded appropriately then there is something not clear about the exclusion criteria. As stated above, the most egregious of these appear to be in the 266 excluded because they pertain to non-clinically localized patients. Many on that list do in fact report clinically localized patients. There appear to be several other important articles excluded in multiple other categories that seem to be mistakenly classified as exclusions as well. Otherwise methods are fine.	We conducted an audit of 10% of the articles on the exclusions list. They consistently were found to have valid reasons for exclusion. A problem we face with this literature and discussed in the report is that numerous studies do not explicitly mention the clinical stage of the patients included but rather provide pathological stage which includes a significant number of stage III or greater patients. We cannot be sure that they are clinically localized patients and for this reason the study is excluded to satisfy the pre-established exclusion criteria.
TEP #2	Results	helpful results	Thank you
TEP #3	Results	The amount of detail presented in the results section is appropriate. The characteristics of the studies are clearly describe. The key messages explicit and applicable.	Thank you
TEP #4	Results	Clearly described; no changes.	Thank you
TEP #5	Results	The amount of detail presented in the results section appropriate. The characteristics of the studies clearly described. The key message explicit; figures, tables and appendices are adequate. Although there are not many key messages that are readily applicable in terms of guiding/changing current clinical practice, I think it accurately reflects the knowledge we have regarding management of localized renal masses.	Thank you.

<p>Public Reviewer #3 (Ithaar H. Derweesh, Moores UCSD Cancer Center)</p>	<p>Results</p>	<p>Pages 80/115: Renal Functional Outcomes In the comparisons between partial nephrectomy and radical nephrectomy, the delta eGFR difference between radical and partial nephrectomy is lower than what would be expected.</p> <p>Furthermore, the calculated rate of CKD for radical nephrectomy (29%) is lower than which is quoted by most series (Huang et al 71%, Malcolm et al, on the lower end of the spectrum at 43-44%)? The panel should re-visit the selected/excluded/reviewed papers and make suggestions as to what other, if any papers to include in the analysis.</p> <p>Page 106: Meta Analysis for Radical vs. Partial Nephrectomy—we should include only patients with Cancer for Oncological Outcomes. For example the data included for the Van Poppel et al. study includes non-cancer patients, and this (and any other study) from which non-cancer patients were abstracted for the meta-analysis should be modified accordingly—if this is not possible, as part of the limitations of the analysis, the fact that in some studies, the granular data was not available to make such a distinction should be acknowledged.</p>	<p>The delta eGFR between radical nephrectomy and partial nephrectomy was likely not as prominent as expected due to the influence of the large study by Woldu from 2014 (contributed 79% of the weight to the meta analysis). We have added the following statement to the executive summary: “Individual decline in estimated glomerular filtration rate and risk of chronic kidney disease can be more or less extreme than this (range of average decline seen across studies was 0.7 to 37.5 ml/min/1.73 m2 more for radical nephrectomy)</p> <p>We agree that there is significant heterogeneity in the risk of chronic kidney disease (CKD), and have added the statement “The incidence of chronic kidney disease was estimated at 29-32 percent for radical nephrectomy, with range of average incidence across studies 2-70%.”</p> <p>The focus of the review is on the outcomes after treatment of clinically localized small renal masses. Therefore, we do expect a portion of these tumors to be benign on final pathology. The majority of studies limit analysis to pathologically confirmed cancers, especially for partial nephrectomy and radical nephrectomy. EORTC is an exception due to its randomized control trial design with the most valid analysis being intention to treat. However, data for active surveillance purposefully includes some benign disease (whether known or not) and some thermal ablation studies also include benign disease if no pathologic data from pre-ablation biopsy is available. In the case of ablation, almost all institutional studies only included biopsy-confirmed cancers while not all patients treated with thermal ablation had confirmed pathology in SEER studies. We tabulated oncologic outcomes based on how individual studies reported the data</p>
---	----------------	--	--

