Diagnostic Accuracy of Noncontrast Computed Tomography for Appendicitis in Adults: A Systematic Review

Veronica Hlibczuk, MD
Judith A. Dattaro, MD
Zhezhen Jin, PhD
Louise Falzon, BA
Michael D. Brown, MD, MSc

From the Division of Emergency Medicine, Columbia University Medical Center (Hlibczuk), and Division of Emergency Medicine, Weill Cornell Medical Center (Dattaro), New York–Presbyterian Hospital, New York, NY; the Department of Biostatistics, Mailman School of Public Health, Columbia University, New York, NY (Jin); Mount Sinai Heart, Mount Sinai Medical Center, New York, NY (Falzon); and the Division of Emergency Medicine, Michigan State University College of Human Medicine, Grand Rapids, MI (Brown).

Study objective: We seek to determine the diagnostic test characteristics of noncontrast computed tomography (CT) for appendicitis in the adult emergency department (ED) population.

Methods: We conducted a search of MEDLINE, EMBASE, the Cochrane Library, and the bibliographies of previous systematic reviews. Included studies assessed the diagnostic accuracy of noncontrast CT for acute appendicitis in adults by using the final diagnosis at surgery or follow-up at a minimum of 2 weeks as the reference standard. Studies were included only if the CT was completed using a multislice helical scanner. Two authors independently conducted the relevance screen of titles and abstracts, selected studies for the final inclusion, extracted data, and assessed study quality. Consensus was reached by conference, and any disagreements were adjudicated by a third reviewer. Unenhanced CT test performance was assessed with summary receiver operating characteristic curve analysis, with independently pooled sensitivity and specificity values across studies.

Results: The search yielded 1,258 publications; 7 studies met the inclusion criteria and provided a sample of 1,060 patients. The included studies were of high methodological quality with respect to appropriate patient spectrum and reference standard. Our pooled estimates for sensitivity and specificity were 92.7% (95% confidence interval 89.5% to 95.0%) and 96.1% (95% confidence interval 94.2% to 97.5%), respectively; the positive likelihood ratio=24 and the negative likelihood ratio=0.08.

Conclusion: We found the diagnostic accuracy of noncontrast CT for the diagnosis of acute appendicitis in the adult population to be adequate for clinical decisionmaking in the ED setting. [Ann Emerg Med. 2010;55:51-59.]

Please see page 52 for the Editor’s Capsule Summary of this article.

INTRODUCTION

Acute appendicitis is frequently in the differential diagnosis of patients presenting to the emergency department (ED) with right lower quadrant abdominal pain. Unfortunately, the diagnosis is difficult, with the classic presentation of periumbilical pain followed by nausea and vomiting, with migration of pain to the right lower quadrant occurring in only 50% to 60% of patients with appendicitis.\(^1\) An increased leukocyte count occurs in 70% to 90% of cases but is nonspecific because it can occur in other disease processes.\(^2\) Although there has been some conflicting evidence, the use of computed tomography (CT) has been reported to significantly reduce the negative appendectomy rate in many studies.\(^3,4-6\) Institutions use different combinations of oral, intravenous, and rectal contrast, as well as noncontrast protocols. Noncontrast CT is particularly appealing in today’s crowded EDs because there is no delay caused by waiting for oral contrast transit, no risk of contrast-induced nephropathy, and no risk of allergic reaction. However, the use of noncontrast CT has been controversial because of concerns about diagnostic accuracy and the need for repeated scanning with contrast in a subset of patients when the interpretation is inconclusive.\(^7,8\) The goal of this systematic review was to assess the evidence for the use of noncontrast CT for the diagnosis of acute appendicitis among adults.

MATERIALS AND METHODS

The clinical question addressed in this systematic review is, in adult patients presenting to the ED with acute abdominal pain and suspected of having acute appendicitis, what are the test characteristics of noncontrast helical CT?
Noncontrast CT for Diagnosis of Adult Appendicitis

Hlibczuk et al

Editor’s Capsule Summary

What is already known on this topic
Abdominal computed tomography (CT) is widely used for evaluation of suspected appendicitis, but the utility of CT without contrast remains controversial.

What question this study addressed
What is the diagnostic accuracy of noncontrast CT in the emergency department (ED) evaluation of adults with suspected appendicitis?

