Comparative Effectiveness Review Disposition of Comments Report

Research Review Title: *Imaging Tests for the Staging of Colorectal Cancer*

Draft review available for public comment from October 24, 2013 to November 20, 2014

Research Review Citation: Bruening W, Sullivan N, Carter Paulson E, Zafar H, Mitchell M, Treadwell J, Schoelles K. Imaging Tests for the Staging of Colorectal Cancer. Comparative Effectiveness Review No. 142. (Prepared by the ECRI Institute-Penn Medicine Evidence-based Practice Center under Contract No. 290-2012-00011-I.) AHRQ Publication No. 14-EHC046-EF. Rockville, MD: Agency for Healthcare Research and Quality; September 2014. 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 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: www.effectivehealthcare.ahrq.gov/search-for-guides-reviews-and-reports/?pageaction=displayproduct&productID=1962
Published Online: September 10, 2014

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer 1	Introduction	b. Introduction: The introductory material is generally well done. The use of the term “fiber-optic” when referring to colonoscopes is not current. These instruments have utilized video technology for decades, and fiber-optic scopes are a relic of the past. A better term would be “flexible endoscopes”	Thank you for noticing this error. We have made the suggested change.
Peer Reviewer 2	Introduction	b. Introduction: OK	Thank you.
Peer Reviewer 3	Introduction	b. Introduction: This is a very technical paper. The background is easy to read.	Thank you.
Peer Reviewer 4	Introduction	b. Introduction: Provided good background information on colorectal cancer, staging and imaging modalities.	Thank you.
Peer Reviewer 5	Introduction	b. Introduction: Intro does explain why this was done.	Thank you.
TEP Reviewer 1	Introduction	b. Introduction: Appropriate. Page 36. Should 3-D ultrasound be mentioned?	Thank you for noticing this omission. We have added the following sentence: “Advanced software programs that reconstruct multiple ultrasound images into three-dimensional (3D) images that can be viewed in real-time or studied later are coming into more common use and may improve recognition of the gastrointestinal anatomy and pathological lesions.”
TEP Reviewer 1	Introduction	Agree with the general groupings to include all colon and rectal cancers together as “colorectal cancer”. Also agree that colon and rectal cancer are somewhat different diseases. However, the following statement is not as well substantiated. “specifically in regards to staging, rectal cancer tends to spread locally, whereas colon cancer tends to spread via distant metastases. Therefore, for accurate staging, colon cancer imaging should focus more on identifying metastases as well as on tumor size and extent, while for rectal cancer imaging of distant metastases is not as important as is gauging tumor depth and local spread.”	This statement is based primarily on textbook descriptions of how rectal cancer tends to progress by progressive penetration of the bowel wall, and generally only spreads after complete penetration of the muscularis mucosae, vs. colon cancers which can invade transmurally into the lymphatic system. Thus determining the depth of invasion of rectal cancer would seem to be of prime importance. However, upon checking, we do note that no references are provided in the textbook to support these statements. Accordingly, we have deleted the sentence in question.