			and they otherwise consistently reported on surgical pathology confirmed cancer for partial nephrectomy and radical nephrectomy. The limitation of thermal ablation is more specifically mentioned in the appropriate KQ3 sections in ES and main text.
Peer Reviewer #1	Results	<p>Results: 1. Key question 1: The relevance of the increased risk associated with certain tumor characteristics could be enhanced by providing some illustrative examples that use absolute numbers. In a nutshell, it would appear that every centimeter increase in size corresponds to an approximately 30% increased risk of malignancy. how do that information play out in practice? If we had a patient with a 2 cm enhancing real mass that we thought had a 60% pretest probability, would an otherwise comparable patient with a 3 cm mass have an 80% chance of harboring a malignant mass? When framed in absolute numbers, I hesitate to believe these numbers. I also wonder whether it is appropriate to treat tumor size as a continuous variable, meaning does a size increase from 2 cm to 3 cm, from 3 cm or 4 cm, and from 4 cm to 5 cm etc... All correspond to a 1.3 times increased risk?</p> <p>> 2. Key questions 2a and 2b: I suspect readers will have the greatest interest in the biopsy studies that were judged to be at low risk of bias. Please cite the studies in the executive summary so that they are readily identifiable.</p> <p>> 3. Paragraph 2 leads with a comparison of core biopsies and fine needle aspiration. I would find it preferable to report the diagnostic accuracy of each of them and then to follow up with her comparison. I also have the impression that information about fine needle biopsies may be insufficient for making a judgment.</p> <p>> 4. I would take the opportunity to label the issue of verification bias which severely hampers the interpretation of these results. While the challenges of performing high quality studies are well understood, in the context of this review, this is a major limitation.</p>	<p>1. We are only able to report on the literature and summarize findings from other studies. The included studies tend to treat tumor size as a continuous variable and therefore the results are presented in a “per cm” basis by performing a meta-analysis on similarly reported outcomes. The OR of 1.3 is a relative finding, which presumes the continuous spectrum of tumor size is uniform. The assumption may not be completely accurate, but the interpretation of the OR as “increased risk” should also be avoided. The main point is that larger tumors are associated with malignancy, and that the OR of 1.3 gives a relative magnitude of the relationship.</p> <p>2. The low risk of bias studies are now clearly identified in the Risk of Bias section in the main text. The executive summary does not reference individual studies.</p> <p>3. We have made this change in the ES. We have added mention in the Results of the Abstract. The first sentence simply states the FNA diagnostic accuracy and second sentence mentions core biopsy accuracy. A sentence is added at the end of the paragraph to note the limitation of only one FNA study, which is from early in our analyzed time frame (1997).</p> <p>4. Verification bias is now clearly stated in the ES with insertion of the following language: Verification bias exists in these studies as benign or non-diagnostic biopsies do not necessarily proceed to surgical extirpation, limiting the analysis and making the exact false negative rate difficult to ascertain. In addition, there is bias in</p>

Source: <https://www.effectivehealthcare.ahrq.gov/search-for-guides-reviews-and-reports/?pageaction=displayproduct&productid=2186>

