



# Effective Health Care Program

## Imaging Tests for the Staging of Colorectal Cancer

### Executive Summary

#### Background

##### Colorectal Cancer

In the United States each year colon cancer is diagnosed in approximately 100,000 people and rectal cancer is diagnosed in another 50,000.<sup>1</sup> Colorectal cancer usually affects older adults, with 90 percent of cases diagnosed in individuals 50 years of age and older.<sup>2</sup> Colorectal cancer is often fatal, with approximately 50,000 deaths attributed to it each year in the United States.<sup>1</sup> As such, it is both the third most common type of cancer and the third most common cause of cancer-related death for both men and women. Health care costs associated with care of these cancers is high, second only to breast cancer.<sup>3,4</sup>

Colorectal cancers may be diagnosed during screening of asymptomatic individuals or after a person has developed symptoms. Colon cancer symptoms include abdominal discomfort, change in bowel habits, anemia, and weight loss. Rectal cancer symptoms include bleeding, diarrhea, and pain. The U.S. Preventive Services Task Force currently recommends screening for colorectal cancer in asymptomatic normal-risk individuals using fecal occult blood testing, sigmoidoscopy, or colonoscopy, beginning at age 50 years and continuing until age 75 years.<sup>5</sup> Diagnosis is usually established through histopathologic

#### Effective Health Care Program

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

The full report and this summary are available at [www.effectivehealthcare.ahrq.gov/reports/final.cfm](http://www.effectivehealthcare.ahrq.gov/reports/final.cfm).

examination of tissue samples (most often obtained through biopsies performed during colonoscopy).

#### Staging

Once the diagnosis has been established, patients with colorectal cancer undergo



Agency for Healthcare Research and Quality

Advancing Excellence in Health Care • [www.ahrq.gov](http://www.ahrq.gov)

Effective  
Health Care

testing to establish the extent of disease spread, known as clinical staging. Staging is used primarily to determine appropriate initial treatment strategies. For colorectal cancer, the American Joint Committee on Cancer (AJCC) endorses the widely accepted “TNM” staging system. The AJCC system aims to characterize the anatomic extent of colorectal cancer based on three tumor characteristics: the extent of tumor infiltration into the bowel wall (tumor stage, designated as “T”), the extent of local or regional lymph node spread (nodal stage, designated as “N”), and the presence of distant metastatic lesions (metastatic spread, designated as “M”).

Treatment options for colorectal cancer differ depending on the clinical stage of disease at diagnosis. For example, tumors confined to the rectal wall can be treated primarily by upfront surgical resection, but tumors that have penetrated the bowel wall usually require preoperative chemotherapy and radiation (neoadjuvant therapy) prior to definitive surgical resection. Clinical stage is not the only determinant of treatment options; patient comorbidities and preferences and clinician and institution preferences are also used in decisionmaking. However, stage is the key determinant of the management strategy. Staging is also used to inform patient prognosis and identify patients at higher risk of relapse or cancer-related mortality.

Clinical staging is performed at two distinct timepoints in the management of colorectal cancer. The first is immediately after diagnosis, before any treatment has been given. Imaging and clinical examination are used to assign a clinical stage, which is used to make decisions about primary treatment and management. The second timepoint (interim restaging) applies only to patients who, on the basis of their primary staging, were treated with neoadjuvant chemotherapy or radiotherapy instead of by immediate surgery. Chemotherapy/radiotherapy affects the metabolism and structure of the tissues such that some forms of imaging may be less accurate for restaging than in the pretreatment setting. Also, the role of imaging at each of these two timepoints is very different, and for these reasons they are addressed in separate Key Questions in this review.

## Objectives of This Review

The primary objective of this review is to synthesize the available information on the comparative accuracy and effectiveness of imaging for staging of colorectal cancer. The availability of this information will assist clinicians in selecting protocols for staging, may reduce variability across treatment centers in staging protocols, and may

improve patient outcomes. A secondary objective is to identify gaps in the evidence base to inform future research needs.

## Key Questions and Scope

### Key Questions

The Key Questions are listed below.

**Key Question 1:** What is the comparative effectiveness of imaging techniques for pretreatment cancer staging in patients with primary and recurrent colorectal cancer?

- a. What is the test performance of the imaging techniques used (singly, in combination, or in a specific sequence) to stage colorectal cancer compared with a reference standard?
- b. What is the impact of alternative imaging techniques on intermediate outcomes, including stage reclassification and changes in therapeutic management?
- c. What is the impact of alternative imaging techniques on clinical outcomes?
- d. What are the adverse effects or harms associated with using imaging techniques, including harms of test-directed management?
- e. How is the comparative effectiveness of imaging techniques modified by the following factors:
  - i. Patient-level characteristics (e.g., age, sex, body mass index)?
  - ii. Disease characteristics (e.g., tumor grade)?
  - iii. Imaging technique or protocol characteristics (e.g., use of different tracers or contrast agents, radiation dose of the imaging modality, slice thickness, timing of contrast)?

**Key Question 2:** What is the comparative effectiveness of imaging techniques for restaging cancer in patients with primary and recurrent colorectal cancer after initial treatment?

- a. What is the test performance of the imaging techniques used (singly, in combination, or in a specific sequence) to restage colorectal cancer compared with a reference standard?
- b. What is the impact of alternative imaging techniques on intermediate outcomes, including stage reclassification and changes in therapeutic management?

- c. What is the impact of alternative imaging techniques on clinical outcomes?
- d. What are the adverse effects or harms associated with using imaging techniques, including harms of test-directed management?
- e. How is the comparative effectiveness of imaging techniques modified by the following factors:
  - i. Patient-level characteristics (e.g., age, sex, body mass index)?
  - ii. Disease characteristics (e.g., tumor grade)?
  - iii. Imaging technique or protocol characteristics (e.g., use of different tracers or contrast agents, radiation dose of the imaging modality, slice thickness, timing of contrast)?