Commentator & Affiliation	Section	Comment	Response
TEP Reviewer 2	Introduction	b. Introduction: The introduction is informative and outlines the goals of this project with appropriate background information. A few minor comments are as follows: 1) page 31/231 line 10. It might be appropriate to state that the screening mentioned is for the general population and that screening frequency and methods differ for hereditary cancers such as HNPCC. In addition screening for African Americans is recommended to begin at an earlier age as well as for family members who have had a relative with colon cancer at a younger age.	Since the topic of the report is not about screening, we only wish to provide a very general overview. We have added "of average-risk adults" to clarify that the USPSTF recommendations do not address higher risk groups.
TEP Reviewer 2	Introduction	2) page 31/231, Staging, line27-28 "...and may require neoadjuvant therapy." This is an understatement since the standard of care for stage II and III rectal cancer is neoadjuvant therapy and is predominately chemoradiation.(also applies to ES-1 line 42-43). 4) page 34/231 line 9 "possibly radiation is the preferred treatment" Possibly is too weak of a word since combined chemoradiation is the standard of care.	We have changed the wording to indicate that the majority of patients will require neoadjuvant therapy.
TEP Reviewer 2	Introduction	3) ES-1(page 9/231) line 57 "biomarker assessment" is currently not used for staging.	The NCCN guidelines recommend assessing CEA levels during initial colorectal cancer workup, and assessing K-RAS, BRAF, MMR, and microsatellite instability. Although these biomarkers may not directly affect TNM staging, they do affect clinical decision making. We have changed the wording of the sentence to indicate the biomarkers and clinical assessment are used in decision making in conjunction with the TNM stage.
TEP Reviewer 3	Introduction	b. Introduction: The authors have performed a comprehensive review of the relevant literature evaluating imaging tests used in the staging of colorectal cancer with the primary objective to synthesize the available information on the use of imaging for staging, and the secondary objective to identify gaps in the evidence base that can inform future research.	Thank you, we agree.
Peer Reviewer 1	Methods	c. Methods: Are the inclusion and exclusion criteria justifiable? Yes Are the search strategies explicitly stated and logical? Yes Are the definitions or diagnostic criteria for the outcome measures appropriate? Yes Are the statistical methods used appropriate? I cannot say as I am not an expert in this type of review.	Thank you.
Peer Reviewer 2	Methods	c. Methods: Ok as written	Thank you.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer 3	Methods	c. Methods: The methods There are some seminal papers on MRI post-treatment that are not cited. While they may not meet the criteria for inclusion, they are widely considered timely by the community who takes care of these patients. The senior authors of several of these papers are Gina Brown and Regina Beets-Tan. I would re-consider including these papers and/or mention why they are not included.	We included, and cite, one study by G. Brown; most of her other publications appear to be more focused on predicting tumor response to chemotherapy rather than on staging, and are therefore outside the scope of this particular assessment. We included one study and excluded two papers authored by R. Beets-Tan; the reasons for exclusion are listed in the Appendixes (one was not a clinical study, and one was a systematic review that did not meet our quality criteria).
Peer Reviewer 4	Methods	c. Methods: Study design criteria for single test performance: Imaging techniques evolve over time and it is more reasonable to include both high-quality systematic reviews and primary studies published after the most recent search dates. Rating of risk of bias was based on explicit criteria of the number of questions with positive or negative answers – will this compass the complexity of the actual studies? As mentioned above, for KQ 1.a., it seems reasonable to include both the high-quality reviews and primary studies published after the search dates of those reviews. For example, for US and some CT reviews, the end dates for search is 01/2008, or 03/2009, quite late for a review conducted in 2013.	Yes, that is true. However, since the focus of the report is on comparative effectiveness, we chose to focus our efforts on primary studies that directly compared imaging studies. In an ideal world we would have the time and resources to evaluate both comparative studies and single-test studies, but since we did not have the time and resources, we chose to provide results from recent systematic reviews that analyzed single-test primary studies.
Peer Reviewer 4	Methods	Strength of Evidence Grading: it is simplistic to use I2 = 50% to judge the consistency of evidence. Bivariate mixed-effects models don't readily produce an I2. Also, the magnitude of I2, as a measure of the between-study heterogeneity, is highly affected by the precision of the study estimates. When studies are big, smaller difference between studies could be detected along with a higher I2 value.	I^2 was only one factor in judging the consistency. The first aspect of consistency was a judgment about whether the studies should be pooled together or not, which was based on a subjective assessment of whether the studies were similar enough in patient characteristics and methodology to be pooled in the first place. After this decision was made, we also used subjective judgment after studying the data—examining the forest plots, for example, to confirm the consistency, and looking across the study author's conclusions. Most of the analyses were odds ratios of errors in staging—not bivariate models—which readily produces an I^2 statistic.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer 4	Methods	Publication bias: Asymmetry of funnel plots indicates the presence of small study effect. Asymmetry of funnel plots could be due to publication bias or other reasons.	Your point is well-taken that asymmetry of funnel plots do not equate with publication bias. We are also aware that funnel plots are poor indicators of publication bias.
Peer Reviewer 4	Methods	Data analysis: The statistical analysis methods are generally sound. 1. It helps to provide an explicit definition of accuracy, over- or under-staging. What are the criteria for the choice of relative risks vs. odds ratios? For the comparative test performance, are under- or over-staging defined consistently across studies?	We have added an extensive discussion of the definitions of accuracy, over- and understaging to the methods section. We have redone the analyses using odds ratios for matched data since the same patients underwent both tests. We have provided the data on which the under and over staging analyses were based.
Peer Reviewer 4	Methods	2. While I understand that the metandi or other commands don't produce forest plots automatically, it is helpful to produce some forest plots for important results to show the data and the heterogeneity among studies.	We have added several forest plots showing the most important results.
Peer Reviewer 4	Methods	3. To compare test performance among imaging modalities, the authors could use the bivariate mixed-effects model to make direct comparison and the models could be fit using SAS, PROC NLMIXED procedure. Specific comments are as following:	Unfortunately the nature of the reported data precluded fitting bivariate models for the most part. We did fit bivariate models (using Stata MP 13.0) wherever possible. Bivariate models do not allow direct comparisons across imaging tests; they fit a model to each imaging test in isolation. Unfortunately most studies do not provide fully cross-classified data, so we were limited in what we could do.
Peer Reviewer 4	Methods	Page 16, one typo -- Al-Sukhni et al., Primary method of analysis: hierarchical summary receiver operating…, not summer.	We have corrected the typo, thank you for noticing it.
Peer Reviewer 4	Methods	Page 18, it is useful to report the heterogeneity measures for each summary estimates in the table in addition to the summary estimates. How do you judge the author's conclusions? Appropriate? This also applies to other imaging modalities.	We have provided the heterogeneity measures in the appendix. We only provided authors' conclusions to the systematic reviews of single imaging modalities, which we did not analyze further. Our focus was on the comparative data, which we did analyze.
Peer Reviewer 4	Methods	Another general comment is the unit of analysis, which could apply to all analyses in the report. For test of performance, the unit of evaluation could be patient, or lesion or other – is this true for the studies included for this review? Is it an issue for this review? How do the authors handle the different units if present? Have the systematic reviews addressed this issue adequately?	Aside from the liver lesions, all of the "units" are patients (i.e., patients being correctly staged or not.) The liver lesions were analyzed on a per-lesion basis. Our clinical experts confirmed that having per-lesion data for the liver was useful clinical information.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer 4	Methods	The comparative studies also provide information on the test performance of each imaging modalities. It would be useful to assess whether the estimates from these studies are comparable with those from studies only evaluating single modalities. Estimates of test performance should incorporate information from all sources.	The estimates from studies of imaging modalities in isolation are compared to the estimates from comparative studies in the "Findings in relationship to what is already known" section in the Discussion chapter.
Peer Reviewer 4	Methods	Table 12: what is a retrospective controlled trial?	The study was a cohort of consecutive patients who underwent either MRI or ERUS, but not both, so in effect a "controlled, nonrandomized" study. Although the publication refers to it as "prospective" cohort, the data appear to have been collected retrospectively on consecutive patients. We have listed it as cohort study in the revised report.
Peer Reviewer 4	Methods	The meaning of HSROC curves is not obvious and it helps to provide interpretation of the plots. Further, HSROC implies a threshold effect – is a threshold effect assumed here?	We have added a figure to the text with a legend explaining what ROC space is and what the graphs mean. Yes, a threshold effect is assumed. It would be difficult not to have a threshold effect with imaging tests that require subjective judgment to interpret - every reader will have a slightly different threshold for interpreting the images.
Peer Reviewer 4	Methods	Table 13, the text in Page 22 says the analysis is "distinguishing between T1/T2 vs. T3/T4". However, this is no such distinction in Table 13? For CT vs. ERUS, it is still helpful to present some information of sensitivity and specificity, for example, the range.	In Table 13, the first column, top two rows, says "distinguishing between T1/T2 and T3/T4." Insufficient information was reported by most of the studies about sensitivity/specificity (or data that allowed the calculation of sensitivity/specificity), and thus a range cannot be presented.
Peer Reviewer 4	Methods	Page 23, paragraph 4 (the paragraph above table 13) – it is generally not advisable to make such indirect comparison to make conclusions about MRI vs. CT – such data are indirect not direct and the patients populations should be very comparable across different sets of studies. Also the involved studies are generally small.	We have removed the conclusion that was based on a single study comparing the 3 imaging modalities and on transitive logic (i.e., if A is better than B and B is better than C, then A is better than C) using data from other studies.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer 4	Methods	MRI vs. CT, “only three of four studies reported sensitivity and specificity” – such sensitivity and specificity could still be pooled using a random effects model (separately) and at least data from sensitivity and specificity should be reported in some way. PROC NLMIXED does not have a restriction on the number of studies and convergence of the model is determined by the data.	The MRI vs. CT data on pretreatment Rectal T staging have now been analyzed in terms of accuracy.
Peer Reviewer 5	Methods	c. Methods: The statistical method looks at relative risk which makes no clinical sense for diagnostic test.	Staging doesn’t fall neatly into the dichotomous model of diagnostic tests; instead of 2x2 tables, the data is most clearly viewed as 4x4 or even 6x6 tables. In order to force the data into the diagnostic paradigm, artificial dichotomy needs to be forced upon the data. Wherever possible we analyzed the data using both standard diagnostic test characteristics (sensitivity, specificity, likelihood ratios, diagnostic odds ratios) and by odds ratios (in our revised analyses) of making errors. Since the primary goal of the review was to directly compare different imaging tests, and there are no standard methods of directly comparing sensitivity and specificity of tests, we chose to focus on the odds ratios of making errors of over- or under-staging. We feel indicating which test has a higher risk of making errors in staging is actually more clinically useful for this particular clinical situation than presenting pairs of sensitivity/specificity that can only be compared by visual inspection. We have added a more thorough explanation of the odds ratios and what they mean to the methods section.
TEP Reviewer 1	Methods	c. Methods: The criteria and search strategies are well defined (and necessarily narrow). For this topic, outcomes are somewhat hard to define.	Thank you, we agree.
TEP Reviewer 2	Methods	c. Methods: Overall the methods are sound including the inclusion and exclusion criteria and standard search strategies. Data abstraction, study quality evaluation including the listed critical questions, grading system, data analysis and synthesis all are appropriate. Comment: page 35/231 line 35-39, Color Doppler imaging is not widely available and such should be mentioned.	Yes, none of the included studies reported using color Doppler imaging; we have made the suggested change.

Commentator & Affiliation	Section	Comment	Response
TEP Reviewer 2	Methods	Page 37/231, line14-16 states that “iodinated contrast agents are sometimes used “when in fact they are usually used and recommended unless there is a contraindication in most clinical situations.	Fewer than 60% of the included studies of CT reported using intravenous iodinated contrast agents. Although we rated the evidence as insufficient to support an evidence-based conclusion, the data from the included studies suggested that using IV contrast did not improve the accuracy of CT for rectal T or N staging.

