

Evidence-based Practice Center Systematic Review Protocol

Project Title: Diagnosis of Right Lower Quadrant Pain (Suspected Acute Appendicitis)

I. Background and Objectives for the Systematic Review

Nature and burden of the condition

Abdominal pain is a common presenting complaint for patients seeking care at emergency departments, with the number of cases in the USA estimated at approximately 3.4 million per year.¹ Appendicitis is a common etiology of abdominal pain, caused by acute inflammation of the appendix, and occurs in approximately 8-10% of the population (over a lifetime).^{2,3} Appendicitis is most common between the ages of 10 and 30 years. The ratio of incidence in men and women is 3:2 through the mid-20s and then equalizes after age 30. Appendicitis is the most common abdominal surgical emergency, with over 250,000 appendectomies performed annually in the U.S. Risk for development of acute appendicitis in pregnant women is similar to that of the general population, making acute appendicitis the most common non-obstetric emergency during pregnancy.^{4,5} Untreated appendicitis can lead to perforation of the appendix, which typically occurs within 24 to 36 hours of the onset of symptoms. Perforation of the appendix can lead to intra-abdominal infection, sepsis, the formation of intraperitoneal abscesses, and rarely death (in approximately 3% of cases with perforation).⁴

Diagnosis of right lower quadrant pain/ suspected acute appendicitis

Guidelines suggest that when a diagnosis of acute appendicitis can be made on clinical grounds surgical consultation should be sought without delay for additional diagnostic testing.⁶ Several clinical signs and symptoms have been described as suggestive of appendicitis, including central abdominal pain migrating to the right iliac fossa, fever and nausea/vomiting, signs of peritoneal irritation (rebound tenderness, guarding, rigidity), and classic signs elicited by clinical examination (e.g., the McBurney, Rovsing, psoas, or obturator signs).⁷⁻⁹ The performance of clinical signs and symptoms for identifying acute appendicitis seems to be variable across studies, and few clinical findings appear to have adequate sensitivity and specificity when used in isolation.^{8,9}

For patients with right lower quadrant (RLQ) pain, when the diagnosis cannot be made on clinical grounds alone, laboratory or imaging tests are often used to attempt to establish a diagnosis and guide treatment. Laboratory evaluations potentially useful for the diagnosis of appendicitis include white blood cell count, granulocyte count, the proportion of polymorphonuclear blood cells, and C-reactive protein concentration.⁸⁻¹⁰ Imaging tests, such as ultrasound (US), computed tomography (CT) with and without contrast, and magnetic resonance imaging (MRI), are also used extensively for the diagnosis of appendicitis.¹¹⁻¹⁷ Imaging tests can be used alone or in combination. For example, US is sometimes used as a triage test to separate patients in whom sonography is adequate to establish a diagnosis from those who require further imaging with CT.⁶ Different factors may affect the performance of alternative tests and their impact on

clinical outcomes. For example, US examination is considered to be operator dependent¹⁸ and is technically challenging in obese patients or women in late pregnancy. CT scanning can be performed with or without the use of contrast agents, and contrast can be administered orally, rectally, intravenously, or via combinations of these routes.⁶ It has been suggested that low body mass index (BMI), a marker for lack of sufficient mesenteric fat (which helps visualize periappendiceal fat stranding, a radiological sign of appendicitis), may affect the relative test performance of CT performed with or without contrast (contrast being more useful in individuals with low BMI and children).⁶

Clinical signs and symptoms, along with the results of laboratory or imaging tests, can be combined into clinical prediction tools, i.e. algorithms that synthesize the findings of different investigations to determine the most likely diagnosis.¹⁹ In adults, the most commonly used clinical prediction rule for appendicitis is the Alvarado score,²⁰ which separates patients into 3 groups of increasing probability of appendicitis (the score is based on 8 items: pain migration, anorexia, nausea, tenderness in RLQ, rebound pain, elevated temperature, leukocytosis, and shift of white blood cell count to the left).²¹ The Alvarado score is also used in pediatric populations.^{22, 23} The Pediatric Appendicitis Score has also been developed and validated for use in children.²⁴ It is based on 9 items (migration of pain, anorexia, nausea/vomiting, fever, cough/percussion tenderness, hopping tenderness, RLQ tenderness, leukocytosis, polymorphonuclear neutrophilia) and classifies children into two groups (high vs. low probability of appendicitis).²⁴

Finally, diagnostic laparoscopy is also used for the evaluation of patients with RLQ pain/ suspected acute appendicitis. Although diagnostic laparoscopy is generally considered safe, studies have reported variable rates of morbidity and mortality from the procedure.^{25, 26}

In general the diagnostic tests discussed in this section are widely available in the USA. Clinical signs and symptoms can be evaluated relatively easily and inexpensively. Evidence from the National Hospital Ambulatory Medical Care Survey suggests that CT and complete blood counts are obtained in the majority of patients presenting to the emergency department with abdominal pain. The survey also showed that over time (between 1992 and 2006) the use of CT for both adults and children has been increasing. Over the same period, the use of the complete blood count has increased in adults but decreased in children.^{27, 28} The use of US and MRI is increasing in populations where exposure to ionizing radiation is a particular concern (e.g., children and pregnant women).²⁹⁻³⁷

Importance of accurate diagnosis and impact on outcomes

As with all diagnostic tests, the modalities used in the diagnostic investigation of patients with RLQ pain/suspected appendicitis affect clinical outcomes indirectly, through their impact on clinicians' diagnostic thinking and therapeutic decisionmaking.³⁸ More accurate and timely diagnosis of appendicitis can minimize the time to the indicated intervention (surgery), thus reducing pain and improving clinical outcomes (e.g., reducing bowel perforation and associated infectious complications).³⁹ Conversely, time-consuming or unnecessary imaging (or other diagnostic workup) may delay the indicated treatment and increase the risk of complications or result in false positive results and more "negative" appendectomies. Furthermore, diagnostic testing can impact resource utilization for the management of patients with acute abdominal pain. For example, examination with CT may reduce length of stay by avoiding prolonged

observation in cases where a diagnosis cannot be established clinically or by eliminating the need for additional diagnostic testing.¹⁵

Special Considerations for the Diagnosis of RLQ Pain/Acute Appendicitis

The diagnosis of appendicitis is particularly challenging in some population subgroups, including children, women of reproductive age, pregnant women, and frail or elderly patients.^{6, 40, 41}

Children: Acute appendicitis in children is often diagnosed after perforation has occurred,⁴²⁻⁴⁴ in part because children have a thinner appendiceal wall and less developed omentum (the largest peritoneal fold). Many common childhood illnesses have symptoms similar to those of early acute appendicitis (fever, nausea, and vomiting) making the differential diagnosis more challenging. Young children may have difficulty communicating about their discomfort or describing their symptoms, making the clinical examination less informative and leading to diagnostic delays.⁸ In addition, the use of modalities that involve ionizing radiation (e.g., CT) may entail additional radiation-related risks for children.⁶

Women of reproductive age: Up to a third of women of reproductive age with appendicitis are misdiagnosed.⁴⁵ Establishing a diagnosis in women of reproductive age with RLQ pain/suspected acute appendicitis can be particularly challenging because symptoms of acute appendicitis can mimic those of gynecologic disease (e.g., pelvic inflammatory disease, ectopic pregnancy, etc.).