Published Online: February 24, 2016

		<p>> 5. Overall survival and oncological outcomes: With regards to thermal ablation there is a sentence that reads "... due to the selection of older patients with greater comorbidity to undergo the procedure." to explain the potential for overall survival of these patients. This appears to be speculative. Given the preponderance of observational studies selection bias is a major issue across all these studies.</p> <p>> 6. On several occasions the claim of equivalent outcomes is made, for example for the comparison of radical versus partial and the outcome of cancer specific survival. Such judgment requires a ideally predefined threshold for equivalence. What was that threshold? Such a threshold would also have been necessary for judgments about precision.</p>	<p>who proceeds to surgery as patient or tumor characteristics (i.e. male sex, larger tumors) influence the decision to proceed to surgery. Therefore, the strength of evidence for diagnostic accuracy of renal mass sampling was graded as moderate. .</p> <p>5. Selection bias exists in much of the literature reviewed. However, demographic and clinical data provided demonstrate that patients undergoing thermal ablation are more likely to be comorbid and older than patients undergoing surgery.[Table 19]</p> <p>6. The assessment of equivalent outcomes is based on several factors, among which as mentioned is consistency and precision, which are also important in grading the strength of the evidence (we used standard AHRQ methods and the GRADE approach). Consistency is interpreted on the direction of a finding across studies (for example, if no study comparing partial nephrectomy to radical nephrectomy found a difference in cancer-specific survival; or if ALL studies found a oncologic survival benefit for partial nephrectomy). Studies providing statistical tests for comparison are directly quoted. Precision is based more on sample size and number of studies. We also relied on minimally important differences (MID) depending on the outcomes after summarizing the literature (e.g. MID for oncologic survival was 10% based on the EORTC study, and for cancer-specific survival was selected as 5%). When a MID was not met and it was a consistent finding, we made the assessment that an outcome was equivalent between two modalities. Any specific instances of individual studies that diverged from this finding (and perhaps state a statistically significant association, regardless of the magnitude of difference) are discussed in the main text of the relevant sections.</p>
--	--	---	---

Source: <https://www.effectivehealthcare.ahrq.gov/search-for-guides-reviews-and-reports/?pageaction=displayproduct&productid=2186>

Published Online: February 24, 2016

		<p>> 7. Renal functional outcomes: from the perspective of a clinician reading this report I do not find it helpful to categorize outcomes by their type as either continuous or categorical in nature. It would appear much more meaningful to distinguish them by the degree of how patient important they are. The development of end-stage renal disease and the need for dialysis or the development of chronic kidney disease related complications is clearly patient important. Estimated glomerular filtration rates and serum creatinine levels on the other hand are surrogate outcomes that are not directly patient relevant.</p> <p>> 8. Health-related quality of life: There is a suggestion of potentially better quality of life in patients undergoing radical nephrectomy. I think it would be important to define the time horizon with which the study assessed quality of life. It would be most obvious that this would relate to short-term outcomes, say within it appeared up three months.</p> <p>> 9. Perioperative outcomes and harms: this may not be consistent with AHRQ guidance on formatting of the executive summary. Should either provide all relevant information or reference the tables and figures where appropriate, which it currently does not do.</p> <p>> 10. I would encourage the consistent reporting of effect sizes and associated confidence intervals as well as the strength of evidence ratings; these belong together and none of them should be viewed in isolation.</p> <p>> 11. The figures refer to " incidence of major Clavien complications"; I think the wording could be improved upon. Consider, "rates of major complications (Clavien grade 3 - 5)". As an aside, is the word</p>	<p>7. We thank the reviewer for this comment. We agree that the development of end-stage renal disease with the need for dialysis is what is most important to patients. That is why we included the incidence of end-stage renal disease as one of the main outcomes of interest. Unfortunately, assessment of this outcome is limited by the low power of the studies for detecting an important difference in this infrequent outcome. We thought it was important to report on the measures of renal function that were reported most frequently in the studies, and we categorized the renal function measures by their continuous versus categorical nature to appropriately combine them in quantitative synthesis of the results (as explained in the methods section).</p> <p>8. There are only four studies that evaluate HRQOL. Each study is detailed in the HRQOL section with corresponding follow-up times. Heterogeneity in study design and reported outcomes prevent meta-analysis. Specifically, each study had different study design and follow-up times. Some had pre-defined times at which HRQOL questionnaires were completed and some were cross-sectional studies preventing any meaningful comparison or interpretation of composite data based on time.</p> <p>9. All studies are referenced in Table 39 or the appropriate portion of the text.</p> <p>10. We have confirmed consistency in our reporting of the effect sizes, associated confidence intervals and strength of evidence.</p> <p>11. Thank you. Incidence has been substituted with "rates" in the harms section.</p>
--	--	---	--

Source: <https://www.effectivehealthcare.ahrq.gov/search-for-guides-reviews-and-reports/?pageaction=displayproduct&productid=2186>