Commentator & Affiliation	Section	Comment	Response
TEP Reviewer 3	Methods	The review centered on two key questions: A) What is the comparative effectiveness of imaging techniques for pretreatment cancer staging in patients with primary and recurrent colorectal cancer? B) What is the comparative effectiveness of imaging techniques for restaging cancer in patients with primary and recurrent colorectal cancer after initial treatment? c) Methods: Reference standards used to assess test performance were histopathological findings, intraoperative findings and clinical follow-up. Conventional and accepted outcome measures were utilized: test performance, clinical, adverse effects and the occurrence of stage reclassification and changes in therapy. Thorough searches performed by librarians of the relevant data-bases identified references from 1980-2013. Selection criteria, determined a priori, were applied to the results. Of 4,162 citations initially retrieved, 3965 were excluded most often for lack of relevancy. Of the remaining 197 articles, 6 systematic reviews and 64 primary clinical studies were selected based on the study selection criteria. c. Methods: Reference standards used to assess test performance were histopathological findings, intraoperative findings and clinical follow-up. Conventional and accepted outcome measures were utilized: test performance, clinical, adverse effects and the occurrence of stage reclassification and changes in therapy. Thorough searches performed by librarians of the relevant data-bases identified references from 1980-2013. Selection criteria, determined a priori, were applied to the results. Of 4,162 citations initially retrieved, 3965 were excluded most often for lack of relevancy. Of the remaining 197 articles, 6 systematic reviews and 64 primary clinical studies were selected based on the study selection criteria. Study quality evaluation included a determination of bias with a modified AMSTAR instrument, and strength of evidence grading employed the formal grading system that conforms with the CER methods guide. The majority of the analysis focused on the comparative analysis of the effectiveness of the imaging modalities used in the preoperative and interim T and N staging of rectal cancer, and the preoperative and interim M staging of colorectal cancer. Applicability was determined using a panel of technical experts. Data extraction, analysis and synthesis employed standardized and accepted methods.	We agree. However, applicability of the evidence was described in terms of the populations, interventions, outcomes and settings, not by a technical expert panel.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer 1	Results	d. Results: Is the amount of detail presented in the results section appropriate? It would be easy to get lost in the details or oversimplify. The results presented are “just right” in terms of detail. Are the characteristics of the studies clearly described? Yes Are the key messages explicit and applicable? Yes	Thank you.
Peer Reviewer 2	Results	d. Results: The results are presented appropriately, with the caveats discussed above.	Thank you.
Peer Reviewer 3	Results	d. Results: One of the issues not addressed in this document is the actual methodology for the tests. For instance, with ERUS, there may be some discussion re user dependency re accuracy as well as the method of the ERUS (flexible vs rigid w or w/o proctoscopy) would be helpful. With respect to MRI, more clarity is required with respect to the type of MRI, number of series and whether this is a pelvic MRI vs rectal MRI would be very helpful. In providing this type of detail it would be easier to implement these recommendations to various groups that use these tests for staging.	Yes, we attempted to address these issues of methodology in Key Questions 1.e.iii and 2.e.iii; however, we found insufficient numbers of studies to come to many conclusions about the impact of varying the methodology. Studies of endorectal MRI were excluded on the grounds that endorectal coils are obsolete technology (see the inclusion/exclusion criteria).
Peer Reviewer 4	Results	d. Results: The summary tables provide some of the combined estimates. While I understand it is always a challenge to present the evidence in a clear and logical way for a big report, the information about individual studies and some combined estimates are presented in Tables Cs and Ds, which seems to be fragmented for some KQs. More summary estimates could be presented in the text. Also it helps a lot to provide forest plots to show important data and summary estimates in a concise way.	We have added a number of forest plots.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer 4	Results	In general, not much study characteristics are presented. Heterogeneity was not discussed in the text. No evaluation of risk of bias or other methodological features on the results. The estimates of test performance are explicitly described, but the report could benefit from some more interpretation of data (e.g., whether the imaging modality is adequate for the condition). There are some over-statements in results.	Statistical heterogeneity was discussed in the text for analyses where it was detected. Heterogeneity of the imaging protocols was also discussed, and meta-regressions were performed with some of these differences as covariates. Populations were not always well described, but we did not think that differences in populations were significant contributors to heterogeneity of the data. The data was, for the most part, fairly consistent. The risk of bias was presented in many places, and the strength of evidence ratings incorporated it. Most of the conclusions are graded as "low" strength primarily due to risk of bias. The column "consistency" in the summary tables indicates whether there was heterogeneity or not. We have responded to your concern about overstatement of results below where you are more specific.
Peer Reviewer 4	Results	Page 25, comparison of the overlapping of CIs is not a reliable method to make group comparisons. As mentioned above, formal statistical comparison of results should be made using the bivariate mixed-effects model. This applies equally to other such indirect comparisons.	Bivariate models do not allow direct comparisons across imaging tests; they fit a model to each imaging test in isolation. We were limited in part by the absence of fully cross-classified data for the imaging tests being compared. Because this report is about staging rather than diagnosis, we decided to focus on the types of staging inaccuracy (under or overstaging), and compared modalities using odds ratios for matched comparisons.
Peer Reviewer 4	Results	Page 25, first paragraph, comparison of MRI vs. ERUS – probably it is more accurate not to state there is a trend in favor of MRI. Page 26, first paragraph, -- I would not take the leap to make such statements. Overall, the evidence base is weak with smaller studies. For overstaging, the directions of comparison are consistent across the imaging modalities. There is not enough evidence to make any conclusions.	We disagree with the reviewer's statement that no conclusions can be made. For the MRI vs. ERUS comparison we concluded there was no statistically significant difference, so we do not conclude that there is a trend favoring MRI. We have removed the statements on page 26 that were based on transitive logic.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer 4	Results	Table 16, Data of publications, why “before 2000” vs. “After 2005”?	The dates for grouping the studies for the sub-group analysis were chosen for the following reasons: one of the included systematic reviews reported that accuracy of ERUS had been declining over time but then appeared to stabilize around 2000. Results from other systematic reviews suggest that the ERUS literature is affected by publication bias prior to 2005.
Peer Reviewer 4	Results	Page 27, first paragraph, “CT was more accurate than MRI in assessing CRM status” -- statistically significantly more accurate? It seems that in some places, “more” means “statistically significantly more” and in some places, it is not clear. – the use of terms should be clarified and consistent across the report.	As a general rule, the authors of the studies did not perform any statistical analyses. Wherever an outcome was statistically significant, we use the word “statistically significant”; it is very consistently used.
Peer Reviewer 4	Results	Table 17, number of lesions for some studies?	Table 17 provides a general description of the studies—study design, number of enrolled patients. It does not address the number of lesions per patient or per study.
Peer Reviewer 4	Results	Table 18, CT vs. MRI – sensitivity should have been calculated.	Table 18, CT vs MRI—sensitivity could not be calculated due to the nature of the reported data.
Peer Reviewer 4	Results	Page 29, provide explicit definition of “correct treatment”, “under-treatment” and “over-treatment”.	We have added the requested information.
Peer Reviewer 4	Results	Page 31, first paragraph, why 2008 for clinical studies?	We chose to examine the most recent studies for harms data, and chose the previous 5 years to manage the scope since we were looking at studies regardless of indication for the imaging modality..
Peer Reviewer 4	Results	In general, the harms section could benefit from further synthesis with clearer message on main points. The current text includes too much emunurating of “random” events.	We have extensively edited the harms section.
Peer Reviewer 4	Results	Page 40, Blomqvist study – there are only 15 patients and a paragraph was devoted to this study reporting estimates that may be associated with wide CIs. It may be more meaningful to describe low-risk larger studies.	There are no other studies on that topic to discuss.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer 4	Results	Page 40, in general, it helps to provide some evaluation that while the modalities may not be different in test performance, are they good or adequate modalities to use?	It is important for clinicians to consider both the accuracy and risk of harms associated with the imaging modalities. Determining the "adequacy" of the modalities should be made by clinicians after considering the information presented in the report, along with other factors that could impact the decision.
Peer Reviewer 4	Results	Page 41, the first paragraph after Table 21 – "more sensitive": significantly different or not? Another example of using the terms.	The text states that there was no difference in accuracy based on the pooled data, but does describe findings of individual studies illustrating some of the trade-offs that led the study authors to draw similar conclusions.
Peer Reviewer 4	Results	Page 41, the second paragraph after Table 21 – would focus on "no statistically significant differences" instead of the trend.	We concluded there was no statistically significant difference.
Peer Reviewer 4	Results	Page 43, the paragraph after Table 23, results in Table D-12 showed significant differences? Also "not statistically significant" should not be a reason for the grade to be "insufficient".	Thank you for catching that error; we have corrected it. The data were graded as insufficient due to inconsistency, not lack of statistical significance, and there was an entry error in Table D-12.
Peer Reviewer 5	Results	d. Results: I see that multiple numbers are wrongly stated in the paper. Table A: The table shows specificity of T1 stage as 75.8. In the analysis by Puli et al the specificity of T1 stage is higher with specificity of ERUS as 98.3. a. Mistakes like these make the analysis in this paper very questionable.	We thank you for pointing out this error; we have corrected it. The error only occurs in Table A; the number is given correctly elsewhere in the report. We are unaware of any other errors but will correct any that are noticed.
TEP Reviewer 1	Results	d. Results: Yes, very extensive reporting. Clarification needed: Page 78. Data from recent, high-quality systematic reviews were compiled to estimate the accuracy of each individual imaging modality in isolation and summarized the data in Table 19.	We reworded the sentence and hope it is clearer now: "We compiled data from recent, high-quality systematic reviews to estimate the accuracy of each individual imaging modality in isolation (see Table 19 for a summary of these data)."
TEP Reviewer 1	Results	In Table A, row 1; Table F, row 1; Table 19, row 1; and Table 24, row 1, "CRM" is used. If this refers to "circumferential resection margin" someone will need a bit of a help getting this taken care of.	It does refer to circumferential resection margin; we have corrected the definition underneath the tables. Thank you for bringing this error to our attention.