Pregnant women: Diagnosis of suspected acute appendicitis in pregnant women can also be challenging because some symptoms of appendicitis (nausea and vomiting) are common in normal pregnancies and because enlargement of the uterus can alter the location of the appendix, which often moves higher and to the back.⁴⁶ Anatomic changes induced by pregnancy make the clinical examination of pregnant patients with abdominal pain more challenging and result in technical difficulties when using US.³⁵⁻³⁷ Tests involving ionizing radiation (e.g., CT) are also generally avoided during pregnancy.⁶ Finally, obtaining a white blood cell count is generally not helpful in the diagnosis of acute appendicitis in pregnant women because leukocytosis is common during pregnancy. From a decisionmaking perspective, the management of suspected appendicitis in pregnant women is complicated by the need to balance the potential benefits and harms of testing for both the mother and the fetus.

Frail and elderly individuals: The elderly typically present with appendicitis in more advanced stage, when compared to younger patients.⁴⁷ Older patients delay seeking care, and definitive diagnosis is sometimes delayed further because competing etiologies for abdominal pain (e.g., malignancy or diverticulitis) are considered more likely. The test performance of clinical signs and symptoms, laboratory tests, or imaging tests may be modified by patient age (e.g., US has been reported to have higher diagnostic performance in older patients) and by the more advanced disease stage that is common in this age group. Elderly and frail individuals with appendicitis have a higher complication rate and a higher risk of mortality, compared to younger/less-frail patients.

Uncertainty and the Rationale for an Evidence Review

The reliable identification of patients with RLQ pain who need surgical intervention for acute appendicitis can improve clinical outcomes and reduce resource utilization. Our review of guidelines and published systematic reviews indicates a lack of

specific guidance for selecting diagnostic modalities, particularly in patient subgroups in whom the diagnosis is known to be particularly challenging (e.g., children, women of reproductive age, and pregnant women). Existing systematic reviews do not adequately address the comparative effectiveness of alternative diagnostic approaches because they typically assess a single diagnostic modality, do not evaluate the comparative effectiveness of tests, and focus almost exclusively on test performance outcomes (without providing evidence on the impact of tests on intermediate or patient-relevant outcomes). No review has looked comprehensively at all tests of interest or focused on comparisons between alternative strategies.

A review of alternative diagnostic strategies for acute appendicitis could provide a synthesis of the evidence to help clinicians select the optimal diagnostic strategy when evaluating patients with abdominal pain with a suspected diagnosis of appendicitis. Thus, the current project could serve as a basis for clinical practice recommendations from professional societies tasked with preparing guidelines for the diagnosis of acute appendicitis.

II. The Key Questions

With input from clinical experts during Topic Refinement, we have developed the following Key Questions and study eligibility criteria to clarify the focus of the proposed systematic review. Draft Key Questions were posted for public comment (April 17 to May 14, 2013). One individual (a clinical researcher) submitted comments regarding the interventions of interest (specifically, whether clinical decision rules would be addressed by the review). We modified the selection criteria (see below) to reflect that such instruments will be within the review scope. The commenter also provided citations to potentially relevant studies. These have been retained and will be considered for inclusion using the selection criteria listed below.

Following additional discussions with technical experts, we specified the following Key Questions to be addressed by the review:

KQ 1:	<p>What is the performance of alternative diagnostic tests, alone or in combination, for patients with right lower quadrant (RLQ) pain and suspected acute appendicitis?</p> <p>a) What is the performance and comparative performance of alternative diagnostic tests in the following patient populations:</p> <ul style="list-style-type: none"> • Children • Adults • Non pregnant women of reproductive age • Pregnant women • The elderly (age ≥ 65 years) <p>b) What factors modify the test performance and comparative test performance of available diagnostic tests in these populations?</p>
KQ 2:	<p>What is the comparative effectiveness of alternative diagnostic tests, alone or in combination, for patients with RLQ pain and suspected acute appendicitis?</p> <p>a) For the populations listed under Key Question 1a, what is the effect of alternative testing strategies on diagnostic thinking, therapeutic decisionmaking, clinical outcomes, and resource utilization?</p>

KQ 3:

b) What factors modify the comparative effectiveness of testing for patients with RLQ pain and suspected acute appendicitis?

What are the harms of diagnostic tests per se, and what are the treatment-related harms of test-directed treatment for tests used to diagnose RLQ pain and suspected acute appendicitis?

Population(s)

- Patients with acute RLQ abdominal pain (≤ 7 days duration) for whom appendicitis is considered in the differential diagnosis
- Separate analyses will be performed for the following populations:
 - Children (age < 18 years); additional analyses will be performed for younger children (< 2 years and 2-5 years of age)
 - Adults (age ≥ 18 years)
 - Non pregnant women of reproductive age
 - Pregnant women
 - Elderly (age ≥ 65 years)

Interventions

- Diagnostic tests (alone or in combination) for diagnosing appendicitis
 - Clinical signs (e.g., psoas sign, obturator sign, Rovsing sign, McBurney sign)
 - Clinical symptoms (e.g., fever, migrating pain, guarding)
 - Laboratory tests (e.g., white blood cell count, C-reactive protein concentration, left shift)
 - Clinical prediction or decision rules (e.g., Alvarado score, Pediatric Appendicitis Score, other predictive models)
 - Imaging tests (e.g., US; multidetector or helical CT with or without contrast administered orally, rectally, or intravenously; MRI with or without contrast; abdominal X-ray)
 - Nuclear imaging studies
 - Diagnostic laparoscopy

Comparators

- Alternative tests or test combinations (as listed above), clinical observation

Outcomes

- Test performance (e.g., sensitivity, specificity, accuracy, proportion of “negative” appendectomies) using pathology or clinical followup as the reference standard
- Intermediate outcomes
 1. Impact on diagnostic thinking (e.g., change in diagnosis after testing; change in subsequent diagnostic approach after obtaining initial test results)
 2. Impact on therapeutic decisionmaking (e.g., change in treatment plan after testing; time from admission to surgery)
- Final health or patient-centered outcomes
 1. Bowel perforation (ruptured appendix)

2. Fistula formation
 3. Infectious complications (abscess formation, peritonitis, sepsis, stump appendicitis)
 4. Delay in diagnosis (time from presentation to definitive diagnosis; time from presentation to initiation of treatment; time from presentation to resolution of pain)
 5. Length of hospital stay
 6. Fetal/maternal outcomes (for pregnant women; including premature labor, pregnancy loss, fetal morbidity, fetal mortality, maternal morbidity, maternal mortality)
 7. Mortality
- Adverse effects of intervention(s)
 1. Direct harms of testing (e.g., harms from exposure to ionizing radiation, allergic reactions/kidney injury caused by contrast agents)
 2. Harms of test-directed treatment (indirect)

Timing

- Studies will be considered regardless of duration of followup.

Setting

- All health care settings will be considered.

III. Analytic Framework

Figure 1 presents a schematic of the Draft Analytic Framework for this report. The framework provides a visual representation of the clinical logic and preliminary PICO criteria (population, interventions, comparators, harms, intermediate outcomes, and final health outcomes).