### Scope

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

### Populations:

- Adult patients with an established diagnosis of primary colorectal cancer
- Adult patients with an established diagnosis of recurrent colorectal cancer

### Interventions:

Noninvasive imaging using the following tests (alone or in combination) for assessing the stage of colorectal cancer:

- Endoscopic rectal ultrasound (ERUS)
- Computed tomography (CT)
- Magnetic resonance imaging (MRI)
- Positron emission tomography combined with computed tomography (PET/CT)

### Reference Standards To Assess Test Performance:

- Histopathologic examination of tissue
- Intraoperative findings
- Clinical followup

### Comparators:

- Any direct comparisons of the imaging tests of interest
- Any direct comparisons of variations of any of the imaging tests of interest (e.g., diffusion-weighted MRI vs. T2-weighted MRI)

### Outcomes:

- Test performance outcomes
  - Test performance (sensitivity, specificity, accuracy, understaging, overstaging)
- Intermediate outcomes
  - Stage reclassification
  - Changes in therapeutic management
- Clinical outcomes
  - Overall mortality
  - Colorectal cancer–specific mortality
  - Quality of life and anxiety
  - Need for additional staging tests, including invasive procedures
  - Need for additional treatment, including surgery, radiotherapy, or chemotherapy
  - Resource use related to testing and treatment
- Adverse effects and harms
  - Harms of testing per se (e.g., radiation exposure)
  - Harms from test-directed treatments (e.g., overtreatment, undertreatment)

### Timing:

- Primary staging
- Interim restaging

### Setting:

All settings were considered.

## Methods

### Search Strategy

Medical librarians in the Evidence-based Practice Center (EPC) Information Center performed literature searches following established systematic review protocols. We searched the following databases from 1980 through November 2013 using controlled vocabulary and text words: Embase<sup>®</sup>, MEDLINE<sup>®</sup>, PubMed, and the Cochrane Library. The full search strategy is shown in Appendix A of the full report.

Literature screening was performed in duplicate using the database DistillerSR (Evidence Partners, Ottawa, Canada). Initially, we screened literature search results in duplicate (two screeners) for relevancy. We screened relevant abstracts again, in duplicate, against the inclusion and exclusion criteria. Studies that appeared to meet the inclusion criteria were retrieved in full, and we screened

them again, in duplicate, against the inclusion and exclusion criteria. All disagreements were resolved by consensus discussion among the two original screeners and, if necessary, an additional third screener.

## Study Selection

### Criteria for Inclusion and Exclusion of Studies in the Review

The inclusion/exclusion criteria were—

1. *Publication type.* The article must have been published as a full-length, English-language, peer-reviewed study. Abstracts and meeting presentations were excluded.
2. *Single test performance.* For questions about the performance of a single imaging test against a reference standard, we used a two-stage inclusion process. We first included only recent (2009 or later) high-quality systematic reviews. We included primary studies (1980 or later) only if the evidence from systematic reviews was insufficient to support an estimate of test performance for a particular imaging test.
3. *Comparative test performance.* For questions about comparative test performance, we considered studies of any design—randomized, cross-sectional, case-control, or cohort—for inclusion. Both retrospective and prospective studies were considered for inclusion, but retrospective studies must have used consecutive/all enrollment or enrollment of a random sample of participants. Studies must have directly compared the tests with each other and with a reference standard; all tests being compared must have been evaluated by the same reference standard.
4. *Stage reclassification or clinical decision impact.* For questions about stage reclassification or impact on clinician decisionmaking, cross-sectional, cohort, or prospective comparative (randomized or nonrandomized) studies were considered for inclusion.
5. *Clinical outcomes.* For questions about the impact of testing on patient-oriented clinical outcomes, we considered comparative studies (randomized or nonrandomized, prospective or retrospective) for inclusion.
6. *Harms.* The adverse events and harms reported by any studies included to address any of the other questions were used to address questions about harms and adverse events. In addition, we searched specifically for reports of harms and adverse events associated with the use of each specific imaging modality, such as radiation exposure and reactions to contrast agents. Any study design, including modeling, was acceptable for inclusion for questions about harms.
7. *Type of patient.* For inclusion, the study must have reported data obtained from groups in which at least 85 percent of patients were from one of the four patient populations of interest: (1) patients with newly diagnosed colorectal cancer underdoing primary staging, (2) patients with newly diagnosed colorectal cancer undergoing interim restaging, (3) patients with newly diagnosed recurrent colorectal cancer undergoing primary staging, and (4) patients with newly diagnosed recurrent colorectal cancer undergoing interim restaging.
8. *Adults.* Only studies of adult patients (18 years of age and older) were considered for inclusion.
9. *Obsolete technology.* The Technical Expert Panel was consulted a priori about which imaging technologies and variants of imaging technologies are obsolete and not relevant to clinical practice, and these were excluded. Likewise, experimental technologies and prototypes were excluded. The imaging technologies that were determined, after discussion and consensus, to be obsolete for staging colorectal cancer are transabdominal ultrasound, MRI using endorectal coils, nonmultidetector CT, CT arterial portography, CT angiography, CT colonography, and stand-alone PET. The Technical Expert Panel indicated that PET/MRI and PET fused with CT colonography are considered to be experimental. MRI using ultrasmall paramagnetic iron oxide is also considered experimental.
10. *Number of patients.* We included data from timepoints and outcomes reported from groups with at least 10 patients with the condition of interest who represented at least 50 percent of the patients originally enrolled in the study. We included case series, but not individual case reports, in the search for harms.

### Criteria for Key Questions on Harms

While we utilized data from studies meeting the inclusion criteria above for questions about harms, we supplemented this information with information from narrative reviews and other sources, such as U.S. Food and Drug Administration (FDA) alerts. Additionally, we systematically searched for information on harms related

to the various imaging modalities of interest (regardless of condition or disease state). Our search strategy is shown in Appendix A.

Our inclusion criteria for the supplemental harms searches were—

1. Articles must have been published in English.
2. Articles must have specifically focused on adverse events from ERUS, CT, MRI, or PET/CT, but any patient population or disease was acceptable.
3. Clinical studies had to be published in 2008 or later (to include the most current literature only).
4. Narrative reviews had to be published in 2012 or later.

### Data Abstraction

We abstracted data using the database DistillerSR (Evidence Partners Incorporated, Ottawa, Canada). Data abstraction forms were constructed in Distiller, and we extracted the data into these forms. Duplicate abstraction was used to ensure accuracy.

Elements that were abstracted include general study characteristics, patient characteristics, details of the imaging methodology, risk-of-bias items, and outcome data.

### Individual Study Quality (Risk-of-Bias) Evaluation

We used internal validity rating instruments to evaluate the risk of bias of each individual study. The instruments are shown in Appendix D. Studies were rated as low, medium, or high risk of bias. The ratings were defined by selecting critical questions from a rating scale that must be answered “yes.” We selected the critical questions for these ratings for the review after discussions with the Technical Expert Panel.

As suggested by the “Methods Guide for Effectiveness and Comparative Effectiveness Reviews” (Methods Guide), systematic reviews used to address Key Questions 1a and 2a were evaluated for risk of bias with a modified AMSTAR instrument,<sup>6</sup> which is shown in Appendix C. Systematic reviews were rated as either high quality or not. The rating was defined by selecting critical questions from the rating scale that must be answered “yes.” The critical questions for these ratings for the review were selected after discussions with the Technical Expert Panel. Only high-quality systematic reviews were included to address Key Questions 1a and 2a.