Commentator & Affiliation	Section	Comment	Response
TEP Reviewer 2	Results	d. Results: The results section is comprehensive with appropriate tables. Comment: ES -12, page 20/231, line10 “For detecting colorectal liver metastases, MRI is clearly superior to CT” This is the case for patients with fatty infiltration of the liver; however, portal venous phase contrast enhanced CT scan is an excellent method to detect liver metastases in most other circumstances and is widely used. See additional comments in the discussions/conclusion section as per below.	The conclusion that MRI is better than CT for preoperative detection of colorectal liver metastases is based on a meta-analysis of five studies, none of which reported including patients with fatty infiltration of the liver. Two studies of interim restaging of colorectal liver metastases included patients with fatty liver and found MRI to be better. A third specifically stated that patients did not have fatty liver, and also found that MRI was superior. The studies were small, and we did not attempt a meta-regression on this characteristic to see if it explained statistical heterogeneity in the results. We excluded studies of portal venous phase contrast CT on the advice of the Technical Expert Panel, who indicated it is considered to be an obsolete technology due to its invasiveness (see the study inclusion/exclusion section). Neither the NCCN guidelines nor the ACR appropriateness criteria suggest the use of portal venous phase contrast CT in colorectal M staging.

Commentator & Affiliation	Section	Comment	Response
TEP Reviewer 3	Results	d. Results: Significant findings of the paper includes: 1. For preoperative rectal cancer T (tumor) staging, ERUS is less likely than CT to incorrectly stage (relative risk [RR]=0.58; 95% CI, 0.48 to 0.69), less likely to understage (RR=0.65; 95% CI, 0.42 to 0.10), and less likely to overstage (RR=0.55; 95% CI, 0.36 to 0.85),and strength of evidence low. 2. MRI is also more accurate than CT for preoperative rectal cancer T staging, and strength of evidence low. 3. For preoperative rectal cancer T staging, there is no significant difference in accuracy between MRI and ERUS, strength of evidence low. However, using MRI instead of ERUS for patient management decisions is less likely to lead to under-treatment (RR=0.38; 95% CI, 0.21 to 0.68), and strength of evidence low. 4. For preoperative rectal cancer N (lymph node) staging, there was no significant difference in accuracy across CT, MRI, or ERUS, and strength of evidence low. 5. MRI is more sensitive than CT for detecting colorectal liver metastases (RR=1.1, 95% CI, 1.0 to 1.2), and strength of evidence is moderate. Additional important findings however that merit emphasis: 1. Few studies evaluating the effect of imaging on patient management 2. No studies reporting patient oriented outcomes. The detail is, at times, overwhelming. But this represents a definitive review at the time of its preparation.	We have reanalyzed some of the staging data using matched odds ratios and performed sensitivity analyses in which we varied the correlation between results of one imaging test and the other imaging test in the same patients. Where the results were not robust in sensitivity analyses (a single instance), we did not draw a conclusion. Our revised results are as follows: 1. ERUS is more accurate (less likely to give an incorrect result) (OR=0.36, 95% CI, 0.24 to 0.54), less likely to understage (RR=0.63, 95% CI, 0.44 to 0.89), and less likely to overstage (RR=0.47; 95% CI, 0.28 to 0.80) rectal cancer than CT in the preoperative T staging setting. 2. There is no statistically significant difference in accuracy between MRI and ERUS for preoperative rectal T staging. 3. There is no statistically significant difference in accuracy across CT, MRI, or ERUS for preoperative rectal N staging. 4. While there is no statistically significant difference in accuracy between CT and MRI for rectal N staging, MRI is less likely to overstage. 5. MRI is superior to CT in detecting colorectal liver metastases in the preoperative setting (OR=1.334; 95% CI, 1.012 to 1.761). 6. There is no statistically significant difference in accuracy across MRI, CT, or ERUS for rectal T staging in the interim restaging setting.

Commentator & Affiliation	Section	Comment	Response
(continued)	(continued)	(continued)	All of these conclusions are supported by evidence of low strength except regarding MRI vs. CT for liver metastases, which was based on moderate-strength evidence. As before, there were few studies evaluating the effect of imaging on management and none on health outcomes.
Peer Reviewer 1	Discussion	e. Discussion/ Conclusion: Are the implications of the major findings clearly stated? Yes. Unfortunately, the evidence for any modality is not strong, and there are huge gaps in the available literature. This is appropriately handled. Are the limitations of the review/studies described adequately? Yes In the discussion, did the investigators omit any important literature? No	Thank you.
Peer Reviewer 1	Discussion	Is the future research section clear and easily translated into new research? What future research section? I did not find one. There are implications about the need for future research, but this was not explicitly described - certainly not in a form that lends itself to easy application.	It has been renamed the “research gaps” section. Apparently the name has not been updated in the structured review form.
Peer Reviewer 2	Discussion	e. Discussion/ Conclusion: See general comments	No response necessary.
Peer Reviewer 3	Discussion	e. Discussion/ Conclusion: I think that the issue about technical variation that may occur with MRI or ERUS needs to be discussed in the manuscript. Moreover, the concept that CT is as good as ERUS and/or MRI for nodes is flawed.	Yes, we attempted to address technical variations in Key Questions 1.e.iii and 2.e.iii; however, we found insufficient numbers of studies to come to many conclusions about the impact of technical variation.
Peer Reviewer 3	Discussion	I liked the discussion about ERUS vs MRI vs both was useful and would highlight the importance of studying in the future.	Thank you.
Peer Reviewer 4	Discussion	e. Discussion/ Conclusion: For the implication of results, the introduction has a lot of discussion on the current guidelines. It helps to discuss how the results of this review relate the current guidelines. It is useful to further elaborate the relationship between pattern of care and the results of this review. Two major limitations: assumption of reference standard and evidence base was discussed. The research gap seems to clear.	We have described our findings in the context of other systematic reviews and clinical practice guidelines. We have pointed out that the evidence for the role of PET/CT in staging is much more limited than current utilization patterns would suggest.
Peer Reviewer 5	Discussion/Conclusion	e. Discussion/ Conclusion: The numbers do not make sense to me since they are given in relative risk terms.	We have reanalyzed the staging data using odds ratios for matched data, since the same patients are undergoing both tests. We did this to further define the nature of the inaccuracies of staging. Staging is not as simple as presence/absence of disease.

Commentator & Affiliation	Section	Comment	Response
TEP Reviewer 1	Discussion	e. Discussion/ Conclusion: Implications will be widespread, especially because the findings are not based on high levels of evidence to inform future decision making. The future research section is quite short. f. Clarity and Usability: Yes, yes, and yes.	We agree.
TEP Reviewer 2	Discussion	e. Discussion/ Conclusion: The discussion and conclusions are clearly stated and summarized nicely in table format. There is clear mention that the level of evidence is low which represents a major problem and thus should drive future research.	Thank you.
Peer Reviewer 1	Conclusion	Can the conclusions be used to inform policy and/or practice decisions? Unfortunately, I think there is little information of clinical utility in the report. The strength of evidence levels are all relatively low and there is not enough evidence to even evaluate PET/CT. The use of combination modalities is also clinically important, but not able to be addressed statistically. In clinical practice at a University Medical Center, ERUS and PET/CT are leading staging tools -- ERUS for T stage and nodes and PET/CT for metastases and nodes. However, CT and MRI are also used. Some patients get all of these modalities. Based on this report, I would not be able to create effective guidelines. I want to be clear, though, that I think the source evidence is the issue, not the report or its methodology. The glaring truth of the report is that more research needs to be performed to answer these amazingly important questions.	We agree.
Peer Reviewer 3	Conclusion	The second issue in the manuscript is the conclusion that CT is as good as MRI or ERUS for the assessment of primary nodal disease. Even though the authors acknowledge that this is of low strength, this does not have clinical face validity.	We concluded there was no statistically significant difference; in judging the strength of evidence as low, we acknowledge that publication of additional studies may change the results. It is possible that with new studies, there will be sufficient statistical power for a difference to emerge. However, the prior systematic reviews we included also support the conclusion that there is little difference across the three modalities, reporting approximately the same sensitivity for ERUS, CT, and MRI for nodal staging.