Figure 1: Analytic framework for the diagnosis of acute appendicitis

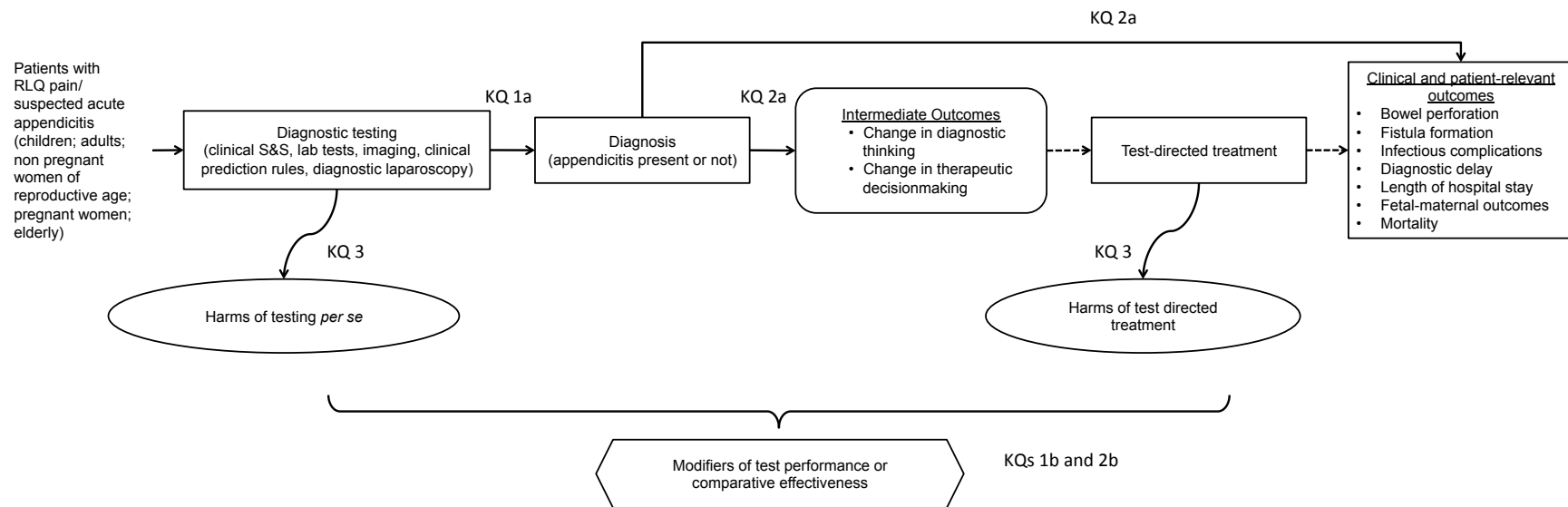


Figure 1. Key Questions are shown within the context of the PICO (Population, Intervention, Comparators, and Outcomes) criteria. Diagnostic strategies are compared in relevant clinical populations (patients with acute RLQ/ suspected acute appendicitis) with regards to intermediate outcomes (e.g., change in diagnostic thinking, change in therapeutic decisionmaking), clinical and patient relevant outcomes (bowel perforation, fistula formation, infectious complications, diagnostic delay, etc.), or adverse events. The treatment effect may be modified by several patient level factors (e.g., patient characteristics, setting of test use, operator experience, etc.). KQ = Key Question; RLQ = right lower quadrant; S & S = signs and symptoms. Please see the preceding section for a detailed description of the populations, interventions, and outcomes of interest.

IV. Methods

A. Criteria for Inclusion/Exclusion of Studies in the Review

Based on a random sample of 1000 double-screened abstracts, using the PICOTS criteria, we estimated that 12% of citations retrieved by our search strategy (search performed in PubMed on December 7, 2012) will have to be retrieved and reviewed in full text (the total corpus comprises >20,00 citations). Given the large expected number of potentially relevant studies (>2400) to be reviewed in full text, we believe that the scope of the project will need to be constrained operationally to ensure feasibility. Several approaches that could be used to achieve this aim were discussed with Key Informants during Topic Refinement and the TEP members (in preparation of this protocol). Based on these discussions and preliminary literature scans, we plan to use the following approach:

- *For single index test studies assessing test performance outcomes* (i.e. for a subset of the studies pertaining to Key Question 1), we will rely on existing systematic reviews (when available) to identify relevant studies, abstract specific data items (see section IV-C, below) from the review for each of those studies, and then synthesize those data items with the data from more recent single index test studies identified through literature searches to create a fully up to date review.
- *For single index tests where no relevant systematic review of test performance meeting our selection criteria* (see paragraph below) can be identified, we will perform a *de novo* review.
- *For studies that directly compare alternative tests* (for all outcomes of interest) and *for studies* (comparative or non-comparative) *reporting outcomes other than test performance* (e.g., change in diagnostic thinking, impact on therapeutic decisionmaking, clinical outcomes, and harms) we will perform a *de novo* review, because we believe that these topics are not addressed adequately by existing reviews.

We will use existing systematic reviews to identify single index test studies with test performance outcomes (Key Question 1). Systematic reviews will be considered as potential sources of eligible studies if they meet the following criteria:

- Report the bibliographic databases searched and any additional sources of included studies.
- Have used explicit criteria for selecting primary studies of the populations and index tests of interest (as described in section II, above).
- Have examined test performance outcomes.
- Provide a list of included studies that allows the retrieval of the corresponding full text publications.

Based on database searches for primary studies and the lists of studies included in previously published systematic reviews, we will compile a list of potentially eligible studies, to which we will apply the PICOTS eligibility criteria, listed in Section II. Inclusion/exclusion criteria will vary by Key Question in order to optimize the scope of the review and will be based on the for population, intervention, comparator, outcomes, timing, study design, and setting (PICOTS) criteria listed above. **Table 1** summarizes the selection criteria that will be applied to all potentially relevant studies (both those

identified through existing reviews and those identified through primary literature searches). Studies meeting these criteria will be included regardless of the specific role of testing evaluated (replacement, add-on, triage).

Table 1: Selection criteria for primary studies (all KQs), regardless of their source (literature searches for primary studies and previously published systematic reviews)

PICOTS item	Selection criteria
Population(s)	Patients with acute RLQ abdominal pain with suspected appendicitis
Interventions	Diagnostic tests (alone or in combination) for diagnosing appendicitis: clinical signs, laboratory tests, clinical prediction or decision rules, imaging tests, nuclear imaging studies, diagnostic laparoscopy.
Comparators	Alternative tests or test combinations (as listed above), clinical observation.
Outcomes	(1) Test performance using pathology or clinical followup as the reference standard; (2) Intermediate outcomes (impact on diagnostic thinking; impact on therapeutic decisionmaking); (3) Final health or patient-centered outcomes (bowel perforation; fistula formation; infectious complications; delay in diagnosis; length of hospital stay; fetal/maternal outcomes; mortality); (4) Adverse effects of intervention(s) (direct harms of testing; harms of test-directed treatment).
Timing	Studies will be considered regardless of duration of followup.
Setting	All health care settings will be considered.
Study design	(1) Prospective or retrospective diagnostic cohort studies including at least 10 patients. (2) Randomized controlled trials of alternative testing strategies including at least 10 patients per arm. (3) Nonrandomized comparative studies of alternative testing strategies including at least 100 patients per arm. (4) Single group (pre-post or time-series) studies evaluating the implementation of a diagnostic strategy. (5) We will require that at least 100 patients have been included in each phase of these studies. We will exclude: (1) primary research studies on animals; (2) editorials; (3) commentaries; (4) narrative reviews; (5) letters to the editor; (6) other manuscripts not reporting primary research findings.

KQ = Key Question; PICOTS = populations, interventions, comparators, outcomes, timing, study designs, and setting; RLQ = right lower quadrant.