### Strength-of-Evidence Grading

We used a formal grading system that conforms with the Methods Guide recommendations on grading the strength of evidence.<sup>7-9</sup>

The overall strength of evidence supporting each major conclusion was graded as high, moderate, low, or insufficient. The grade was developed by considering four important domains: study limitations (based on the risk of bias of the individual studies addressing a question), consistency of the findings, precision of the results, and directness of the evidence. The grades are defined as follows:

- *High.* We are very confident that the estimate of effect lies close to the true effect for this outcome. The body of evidence has few or no deficiencies. We believe that the findings are stable—that is, another study would not change the conclusions.
- *Moderate.* We are moderately confident that the estimate of effect lies close to the true effect for this outcome. The body of evidence has some deficiencies. We believe that the findings are likely to be stable, but some doubt remains.
- *Low.* We have limited confidence that the estimate of effect lies close to the true effect for this outcome. The body of evidence has major or numerous deficiencies (or both). We believe that additional evidence is needed before concluding either that the findings are stable or that the estimate of effect is close to the true effect.
- *Insufficient.* We have no evidence, we are unable to estimate an effect, or we have no confidence in the estimate of effect for this outcome. No evidence is available or the body of evidence has unacceptable deficiencies, precluding reaching a conclusion.

We did not grade the strength of evidence from published systematic reviews on the accuracy of individual imaging tests.

### Applicability

The applicability of the evidence involves four key aspects: patients, tests/interventions, comparisons, and settings. After discussions with the Technical Expert Panel, we concluded that age and sex of patients are unlikely to affect staging accuracy, but other patient characteristics, such as race, obesity, genetic syndromes predisposing to colorectal cancer, and enrollment of populations with high rates of comorbid conditions, could affect the applicability of study findings, particularly with regard to patient-

oriented outcomes. To improve the applicability of the findings regarding specific tests and comparisons, we excluded obsolete and experimental imaging tests.

## Data Analysis and Synthesis

For questions addressing individual test performance (accuracy), we used evidence from earlier systematic reviews. As recommended by the Methods Guide, we summarized all relevant high-quality reviews.<sup>6</sup> (See above for a definition of high-quality systematic reviews.)

For comparative questions, we synthesized the evidence from the primary studies themselves. We performed meta-analysis wherever appropriate and possible. Decisions about whether meta-analysis was appropriate were based on the judged clinical homogeneity of the different study populations, imaging and treatment protocols, and outcomes. When meta-analysis was not possible (because of limitations of reported data) or was judged to be inappropriate, the data were synthesized using a descriptive approach.

Consistency of the evidence was assessed by considering study populations, imaging and treatment protocols, study designs, and outcomes, in addition to statistical heterogeneity. We rated the consistency of conclusions supported by random-effects meta-analyses with the statistic  $I^2$ . For qualitative comparisons, we rated conclusions as consistent if the effect sizes were all in the same direction.

For studies of clinical outcomes and analyses of accuracy, overstaging, and understaging, we computed effect sizes (odds ratios [ORs] of making errors) and measures of variance using standard methods and performed DerSimonian and Laird random-effects meta-analyses using Comprehensive Meta-Analysis (CMA) software (Biostat, Inc., Englewood, NJ). Because the same patients underwent both tests being compared and studies did not report the correlations among tests, we assumed a correlation of 0.5 and performed sensitivity analyses using correlations of 0.1 and 0.9.

To analyze diagnostic test characteristics, the data must first be dichotomized. For N staging, dichotomization is straightforward: the lymph nodes are affected (N1, N2) or are not affected (N0). For M staging, the situation is similar. For T staging, dichotomization is not as straightforward; however, after considering the clinical situation, a clinically relevant dichotomization is apparent: groups T1/T2 together and T3/T4 together. This dichotomization is clinically relevant because treatment of T1/T2 colorectal cancer is similar, treatment

of T3/T4 is similar, and treatment of T1/T2 versus T3/T4 is very different. After dichotomization, for studies of test performance (sensitivity, specificity), we meta-analyzed the data reported by the studies using a binomial-bivariate random-effects regression model, as described by Harbord et al.<sup>10</sup> All such analyses were computed by the STATA 13.0 statistical software package using the `metandi` command.<sup>11</sup> In cases in which a bivariate binomial regression model could not be fit to the available data, we meta-analyzed the diagnostic data using a random-effects model and the software package Meta-Disc (freeware developed by the Unit of Clinical Biostatistics, Ramón y Cajal Hospital, Madrid, Spain).<sup>12</sup>

Wherever possible, we performed calculations of standard diagnostic test characteristics (sensitivity, specificity) and also calculations of accuracy, understaging, and overstaging. If the two different approaches to analysis produced different conclusions about which test is to be preferred for that situation, the data were categorized as inconsistent/heterogeneous.

We explored possible causes of heterogeneity with subgroup analysis. Covariates included population descriptors, tumor site and type, country and setting of care, variations in imaging technology, and publication date.

## Peer Review and Publication

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

## Results

### Evidence Base

The literature searches identified 4,683 citations. After review of the abstracts of these articles in duplicate, 4,473 were excluded. The most common reason for exclusion was lack of relevancy to the questions. Some of the excluded narrative reviews and patterns-of-care articles were used to inform the background section and the patterns-of-care discussion in the final chapter of the full report. In all, 210 articles were retrieved in full: 31 were screened against the systematic review inclusion criteria, and 179 were screened against the clinical study inclusion criteria. See the Methods section for lists of the

inclusion criteria. After screening the articles in duplicate, we included 8 systematic reviews and 65 primary clinical studies. See Appendix B for a list of the excluded studies.

**Key Question 1. What is the comparative effectiveness of imaging techniques for pretreatment cancer staging in patients with primary and recurrent colorectal cancer?**

**Key Question 1a. What is the test performance of the imaging techniques used (singly, in combination, or in a specific sequence) to stage colorectal cancer compared with a reference standard?**

Seven recent (2009 or later) high-quality systematic reviews and 38 primary comparative studies met the inclusion criteria for this question. We compiled data from the recent high-quality systematic reviews to estimate the accuracy of each individual imaging modality in isolation. These data are summarized in Table A. One of the seven systematic reviews evaluated only a particular type of ERUS (miniproboscopes), so we did not include information from it in Table A due to concerns about generalizability. Because there were insufficient data on PET/CT from systematic reviews, we examined the studies of PET/CT addressing the comparative questions in this report to obtain an estimate of accuracy.