Commentator & Affiliation	Section	Comment	Response
TEP Reviewer 2	Conclusion	- Issues with the conclusions are as follows: --"No significant difference exists in accuracy between MRI and ERUS for preoperative rectal T staging" Although this is generally true there should be mention that ERUS is operator dependent which can influence the accuracy. This also should be emphasized in the appropriate other sections of the review.	We have added mention of ERUS being operator-dependent.
TEP Reviewer 2	Conclusion	--" No significant difference exists in accuracy across CT, MRI or ERUS for preoperative rectal N staging" This is questionable since ERUS will not detect more distant lymph nodes such as the iliac nodes	We concluded there was no statistically significant difference; publication of additional studies may lend sufficient statistical power for a difference to emerge. However, the prior systematic reviews we included also support the conclusion that there is little difference across the three modalities, reporting approximately the same sensitivity for ERUS, CT, and MRI for nodal staging. The sensitivity of all three modalities for detecting affected lymph nodes is quite low—our comparative evidence base suggested that none of them are much better than guessing. While we would not expect ERUS to detect distant lymph nodes, neither CT nor MRI are able to do so with any degree of accuracy. Sensitivity of MRI was 49.5% (36.0% to 63.1%), of ERUS was 53.0% (39.7% to 65.5%), and for CT was 39.6% (28.1% to 52.4%). The studies of rectal N staging were primarily concerned with identification of regional lymph nodes.

Commentator & Affiliation	Section	Comment	Response
TEP Reviewer 2	Conclusion	--"MRI is superior to CT in detecting colorectal liver metastases in the preoperative setting." This is true for individuals who have fatty infiltration of the liver; however, the portal venous phase contrast enhanced CT is an excellent modality for detecting liver metastases otherwise.	The conclusion that MRI is better than CT for preoperative detection of colorectal liver metastases is based on a meta-analysis of five studies, none of which reported including patients with fatty infiltration of the liver. Two studies of interim restaging of colorectal liver metastases included patients with fatty liver and found MRI to be better. A third specifically stated that patients did not have fatty liver, and also found that MRI was superior. The studies were small, and we did not attempt a meta-regression on this characteristic to see if it explained statistical heterogeneity in the results. We excluded studies of portal venous phase contrast CT on the advice of the Technical Expert Panel, who indicated it is considered to be an obsolete technology due to its invasiveness (see the study inclusion/exclusion section). Neither the NCCN guidelines nor the ACR appropriateness criteria suggest the use of portal venous phase contrast CT in colorectal M staging.

Commentator & Affiliation	Section	Comment	Response
TEP Reviewer 2	Conclusion	-"No significant difference this in accuracy across MRI, CT or ERUS for rectal T staging in the interim restaging setting." This also is questionable since CT scan is less accurate then MRI and ERUS can be limited by operator experience.	This conclusion for interim restaging is based on a meta-analysis of two studies that directly compared all three modalities. We concluded there was no statistically significant difference; however, publication of additional studies may lend sufficient statistical power for a difference to emerge. Examination of the data reported by the two individual studies suggests that all three modalities are rather inaccurate for interim re-staging. One of the two studies found ERUS to be slightly more accurate than either CT or MRI, but neither study found that CT was less accurate than MRI. Neither study reported any information on operator experience, but we recognize that operator experience may affect ERUS findings, and interpreter experience may affect interpretation of all three imaging modalities.
TEP Reviewer 2	Conclusion	-"Intravenously administered contrast agent does not improve the accuracy of MRI for preoperative rectal T and N staging." This conclusion is also questionable since contrast enhancement improves the ability to detect liver metastases for example both by MRI or CT and also improves the ability to detect lymph nodes.	The statement about contrast agents only applies to T and N staging, not liver metastases. It is based on the findings of three studies that directly compared contrast-enhanced and non-enhanced images for rectal T and N staging, and did not find that contrast-enhancement improved the ability to detect affected lymph nodes or perform T staging. This was not one of our "critical outcomes" for evidence grading, and it was based on conclusions of the three studies rather than our own reanalysis of the data.

Commentator & Affiliation	Section	Comment	Response
TEP Reviewer 3	Conclusion	e. Discussion/ Conclusion: Conclusion statements I am concerned that the conclusion statements can be misinterpreted if there isn't more detail provided. For example, the statement "Moderate strength of evidence suggests MRI is the preferred modality for detecting colorectal liver metastases." could be read as meaning it is preferred over PET/CT as well. In general I prefer the wording in the executive summary for each Key Question and suggest that it is repeated in the conclusion: 1. For rectal T staging, ERUS and MRI appear to not be statistically significantly different in accuracy, and both are more accurate than CT 2. For rectal N staging, ERUS, MRI, and CT are not significantly different in accuracy, but they all have such low sensitivity for detecting affected lymph nodes that it may be fairer to say they are all equally inaccurate for rectal N staging. 3. For rectal staging overall, MRI may be superior to ERUS. 4. For detecting colorectal liver metastases, MRI is clearly superior to CT. 5. There is no significant difference in accuracy across ERUS, CT, and MRI for interim rectal T-staging, and that there is a nonsignificant trend for MRI to be more accurate than CT for detecting colorectal liver metastases during restaging. 6. There was insufficient evidence to come to any evidence-based conclusions about the use of PET/CT for colorectal cancer staging.	We agree and have modified the statements.
Peer Reviewer 1	Figures	Are figures, tables and appendices adequate and descriptive? Actually, I was quite impressed by the figures and tables. There is a huge risk of their being unwieldy or overdetailed. Again, the materials presented were "just right"	Thank you.
Peer Reviewer 1	References	Did the investigators overlook any studies that ought to have been included or conversely did they include studies that ought to have been excluded? No	Thank you.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer 1	General Comments:	Recommendation: Comments: (There are no comments.) Additional Questions: Quality of the Report: Superior Number of Hours Spent to Review the Report: 3 a. General Comments: This is a topic of critical clinical importance. Having colon cancer is literally a life and death situation. Mistakes of over- or under-staging can have a major impact on expense, quality of life, length of life, and survival. The target population and audience are well defined. The key questions are appropriate and get right to the important point. f. Clarity and Usability: Is the report well structured and organized? Actually, I was really impressed with the structure and organization of the report. It was logical yet readable. Are the main points clearly presented? Yes	Thank you, we agree.
Peer Reviewer 2	1. General Comment	Recommendation: Comments: (There are no comments.) Additional Questions: Quality of the Report: Good Number of Hours Spent to Review the Report: 3 a. General Comments: This is overall a good report, but there are some problems. First is that in the title and in many places in the text this is referred to as evaluating colorectal staging. However, except for the section on liver metastases, this is a study of rectal cancer staging and is not relevant to colon cancer.	It is true that very few studies of colon cancer staging met the inclusion criteria; however, that is an important finding in and of itself. We will add a sentence pointing this out in the Research Gaps section.
Peer Reviewer 2	2. General Comment:	Second, the report is supposedly for patients as well as others. However, I do not think this would be an appropriate report for the vast majority of patients.	We agree that most patients will not be interested in reading this report; however, AHRQ's comparative effectiveness reviews can be condensed into concise reports for clinicians, patients and policymakers.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer 2	3. General Comment:	Third, and my biggest substantive problem, is that there is a complete lack of clarity on the critical issues in T-staging of rectal cancer patients. Specifically, most of the analyses on T-stage did not distinguish the strengths of ultrasound vs MRI. Although these tests are listed as being equal, they are not. Ultrasound is better at distinguishing between early T-stage extent, while MRI is better at determining late T-stage extent. The treating physician is often interested whether a tumor is T1 vs T2, where MRI is probably not as good as ultrasound.	Most of the studies, and all of the experts we consulted, stated that the most important clinical issue was distinguishing between T1/T2 and T3/T4, and on that issue, as indicated in our report, MRI and ERUS appear to be equal in accuracy. It appears that the idea that ultrasound is better for early stages and MRI for later stages came about because of the findings of one study. Of the six studies we included, only one, Yimea et al., came to that particular conclusion.
Peer Reviewer 2	4. General Comment:	However, at other times there is great interest in determining the distance to the mesorectal resection margin. Although this is not a category in the T-stage, it can have great importance in determining therapy, and this is not captured at all in the report. I think all clinicians would agree that MRI is far superior in this regard. So the test of choice is heavily dependent on the clinical situation. This fact needs to be captured in some manner in the discussion and conclusions.	We do discuss, analyze, and report conclusions about the use of MRI for predicting whether the circumferential resection margin (CRM) (also known as the mesorectal resection margin) will be involved or not. There is a paragraph in the background section introducing the topic: "Besides the factors considered in the TNM system, the circumferential resection margin is an important indicator of prognosis and essential information for treatment planning for rectal cancer. The circumferential resection margin is defined as the distance from the edge of the tumor to the margin of the resected specimen. Imaging technologies such as magnetic resonance imaging (MRI) are capable of predicting tumor involvement of the surgical circumferential resection margin. Patients with positive margins are at much higher risk of recurrence (19 percent to 22 percent vs. 3 percent to 5 percent risk for those with negative margins)."