B. Searching for the Evidence: Literature Search Strategies for Identification of Relevant Studies to Answer the Key Questions

Appendix 1 describes our proposed literature search strategy. This search will be conducted in MEDLINE®, EMBASE®, the Cochrane Central Register of Controlled Trials, and the Cumulative Index to Nursing and Allied Health Literature (CINAHL®) database, to identify primary research studies meeting our criteria. We will also use the MEDLINE® search results to identify systematic reviews of the tests of interest. We will not restrict searches by year of publication.

A common set of 200 abstracts (in 2 pilot rounds, each with 100 abstracts) will be screened by all reviewers, and discrepancies will be discussed in order to standardize screening practices and ensure understanding of screening criteria by all team members. The remaining citations will be split into nonoverlapping sets, each screened by two reviewers independently. Discrepancies will be resolved by consensus involving a third investigator.

Potentially eligible citations (i.e., abstracts considered potentially relevant by at least one reviewer) will be obtained in full text and reviewed for eligibility on the basis of the predefined inclusion criteria. Full-text articles will be screened independently by two reviewers for eligibility. Disagreements regarding article eligibility will be resolved by

consensus involving a third reviewer. We plan to include only English-language studies during full text review because our preliminary searches indicate that non-English-language studies are few and have small sample sizes; as such, they are unlikely to affect our conclusions. We may reconsider this decision if large relevant studies are identified during full text screening. To accommodate this potential modification of our inclusion criteria, we will not use language of publication as a criterion at the abstract screening stage (instead, we will evaluate the language of publication only at the full text review stage). We will exclude studies published exclusively in abstract form (e.g., conference proceedings) because they are typically not peer reviewed, only partially report results, and may change substantially when fully published. We will generate a list of reasons for exclusion for all studies excluded at the full text screening stage.

We will ask the TEP to provide citations of potentially relevant articles. Additional studies will be identified through the perusal of reference lists of eligible studies, published clinical practice guidelines, relevant narrative and systematic reviews, conference proceedings, Scientific Information Packages from manufacturers, and a search of U.S. Food and Drug Administration databases. All articles identified through these sources will be screened for eligibility against the same criteria as for articles identified through literature searches. If necessary, we will revise the search strategy so that it can better identify articles similar to those missed by our current search strategy. We will also ask the TEP to review the final list of included studies to ensure that no key publications have been missed.

Following submission of the draft report, an updated literature search (using the same search strategy) will be conducted. Abstract and full-text screening will be performed as described above. Any additional studies that meet the eligibility criteria will be added to the final report.

C. Data Abstraction and Data Management

Previously published reviews will be used as sources of eligible single index test studies of test performance, and as sources of data for objective data elements from these studies (bibliographic study information, characteristics of included populations, counts of individuals stratified by diagnostic test result and disease status). For all studies, EPC investigators will extract data elements that require the use of standardized operational definitions (e.g., elements of study design, risk of bias assessment) from the full text of primary study publications.

Data will be extracted into electronic forms using the Systematic Review Data Repository (SRDR, <http://srdhr.ahrq.gov/home/index>). The basic elements and design of these forms will be the similar to those we have used for other reviews of diagnostic tests and will include elements that address population characteristics, sample size, study design, descriptions of the index and reference standard tests of interest, analytic details, and outcome data. Prior to extraction, forms will be customized to capture all elements relevant to the Key Questions. We will use separate sections in the extraction forms for Key Questions related to intermediate outcomes, terminal outcomes, or adverse events, and for factors affecting (modifying) test performance (and other outcomes) among subgroups of patients. We will pilot test the forms on several studies extracted by all team members to ensure consistency in operational definitions. If necessary, forms will be revised before full data extraction.

A single reviewer will extract data from each eligible study. The extracted data will be reviewed and confirmed by at least one other team member (data verification). Disagreements will be resolved by consensus including a third reviewer.

We will contact authors (a) to clarify information reported in the papers that is hard to interpret (e.g., inconsistencies between tables and text); (b) to obtain missing data on key subgroups of interest when not available in the published reports (e.g., pregnant women, women of reproductive age, children); and (c) to verify suspected overlap between study populations in publications from the same group of investigators. Author contact will be by email (to the corresponding author of each study), with a primary contact attempt (once all eligible studies have been identified) and up to two reminder emails (approximately 2 and 4 weeks after the first attempt).

D. Assessment of Methodological Risk of Bias of Individual Studies

We will assess the risk of bias for each individual study using the assessment methods detailed by the Agency for Healthcare Research and Quality in its Methods Guide for Effectiveness and Comparative Effectiveness Review hereafter referred to as the Methods Guide. We will use the updated QUADAS 2 instrument to assess the risk of bias (methodological quality or internal validity) of the diagnostic test studies included in the review (these studies will comprise the majority of the available studies).⁴⁹⁻⁵² The tool assesses four domains for risk of bias related to patient selection, index test, reference standard test, and patient flow and timing. For studies of other designs, we will use appropriate sets of items to assess risk of bias: for nonrandomized cohort studies we will use the Newcastle-Ottawa scale;⁵³ for randomized controlled trials we will use the Cochrane Risk of Bias tool.⁵⁴

We will not calculate “composite” quality scores. Instead, we will assess and report each methodological quality item (as Yes, No, or Unclear/Not Reported) for each eligible study. We will rate each study as being of low, intermediate, or high risk of bias on the basis of adherence to accepted methodological principles. Generally, studies with low risk of bias have the following features: lowest likelihood of confounding due to comparison to a randomized controlled group; a clear description of the population, setting, interventions, and comparison groups; appropriate measurement of outcomes; appropriate statistical and analytic methods and reporting; no reporting inconsistencies; clear reporting of dropouts, and a dropout rate less than 20 percent; and no apparent bias. Studies with moderate risk of bias are susceptible to some bias but not sufficiently to invalidate results. They do not meet all the criteria for low risk of bias owing to some deficiencies, but none are likely to introduce major bias. Studies with moderate risk of bias may not be randomized or may be missing information, making it difficult to assess limitations and potential problems. Studies with high risk of bias are those with indications of bias that may invalidate the reported findings (e.g., observational studies not adjusting for any confounders, studies using historical controls, or studies with very high dropout rates). These studies have serious errors in design, analysis, or reporting and contain discrepancies in reporting or have large amounts of missing information. We discuss the handling of high risk of bias studies in Sections E and F.

In quantitative analyses, we will consider performing subgroup analyses to assess the impact of each risk of bias item on the meta-analytic results. The grading will be outcome specific, such that a given study that reports its primary outcome well but did an

incomplete analysis of a secondary outcome would be graded of different quality for the two outcomes. Studies of different designs will be graded within the context of their study design. Thus, randomized controlled trials will be graded as having a high, medium, or low risk of bias, and observational studies will be separately graded as having a high, medium, or low risk of bias.

E. Data Synthesis

We will summarize included studies qualitatively and present important features of the study populations, designs, interventions, outcomes, and results in summary tables. Population characteristics of interest include age, sex, duration of symptoms, and clinical presentation at enrollment. Design characteristics include methods of population selection and sampling, and follow-up duration. Test characteristics include aspects specific to each diagnostic test of interest (e.g., the use and route of administration of contrast agents (for imaging tests), the specific definitions of clinical signs, the components and their weights for clinical prediction rules, the surgical approach for diagnostic laparoscopy, etc.). We will present information on test performance, harms, intermediate and terminal outcomes, and resource utilization. Of note, studies evaluating the test performance of (the same) single index test will be synthesized jointly, regardless of their source (our own literature searches or previously published reviews).