What this study adds to our knowledge
This 7-study systematic review of noncontrast CT yielded pooled estimates of sensitivity 93%, specificity 96%, positive likelihood ratio 24, and negative likelihood ratio 0.08, which are comparable to findings of previously published reviews.

How this might change clinical practice
Noncontrast CT has high sensitivity and specificity in the ED diagnosis of adult appendicitis and is a reasonable alternative to contrast CT for this indication.

review protocol was developed to directly address this clinical question and was reviewed and agreed upon by all authors a priori.

All relevant electronic databases were searched for studies assessing the accuracy of noncontrast helical ( multislice) CT to diagnose acute appendicitis in adults. No age, language, or publication limits were applied. Databases searched were MEDLINE (Ovid 1950 to March, week 3, 2008), EMBASE, the Cochrane Library (including the Cochrane Central Register of Controlled Clinical Trials, the Database of Abstracts of Review of Effects, and the Health Technology Assessment Database), the American College of Physicians Journal Club, and Emergency Medicine Abstracts (available online at http://ccme.org/). The MEDLINE search used a combination of the following MeSH headings and free-text terms: “appendicitis,” “appendectomy,” “append$,” “tomography,” “x-ray computed,” “contrast media,” “computed tomography,” “ct,” “helical,” “enhanced,” “unenhanced,” “contrast$,” and “noncontrast.” Similar search terms were then adapted for each database. The complete MEDLINE search strategy is provided as Appendix E1 (available online at http://www.annemergmed.com). The MEDLINE search was updated weekly for any new relevant citations until March 2009. The bibliographies of studies meeting the final inclusion criteria and those of previous systematic reviews9–12 were also searched.

An appropriate patient spectrum and reference standard were deemed essential to the validity of the systematic review. Therefore, specific inclusion and exclusion criteria for selecting studies were applied to ensure that any inferences made according to the results of the meta-analysis were appropriate and applicable to the ED setting.13 Only studies with patients who presented or were referred to the ED with acute abdominal pain suspicious for acute appendicitis, but not immediate candidates for surgery, were included. Studies with a mix of adult and pediatric patients were excluded unless data for those aged 16 years or older were clearly reported or made available from authors who were contacted. Studies were included if the CT was completed using a multislice helical scanner. Studies aimed at assessing the accuracy of a single-row CT were excluded because this technology is no longer used in the United States. Studies were included only if an appropriate reference standard for diagnosis or exclusion of acute appendicitis was used in all patients; in other words, partial verification bias was an exclusion criterion. The diagnosis of appendicitis could be determined at surgery by the surgeon or on assessment of the specimen by a pathologist. Given our requirement that studies include an appropriate patient spectrum, it was expected that not all patients would undergo laparotomy. Therefore, uneventful clinical follow-up for a minimum of 2 weeks was also considered an acceptable reference standard for the exclusion of clinically significant acute appendicitis. Studies were not included if length of clinical follow-up was not reported or could not be confirmed by author contact.

Data Collection and Processing

Two reviewers (V.H., J.A.D.) independently performed a relevance search; both reviewers examined the titles and abstracts of all references identified to determine whether the article was relevant to the research question. The 2 reviewers then compared their inclusion and exclusion logs, and the level of agreement was calculated with the κ statistic. Any disagreements were adjudicated by a third investigator (M.D.B.).

A data collection form was used to extract data from each study that appeared to satisfy the inclusion criteria. An attempt was made to contact the authors of studies that required data clarification or supplementation. Again, 2 reviewers independently abstracted the data, and consensus was reached by conference between the 2 reviewers (V.H., J.A.D.); disagreements were resolved by the third party (M.D.B.).

As described under the inclusion criteria, an appropriate reference standard was required for a study to be included in the systematic review. Given that an appropriate patient spectrum was required for inclusion, it was recognized that differential verification bias would likely be present among the included studies (ie, some patients would have clinical follow-up as the reference standard for the exclusion of acute appendicitis). Studies were not included if length of clinical follow-up was not reported or could not be confirmed by author contact.

We elected to be restrictive by including only higher-quality studies based on the 2 key quality measures discussed under the inclusion criteria above; studies had to include an appropriate patient spectrum and use an acceptable reference standard.14,15 We then assessed the quality of each study according to the most relevant items from the QUADAS tool.14 QUADAS was developed as an evidence-based quality assessment tool to be used in systematic reviews to assess the quality of primary

Abdominal computed tomography (CT) is widely used for evaluation of suspected appendicitis, but the utility of CT without contrast remains controversial.