Commentator & Affiliation	Section	Comment	Response
			We present the sensitivity and specificity of MRI for this purpose as reported by a recent systematic review (Sensitivity: 77% [57 to 90%]; Specificity: 94% [88 to 97%]), and we searched for studies that compared various modalities for this purpose. We identified only one study each of preoperative circumferential resection margin (CRM) status (MRI vs. CT) and colorectal T staging (CT vs. PET/CT). However, we also found two studies that reported on patient management based on MRI or ERUS for preoperative rectal staging. Both studies used a similar design: for each patient, the investigators devised a theoretical treatment strategy based solely on MRI information. They devised another theoretical treatment strategy based solely on ERUS information, and then they used a third strategy based on clinical information, MRI, and ERUS data to actually treat the patient. The histopathology after surgery was used to define the "correct" treatment strategy that should have been used. We pooled the results from both studies in a random-effects meta-analysis. We analyzed the outcomes "correct treatment," "undertreatment," and "overtreatment." All three analyses favored MRI as the more accurate modality for treatment planning, but none reached statistical significance. It is possible that information on CRM status, although not addressed explicitly, may have contributed to this trend.
Peer Reviewer 2	5. General Comment:	f. Clarity and Usability: Structure is good.	Thank you

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer 3	General Comments:	Recommendation: Comments: (There are no comments.) Additional Questions: Quality of the Report: Good Number of Hours Spent to Review the Report: 5 a. General Comments: The authors should be commended for a thorough review of a very complex topic. Tables in the manuscript are very worthwhile in distilling a significant amount of data. The questions are well thought through and the organization of the sections around each question was helpful. It is a very “methods” dense paper, making it somewhat intimidating for an end user and when desseminating to specialty journals the methods may need to be diluted down a little. f. Clarity and Usability: It’s an understandably dense document. It is possible to be used by government organizations although it is not very “accessible” to the specialty societies and/or journals. The document has several strengths but i worry that it may get dismissed given its conclusions re CT as well as absence of a discussion re some of the newer work in imaging for restaging.	If the report is selected for derivative publications for clinicians, patients, or policymakers, we will work with the writers to simplify the language regarding methods.
Peer Reviewer 4	4. General Comment:	Recommendation: Comments: (There are no comments.) Additional Questions: Quality of the Report: Fair Number of Hours Spent to Review the Report: 12 a. General Comments: Generally Yes. f. Clarity and Usability: Generally Yes.	No response necessary.
Peer Reviewer 5	General Comments	Recommendation: Comments: (There are no comments.) Additional Questions: Quality of the Report: Poor Number of Hours Spent to Review the Report: 12 a. General Comments: I think the clinical question of meaningful. It’s a meta-analysis. The question is not well answered. f. Clarity and Usability: The conclusions cannot be used clinically since it’s in relative risk terms.	We agree that many of the questions were not answered well or at all; however, that is due to deficiencies in the evidence base. Often identifying research gaps is the most important clinical contribution a systematic review can make.

Commentator & Affiliation	Section	Comment	Response
TEP Reviewer 1	General Comments	Recommendation: Comments: (There are no comments.) Additional Questions: Quality of the Report: Superior Number of Hours Spent to Review the Report: 3 a. General Comments: Yes. The scope of the report is well defined; the results of the review are clinically meaningful to the defined population.	Thank you.
TEP Reviewer 2	General Comments	Recommendation: Comments: (There are no comments.) Additional Questions: Quality of the Report: Superior Number of Hours Spent to Review the Report: 5 a. General Comments: This report addresses a clinically important question as to the optimal staging methodology for colon and rectal cancer using imaging. The key questions are all appropriate. Unfortunately the literature is not robust as the authors point out in drawing their conclusions. The hope is that this report will help develop further research in this area including comparative effectiveness. The lack of standardization of imaging techniques and reporting will be a drawback for future studies and should be addressed. -- f. Clarity and Usability: The report as well structure and organized and the main points are very clearly presented. Since the body of evidence is so low in assessing imaging tests for the staging of colorectal cancer, there is clear need for future research in this area as emphasized by the authors. In fact in the final conclusion the authors might add that there is need for research in this area.	Thank you for your comments. We have expanded the discussion of research gaps, which we hope will be useful for discussions about potential future research.

Commentator & Affiliation	Section	Comment	Response
TEP Reviewer 3	General Comments	Recommendation: Comments: (There are no comments.) Additional Questions: Quality of the Report: Superior Number of Hours Spent to Review the Report: 10 a. General Comments: General comments The authors have undertaken a comprehensive and exhaustive review of the literature evaluating the use of imaging in the staging of cancers of the colon and rectum. In my opinion, the two most important roles imaging plays in the staging and subsequen t management of colorectal cancers, and correspondingly the two areas most in need of clarity on appropriate use imaging are: 1. The preoperative T and N classification of rectal cancer 2. The preoperative determination of metastatic disease for both colon and rectal cancer. The literature search yielded significantly more references to address the first role of imaging than for the second role. Yet the data was fairly weak in being able to provide definitive guidance. Most striking was the finding that there were too few references to provide sufficient data for the assessment of PET/CT versus CT in the preoperative determination of metastatic disease in CRC patient; this is particularly disturbing as this represents an area of inconsistency and optimal pre-operative imaging can have a tremendous impact on patient management, and yet the use of the more expensive modality is very prevalent. This represents limitations of the field and not of the methodology. f. Clarity and Usability: The report's limitations are clearly defined, and affect the usability of the results. This is very clearly presented.	No response necessary.