**Table A. Accuracy of imaging tests as reported by recent systematic reviews**

Staging	ERUS	CT	MRI	PET/CT
Rectal T	<p><u>For identifying T1:</u> Sensitivity: 87.8% (85.3% to 90.0%) Specificity: 98.3% (97.8% to 98.7%)</p> <p><u>For identifying T2:</u> Sensitivity: 80.5% (77.9% to 82.9%) Specificity: 95.6% (94.9% to 96.3%)</p> <p><u>For identifying T3:</u> Sensitivity: 96.4% (95.4% to 97.2%) Specificity: 90.6% (89.5% to 91.7%)</p> <p><u>For identifying T4:</u> Sensitivity: 95.4% (92.4% to 97.5%) Specificity: 98.3% (97.8% to 98.7%)</p>	<p><u>For distinguishing T1/T2 from T3/T4:</u> Sensitivity: 86% (78% to 92%) Specificity: 78% (71% to 84%)</p>	<p><u>For distinguishing T1/T2 from T3/T4:</u> Sensitivity: 87% (81% to 92%) Specificity: 75% (68% to 80%)</p> <p><u>For identifying affected CRM:</u> Sensitivity: 77% (57% to 90%) Specificity: 94% (88% to 97%)</p>	Not reported
Rectal N	<p><u>For identifying affected nodes:</u> Sensitivity: 73.2% (70.6% to 75.6%) Specificity: 75.8% (73.5% to 78.0%)</p>	<p><u>For identifying affected nodes:</u> Sensitivity: 70% (59% to 80%) Specificity: 78% (66% to 86%)</p>	<p><u>For identifying affected nodes:</u> Sensitivity: 77% (69% to 84%) Specificity: 71% (59% to 81%)</p>	<p><u>For identifying affected nodes:</u> Sensitivity: 61% Specificity: 83%</p>
Colorectal T	Not reported	Not reported	Not reported	Accuracy: 95%

**Table A. Accuracy of imaging tests as reported by recent systematic reviews (continued)**

Staging	ERUS	CT	MRI	PET/CT
Colorectal N	Not reported	Not reported	Not reported	<u>For identifying affected nodes:</u> Sensitivity: 34.3% Specificity: 100%
Colorectal M	Not indicated	<u>For identifying liver metastases:</u> Sensitivity: 83.6%	<u>For identifying liver metastases:</u> Sensitivity: 88.2%	<u>For identifying liver metastases:</u> Sensitivity: 72.0% to 97.9%

Note: The 95% confidence intervals are shown in parentheses.

CRM = circumferential resection margin; CT = computed tomography; ERUS = endorectal ultrasound; M = metastases stage; MRI = magnetic resonance imaging; N = nodal stage; PET/CT = positron emission tomography/computed tomography; T = tumor stage.

To determine the comparative effectiveness of the different modalities, we examined studies that directly compared modalities with each other and verified the results with a reference standard (usually histopathology/intraoperative findings).

We identified 23 studies of preoperative rectal T staging. Six studies compared MRI with ERUS, 13 compared

CT with ERUS, 3 compared MRI with CT, and 1 study compared CT, MRI, and ERUS. If possible, we fit a binomial-bivariate normal regression model to diagnostics accuracy data, and we performed random-effects meta-analyses on the measures of accuracy, overstaging, and understaging. The results of our calculations are shown in Table B.

**Table B. Summary results for primary preoperative rectal T staging**

Test Characteristics	MRI vs. ERUS	ERUS vs. CT	MRI vs. CT
Sensitivity (95% CI) of T1/T2 vs. T3/T4	<u>MRI</u> : 88.9% (79.0% to 94.4%) <u>ERUS</u> : 88.0% (80.0% to 93.1%)	Not calculated due to insufficient data reported	Not calculated
Specificity (95% CI) of T1/T2 vs. T3/T4	<u>MRI</u> : 85.3% (70.6% to 93.4%) <u>ERUS</u> : 85.6% (65.8% to 94.9%)	Not calculated due to insufficient data reported	Not calculated
Accuracy: OR of getting an incorrect result (95% CI) <sup>a</sup>	1.24 (0.835 to 1.84)	0.359 (0.238 to 0.541)	0.317 (0.056 to 1.784) <sup>b</sup>
Understaging OR (95% CI) <sup>a</sup>	1.571 (0.605 to 4.083)	0.626 (0.438 to 0.894)	0.317 (0.027 to 3.646) <sup>b</sup>
Overstaging OR (95% CI) <sup>a</sup>	1.05 (0.518 to 2.16)	0.472 (0.28 to 0.798)	0.317 (0.028 to 3.653) <sup>b</sup>
Favors	No statistically significant difference	ERUS	No statistically significant difference

<sup>a</sup>OR < 1 indicates a lower risk of error in the first imaging modality listed in the column header; OR > 1 indicates a higher risk of error in the first imaging modality listed in the column header.

<sup>b</sup>Study with 0.15T magnet excluded from analyses.

CI = confidence interval; CT = computed tomography; ERUS = endorectal ultrasound; MRI = magnetic resonance imaging; OR = odds ratio; T = tumor stage.

We identified 19 studies that reported data on rectal N staging. One study compared MRI with PET/CT, five compared MRI with ERUS, nine compared CT with ERUS, and four compared MRI with CT. If possible, we fit a binomial-bivariate normal regression model to diagnostics accuracy data, and we performed random-

effects meta-analyses on the measures of accuracy, overstaging, and understaging. The results of our calculations are shown in Table C. The MRI versus PET/CT comparison (single study) was not statistically significant (0.467; confidence interval [CI], 0.193 to 1.130).

**Table C. Summary results for primary preoperative rectal N staging**

Test Characteristics	MRI vs. ERUS	CT vs. ERUS	MRI vs. CT
Sensitivity (95% CI)	<u>MRI</u> : 49.5% (36.0% to 63.1%) <u>ERUS</u> : 53.0% (39.7% to 65.5%)	<u>CT</u> : 39.6% (28.1% to 52.4%) <u>ERUS</u> : 49.1% (34.9% to 63.5%)	Not calculated
Specificity (95% CI)	<u>MRI</u> : 69.7% (51.9% to 83.0%) <u>ERUS</u> : 73.7% (43.6% to 91.0%)	<u>CT</u> : 93.2% (58.8% to 99.2%) <u>ERUS</u> : 71.7% (56.2% to 83.4%)	Not calculated
Accuracy: OR of getting an incorrect result (95% CI) <sup>a</sup>	0.882 (0.542 to 1.408)	1.13 (0.85 to 1.503)	1.316 (0.709 to 2.443)
Understaging OR (95% CI) <sup>a</sup>	0.972 (0.563 to 1.679)	1.453 (0.854 to 2.473)	1.743 (1.028 to 2.957); not robust in sensitivity analysis
Overstaging OR (95% CI) <sup>a</sup>	0.752 (0.457 to 1.237)	1.015 (0.571 to 1.801)	0.498 (0.308 to 0.806)
Favors	No statistically significant difference	No statistically significant difference	MRI favored for avoiding overstaging

<sup>a</sup>OR < 1 indicates a lower risk of error in the first imaging modality listed in the column header; OR > 1 indicates a higher risk of error in the first imaging modality listed in the column header.