What question this study addressed
What is the diagnostic accuracy of noncontrast CT in the emergency department (ED) evaluation of adults with suspected appendicitis?

What this study adds to our knowledge
This 7-study systematic review of noncontrast CT yielded pooled estimates of sensitivity 93%, specificity 96%, positive likelihood ratio 24, and negative likelihood ratio 0.08, which are comparable to findings of previously published reviews.

How this might change clinical practice
Noncontrast CT has high sensitivity and specificity in the ED diagnosis of adult appendicitis and is a reasonable alternative to contrast CT for this indication.

Abdominal computed tomography (CT) is widely used for evaluation of suspected appendicitis, but the utility of CT without contrast remains controversial.

What question this study addressed
What is the diagnostic accuracy of noncontrast CT in the emergency department (ED) evaluation of adults with suspected appendicitis?

What this study adds to our knowledge
This 7-study systematic review of noncontrast CT yielded pooled estimates of sensitivity 93%, specificity 96%, positive likelihood ratio 24, and negative likelihood ratio 0.08, which are comparable to findings of previously published reviews.

How this might change clinical practice
Noncontrast CT has high sensitivity and specificity in the ED diagnosis of adult appendicitis and is a reasonable alternative to contrast CT for this indication.
studies of diagnostic accuracy. Each item was rated as yes, no, or unclear, as defined by the authors of QUADAS. Although results were reported for all 11 items of the QUADAS instrument, there were 6 items considered particularly relevant to the clinical question: (1) did patients receive the same reference standard regardless of the index test result (ie, was differential verification bias avoided)?; (2) were the index test results interpreted without knowledge of the results of the reference standard?; (3) were the reference standard results interpreted without knowledge of the results of the index test?; (4) were the same clinical data available when test results were interpreted as would be available when the test is used in practice (ie, was the test used in clinically applicable circumstances)?; (5) were uninterpretable/indeterminate test results reported?; and (6) were withdrawals from the study explained? No attempt was made to combine the quality assessment into an overall score; instead, the results of the quality assessment were reported in a summary figure.

**Primary Data Analysis**

Unenhanced CT test performance was assessed with the traditional summary receiver operating characteristic (SROC) curve analysis, with independently pooled sensitivity and specificity values across studies using a random-effects model (for technical details, see Appendix E2, available online at http://www.annemergmed.com). A subgroup analysis based on CT technology to detect any difference in test performance with multislice scanners (ie, 4, 16, or 64 slice) versus dual-slice scanners was planned. In addition, a sensitivity analysis was planned with a bivariate random-effects regression model that focuses on estimating the average sensitivity and specificity and also allows for adjustment for sources of bias and variability among studies. The statistical analysis was performed with RevMan 5.0.16, SAS statistical application program (version 9.1; SAS Institute, Inc., Cary, NC), and S-Plus 8 Enterprise (TIBCO Software Inc., Palo Alto, CA). The precision for the test characteristics of noncontrast CT test was reported with 95% confidence intervals (CIs).

The validity of using funnel plots or statistical models to detect publication bias for diagnostic test meta-analysis has not been established, so we did not formally test for the presence of publication bias.

**RESULTS**

There was good agreement between the 2 reviewers (V.H., J.A.D.) for the relevance screen of the 1,161 titles identified in the MEDLINE, Cochrane, and EMBASE databases ($\kappa=0.62$). The comprehensive search yielded a total of 1,258 publications (Figure 1). After adjudication of the relevance search, a full article review was completed on the remaining 32 articles. Upon full article review, 15 studies did not meet our inclusion criteria for various reasons such as inappropriate patient spectrum, inadequate data available for extraction, or the use of contrast. A further 10 studies were excluded.
Table 1. Summary of exclusion of studies.