For each comparison of interest, we will judge whether the eligible studies are sufficiently similar to be combined in a meta-analysis on the basis of clinical heterogeneity of patient populations and testing strategies, as well as methodological heterogeneity of study designs and outcomes reported. We will perform analyses appropriate for the specific role of testing evaluated in each study (replacement, triage, add-on), whenever possible.⁴⁸ However, the complexity of the differential diagnosis of RLQ pain, and limited reporting of relevant information in published studies, may limit our ability to distinguish between alternative test roles.

On the basis of discussions with local clinical experts and our preliminary review of the literature, we expect that eligible studies will have employed a variety of different diagnostic methods (e.g., different imaging modalities, clinical signs and symptoms, laboratory measurements, and combinations thereof). We will base our judgments on the similarity of available tests on technical descriptions of the modalities used in each study (e.g., whether studies used similar imaging technologies or similar clinical examination protocols). We will seek input from TEP members to define groups of “sufficiently similar” studies for synthesis (including meta-analysis) during later stages of the review if questions arise. Of note, the material used to solicit TEP input will not include any data on outcome results extracted from the studies (to limit the potential for bias). The determination on the appropriateness of meta-analysis will be made before any data analysis. We will not base the decision to perform a meta-analysis on statistical criteria for heterogeneity. Such criteria are often inadequate (e.g., low power when the number of studies is small) and do not account for the ability to explore and explain heterogeneity by examining study-level characteristics. Instead, we will use clinical criteria to assess exchangeability (e.g., we will consider whether studies enrolled populations selected using similar inclusion criteria, with comparable baseline risk of appendicitis, and assessed using similar imaging technologies or other tests). Main analyses will include all relevant studies.

Analyses will be performed separately for the following patient populations: children, women of reproductive age, pregnant women, and the elderly. Subgroup analyses (e.g., by clinical presentation at diagnosis, duration of symptoms, BMI, etc.) will also be performed. The concordance of findings across subgroup analyses will be evaluated qualitatively (in all instances) and quantitatively (using meta-regression, when the data allow). We will consider the following potential modifiers of test performance or other outcomes in meta-regression analyses: patient characteristics (e.g., age, sex, clinical presentation at enrollment, BMI), test characteristics (e.g., number of detectors for CT scanning, extent of imaging field, use of contrast agents and route of administration), clinician and facility factors (e.g., training of the operator, setting of test use), and date of publication. We will also perform subgroup analyses by individual risk of bias items to assess the impact of each risk of bias item on the results of the meta-analysis. We will evaluate the robustness of our findings in sensitivity analyses that exclude studies at high risk of bias. We will perform additional sensitivity analyses including leave-one-out meta-analysis and all-subsets meta-analysis.^{55, 56}

For studies reporting on test performance outcomes statistical analyses will be conducted using methods currently recommend for use in Comparative Effectiveness Reviews of diagnostic tests.^{57, 58} For parallel arm studies comparing alternative test strategies, meta-analyses will be undertaken when there are more than three unique studies evaluating the same intervention and comparator and reporting the same outcomes. All meta-analyses will be based on random effects models.⁵⁹ Sensitivity analyses (including leave-one-out analyses, analyses assuming a fixed effects model, and reanalyses after excluding a group of studies) may be undertaken if considered appropriate (e.g., in the presence of studies with outlying effect sizes or evidence of temporal changes in effect sizes). For all statistical tests, except those for heterogeneity, statistical significance will be defined as a two-sided p -value where $P < 0.05$. Heterogeneity will be considered statistically significant when the p -value of the Q statistic is $P < 0.1$ to account for the low statistical power of the test.⁶⁰ We will attempt to explore between-study heterogeneity using subgroup and meta-regression analyses.⁶¹

In cases when only a subset of the available studies can be quantitatively combined (e.g., when some studies are judged to be so clinically different from others as to be excluded from meta-analysis), we will synthesize findings across all studies qualitatively by taking into account the magnitude and direction of effects and estimates of performance.

F. Grading the Strength of Evidence (SOE) for Individual Outcomes

We will follow the Methods Guide⁶² to evaluate the strength of the body of evidence for each Key Question with respect to the following domains: risk of bias, consistency, directness, precision, and reporting bias.^{62, 63}

Briefly, we will define the risk of bias (low, medium, or high) on the basis of the study design and the methodological quality of the studies. Generally, lack of studies at low risk of bias or inconsistencies among groups of studies at different risk of bias will lead to downgrading the strength of the evidence. We will rate the consistency of the data as no inconsistency, inconsistency present, or not applicable (if there is only one study available). We do not plan to use rigid counts of studies as standards of evaluation (e.g., four of five studies agree, therefore the data are consistent); instead, we will assess the

direction, magnitude, and statistical significance of all studies and make a determination. We will describe our logic where studies are not unanimous. We will assess directness of the evidence (“direct” vs. “indirect”) on the basis of the use of surrogate outcomes or the need for indirect comparisons. We will assess the precision of the evidence as precise or imprecise on the basis of the degree of certainty surrounding each effect estimate. A precise estimate is one that allows for a clinically useful conclusion. An imprecise estimate is one for which the confidence interval is wide enough to include clinically distinct conclusions and that therefore precludes a conclusion.

We anticipate that the majority of studies to be included in this review will be observational cohorts reporting on outcomes of test performance, utilizing one or more index tests on all study participants. However, we also expect to find a small number of parallel group, randomized or non-randomized, comparative studies of alternative test strategies (e.g., reporting comparisons between alternative tests). We will not combine the results of randomized and non-randomized studies statistically. Instead, we will qualitatively evaluate similarities and differences in study populations, diagnostic methods, and outcomes among study designs. We will use these comparisons to inform our judgments on applicability of study findings to clinical practice (see also section G).

The potential for *reporting bias* (“suspected” vs. “not suspected”) will be evaluated with respect to publication bias, selective outcome reporting bias, and selective analysis reporting bias. For reporting bias, we will make qualitative dispositions rather than perform formal statistical tests to evaluate differences in the effect sizes between more precise (larger) and less precise (smaller) studies. Although these tests are often referred to as tests for publication bias; reasons other than publication bias can lead to a statistically significant result, including “true” heterogeneity between smaller and larger studies, other biases, and chance, rendering the interpretation of the tests nonspecific and the tests noninformative.^{64, 65} Therefore, instead of relying on statistical tests, we will evaluate the reported results across studies qualitatively, on the basis of completeness of reporting (separately for each outcome of interest), number of enrolled patients, and numbers of observed events. Judgment on the potential for selective outcome reporting bias will be based on reporting patterns for each outcome of interest across studies. We acknowledge that both types of reporting bias are difficult to reliably detect on the basis of data available in published research studies (i.e., without access to study protocols and detailed analysis plans). Because such assessments are inherently subjective, we will explicitly present all operational decisions and the rationale for our judgment on reporting bias in the Draft Report.

Finally, we will rate the body of evidence using four strength of evidence levels: high, moderate, low, and insufficient.⁶² These will describe our level of confidence that the evidence reflects the true effect for the major comparisons of interest.

G. Assessing Applicability

We will follow the Methods Guide⁶² to evaluate the applicability of included studies to patient populations of interest. We will evaluate studies separately by important clinical subgroups: children, women of reproductive age, pregnant women, and the elderly. Applicability to the population of interest will also be judged separately on the basis of duration of symptoms before enrollment, outcomes (e.g., test performance, impact on diagnostic thinking and clinical decisionmaking, clinical outcomes), and setting of care (e.g., whether patients were recruited in an academic, tertiary, or primary care setting).