Published Online: February 24, 2016

		<p>"incidence" needed?</p> <p>> 12. I would find it helpful if the figures included in the executive summary included numerical information about the effect sizes in the associated confidence intervals.</p> <p>> 13. Data synthesis (page 3): cancer specific survival is a time to event outcome; the definition could be more clear and refer to time.</p> <p>> 14. What was the denominator for assessing the diagnostic accuracy of determining tumor grade. specifically, how did you handle the nondiagnostic biopsies; removing them from the denominator they overestimate diagnostic accuracy.</p> <p>> 15. This report includes a single randomized controlled trial. While I recognize that this is not AHRQ standard, it would seem very reasonable to reach out to the authors of this trial could clarify methods. Given that this was an EORTC trial, I suspect that the methodology for the random sequence generation and allocation concealment were appropriate, but of course that needs to be verified. I believe you can safely judge that patients and study personnel were not blinded. The method section would also benefit from a clearer distinction of the importance of blinding with regards to performance bias versus detection bias. Blinding of outcome assessors is important for cancer specific survival but should not matter or overall survival.</p> <p>> 16. I failed to find any ROC curves or other graphic forms of reporting the results of the DTA review.</p>	<p>12. Thank you. We have added the numerical information</p> <p>13. Thank you. This has been clarified in the section. (Methods, Data Synthesis)</p> <p>14. For tumor grade evaluation, we clearly indicated the proportion of patients with biopsy results showing RCC, and then took the percentage of those for which grade could be assigned (shown to be 67.3%) to show that a proportion of RCCs did not have grade assigned at biopsy. Then, among patients with surgical pathology available, we compared final tumor grade to biopsy grade to assess upgrading (from 1-2 to 3-4). We only looked at the subset of biopsies showing RCC (nondiagnostic biopsies excluded) as those are the only ones where grade can be assigned (Results, KQ2, Fuhrman Grade).</p> <p>15. We reached out to the authors of the EORTC study for clarification of methodology. This methodology is not detailed in the text of any manuscript associated with the study. The response helped determine that there was center level variation in allocation concealment (so it remained unclear) and appropriate random sequence generation was utilized.</p> <p>16. No study included ROC curves given that the diagnostic test being evaluated is uniformly dichotomous across studies without varying thresholds for positive/negative results. Furthermore, we could not pool the data in ROC curves given the lack of surgical pathology for most patients with a benign biopsy (when there is no surgical pathology for benign biopsies, the true negative and false negative numbers essentially become zero or near zero and no ROC analysis is</p>
--	--	---	--

Source: <https://www.effectivehealthcare.ahrq.gov/search-for-guides-reviews-and-reports/?pageaction=displayproduct&productid=2186>

Published Online: February 24, 2016

			possible).
Peer Reviewer #2	Results	The level of detail in each result section is extensive and appropriate. Studies are clearly described. Key messages are applicable and the figures and tables are very good. The evaluation appears exhaustive.	Thank you
Peer Reviewer #3	Results	Results are presented well, balanced, clear and appropriately concise. Tables are clear. Multiple studies overlooked as above. None included that should have been excluded.	We performed an audit to confirm that studies were not excluded inappropriately.
TEP #2	Discussion/Conclusion	Good points. Bias still permeates as it comes from bias sources and authors. Overall the project tries to weed out the bias, but unfortunately, it still maintains some of the old beliefs. For example, the negative biopsy is never taken in a vacuum. If the biopsy result is normal fat or normal kidney in the setting of a solid renal mass, then we know that the biopsy "missed." Clinically, we know the biopsy needs to be repeated. The propagation of the concept that biopsy is fraught with a high non-diagnostic rate is not a current, realistic concept.	We appreciate the reviewer's comment. We do not count normal fat or kidney as a diagnostic biopsy; instead it is labeled "non-diagnostic" and not included in the accuracy measurements. (Methods, Data Synthesis; and Results, KQ2a). We merely report the non-diagnostic rates in the current body of literature. A biopsy that does not yield a definitive diagnosis is non-diagnostic. Further we required verification by surgical pathology for calculations of performance characteristics. We hope this clarifies this point. This report reviews the existing literature and the non-diagnostic rate is an accurate measure of what is published.
TEP #3	Discussion/Conclusion	The discussion section is well written. The Conclusion section should include a statement saying core biopsy offers improved diagnostic abilities over fine needle aspiration since the strength of evidence is moderate and a lot of readers would emphasize their reading of Conclusion paragraphs.	This was added to the ES conclusions section.
TEP #4	Discussion/Conclusion	Implications are clearly stated; no changes.	Thank you
TEP #5	Discussion/Conclusion	The implications of the major findings clear stated. The limitations of the reviews are described adequately. Regarding the research gaps, although this report points out many areas that clearly needs future studies, some inherent difficulties may hinder an easy translation into new research, for example, the high risk of bias of renal mass biopsy because of a lack of follow-up surgical pathology for negative or non-diagnostic biopsies. In Page 30, line 6. Should it be "a greater decrease in GFR" or "a great increase in serum creatinine"?	We agree that the lack of follow-up pathology is a problem for interpreting the results of the biopsy studies, and we have emphasized that limitation of the evidence. Thank you, this has been edited.