Commentator & Affiliation	Section	Comment	Response
Public Reviewer 1 (IQWiG)	References	A) Regarding effectiveness In the present report, the study by Ruers et al. (J Nucl Med 2009; 50(7): 1036-41) is listed among the excluded studies because the study failed to "report on one of the test comparisons of interest". Although the decision to exclude this study appears correct, in our opinion the study deserves more attention and discussion because of its randomized controlled design. In the study by Ruers et al. patients were randomly assigned to CT imaging or combined PET and CT imaging. Theoretically, this study therefore could be extremely valuable for a relative effectiveness assessment of PET imaging. The idea of the trial was that patients scheduled for resection of liver metastases could be spared futile surgery if PET imaging showed additional metastases. However, the trial's primary outcome measure was changed, which becomes evident when comparing trial registration and publication. We therefore contacted the authors' group and learned that "the changed outcome was actually not sought for, but the direct consequence of the position of the ethical review board, who considered it a step too far to refrain from surgery with PET serving as the decisive imaging modality." This explanation by Prof. W. J. G. Oyen, Radboud University Nijmegen Medical Centre (written communication dated December 11th, 2011) showed that the trial was essentially of little value because imaging results were not used for clinical decision-making. Thus, the trial was unable to assess effectiveness in terms of futile laparotomies, overall survival or other outcomes. The results regarding the futile laparotomy rate obviously describe only the hypothetical changes in patient management. We believe that these special aspects are worth mentioning in the AHRQ report. Furthermore, we suggest mentioning the PETCAM trial, as this randomized controlled trial seems to lack the problems found in the trial by Ruers et al., 2009. Unfortunately, trial results from PETCAM are still awaiting publication. So far, only an abstract presentation is available (Moulton C, et al. Survival analysis of PETCAM: a multicenter randomized controlled trial of PET/CT versus no PET/CT for patients with resectable liver colorectal adenocarcinoma metastases. J Clin Oncol 2012; 30(Suppl): Abstract 390; available at: http://meetinglibrary.asco.org/content/88958-115).	The study did not use a combined, integrated PET/CT device and was excluded for that reason. Its results are unlikely to be applicable to current practice in the United States, which uses integrated devices. The PETCAM trial has not been published yet. Bellomi et al., Kitajima et al., Liu et al. and Arulampalam are not about staging—they are about diagnosis of suspected recurrences. One of our inclusion criteria was the study had to have enrolled patients already diagnosed with colorectal cancer, either primary or a recurrence of it. Selzner et al. is using CT arterial portography, which was identified as an obsolete technology by the Technical Expert Panel.

Commentator & Affiliation	Section	Comment	Response
(continued)	(continued)	B) Regarding test accuracy Our own report was restricted to test accuracy studies that examined comparative test performance, and we excluded retrospective studies and those with an inadequate reference standard. Of note is the fact that the AHRQ report included only studies on PET/CT, thereby excluding older studies on stand-alone PET imaging. There are nevertheless several primary studies that were included in the IQWiG report, but were not included in the AHRQ report. We suggest assessing these discrepancies. We suggest assessing the following studies for possible inclusion, as these studies were included by IQWiG but are not mentioned in the AHRQ report: a. Bellomi M, et al.: Role of multidetector CT and FDG-PET/CT in the diagnosis of local and distant recurrence of resected rectal cancer. Radiol Med 2007; 112(5): 681-90. included by IQWiG but are not mentioned in the AHRQ report: b. Kitajima K, et al.: Performance of integrated FDG PET/contrast-enhanced CT in the diagnosis of recurrent colorectal cancer: comparison with integrated FDG PET/non-contrast-enhanced CT and enhanced CT. Eur J Nucl Med Mol Imaging 2009; 36(9): 1388-1396. c. Liu FY, et al.: Utility of 2-fluoro-2-deoxy-D-glucose positron emission tomography in managing patients of colorectal cancer with unexplained carcinoembryonic antigen elevation at different levels. Dis Colon Rectum 2005; 48(10): 1900-1912. d. Selzner M, et al.: Does the novel PET/CT imaging modality impact on the treatment of patients with metastatic colorectal cancer of the liver? Ann Surg 2004; 240(6): 1027-1034. All of these studies used PET/CT, compared test accuracy of PET/CT vs. CT only, and apparently fulfilled all other inclusion criteria of the AHRQ report (e.g. sample size > 10 patients, adequate reference standard, etc.) 2. The study by Arulampalam et al. (Eur J Nucl Med 2001; 28(12): 1758-65) was excluded in the AHRQ report because patients were “not diagnosed with cancer before enrollment”. In the article, however, Arulampalam et al. report that they included “patients previously treated for CRC”, who now had a suspected recurrence. Was this study excluded because PET (and not PET/CT) was applied as an imaging technique? We suggest addressing this issue. We hope that these comments help to further increase the quality of this already very fine report.	(continued)

Commentator & Affiliation	Section	Comment	Response
Public Reviewer 1 (IQWiG)	General Comments	Dear Colleagues, We have read the AHRQ draft report on imaging in colorectal cancer with great interest, as we published a similar report in 2012. However, the report prepared by IQWiG was published only in German (with an English summary available at https://www.iqwig.de/download/D06-01C_Executive-summary-of-final-report_PET_PETCT-in-recurrent-colorectal-cancer.pdf). Furthermore, our report was restricted to an assessment of PET and PET/CT in the diagnosis of recurrent colorectal cancer. Therefore, we limit our comments to those topics that are related to this type of imaging and this indication.	No response necessary.

Commentator & Affiliation	Section	Comment	Response
Public Reviewer 2 (Medical Imaging and Technology Alliance)	Methods	Imaging modalities have varied functions and uses in a clinical setting. As such, comparative analyses of modalities are of limited value, especially when removed from the particular clinical setting and circumstances of the individual patient. Medical imaging includes multiple modalities and each modality provides unique value in the clinical setting. In fact, outside the context of a particular episode of clinical care, comparisons of modalities do not appropriately value the contribution of each modality to healthcare. Rather imaging modalities should be considered in the context of the clinical setting and their value in informing appropriate care for the individual patient. Access to appropriate imaging is necessary to inform clinical decisions related to the diagnosis and treatment of disease. In order to inform the use and appropriateness of imaging, physicians' societies have developed appropriate use criteria and practice guidelines specific to individual clinical indications. These clinical decision-support tools are based on research and evidence, and aid physicians to determine the appropriate scans for specific clinical indications for their patients. The American College of Radiology (ACR) has appropriate use criteria on Pretreatment Staging of Colorectal Cancer.4 These guidelines evaluate staging for 1) rectal cancer of small lesion size, 2) rectal cancer of large lesion size, and 3) colon cancer, which is defined as cancer of the colon present in areas other than the rectum. For each indication, the guidelines rank specific value and appropriateness. For example, CT of the abdomen and pelvis is recommended for colorectal cancer staging due to its "ability to obtain a rapid global evaluation and demonstrate complications (perforation, obstruction, etc.) that may not be clinically apparent,"5 but for large lesion rectal cancers in particular, the criteria note that PET/CT has been shown to alter staging as compared to CT alone. In other examples, for small lesion rectal cancer ultrasound is recommended for assessment of rectal wall involvement, and MRI of the abdomen or pelvis are recommended to different degrees depending on the size of lesion in rectal cancers.	Yes, we are aware of the ACR criteria, but regular assessments of the literature are valuable inputs during development and revision of clinical practice guidelines. The fact that the NCCN guidelines conflict in several areas with the ACR guidelines suggests that there is still uncertainty about which modalities are most appropriate. The report does address changes in management as one of the key questions. While change in management is an easier endpoint to measure than clinical outcomes, it is only helpful if the change is shown to be an appropriate change. Changing management to an inappropriate treatment course due to inaccurate imaging is potentially very harmful to patients. Also, the well-known and accepted Fryback and Thornbury approach to diagnostic evaluations lists changes in management as an intermediate outcome—level 4 in a 6-level framework. We appreciate the efforts of industry and imaging facilities to reduce radiation exposure from diagnostic imaging studies.