V. References

1. Paulson EK, Kalady MF, Pappas TN. Clinical practice. Suspected appendicitis. *N Engl J Med*. 2003 Jan; 348(3):236-42. [PMID: 12529465]
2. Addiss DG, Shaffer N, Fowler BS, et al. The epidemiology of appendicitis and appendectomy in the United States. *Am J Epidemiol*. 1990 Nov; 132(5):910-25. [PMID: 2239906]
3. Ferri F. Appendicitis. In: Ferri F. *Ferri's Clinical Advisor*, 12th ed. Philadelphia, Elsevier Mosby; 2012. p. 94.
4. Parks NA, Schroepel TJ. Update on imaging for acute appendicitis. *Surg Clin North Am*. 2011 Feb; 91(1):141-54. [PMID: 21184905]
5. Wolfe J, Henneman P. Acute Appendicitis. In: Marx J, Hockberger R, Walls R, eds. *Rosen's Emergency Medicine*, 7th ed. Philadelphia, Elsevier Mosby; 2010. p. 1193-9.
6. Howell JM, Eddy OL, Lukens TW, et al. Clinical policy: Critical issues in the evaluation and management of emergency department patients with suspected appendicitis. *Ann Emerg Med*. 2010 Jan; 55(1):71-116. [PMID: 20116016]
7. Humes DJ, Simpson J. Clinical presentation of acute appendicitis: clinical signs - laboratory findings - clinical scores, Alvarado score and derivative scores. In: Keyzer C, Gevenois PA, eds. *Imaging of Acute Appendicitis in Adults and Children (Medical Radiology/Diagnostic Imaging)*. Dusseldorf, Springer-Verlag; 2011.
8. Bundy DG, Byerley JS, Liles EA, et al. Does this child have appendicitis? *JAMA*. 2007 Jul; 298(4):438-51. [PMID: 17652298]
9. Andersson RE. Meta-analysis of the clinical and laboratory diagnosis of appendicitis. *Br J Surg*. 2004 Jan; 91(1):28-37. [PMID: 14716790]
10. Hallan S, Asberg A. The accuracy of C-reactive protein in diagnosing acute appendicitis--a meta-analysis. *Scand J Clin Lab Invest*. 1997 Aug; 57(5):373-80. [PMID: 9279962]
11. Al-Khayal KA, Al-Omran MA. Computed tomography and ultrasonography in the diagnosis of equivocal acute appendicitis. A meta-analysis. *Saudi Med J*. 2007 Feb; 28(2):173-80. [PMID: 17268692]
12. Barger RL, Jr., Nandalur KR. Diagnostic performance of magnetic resonance imaging in the detection of appendicitis in adults: a meta-analysis. *Acad Radiol*. 2010 Oct; 17(10):1211-6. [PMID: 20634107]
13. Doria AS, Moineddin R, Kellenberger CJ, et al. US or CT for Diagnosis of Appendicitis in Children and Adults? A Meta-Analysis. *Radiology*. 2006 Oct; 241(1):83-94. [PMID: 16928974]

14. Krajewski S, Brown J, Phang PT, et al. Impact of computed tomography of the abdomen on clinical outcomes in patients with acute right lower quadrant pain: a meta-analysis. *Can J Surg*. 2011 Feb; 54(1):43-53. [PMID: 21251432]
15. Terasawa T, Blackmore CC, Bent S, et al. Systematic review: computed tomography and ultrasonography to detect acute appendicitis in adults and adolescents. *Ann Intern Med*. 2004 Oct 5;141(7):537-46. [PMID: 15466771]
16. Blumenfeld YJ, Wong AE, Jafari A, et al. MR imaging in cases of antenatal suspected appendicitis--a meta-analysis. *J Matern Fetal Neonatal Med*. 2011 Mar; 24(3):485-8. [PMID: 20695758]
17. Greenhalgh R, Punwani S, Taylor SA. Is MRI routinely indicated in pregnant patients with suspected appendicitis after equivocal ultrasound examination? *Abdom Imaging*. 2008 Jan-Feb; 33(1):21-5. [PMID: 17874265]
18. Carroll PJ, Gibson D, El-Faedy O, et al. Surgeon-performed ultrasound at the bedside for the detection of appendicitis and gallstones: systematic review and meta-analysis. *Am J Surg*. 2012 Jun. [PMID: 22748292]
19. Liu JL, Wyatt JC, Deeks JJ, et al. Systematic reviews of clinical decision tools for acute abdominal pain. *Health Technol Assess*. 2006 Nov; 10(47):1-167, iii-iv. [PMID: 17083855]
20. Alvarado A. A practical score for the early diagnosis of acute appendicitis. *Ann Emerg Med*. 1986 May;15(5):557-64. [PMID: 3963537]
21. Ohle R, O'Reilly F, O'Brien KK, et al. The Alvarado score for predicting acute appendicitis: a systematic review. *BMC Med*. 2011; 9:139. [PMID: 22204638]
22. Kulik DM, Uleryk EM, Maguire JL. Does this child have appendicitis? A systematic review of clinical prediction rules for children with acute abdominal pain. *J Clin Epidemiol*. 2013 Jan; 66(1):95-104. [PMID: 23177898]
23. Escriba A, Gamell AM, Fernandez Y, et al. Prospective validation of two systems of classification for the diagnosis of acute appendicitis. *Pediatr Emerg Care*. 2011 Mar; 27(3):165-9. [PMID: 21346681]
24. Samuel M. Pediatric appendicitis score. *J Pediatr Surg*. 2002 Jun; 37(6):877-81. [PMID: 12037754]
25. Korndorffer JR, Jr., Fellingner E, Reed W. SAGES guideline for laparoscopic appendectomy. *Surg Endosc*. 2010 Apr; 24(4):757-61. [PMID: 19787402]
26. Society for American Gastrointestinal and Endoscopic Surgeons. Guidelines for Diagnostic Laparoscopy. Available at: <http://www.sages.org/publications/guidelines/guidelines-for-diagnostic-laparoscopy>. Accessed April, 2013.