Public Reviewer #3 (Ithaar H. Derweesh, Moores UCSD Cancer Center)	Discussion/Conclusion	<p>Page 192, Discussion Section on Key Question 1:</p> <p>I have a problem with "Guidelines from the AUA identified tobacco and obesity as risk factors for renal cell carcinoma. The AUA also suggested that hypertension may increase the risk of renal cell carcinoma, and certain diets may protect against renal cell carcinoma. There was little evidence regarding hypertension, smoking status or dietary habits in preoperative composite models. While these patient characteristics may demonstrate a relationship to renal cell carcinoma in epidemiological studies, they do not demonstrate a relationship to malignancy in the preoperative composite models in this systematic review." I have a problem with us commenting on this issue, as the studies being used for this analysis were not powered to look at risk factors for malignancy on an epidemiological level, which are really more population-based types of studies with larger numbers. My fear is that this sends the wrong message, and we are wading into murky waters on this. The best that can be said, from studies in patients with localized renal cortical tumors suspicious for RCC, risk factors such as HTN, smoking status or dietary habits did not demonstrate an increased risk for malignant histology and leave it at that.</p>	<p>Thank you. We changed a sentence to be clear according the reviewer's comments: "While these patient characteristics may demonstrate a relationship to renal cell carcinoma in epidemiological studies, they do not demonstrate an increased risk of malignant histology in patients with a localized renal mass based on this systematic review."</p>
Peer Reviewer #1	Discussion/Conclusion	<p>Page 13 ES. Key question 2: accuracy and a figure C of renal mass biopsy of the diagnosis of renal mass suspicious for localized renal cell carcinoma: in my opinion this deserves a clear discussion. Based on current guidelines and assuming a pretest probability of 70 to 80% for say a 4 cm enhancing renal mass it is the negative likelihood ratio and the resulting negative predictive value that is really the most important measure.</p> <p>Please see Michael D. Bell et al, BJU International (2013) for a type of discussion that I would perceive to be helpful in putting the results into context.</p> <ol style="list-style-type: none"> 1. there is a sentence that reads " for example, renal mass biopsy could be performed prior to thermal ablation.."; is that not already the recommendation based on AUA guidelines? 2. Key question 3B, paragraph 3: the first sentence states that the guidelines recommended thermal ablation as a treatment option; I don't think the way this is stated is the same as saying that thermal ablation was changed to be an " option". See also above. 3. Applicability: I'm not sure that the sentence in the second paragraph that talks about studies without a surgical reference standard fits into the applicability section. The main issue is that of verification bias. if you believe this should be discussed here, also consider that patients enrolled in studies that all went on to undergo surgery may also systematically differ from the general population that you are seeking to 	<p>The section regarding diagnostic and non-diagnostic biopsies has been most scrutinized by the reviewers. This section has undergone extensive editing to ensure accurate methods and clarity of reported information. We hope the revised section suffices to address these concerns.</p> <ol style="list-style-type: none"> 1. We revised the wording. Renal mass biopsy is recommended by the AUA and EAU prior to ablative therapy. This is addressed in the discussion section. 2. Thank you. We edited the sentence. 3. Thank you. This point has been addressed throughout the report.