CI = confidence interval; CT = computed tomography; ERUS = endorectal ultrasound; MRI = magnetic resonance imaging; N = nodal stage; OR = odds ratio.

We identified nine studies of preoperative colorectal M staging. Four compared PET/CT with CT, and five compared MRI with CT. Where possible, we fit a binomial-bivariate normal regression model to diagnostics accuracy data, and we performed random-effects meta-analyses on the measures of accuracy, overstaging, and understaging. The results of our calculations are shown

in Table D. The statistical heterogeneity of the PET/CT data makes it difficult to draw any conclusions about the comparison with CT, and in fact, the conclusions drawn by the individual study authors ranged from no difference, to superiority of CT, to superiority of PET/CT for this purpose.

**Table D. Pooled random-effects meta-analyses of preoperative colorectal M staging (per lesion basis)**

Measure	MRI vs. CT	PET/CT vs. CT
Sensitivity	Not calculated	CT: 83.6% (95% CI, 78.1% to 88.2%) PET/CT: 60.4% (95% CI, 53.7% to 66.9%)
Summary OR for lesion detection <sup>a</sup>	1.334 (95% CI, 1.012 to 1.761)	Not calculated
I <sup>2</sup>	12.4%	CT: 0.0% PET/CT: 95.1%
Favors	MRI	Insufficient evidence

<sup>a</sup>OR > 1 indicates a higher likelihood of detecting metastatic lesions by the first imaging modality listed in the column header. CI = confidence interval; CT = computed tomography; M = metastases stage; MRI = magnetic resonance imaging; OR = odds ratio; PET/CT = positron emission tomography/computed tomography.

We identified only one study each of preoperative circumferential resection margin (CRM) status (MRI vs. CT) and colorectal T staging (CT vs. PET/CT).

We did not identify any studies of staging enrolling only patients who had colon cancer (i.e., results not combined with those for patients who had rectal cancer) that met the inclusion criteria.

**Key Question 1b. What is the impact of alternative imaging techniques on intermediate outcomes, including stage reclassification and changes in therapeutic management?**

We identified seven primary comparative studies that addressed this question.

Two studies 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.

Two studies that met the inclusion criteria reported the impact of adding PET/CT results to CT results for preoperative staging of colorectal cancer. One study did not measure whether the changes were appropriate. The other study reported that adding PET/CT to CT results changed management for 17.5 percent of patients, but after treatment, surgery, and followup, results indicated that only half of the changed treatment plans were the appropriate choice.

Two studies that met the inclusion criteria reported the impact of adding ERUS information to CT results, and one study reported the impact of adding PET/CT to MRI and CT for preoperative staging of rectal cancer. However, none of these studies verified whether the changes were appropriate.

**Key Question 1c. What is the impact of alternative imaging techniques on clinical outcomes?**

We did not identify any studies that addressed this question.

**Key Question 1d. What are the adverse effects or harms associated with using imaging techniques, including harms of test-directed management?**

To address this question, we abstracted data about harms reported by the included studies to address the questions on comparative accuracy in this report. We supplemented this information with information from narrative reviews and other sources (e.g., FDA alerts). Additionally, we systematically searched for information on harms related

to the various imaging modalities of interest (regardless of condition or disease state). Our search strategy is shown in Appendix A. Our supplemental searches identified 1,961 abstracts; after review of these abstracts, we selected 66 articles to review in full text, of which 32 were selected for inclusion.

Ultrasound is generally considered to be extremely safe. For rectal imaging, an additional consideration is the fact that an endorectal probe is used; the probe is inserted into the rectum. Possible complications include perforation, bleeding, and pain. The majority of included studies of ERUS did not report any complications; whether this means that none occurred is unclear. Six studies reported adverse events such as pain and minor rectal bleeding. Four studies reported failure to complete the procedure because of stenosis or strictures. No studies reported any cases of perforation.

The supplemental harms searches identified one review of endoscopic ultrasound–related adverse events that included information on complications of ERUS. The authors reported that a large multicenter prospective German registry of endoscopic ultrasound procedures reported one perforation related to ERUS.

None of the included studies reported any adverse events related to CT or PET/CT. The supplemental harms searches identified reports of reactions to intravenous contrast agents. CT and PET/CT scans also expose the body to x rays. A typical abdominal CT scan exposes the body to approximately 10 mSv of radiation, and a typical PET/CT scan exposes the body to 18 mSv.

Only two of the included studies reported adverse events due to MRI, and both were reports of patients refusing the procedure because of severe claustrophobia. The supplemental harms searches identified the possibility of adverse events due to intravenous contrast agents, such as allergic reactions and nephrogenic systemic fibrosis, a scleroderma-like fibrosing condition that occurs in patients with renal failure and can be fatal. Labeling for gadolinium-based contrast agents now includes a warning regarding the risk of nephrogenic systemic fibrosis in patients with severe kidney insufficiency, patients just before or just after liver transplantation, or individuals with chronic liver disease.

### **Key Question 1e. How is the comparative effectiveness of imaging techniques modified by the following factors:**

- i. Patient-level characteristics (e.g., age, sex, body mass index)?**
- ii. Disease characteristics (e.g., tumor grade)?**
- iii. Imaging technique or protocol characteristics (e.g., use of different tracers or contrast agents, radiation dose of the imaging modality, slice thickness, timing of contrast)?**

We identified 16 primary comparative studies that addressed this question.

Nine studies reported factors affecting MRI’s accuracy for colorectal staging. Most of these studies reported on different factors; however, three studies reported that contrast enhancement did not improve MRI’s accuracy for rectal T and N staging.

Five studies reported factors affecting the accuracy of ERUS for colorectal staging, and three studies reported factors affecting CT’s accuracy for colorectal staging, but they reported on different factors, making it difficult to determine how any specific factors impact accuracy.