<table>
<thead>
<tr>
<th>Authors</th>
<th>Reason for Exclusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bendek et al</td>
<td>Includes pediatric patients</td>
</tr>
<tr>
<td>Brontein et al</td>
<td>All patients underwent appendectomy</td>
</tr>
<tr>
<td>D’Ippolito et al</td>
<td>Includes pediatric patients</td>
</tr>
<tr>
<td>Kailidou et al</td>
<td>Results combined contrast with noncontrast CT scans</td>
</tr>
<tr>
<td>Keyzer et al</td>
<td>Could not extract data from low-dose and standard-dose CT scans</td>
</tr>
<tr>
<td>Malone et al</td>
<td>Includes pediatric patients</td>
</tr>
<tr>
<td>Morris et al</td>
<td>Length of clinical follow-up not stated</td>
</tr>
<tr>
<td>Torbati et al</td>
<td>Includes pediatric patients</td>
</tr>
<tr>
<td>Vadeboncoeur et al</td>
<td>Age range of patients not reported</td>
</tr>
<tr>
<td>Van Breda Vriesman et al</td>
<td>Includes pediatric patients</td>
</tr>
<tr>
<td>Yüsekkekaya et al</td>
<td>Includes pediatric patients</td>
</tr>
<tr>
<td>Yetkin et al</td>
<td>Could not verify length of clinical follow-up</td>
</tr>
<tr>
<td>Yeung et al</td>
<td>Includes pediatric patients</td>
</tr>
<tr>
<td>Wise et al</td>
<td>Definition of unenhanced CT scan included oral contrast</td>
</tr>
</tbody>
</table>

Exclusion of 15 articles after full review

<table>
<thead>
<tr>
<th>Authors</th>
<th>Reason for Exclusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bendek et al</td>
<td>Includes pediatric patients</td>
</tr>
<tr>
<td>Brontein et al</td>
<td>Age range of patients not reported</td>
</tr>
<tr>
<td>D’Ippolito et al</td>
<td>Includes pediatric patients</td>
</tr>
<tr>
<td>Kailidou et al</td>
<td>Results combined contrast with noncontrast CT scans</td>
</tr>
<tr>
<td>Keyzer et al</td>
<td>Could not extract data from low-dose and standard-dose CT scans</td>
</tr>
<tr>
<td>Malone et al</td>
<td>Includes pediatric patients</td>
</tr>
<tr>
<td>Morris et al</td>
<td>Length of clinical follow-up not stated</td>
</tr>
<tr>
<td>Styrud et al</td>
<td>Age range of patients not reported</td>
</tr>
<tr>
<td>Torbati et al</td>
<td>Includes pediatric patients</td>
</tr>
<tr>
<td>Vadeboncoeur et al</td>
<td>Age range of patients not reported</td>
</tr>
<tr>
<td>Van Breda Vriesman et al</td>
<td>Includes pediatric patients</td>
</tr>
<tr>
<td>Yüsekkekaya et al</td>
<td>Includes pediatric patients</td>
</tr>
<tr>
<td>Yetkin et al</td>
<td>Could not verify length of clinical follow-up</td>
</tr>
<tr>
<td>Yeung et al</td>
<td>Includes pediatric patients</td>
</tr>
<tr>
<td>Wise et al</td>
<td>Definition of unenhanced CT scan included oral contrast</td>
</tr>
</tbody>
</table>

Exclusion of 10 articles/abstracts after potential inclusion

<table>
<thead>
<tr>
<th>Authors</th>
<th>Reason for Exclusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cakirer et al</td>
<td>Could not verify length of clinical follow-up</td>
</tr>
<tr>
<td>Christopher et al</td>
<td>Includes pediatric patients</td>
</tr>
<tr>
<td>Gamanagatti et al</td>
<td>Includes pediatric patients</td>
</tr>
<tr>
<td>Heaston et al</td>
<td>Age range of patients could not be verified</td>
</tr>
<tr>
<td>Hershko et al</td>
<td>Could not verify length of clinical follow-up</td>
</tr>
<tr>
<td>Lane, Katz et al</td>
<td>Includes pediatric patients</td>
</tr>
<tr>
<td>Lane, Liu et al</td>
<td>Could not verify length of clinical follow-up</td>
</tr>
<tr>
<td>Peck et al</td>
<td>Includes pediatric patients</td>
</tr>
<tr>
<td>Poortman et al</td>
<td>Could not verify length of clinical follow-up</td>
</tr>
<tr>
<td>Togawa et al</td>
<td>Could not verify length of clinical follow-up</td>
</tr>
</tbody>
</table>