Source: <https://www.effectivehealthcare.ahrq.gov/search-for-guides-reviews-and-reports/?pageaction=displayproduct&productid=2186>

Published Online: February 24, 2016

		<p>generalize to.</p> <p>4. Research gaps, key question 1: in my opinion this section describes an overly optimistic tone with regards to the promise offered by new biomarkers and/or imaging techniques, not only do these need to be validated (which seems like a second step) but they need to be shown to have a value as independent predictors and well-designed, ideally prospective studies.</p> <p>5. Key question 3a: I think the notion of case volume as a surrogate measure for surgical proficiency is not a good one and should not be suggested here.</p> <p>6. What is the basis for recommending renal function data at one and 12 months; that it may take most patients undergoing a partial nephrectomy and extended period of time to realize a true benefit (if it exists) and we should look for long-term data at 3, 5 and 10 years. I think there is an important role for AHRQ reports for pointing out where important information to justify clinical practice today is lacking. In general, I would also value the discussion about time horizons and the lack of good long-term data in general on this topic.</p>	<p>4. Research gaps for KQ1 have been revised to reflect the reviewer's comments.</p> <p>5. We have changed the Research Gap to read: "A standardized definition of surgical competency or expertise is needed. While surgical or procedural case volume may serve as a surrogate measure of experience, careful review of perioperative metrics and long-term outcomes may provide a more rationale definition of expertise. Defining surgical or technical proficiency will be an ongoing challenge and standardizing how this is defined is paramount to comparative studies."</p> <p>6. This is an excellent point, with which we agree. Although long-term renal functional data does not exist, data from the Cleveland Clinic demonstrates that GFR 1-3 months following partial nephrectomy reflects long-term GFR. In the discussion section, we acknowledge the need for more data on long-term renal function outcomes.</p>
Peer Reviewer #1	Conclusion	<p>Conclusions: I believe characterizing renal mass sampling as "safe" is an overstatement; how about referring to it as having a low risk and rear serious complications. In this context, consider making a clear distinction between any complication and serious complication. if there is a 3 to 5% risk of major issues such as pneumothorax and retroperitoneal bleeding requiring admission, this would give me reasons to pause.</p> <p>I would expected to have seen a sensitivity analysis that focuses on the biopsy studies that were perceived to be at low risk of bias because they used an appropriate reference standard. If it was done, it is not featured prominently.</p> <p>Tumor seeding is a rare yet potentially catastrophic complication. Even if the authors did not find any events in the series of studies reviewed, it</p>	<p>Thank you. We changed "Safe" to low risk as suggested.</p> <p>Among core biopsy studies that were determined to be of low risk of bias, diagnostic accuracy estimates were similar to those reported overall (added to "Sensitivity, Specificity, and Positive and Negative Predictive Value" section of KQ2).</p> <p>None of the included studies had a case of tumor seeding. The table in the section records a "0" for</p>

Source: <https://www.effectivehealthcare.ahrq.gov/search-for-guides-reviews-and-reports/?pageaction=displayproduct&productid=2186>