### **Conclusions for Key Question 1**

For rectal T staging, ERUS and MRI appear to not be statistically significantly different in accuracy, and ERUS is more accurate than CT. There were no statistically significant differences in accuracy between MRI and CT for rectal T staging. The evidence was insufficient for drawing conclusions about the accuracy of PET/CT compared with either MRI or CT for rectal T staging.

For rectal N staging, ERUS, MRI, and CT are not significantly different in accuracy, but they all have low sensitivity for detecting affected lymph nodes. MRI is less likely to overstage and CT may be less likely to understage N status (although the latter conclusion was not robust in sensitivity analyses). The evidence was insufficient for drawing conclusions about the accuracy of PET/CT compared with either MRI or CT for rectal N staging.

For detecting colorectal liver metastases, MRI is superior to CT. The evidence was insufficient for drawing conclusions about the accuracy of PET/CT compared with either MRI or CT for colorectal M staging.

The evidence base is characterized by a lack of studies reporting patient-oriented outcomes. Seven studies reported on the impact of imaging on patient management, but only three of these studies confirmed whether the change in management was appropriate. In general, the included studies reported only on diagnostic accuracy. They were all rated as either low or moderate risk of bias.

A systematic review published in 2005 (thus not included to address the Key Questions) concluded that “the performance of EUS [endoscopic ultrasound] in staging rectal cancer may be overestimated in the literature due to publication bias.”<sup>13</sup> The review included 41 studies published between 1985 and 2003. The author, Harewood, performed visual analyses of funnel plots and other diagrams, demonstrating that it appeared that few smaller studies found lower accuracy rates for ERUS and that the reported accuracy appeared to be declining over time. Studies published in the surgical literature reported higher accuracies than studies published in other types of journals.<sup>13</sup>

Puli et al. also analyzed the reported accuracy of ERUS over time and found that the reported accuracy had declined significantly from the 1980s through 2000 and had stabilized or only declined slightly since then.<sup>14</sup> Puli also stated that he found no evidence of publication bias in the ERUS literature in 2009.<sup>14</sup>

Nielke et al. reported no evidence of publication bias for M staging with CT,<sup>15</sup> but Dighe et al. reported that, for N staging with CT, evidence existed that smaller studies were reporting higher accuracies (suggesting publication bias), and a nonsignificant trend showed the same result for T staging.<sup>16</sup>

Nielke et al. reported that the MRI staging literature contained no evidence of publication bias.<sup>15</sup>

Too few studies are available for most of the evidence bases in this review to allow a statistical analysis of the possibility of publication bias. However, because of reports that the ERUS literature, in particular, may be affected by publication bias, we prepared funnel plots for the two larger ERUS evidence bases and also ran a metaregression against publication date. Although visual inspection of funnel plots is of limited value in determining the presence of publication bias, the plots look fairly symmetrical, and there does not appear to be any pattern by date in the ERUS-versus-CT evidence base. There may be a tendency to report higher accuracy in older studies in the MRI-versus-ERUS evidence base, but the number of studies in that evidence base is too small to allow us to reach any firm conclusion.

## **Key Question 2. What is the comparative effectiveness of imaging techniques for restaging cancer in patients with primary and recurrent colorectal cancer after initial treatment?**

### **Key Question 2a. What is the test performance of the imaging techniques used (singly, in combination, or in a specific sequence) to restage colorectal cancer compared with a reference standard?**

As noted previously, interim restaging takes place after neoadjuvant chemotherapy and/or radiotherapy and, in some cases, surgery. We identified only one recent (2009 or later) high-quality systematic review of interim restaging. Therefore, we searched for older high-quality systematic reviews of interim restaging but did not identify any that met the inclusion criteria. We identified nine primary comparative studies of interim restaging.

The one systematic review of interim restaging studied CT, MRI, and PET/CT for detecting liver metastases after neoadjuvant chemotherapy. The review authors concluded that MRI was more sensitive for this purpose than the other two modalities, but even for MRI the sensitivity was very low, possibly too low to be clinically useful (69.9%; 95% CI, 65.6% to 73.9%).

We identified four studies of interim rectal T staging. One study compared CT with MRI, one compared CT with ERUS, and two compared MRI, ERUS, and CT. Considering all the evidence in a qualitative fashion, the evidence seems to consistently support the conclusion that no significant difference in accuracy exists across ERUS, CT, and MRI for interim rectal T staging.

We identified three studies of interim rectal N staging. One study compared ERUS with CT, and two studies compared ERUS, CT and MRI. There were no statistically significant differences across the modalities, but there was a nonsignificant trend for ERUS to be more accurate than MRI and CT and for MRI to be more accurate than CT.

We identified four studies of interim colorectal M staging. Three compared MRI with CT, and one compared PET/CT with CT. We pooled the data reported by the three studies of MRI compared with CT for detecting liver metastases in a random-effects meta-analysis. The results indicated a nonsignificant trend toward MRI being more accurate in detecting colorectal liver metastases than CT.

No studies meeting inclusion criteria reported on interim colon cancer restaging separately (i.e., without mixing rectal cancer cases into the enrolled group), and no studies identified interim colorectal T and N restaging or interim

rectal M restaging. We identified only one study of interim rectal CRM status.

**Key Question 2b. What is the impact of alternative imaging techniques on intermediate outcomes, including stage reclassification and changes in therapeutic management?**

No studies that met the inclusion criteria addressed this question.

**Key Question 2c. What is the impact of alternative imaging techniques on clinical outcomes?**

No studies that met the inclusion criteria addressed this question.

**Key Question 2d. What are the adverse effects or harms associated with using imaging techniques, including harms of test-directed management?**

See the answer to Key Question 1d for harms associated with any use of these imaging tests.

**Key Question 2e. How is the comparative effectiveness of imaging techniques modified by the following factors:**

- i. Patient-level characteristics (e.g., age, sex, body mass index)?
- ii. Disease characteristics (e.g., tumor grade)?
- iii. Imaging technique or protocol characteristics (e.g., use of different tracers or contrast agents, radiation dose of the imaging modality, slice thickness, timing of contrast)?

Only one study of MRI reported on factors affecting accuracy of interim N restaging, and only one study of MRI reported on factors affecting accuracy of interim M restaging.

## **Conclusions for Key Question 2**

The one included systematic review reported that CT and PET/CT had sensitivities of approximately 50 percent for detecting colorectal liver metastases in the interim restaging setting, and MRI's sensitivity in this setting, although slightly better, is still quite low (69.9%; 95% CI, 65.6% to 73.9%).