Reasons for excluding 15 studies after the full review and another 10 studies after attempts at author contact are provided in Table 1. Seven studies were included in the final analysis, providing a sample of 1,060 subjects.44–50

The prevalence of appendicitis among patients presenting or referred to the ED with acute abdominal pain who were suspected of having appendicitis, but were not immediate candidates for surgery, ranged from 20.1% to 84.5%, with a median of 39.3% (Table 2). As expected, in the majority of studies (6/7) there were patients who did not undergo surgery and were followed up clinically. In all 6 of these studies, none of the patients who were followed up were later diagnosed with appendicitis. The single study that did not rely on clinical follow-up (in’t Hof et al)47 used laparoscopy performed by a surgeon blinded to the CT result as the reference standard for all patients.19 Although enrollment was described as including consecutive ED patients suspected of appendicitis, the prevalence of appendicitis in this study was quite high (84.5%), indicating that the results were likely biased toward patients with a more severe presentation compared with the typical ED population presenting with abdominal pain. The sensitivity and specificity of each study are presented in Table 2 and Figure 2.

The strict inclusion criteria required that all 7 studies be rated “yes” as having a representative patient spectrum and acceptable reference standard (Figure 3). All 7 studies were also rated “yes” for an acceptable delay between the CT and the measurement of the reference standard; after noncontrast CT, patients were either operated on in a timely manner or had adequate follow-up of at least 2 weeks. Except for the study by in’t Hof et al,47 in which all patients had laparoscopy regardless of the CT result, the other studies were rated “no” with respect to avoiding differential verification bias. Differential verification bias was unavoidable in studies that used clinical follow-up as a reference standard when the CT result was negative. Indeterminate (uninterpretable) results were clearly reported in 2 studies, Keyzer et al48 and Tamburrini et al.50 The patients with indeterminate results in the study by Tamburrini et al50 were excluded from their primary analysis and underwent a repeat CT with contrast; retrospective review of the indeterminate noncontrast CT results agreed with the prospective reading of the contrast CT in 73% of observations (κ = 0.322).50 The study by Keyzer et al48 reported 6 CT interpretations as inconclusive and included these in the negative CT result group.48 Although the other 5 studies did not separately report the number of indeterminate CT readings, they did include indeterminate scans among their negative CT results. Patient withdrawals were adequately reported in only 2 studies.46,50

Sensitivity Analyses

According to the traditional SROC curve analysis,51 there was no evidence of a significant threshold effect (β = −0.3; 95% CI −1.9 to 1.3). The random-effects pooled estimates for sensitivity and specificity were 92.7% (95% CI 89.5% to 95.0%) and 96.1% (95% CI 94.2% to 97.5%), respectively. According to these summary estimates, the positive likelihood ratio = 24 and the negative likelihood ratio = 0.08. The SROC curve plot is presented in Figure 4. The small size and number of studies precluded any meaningful subgroup analysis, and the sensitivity analysis using the bivariate random-effects method failed to converge.
LIMITATIONS

There were a number of studies that were omitted from our systematic review after we were unable to confirm length of follow-up with the authors.\textsuperscript{13,35,39,40} If we had been able to confirm adequate follow-up, the addition of these studies to our meta-analysis may have affected our overall estimates of sensitivity and specificity.

The inconsistency of reporting inconclusive CT results is an important limitation because at least 1 study has demonstrated a 41% incidence of acute appendicitis with equivocal CT interpretations.\textsuperscript{41} In the only study within our systematic review that separately reported data on inconclusive CT scans, 24% were considered inconclusive with an associated likelihood ratio = 0.9, indicating that a CT result interpreted by the radiologist as equivocal should neither increase nor decrease the clinician’s pretest probability of disease. The important clinical question addressing how to treat the patient with an inconclusive CT is an area ripe for further investigation.