Published Online: February 24, 2016

		<p>might be helpful to provide the reader with an estimate of its incidence, if only to emphasize that it is extremely rare. However at the same time I would expect this outcome to be highly prone to selective outcome reporting.</p> <p>On several occasions single studies are mentioned (for example on page 55) with no reported results, for example the survival rates in each group, but nevertheless a P value is cited. I find p-values to be unhelpful and unnecessary.</p> <p>Page 68: risk of bias: I think there is a mistake in the last sentence of the first paragraph. It would appear that the authors want to say that "the majority of studies were rated.." at high risk for bias.</p> <p>See above. When overall survival is the outcome, blinding of outcome assessors should not be important since death irrespective of cause can be reliably determined.</p> <p>I was surprised to see little discussion of different histological types of renal cell carcinoma.</p>	<p>all studies directly stating no tumor seeding was observed. Tumor seeding is generally a historic, case-reportable complication. We have added language regarding seeding to the discussion.</p> <p>Space limitations make it difficult to report the full results of each individual study. However, the general findings are relayed and the p-values are only to note whether a significant difference was found and not meant to relay a magnitude of effect.</p> <p>Thank you. All risk of bias sections have been updated and edited for new studies.</p> <p>We acknowledge that blinding of outcome assessors may be more important for some outcomes than for others, but the lack of assessor blinding could be a problem even for relatively objective outcomes when assessments are based on retrospective review of available records. It was a conscious decision not to evaluate outcomes by histologic subtype for a number of reasons, most importantly, the majority of clinically localized tumors are clear-cell RCC and outcomes are rarely reported by subtype in comparative studies (Table 20).</p>
Peer Reviewer #2	Discussion/Conclusion	The Discussion is clear and concise. Major findings are clearly stated as are the limitations. The future research section identifies the research gaps clearly.	Thank you
Peer Reviewer #3	Discussion/Conclusion	<p>Very well done and nice review of major findings.</p> <p>Careful study of the limitations and research gaps sections will allow the reader to determine important future work. However, these could be more explicitly stated.</p> <p>Overall amazing work.</p>	Thank you
TEP #2	Clarity and Usability	Too long. Should be presented in a shorter format.	The report follows AHRQ publication guide. It starts with abstract and ES and the full report for the transparency of the analysis and assessment.
TEP #3	Clarity and Usability	It should be made clear in the Conclusion Section that the majority of non-diagnostic biopsies that undergo surgical management are found to be malignant implying that further workup or intervention may be needed for a non-diagnostic biopsy.	Thank you

TEP #4	Clarity and Usability	Well-defined and structured; no changes	Thank you
TEP #5	Clarity and Usability	The report is very well structured and organized. The main points are clearly presented. The conclusions are helpful to inform future research directions. The usefulness regarding informing policy and/or practice is relatively limited because of a lack of evidence with high strength.	Thank you
Peer Reviewer #1	Clarity and Usability	<p>1. In the conclusion, I would place greater emphasis on the strength of evidence rather than the risk of bias since study limitations are only one aspect that affects our confidence in the estimates of effect.</p> <p>> 2. I would find a Cochrane type risk of bias summary figure to be more helpful, informative and transparent than the currently chosen risk of bias graphs. See Cochrane handbook for clarification of what I am referring to: http://handbook.cochrane.org/chapter_8/figure_8_6_c_example_of_a_risk_of_bias_summary_figure.htm</p> <p>> 3. I think the methods used to handle non-diagnostic biopsies is important enough to be mentioned in the executive summary. Results should also be labeled as to what analysis they were derived from. Also I'm not sure that the exclusion of non-diagnostic biopsies should be the primary way that these results should be analyzed; non-diagnostic biopsies of a real and important issue that affect the clinical utility of percutaneous biopsies.</p>	<p>Thank you for the comments. Strength of evidence has been added to the conclusions.</p> <p>Thank you. We have added the graphs in the appendix.</p> <p>Thank you. More detailed explanation on non-diagnostic biopsies has been clarified in the ES and report (results, KQ2a).</p>
Peer Reviewer #2	Clarity and Usability	The structure of the report is somewhat hard to digest because the executive summary is as long as most research papers and has its own reference list all of which are then followed by the actual manuscript which is very granular in its detail. I was initially fooled by this when reviewing the manuscript.	. The report follows AHRQ publication guide. It starts with abstract and ES and the full report for the transparency of the analysis and assessment
Peer Reviewer #3	Clarity and Usability	Unfortunately, the report highlights the relative paucity of work and lack of high quality data in CER for renal cancers. While a fairly good review of the literature, I fear that the limitations of the data preclude definitive statements for policy or practice considerations. Specifically, there is no evidence presented that would allow policy decisions of one treatment over another for a particular patient regardless of the key question/outcome measure. However, this does not diminish the importance of reporting the current state.	This comment is consistent with the limitations we have mentioned in the report.
TEP # 2	General	helpful summary	Thank you.