We found no significant difference in accuracy across ERUS, CT, and MRI for interim rectal T staging and a nonsignificant trend for MRI to be more accurate than CT for detecting colorectal liver metastases during restaging.

The primary conclusion to be reached for Key Question 2 is that there are gaps in the research that has been published. The evidence base is small and limited. Only 10 studies addressed Key Question 2, all of which were rated as being at low to moderate risk of bias. The risk-of-bias rating by key factors is provided in Appendix D. There were too few studies to allow assessment of the possibility of publication bias.

## **Discussion**

### **Key Findings and Strength of Evidence**

Our major conclusions about comparative effectiveness are listed in Table E, along with the strength-of-evidence grade. We have moderate confidence in one conclusion and low confidence in several other conclusions, but the evidence was insufficient for the majority of the questions posed in this review.

**Table E. Summary of major conclusions**

Conclusion Statement	Strength of Evidence
ERUS is less likely to give an incorrect result (OR = 0.36; 95% CI, 0.24 to 0.54), less likely to understage (OR = 0.63; 95% CI, 0.44 to 0.89), and less likely to overstage (OR = 0.47; 95% CI, 0.28 to 0.80) rectal cancer than CT in the preoperative T staging setting. <sup>a</sup>	Low
MRI and ERUS are similar in accuracy for preoperative rectal T staging.	Low
CT, MRI, and ERUS are similar in accuracy for preoperative rectal N staging. MRI is less likely than CT to overstage (OR = 0.498; 95% CI, 0.308 to 0.806).	Low
MRI is superior (more likely to detect lesions) to CT in detecting colorectal liver metastases in the preoperative setting (OR = 1.334; 95% CI, 1.012 to 1.761). <sup>b</sup>	Moderate
MRI, CT, and ERUS are similar in accuracy for rectal T staging in the interim restaging setting.	Low

<sup>a</sup>OR < 1 indicates a lower risk of error; OR > 1 indicates a higher risk of error.

<sup>b</sup>OR > 1 indicates a higher likelihood of detecting metastatic lesions.

CI = confidence interval; CT = computed tomography; ERUS = endorectal ultrasound; MRI = magnetic resonance imaging; N = nodal stage; OR = odds ratio; T = tumor stage.

For harms, in general, all four imaging modalities appear to be reasonably safe. For ERUS, the most common adverse event appears to be pain and minor bleeding; in theory, the major adverse event of bowel perforation could occur, but no included studies reported such an event. Our supplementary harms searches identified a narrative review of complications of endoscopic ultrasound, including ERUS.<sup>17</sup> The authors noted that only one case had been reported in a prospective registry of the German Society of Ultrasound in Medicine but did not report the number of ERUS procedures in the registry.

Harms from CT include contrast agent reactions and radiation exposure. Many of the included studies did not use intravenous contrast, and limited data suggest that using intravenous contrast does not improve the accuracy of CT for colorectal T or N staging. Not surprisingly, there were no studies comparing M staging by CT with and without contrast.

Harms from MRI appear to be limited to contrast agent reactions. Many of the included studies did not use intravenous contrast, and data suggest that the use of intravenous contrast does not improve MRI's accuracy for rectal T or N staging.

The major harm from PET/CT is radiation exposure. A single PET/CT examination exposes the patient to approximately 18 mSv, with the majority coming from

the radiotracer for the PET component. Some experts believe this is a significant exposure; however, in 2010, the Health Physics Society published a position statement recommending against quantitative estimates of health risks below an individual dose of 5 rem per year (approximately 50 mSv) or a lifetime dose of 10 rem in addition to natural background radiation.<sup>18</sup> However, if a patient undergoes a PET/CT scan for staging, has surgical treatment, and then has regular CT scans for surveillance, the accumulated radiation dose could approach or exceed these limits.

Indirect harms of imaging primarily consist of harms related to incorrect treatment decisions based on inaccurate staging.

### Limitations of the Evidence Base

The evidence base is quite limited. Very few studies reported on any outcomes other than staging accuracy. Among studies reporting only accuracy outcomes, we did not find complete cross-classified data (i.e., numbers of patients correctly staged, understaged, and overstaged for each stage for all modalities and the reference standard). Many of the studies that reported on staging accuracy were quite small and provided limited information on patient characteristics. In particular, the evidence base for Key Question 2, interim restaging, is very sparse even for staging accuracy outcomes.

A few studies reported on how imaging modalities affected patient management, but few of these reported whether management changes were deemed appropriate. No studies reported on patient-oriented outcomes such as survival and quality of life.

### **Applicability**

Judging the applicability of the results is difficult. The majority of studies reported very little information about patient characteristics. Most of the studies were set in university-based academic or teaching hospitals, which may limit the applicability of the results to community-based general hospitals. Another area of concern about applicability is the inclusion of many older studies that may have used technology that is now obsolete. During the topic refinement process, experts agreed that using an arbitrary publication cutoff date would introduce bias, so our literature searches went back to 1980.

### **Research Gaps**

The majority of the evidence gaps on the questions in this review fall into the category of insufficient information.

There is practically no literature on interim restaging of either colon or rectal cancer, and very few studies of staging of colon cancer; most of the literature identified was about rectal cancer. This likely reflects the relatively greater importance of clinical locoregional staging in rectal versus colon cancer. Specifically, most studies of staging in colon cancer seemed to focus on looking for metastases, particularly to the liver.

Few studies examined the impact of combining different imaging modalities on pretreatment and interim staging assessments, which may provide more clinically relevant results than studies that examine the accuracy of one imaging modality in isolation. Given that patients often undergo multiple imaging studies for staging purposes, such information would be valuable.

Few studies addressed variations in imaging protocols that could affect study accuracy. Reviewers pointed out particular interest in factors that could affect accuracy of ERUS, such as the types of probes used and the experience of the individual performing the examination.

Very few studies of PET/CT are available; this is a concern because many experts appear to believe its addition to staging leads to useful changes in management. Also, its use for primary and interim clinical staging of patients is on the rise, despite the lack of convincing evidence to support its widespread adoption. We identified one study

of changes in management after addition of PET/CT that concluded that only half of the changes in management triggered by PET/CT were appropriate, suggesting that using PET/CT for staging may result in significant patient harm.<sup>19</sup> Further study on this topic needs to be performed before any firm conclusions about the accuracy and clinical usefulness of PET/CT can be drawn.

Not having the right information is another consideration. Insufficient information is available about changes in management triggered by imaging studies and about patient-oriented outcomes downstream of staging. Ideally, randomized controlled trials would be designed to test different staging and management strategies, capturing health outcomes that occur following treatment.