27. Fahimi J, Herring A, Harries A, et al. Computed tomography use among children presenting to emergency departments with abdominal pain. *Pediatrics* 2012 Nov;130(5):e1069-75. [PMID: 23045569]
28. Tsze DS, Asnis LM, Merchant RC, et al. Increasing computed tomography use for patients with appendicitis and discrepancies in pain management between adults and children: an analysis of the NHAMCS. *Ann Emerg Med.* 2012 May;59(5):395-403. [PMID: 21802777]
29. Drake FT, Florence MG, Johnson MG, et al. Progress in the diagnosis of appendicitis: a report from Washington State's Surgical Care and Outcomes Assessment Program. *Ann Surg.*2012 Oct; 256(4):586-94. [PMID: 22964731]
30. Bachur RG, Hennelly K, Callahan MJ, et al. Advanced radiologic imaging for pediatric appendicitis, 2005-2009: trends and outcomes. *J Pediatr.* 2012 Jun;160(6):1034-8. [PMID: 22192815]
31. Mittal MK, Dayan PS, Macias CG, et al. Performance of Ultrasound in the Diagnosis of Appendicitis in Children in a Multicenter Cohort. *Acad Emerg Med.* 2013; 20(7):697-702.
32. Hryhorczuk AL, Mannix RC, Taylor GA. Pediatric Abdominal Pain: Use of Imaging in the Emergency Department in the United States from 1999 to 2007. *Radiology.* 2012; 263(3):778-85. [PMID: 22535565]
33. Dingemann J, Ure B. Imaging and the use of scores for the diagnosis of appendicitis in children. *Eur J Pediatr Surg.* 2012 Jun; 22(3):195-200. [PMID: 22767172]
34. Jaffe TA, Miller CM, Merkle EM. Practice patterns in imaging of the pregnant patient with abdominal pain: a survey of academic centers. *AJR Am J Roentgenol.* 2007 Nov;189(5):1128-34. [PMID: 17954650]
35. Wallace GW, Davis MA, Semelka RC, et al. Imaging the pregnant patient with abdominal pain. *Abdom Imaging.* 2012 Oct; 37(5):849-60. [PMID: 22160283]
36. Katz DS, Klein MA, Ganson G, et al. Imaging of abdominal pain in pregnancy. *Radiol Clin North Am.*2012 Jan; 50(1):149-71. [PMID: 22099493]
37. Long SS, Long C, Lai H, et al. Imaging strategies for right lower quadrant pain in pregnancy. *AJR American Journal of Roentgenol.* 2011 Jan;196(1):4-12. [PMID: 21178041]
38. Fryback DG, Thornbury JR. The efficacy of diagnostic imaging. *Med Decis Making.* 1991 Apr-Jun;11(2):88-94. [PMID: 1907710]
39. Collaborative S, Cuschieri J, Florence M, et al. Negative appendectomy and imaging accuracy in the Washington State Surgical Care and Outcomes Assessment Program. *Ann Surg.* 2008 Oct; 248(4):557-63. [PMID: 18936568]

40. Clinical policy: critical issues for the initial evaluation and management of patients presenting with a chief complaint of nontraumatic acute abdominal pain. *Ann Emerg Med.* 2000 Oct; 36(4):406-15. [PMID: 11020699]
41. Flum DR, Morris A, Koepsell T, et al. Has misdiagnosis of appendicitis decreased over time? A population-based analysis. *JAMA.* 2001 Oct; 286(14):1748-53. [PMID: 11594900]
42. Nance ML, Adamson WT, Hedrick HL. Appendicitis in the young child: a continuing diagnostic challenge. *Pediatr Emerg Care.* 2000 Jun;16(3):160-2. [PMID: 10888451]
43. Ponsky TA, Huang ZJ, Kittle K, et al. Hospital- and patient-level characteristics and the risk of appendiceal rupture and negative appendectomy in children. *JAMA.* 2004 Oct; 292(16):1977-82. [PMID: 15507583]
44. Newman K, Ponsky T, Kittle K, et al. Appendicitis 2000: variability in practice, outcomes, and resource utilization at thirty pediatric hospitals. *J Pediatr Surg.* 2003 Mar; 38(3):372-9. [PMID: 12632352]
45. Rothrock SG, Green SM, Dobson M, et al. Misdiagnosis of appendicitis in nonpregnant women of childbearing age. *J Emerg Med.* 1995 Jan-Feb;13(1):1-8. [PMID: 7782616]
46. Brown JJ, Wilson C, Coleman S, et al. Appendicitis in pregnancy: an ongoing diagnostic dilemma. *Colorectal Dis.* 2009 Feb;11(2):116-22. [PMID: 18513191]
47. Kraemer M, Franke C, Ohmann C, et al. Acute appendicitis in late adulthood: incidence, presentation, and outcome. Results of a prospective multicenter acute abdominal pain study and a review of the literature. *Langenbecks Arch Surg.* 2000 Nov; 385(7):470-81. [PMID: 11131250]
48. Hayen A, Macaskill P, Irwig L, et al. Appropriate statistical methods are required to assess diagnostic tests for replacement, add-on, and triage. *J Clin Epidemiol.* 2010 Aug; 63(8):883-91. [PMID: 20079607]
49. Whiting PF, Rutjes AW, Westwood ME, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med.* 2011 Oct;155(8):529-36. [PMID: 22007046]
50. Whiting P, Rutjes AW, Reitsma JB, et al. The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. *BMC Med Res Methodol.* 2003 Nov; 3:25. [PMID: 14606960]
51. Whiting P, Rutjes AW, Dinnes J, et al. Development and validation of methods for assessing the quality of diagnostic accuracy studies. *Health Technol Assess.* 2004 Jun;8(25):iii, 1-234. [PMID: 15193208]

52. Whiting PF, Weswood ME, Rutjes AW, et al. Evaluation of QUADAS, a tool for the quality assessment of diagnostic accuracy studies. *BMC Med Res Methodol*. 2006; 6:9. [PMID: 16519814]
53. Ottawa Hospital Research Institute. Our Research Web page. Research Programs: Clinical Epidemiology: The Newcastle-Ottawa Scale (NOS) for assessing the quality of nonrandomised studies in meta-analyses. Available at: http://www.ohri.ca/programs/clinical_epidemiology/oxford.asp. Accessed April 29, 2013.
54. Higgins JP, Altman DG, Gotzsche PC, et al. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *BMJ*. 2011; 343:d5928. [PMID: 22008217]
55. Olkin I. Diagnostic statistical procedures in medical meta-analyses. *Stat Med*. 1999 Sept;18(17-18):2331-41. [PMID: 10474143]
56. Olkin I, Dahabreh IJ, Trikalinos TA. GOSH—a graphical display of study heterogeneity. *Research Synthesis Methods* 2012; 3(3):214-23.
57. Trikalinos TA, Balion CM. Chapter 9: options for summarizing medical test performance in the absence of a "gold standard". *J Gen Intern Med*. 2012 Jun;27 Suppl 1:S67-75. [PMID: 22648677]
58. Trikalinos TA, Balion CM, Coleman CI, et al. Chapter 8: meta-analysis of test performance when there is a "gold standard". *J Gen Intern Med*. 2012 Jun; 27 Suppl 1:S56-66. [PMID: 22648676]
59. DerSimonian R, Laird N. Meta-analysis in clinical trials. *Control Clin Trials*. 1986 Sept;7(3):177-88. [PMID: 3802833]
60. Cochran WG. The combination of estimates from different experiments. *Biometrics*. 1954;10(1):101-29.
61. Thompson SG, Higgins JP. How should meta-regression analyses be undertaken and interpreted? *Stat Med*. 2002 Jun; 21(11):1559-73. [PMID: 12111920]
62. Agency for Healthcare Research and Quality. Effective Health Care Program Website. Methods Guide for Effectiveness and Comparative Effectiveness Reviews. Available at: <http://effectivehealthcare.ahrq.gov/index.cfm/search-for-guides-reviews-and-reports/?pageaction=displayproduct&productid=318>. Accessed November, 2013.
63. Singh S, Chang SM, Matchar DB, et al. Grading a Body of Evidence on Diagnostic Tests. In: Chang SM, Matchar DB, Smetana GW, et al, eds. *Methods Guide for Medical Test Reviews*. Rockville (MD): Agency for Healthcare Research and Quality (US); 2012 Jun. Chapter 7.
64. Lau J, Ioannidis JP, Terrin N, et al. The case of the misleading funnel plot. *BMJ*. 2006 Sept; 33(7568):597-600. [PMID: 16974018]

65. Sterne JA, Sutton AJ, Ioannidis JP, et al. Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. *BMJ*. 2011;343:d4002. [PMID: 21784880]

VI. Definition of Terms

Not applicable.