TEP #3	General	<p>Please consider revising the following sentence on page 39, Key Question 2A to improve clarity.</p> <p>Instead of saying "using fine needle aspiration or core biopsy with cytopathology or surgical pathology," please consider saying "using fine needle aspiration with cytopathology or core biopsy with surgical pathology."</p>	The wording of KQ2 has been adjusted.
Public Reviewer #1(Patrick Tomulty, Valley Medical Center)	General	I did not notice whether comorbidities especially pulmonological ones were weighted against just those patients who presented with just kidney maladies.	Comorbidities were considered as clinical predictors of malignancy (KQ1) and as predictors of comparative effectiveness (KQ3b) but did not have a major impact on any outcomes in question.
TEP #4	General	Well-written; no changes recommended	Thank you
TEP #5	General	This report summarizes the recent literature on the management of localized renal masses, and provides meaningful guidance for evidence-based practice. Although strong evidence to guide clinical practice in this area is still lacking, an accurate summary of current knowledge on the topic is very helpful and will help to guide future research efforts. The targeted populations and audience are explicitly defined. Key questions are appropriate and explicitly defined.	Thank you.
Peer Reviewer #1	General	<p>I worry that this report despite the major effort that has gone into it will have little practice-changing impact.</p> <p>> 2. target population and audience are explicitly defined; no issues</p> <p>> 3. The key questions are appropriate and explicitly stated; no issues.</p>	Thank you.
Peer Reviewer #2	General	<p>The key questions are well stated and appropriate.</p> <p>I question the statement that active surveillance is an option for T2 tumors in the top line of page 16.</p>	Thank you. Active surveillance was not addressed in the AUA Guideline and the sentence has been revised.

Peer Reviewer #3	General	<p>This is a meticulous review. Objectives and key questions are relevant, well defined and explicitly stated. Overall, these questions are answered and the level of scrutiny of studies reviewed is appropriate. The report is well written, clear, and free of bias.</p> <p>Greatest concern is the number of excluded studies. I spent a fair amount of time looking at the many pages of excluded studies. Unfortunately, it seems that there are many that may have been excluded erroneously in nearly every category of exclusions. For example, a large subset of the articles excluded under "clinically non-localized patients" are fairly high quality studies that are actually in clinically localized patients. I am not sure how this occurred.</p> <p>Despite the fact that the resulting review of the literature is incomplete, I doubt the findings and conclusions would be different.</p>	<p>The inclusion and exclusion criteria were rigorously vetted for the highest level of informative data. Despite some merit to large, retrospective single-institution studies, the majority of those studies are single-arm modality studies and not comparative (i.e. lower level of evidence) studies. To address the reviewer's concerns about inappropriate exclusions, we conducted an audit of 10% of the articles on the exclusions list. They consistently were found to have valid reasons for exclusion. A problem we face with this literature and discussed in the report is that numerous studies do not explicitly mention the clinical stage of the patients included but rather provide pathological stage which includes a significant number of stage III or greater patients. We cannot be sure that they are clinically localized patients and for this reason the study is excluded to satisfy the pre-established exclusion criteria for this review.</p>
TEP #1	Quality of the report	Superior	Thank you.
TEP #2	Quality of the report	Superior	Thank you.
TEP #3	Quality of the report	Superior	Thank you.
TEP #4	Quality of the report	Good	Thank you
TEP #5	Quality of the report	Superior	Thank you.
Peer Reviewer #1	Quality of the report	Fair	Thank you.
Peer Reviewer #2	Quality of the report	Superior	Thank you.

Peer Reviewer #3	Quality of the report	Good	Thank you.
---------------------	-----------------------	------	------------