Studies of the impact of imaging on patient management decisions are potentially helpful and can be accomplished in shorter timeframes than studies measuring health outcomes. However, it is critical to confirm whether the changes in management were appropriate; simply reporting that adding information from an imaging modality led to changes in management is insufficient information to be clinically useful.

### **Conclusions**

Low-strength evidence suggests ERUS is more accurate than CT for preoperative rectal cancer T staging and MRI is similar in accuracy to ERUS. Moderate-strength evidence suggests MRI is superior to CT for detecting colorectal liver metastases. There was insufficient evidence to come to any evidence-based conclusions about the use of PET/CT for colorectal cancer staging. Low-strength evidence suggests that CT, MRI, and ERUS are comparable for rectal cancer N staging, but all are limited in sensitivity. Low-strength evidence suggests that they are also comparable for interim rectal cancer T restaging, but both sensitivity and specificity are suboptimal. While all four imaging modalities appear to be reasonably safe, long-range harm from radiation exposure over repeated examinations is particularly of concern with PET/CT.

### **References**

1. American Cancer Society. Cancer Facts & Figures 2012. Atlanta; 2012. [www.cancer.org/acs/groups/content/@epidemiologysurveillance/documents/document/acspc-031941.pdf](http://www.cancer.org/acs/groups/content/@epidemiologysurveillance/documents/document/acspc-031941.pdf).
2. American Cancer Society. Colorectal cancer Facts & Figures 2011-2013. Atlanta; 2011. [www.cancer.org/acs/groups/content/@epidemiologysurveillance/documents/document/acspc-028323.pdf](http://www.cancer.org/acs/groups/content/@epidemiologysurveillance/documents/document/acspc-028323.pdf).

3. Yabroff KR, Lund J, Kepka D, et al. Economic burden of cancer in the United States: estimates, projections, and future research. *Cancer Epidemiol Biomarkers Prev.* 2011 Oct;20(10):2006-14. PMID: 21980008.
4. Mariotto AB, Yabroff KR, Shao Y, et al. Projections of the cost of cancer care in the United States: 2010-2020. *J Natl Cancer Inst.* 2011 Jan 19;103(2):117-28. PMID: 21228314.
5. U.S. Preventive Services Task Force. Screening for Colorectal Cancer. March 2009. [www.uspreventiveservicestaskforce.org/uspstf/uspscolo.htm](http://www.uspreventiveservicestaskforce.org/uspstf/uspscolo.htm). Accessed January 23, 2014.
6. White CM, Ip S, McPheeters M, et al. Using existing systematic reviews to replace de novo processes in conducting Comparative Effectiveness Reviews. Sept. 2009. In: *Methods Guide for Effectiveness and Comparative Effectiveness Reviews*. AHRQ Publication No. 10(14)-EHC063-EF. Rockville, MD: Agency for Healthcare Research and Quality. January 2014. Chapters available at [www.effectivehealthcare.ahrq.gov](http://www.effectivehealthcare.ahrq.gov). PMID: 21433402.
7. Owens DK, Lohr KN, Atkins D, et al. Grading the strength of a body of evidence when comparing medical interventions-Agency for Healthcare Research and Quality and the Effective Health Care Program. *J Clin Epidemiol.* 2010 May;63(5):513-23. PMID: 19595577.
8. Owens D, Lohr K, Atkins D, et al. Grading the strength of a body of evidence when comparing medical interventions. July 2009. In: *Methods Guide for Effectiveness and Comparative Effectiveness Reviews*. AHRQ Publication No. 10(14)-EHC063-EF. Rockville, MD: Agency for Healthcare Research and Quality. January 2014. Chapters available at [www.effectivehealthcare.ahrq.gov](http://www.effectivehealthcare.ahrq.gov).
9. *Methods Guide for Medical Test Reviews*. AHRQ Publication No. 12-EHC017. Rockville, MD: Agency for Healthcare Research and Quality; June 2012. [www://effectivehealthcare.ahrq.gov/reports/final.cfm](http://www.effectivehealthcare.ahrq.gov/reports/final.cfm).
10. Harbord RM, Deeks JJ, Egger M, et al. A unification of models for meta-analysis of diagnostic accuracy studies. *Biostatistics.* 2007 Apr;8(2):239-51. PMID: 16698768.
11. STATA Statistics/Data Analysis. MP Parallel Edition. College Station, TX: StataCorp; 1984-2007. Single user Stata for Windows. [www.stata.com](http://www.stata.com).
12. Zamora J, Abaira V, Muriel A, et al. Meta-DiSc: a software for meta-analysis of test accuracy data. *BMC Med Res Methodol.* 2006;6:31. PMID: 16836745.
13. Harewood GC. Assessment of publication bias in the reporting of EUS performance in staging rectal cancer. *Am J Gastroenterol.* 2005 Apr;100(4):808-16. PMID: 15784023.
14. Puli SR, Bechtold ML, Reddy JB, et al. How good is endoscopic ultrasound in differentiating various T stages of rectal cancer? Meta-analysis and systematic review. *Ann Surg Oncol.* 2009 Feb;16(2):254-65. PMID: 19018597.
15. Niekel MC, Bipat S, Stoker J. Diagnostic imaging of colorectal liver metastases with CT, MR imaging, FDG PET, and/or FDG PET/CT: a meta-analysis of prospective studies including patients who have not previously undergone treatment. *Radiology.* 2010 Dec;257(3):674-84. PMID: 20829538.
16. Dighe S, Purkayastha S, Swift I, et al. Diagnostic precision of CT in local staging of colon cancers: a meta-analysis. *Clin Radiol.* 2010 Sep;65(9):708-19. PMID: 20696298.
17. Jenssen C, Alvarez-Sanchez MV, Napoleon B, et al. Diagnostic endoscopic ultrasonography: assessment of safety and prevention of complication. *World J Gastroenterol.* 2012 Sep 14;18(34):4659-76. PMID: 23002335.
18. Radiation Risk in Perspective. Position Statement of the Health Physics Society. McLean, VA: Health Physics Society; July 2010. [http://hps.org/documents/risk\\_ps010-2.pdf](http://hps.org/documents/risk_ps010-2.pdf)
19. Ramos E, Valls C, Martinez L, et al. Preoperative staging of patients with liver metastases of colorectal carcinoma. Does PET/CT really add something to multidetector CT? *Ann Surg Oncol.* 2011 Sep;18(9):2654-61. PMID: 21431987.

## Full Report

This executive summary is part of the following document: 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](http://www.effectivehealthcare.ahrq.gov/reports/final.cfm).