VII. Summary of Protocol Amendments

No amendments have been made. In the event of protocol amendments, the date of each amendment will be accompanied by a description of the change and the rationale.

VIII. Review of Key Questions

For all EPC reviews, Key Questions were reviewed and refined as needed by the EPC with input from Key Informants and the TEP to assure that the questions are specific and explicit about what information is being reviewed. In addition, for Comparative Effectiveness reviews, the Key Questions were posted for public comment and finalized by the EPC after review of the comments.

IX. Key Informants

Key Informants are the end users of research, including patients and caregivers, practicing clinicians, relevant professional and consumer organizations, purchasers of health care, and others with experience in making health care decisions. Within the EPC program, the Key Informant role is to provide input into identifying the Key Questions for research that will inform healthcare decisions. The EPC solicits input from Key Informants when developing questions for systematic review or when identifying high priority research gaps and needed new research. Key Informants are not involved in analyzing the evidence or writing the report and have not reviewed the report, except as given the opportunity to do so through the peer or public review mechanism.

Key Informants must disclose any financial conflicts of interest greater than \$10,000 and any other relevant business or professional conflicts of interest. Because of their role as end-users, individuals are invited to serve as Key Informants and those who present with potential conflicts may be retained. The TOO and the EPC work to balance, manage, or mitigate any potential conflicts of interest identified.

X. Technical Experts

Technical Experts comprise a multi-disciplinary group of clinical, content, and methodologic experts who provide input in defining populations, interventions, comparisons, or outcomes as well as identifying particular studies or databases to search. They are selected to provide broad expertise and perspectives specific to the topic under development. Divergent and conflicted opinions are common and perceived as health scientific discourse that results in a thoughtful, relevant systematic review. Therefore study questions, design and/or methodological approaches do not necessarily represent the views of individual technical and content experts. Technical Experts provide information to the EPC to identify literature search strategies and recommend approaches to specific issues as requested by the EPC. Technical Experts do not do analysis of any kind nor contribute to the writing of the report and have not reviewed the report, except as given the opportunity to do so through the public review mechanism

Technical Experts must disclose any financial conflicts of interest greater than \$10,000 and any other relevant business or professional conflicts of interest. Because of

their unique clinical or content expertise, individuals are invited to serve as Technical Experts and those who present with potential conflicts may be retained. The TOO and the EPC work to balance, manage, or mitigate any potential conflicts of interest identified.

XI. Peer Reviewers

Peer reviewers are invited to provide written comments on the draft report based on their clinical, content, or methodologic expertise. Peer review comments on the preliminary draft of the report are considered by the EPC in preparation of the final draft of the report. Peer reviewers do not participate in writing or editing of the final report or other products. The synthesis of the scientific literature presented in the final report does not necessarily represent the views of individual reviewers. The dispositions of the peer review comments are documented and will, for CERs and Technical Briefs, be published three months after the publication of the Evidence report.

Potential Reviewers must disclose any financial conflicts of interest greater than \$10,000 and any other relevant business or professional conflicts of interest. Invited Peer Reviewers may not have any financial conflict of interest greater than \$10,000. Peer reviewers who disclose potential business or professional conflicts of interest may submit comments on draft reports through the public comment mechanism.

XII. EPC Team Disclosures

The following team members will be involved:

- The EPC director
- The EPC co-director
- 1 Project Lead
- 3 Research Associates
- 1 Local Clinical Expert
- 1 Project manager
- 1 Program Assistant

All EPC team members have no financial or other conflicts of interest to disclose.

XIII. Role of the Funder

This project is funded under Contract No. HHSA-290-2012-0012-I from the Agency for Healthcare Research and Quality, U.S. Department of Health and Human Services. The Task Order Officer reviews contract deliverables for adherence to contract requirements and quality. The authors of this report are responsible for its content. Statements in the report should not be construed as endorsement by the Agency for Healthcare Research and Quality or the U.S. Department of Health and Human Services.

Appendix A: Search strategies

Clinical Practice Guidelines

(Practice Guideline[pt]) AND ("abdomen, acute"[MeSH] OR "appendicitis"[MeSH] OR "appendectomy"[MeSH] OR "appendix"[MeSH] OR (acute AND (abdome* OR abdomi*) AND pain) OR appendic* OR appendec* OR appendicec* OR appendix OR ((non?specific) AND (abdome* OR abdomi*) AND pain) OR nsap OR RLQ pain OR (right AND lower AND (quarter OR quadrant) AND pain) OR (acute AND abdominal AND pain) OR AAP)

Primary studies

("Computerized tomography" OR "Computed tomography" OR "CT" OR enhancement*) OR (Ultrasonography OR Sonography OR US OR ultrasound OR ultrasound) OR ("MR" OR magnetic resonance OR MRI OR "magnetic resonance imaging"[MeSH]) OR (Radiography[MeSH] OR Tomography, x-ray computed[MeSH] OR Tomography scanners, x-ray computed[MeSH] OR Tomography, spiral computed[MeSH]) OR ("radionuclide imaging"[Subheading] OR (radionuclide* AND imaging)) OR laparoscop* OR "laparoscopy"[MeSH Terms] OR skin temperature OR fever OR temperature OR ((McBurney OR obturator OR psoas OR rovsing*) AND (sign OR point)) OR "acute-phase proteins"[MeSH Terms] OR (urine test OR white blood cell count OR WBC OR leukocyte* OR acute phase proteins) OR (DT OR decision* tools OR decision* support system OR algorithm OR scoring system) OR ((Alvarado OR Mantrels) AND (test OR tests OR score OR scores)) OR checklist* OR algorith* OR (slide rule*) OR calculator* OR (score OR scores) OR (practice AND guideline*) OR (progno* AND (model OR modeling OR models)) OR (decision support system*) OR computer* OR (decision tree*) OR (decision analy*) OR (decision aid*) OR (decision tool*) OR (advisory AND (system OR systems)) OR nomogram* OR expert system\$ OR neural network* OR artificial intellig* OR machine learning OR Bayes* OR "decision support systems, clinical"[MeSH] OR "decision support systems, management"[MeSH] OR "decision support techniques"[MeSH] OR "artificial intelligence"[MeSH] OR "decision making, computer assisted"[MeSH] OR "medical informatics"[MeSH] OR "information systems"[MeSH] OR "decision making"[MeSH] OR "Reminder Systems"[MeSH] OR "Hospital Information Systems"[MeSH] OR "Management Information Systems"[MeSH] OR "Medical Records Systems, Computerized"[MeSH] OR "Computers"[MeSH] OR (information system*) OR informatic*) AND ("abdomen, acute"[MeSH] OR "appendicitis"[MeSH] OR "appendectomy"[MeSH] OR "appendix"[MeSH] OR (acute AND (abdome* OR abdomi*) AND pain) OR appendic* OR appendec* OR appendicec* OR appendix OR ((non?specific) AND (abdome* OR abdomi*) AND pain) OR nsap OR RLQ pain OR (right AND lower AND (quarter OR quadrant) AND pain) OR (acute AND abdominal AND pain) OR AAP)

Systematic reviews

(appendicitis OR appendix OR appendiceal OR appendi* OR appendect*) AND systematic[sb]

Note: *these search strings are designed for use in PubMed. They will be modified as needed for use in other databases we plan to search (see the Methods section of the protocol).*