Commentator & Affiliation	Section	Comment	Response
(continued)	(continued)	These guidelines appropriately acknowledge that clinical value of each imaging modality is determined by how it informs specific clinical care, not how it ranks in comparison to a "reference standard"6 or to other modalities. MITA advocates the development and use of physician-developed appropriateness criteria to guide treatment decisions and training of hospital and imaging facility personnel who perform medical imaging exams. In and adhering to these standards and initiatives, physicians can ensure that patients receive the life-saving benefits of medical imaging technology. 2) Outcomes related to the use of imaging must be defined to reflect the unique contribution of imaging to clinical decisions. The Draft Report points to intermediate outcomes and defines these as stage reclassification and changes in therapeutic management.7 In addition, the Draft Report addresses clinical outcomes. In particular, the AHRQ states that evidence related to clinical outcomes was not found in the studies evaluated for this report. This is cited as a gap in evidence.8 However, we offer that this is not a gap, but rather an inappropriate endpoint to evaluate diagnostic imaging in the context of patient care. One consideration is that it is difficult to isolate the contribution of diagnostic imaging from the larger care paradigm, and in fact, due to the incremental value of diagnostic imaging within the delivery of healthcare, diagnostic imaging's value outside the care paradigm would be of limited meaning. Models that attempt to extract diagnostic imaging from the care that it informs neglect to reflect the reality of healthcare delivery. In fact, in clinical practice, a patient may have multiple diagnostic tests, with additional value from each test used to inform clinical decisions in unique and inimitable ways. In addition, some diagnostics tests are synergistic. For example, a PET scan may be ordered in follow up to a CT scan that shows small indeterminate lesions. Additionally, as the science of cancer staging progresses, diagnostic imaging may inform decision-making in concert with order to provide optimal care and prevent medical errors, physicians and technologists must account for the patient's individual needs. By providing proper training other tests including biomarker identification, genomic studies, and other assays the impact of diagnostic imaging on healthcare.	(continued)

Commentator & Affiliation	Section	Comment	Response
(continued)	(continued)	A recent article on the topic suggests, "The outcomes, or endpoints, appropriate to assessing whether diagnostic interventions are reasonable and necessary are best characterized as "change in clinical management." This is distinct from the outcomes, or endpoints, classically applied in. We agree with the findings that "in general all four imaging modalities appear to be reasonably safe."10 However, the Draft Report also points to radiation dose as a potential harm of CT and PET/CT. In recent years, innovative, dose-lowering technologies assessing whether therapeutic interventions are reasonable and necessary. 3) Innovative, dose-lowering imaging technologies support quality care. have limited dose while maintaining imaging quality. Due to lower dose and high clinical efficacy, the CT and PET/CT benefit-to-risk profiles have improved. Dose efficiency and dose reduction have been important design considerations for CT for many years. The focus on these design considerations has grown and intensified in more recent years, and has yielded a variety of new and innovative hardware and software features that directly help physicians both reduce and monitor dose for CT exams. The CT industry has developed new features that enable both the dose to be displayed prior to scanning, and to alert operators to potentially the patient record. These features are important for both the patient as well as facilities, since they provide facilities with the ability to compare the dose of their CT protocols and establish optimized reference values. The dose monitoring/reduction features described below play a significant role in helping to reduce the dose for CT exams, while maintaining diagnostic higher than expected doses, as well as enabling electronic recording of the CT dose in quality and the capability to report and record dose. For example: • Automatic Exposure Control helps optimize dose for each patient for the given diagnostic task. This feature adjusts the exposure to use only what is needed to maintain a constant image quality. This feature is now standard on CT systems. • Wider coverage detectors minimize the amount of x-ray that falls outside of the active detector region, thereby reducing dose to the patient without impacting image quality. Systems are now available in a range of wide coverage designs.	(continued)

Commentator & Affiliation	Section	Comment	Response
(continued)	(continued)	• "Shutter" modes block unused x-ray at the beginning and end of helical scans and therefore do not degrade image quality. This feature is now standard on many CT systems and is "built in" to each helical acquisition. • Advanced electronics in data acquisition systems result in better imaging performance and less noise, thereby enabling equal performance at a lower dose. First generation CT iterative reconstruction results in a significant dose reduction potential, while maintaining diagnostic image quality, and is well suited to CTC studies. Iterative reconstruction is available on new systems and also as an upgrade to many installed base systems. • More advanced second generation CT iterative reconstruction provides even further dose reduction potential, where some expert users are able to achieve some exams approaching1 mSv levels for combined supine and prone CTC scans, while still maintaining diagnostic image quality. This feature is becoming widely available on new systems. • The DICOM Dose Structured Report allows the exam dose to be electronically captured with the patient record. This feature is now standard on all new CT systems and has also been implemented on newer installed base systems. MITA leads industry efforts to coordinate and establish standards to mitigate radiation dose. Adoption of these standards benefits patient dose. MITA's approach builds upon existing manufacturer safety measures – including equipment safety standards, protocol development, quality and safety checks, provider education programs and physician-developed medical guidelines – to minimize radiation dose as much as possible, and to provide even greater degrees of coordination, transparency and reporting in the delivery of medical radiation. Recent examples of MITA standards which have addressed dose include: NEMA XR 25-2010, Computed Tomography Dose Check. This standard introduced two novel features to assist the imaging team in providing better patient care: dose notifications and dose alerts. Dose notifications are designed to provide a clear indication to health care providers when the parameters for a CT scan will result in a dose higher than the facility's pre-determined dose threshold for routine use.	(continued)

Commentator & Affiliation	Section	Comment	Response
(continued)	(continued)	Dose alerts are designed to prevent dose levels for a complete exam from exceeding pre-determined thresholds that are deemed excessive by the facility. This feature can be configured to prevent equipment operation. These protections help the operator and ultimately the physician to better understand dose implications of protocol choices, and should significantly reduce exposure due to Dose alerts are designed to prevent dose levels for a complete exam from exceeding pre-determined thresholds that are deemed excessive by the facility. This feature can be configured to prevent equipment operation. These protections help the operator and ultimately the physician to better understand dose implications of protocol choices, and should significantly reduce exposure due to inappropriate scan parameter settings. • NEMA standard XR 26 - 2012, Access Controls for Computed Tomography: Identification, Interlocks, and Logs. This standard requires software features that ensure only an authorized operator can alter the controls of CT equipment. This industry-wide standard requires the institutionalization of administrative privileges, access levels, and the recording of clinical protocols to ensure safe and appropriate use. NEMA standard XR 27 - 2012, X-ray Equipment for Interventional Procedures User Quality Control Mode. This standard helps imaging facilities conduct quality testing and monitoring of X-ray equipment used for interventional procedures. • NEMA standard XR 29 - 2013, Standard Attributes on Computed Tomography (CT) Equipment Related to Dose Optimization and Management. This standard, known also as MITA “Smart Dose”, is the fourth dose-related standard to be released by MITA since 2010. This standard includes four components: 1. DICOM Dose Structured Reporting – This enables the recording of post-exam dose information in a standardized electronic format. This information can be included in the patient record, promoting the establishment of diagnostic reference levels, as well as facility dose management and quality assurance. 2. Pediatric and adult reference protocols – These are a set of pre-loaded protocols on a CT system that serve as a baseline for a variety of clinical tests.	(continued)

Commentator & Affiliation	Section	Comment	Response
(continued)	(continued)	3. CT Dose Check – CT Dose Check incorporates two features — dose notifications and dose alerts that can inform operators and physicians when dose exceeds established thresholds. Automatic Exposure Control (AEC) – AEC automatically adjusts the amount of radiation used based on the size, shape and composition of the patient, in order to achieve a specified level of image quality. Studies of AEC procedures	(continued)
Public Reviewer 2 (Medical Imaging and Technology Alliance)	General Comments	The Medical Imaging & Technology Alliance (MITA) is pleased to submit comments on the Agency for Healthcare Research and Quality (AHRQ) draft comparative effectiveness review entitled *Imaging Tests for the Staging of Colorectal Cancer ("Draft Report").*1 MITA has extensive knowledge of the substantial benefits afforded by medical imaging and radiation therapy to the health of Americans due to our role as the leading trade association representing medical imaging, radiation therapy, and radiopharmaceutical manufacturers. We support quality efforts that foster appropriate use of these technologies for the early detection, diagnosis, staging, therapy monitoring, and surveillance of many diseases. Medical imaging encompasses X-ray imaging, computed tomography (CT) scans, diagnostic ultrasound, nuclear imaging (including positron emission tomography (PET)), magnetic resonance imaging (MRI), and related imaging acquisitions. Medical imaging is used to diagnose patients with disease, often reducing the need for costly medical services and invasive surgical procedures.2 In addition, medical imaging equipment often is used to select, guide, and facilitate effective treatment, for example, by using image guidance for surgical or radiotherapeutic interventions.3 MITA's members also develop and manufacture innovative radiotherapy equipment used in cancer treatment. Our comments address three areas in the Draft Report: (1) imaging modalities have varied functions and uses in a clinical setting; (2) outcomes related to the use of imaging must be defined to reflect the unique contribution of imaging to clinical decisions; and (3) innovative, dose-lowering imaging technologies support quality care.	No response necessary.