The quality assessment demonstrated that the majority of studies were prone to differential verification bias, which may have inflated the estimates for test accuracy.\textsuperscript{15} For example, if patients with a negative CT result and uneventful clinical follow-up actually had subclinical appendicitis that resolved without treatment, a false-negative result would have been missed. On the other hand, many would argue that cases of subclinical appendicitis that resolve without treatment are not patient-important outcomes. All 7 studies reported adequate blinding of the radiologist interpreting the CT. Although only 1 study clearly reported blinding of the reference standard interpretation, it is unlikely that single blinding had a significant effect on the overall results.\textsuperscript{15}

As described in the “Materials and Methods,” we did not formally test for the presence of publication bias, given the questionable validity of using funnel plots or statistical models to detect publication bias for diagnostic test meta-analysis.\textsuperscript{13} However, if publication bias were present, it would be expected that our estimates for test sensitivity and specificity would be overstated.
Acute appendicitis remains the most common cause of acute abdominal pain requiring urgent surgery. Despite the high prevalence of appendicitis, the diagnosis is still problematic and perforation can occur within 24 hours of the onset of symptoms. The lifetime incidence of appendicitis in the Western world is 6.7% for females and 8.6% for males, but the lifetime chance of appendectomy is 23.1% and 12.0%, respectively. Traditionally, a high negative appendectomy rate of 10% to 20% has been considered acceptable to minimize the number of missed cases of appendicitis. However, removal of a normal appendix is associated with an early complication rate of 7% to 13% and a late complication rate of 4%; hence, it is not a benign procedure. The clinical presentation of acute appendicitis is often atypical and may mimic other abdominal conditions, confounding its clinical diagnosis and resulting in a clinical diagnostic accuracy of only 60% to 80%. Thus, CT has become the favored diagnostic test for adults presenting to the ED with suspected acute appendicitis. Our pooled summary estimate for specificity would suggest that, when unenhanced CT is used, the acceptable negative appendectomy rate should be only approximately 4%

There are currently many CT protocols for the diagnosis of acute appendicitis, including noncontrast CT, intravenous or oral contrast, rectal contrast, and focused appendiceal imaging versus scanning through the entire abdomen and pelvis. The potential advantages of nonenhanced studies include earlier completion of imaging, thus avoiding delayed diagnoses and allowing more expedient care, which is especially important in the setting of ED crowding; avoidance of exposure to contrast materials; potential cost savings; and decreased patient discomfort and increased patient satisfaction. The usefulness of oral contrast is not clear, and there is considerable variability in oral contrast protocols with respect to type of oral contrast, quantity, and timing of administration. Earlier-generation CT scans may have mandated its use because the longer image acquisition times and artifact caused by respiratory and peristaltic motion led to image degradation. However, helical scanning and multidetector CT with rapid image acquisition times have made the use of oral contrast of questionable value. Factors that may contribute to improved accuracy of noncontrast CT interpretation include presence of intra-abdominal fat, increased severity of disease at image acquisition, thin sections (<5 mm), coronal and sagittal reformatting, and the experience of the interpreting radiologist. Difficulties in interpretation of CT may occur when normal, fluid-filled small bowel has the appearance of a dilated enlarged appendix. In addition, varying positions of the cecum and calcification in the right lower quadrant may also be misinterpreted. Diagnostic accuracy may be further compromised in patients with little abdominal and intrapelvic fat; however, one study that directly addressed this issue did not find a significant difference in CT accuracy with varying body mass index.

Part of the difficulty in comparing contrast and noncontrast CT is the degree of interobserver variability among radiologists. Simple agreement in nonenhanced CT and contrast-enhanced CT interpretation for the diagnosis of acute appendicitis ranges from 80% to 97% and varies with specific radiologic findings (eg, appendicolith versus wall thickening). Studies assessing interobserver variability report fair to excellent agreement according to a κ statistic. Of course, the experience and training of the radiologists may have an effect on the accuracy of interpretation.

Six of the 7 studies in this systematic review included radiologists as investigators and all support the use of noncontrast CT scans for the diagnosis of appendicitis. Ashraf et al state that a “certain level of experience is required for skillful interpretation” of noncontrast CT and have
established an imaging protocol at their institution that includes noncontrast CT for evaluating the appendix. Similarly, Lane et al. mentioned that as a result of their study on noncontrast CT for suspected appendicitis, their institution now considers noncontrast CT as an alternative to ultrasonography for diagnosing appendicitis. Therefore, it appears that there is already some degree of acceptance for using noncontrast CT scans among radiologists for the diagnosis of appendicitis.

Two recently published meta-analyses comparing CT and ultrasonography in the diagnosis of appendicitis reported similar results to ours, with respect to the ability of CT to rule out appendicitis. The negative likelihood ratio of 0.08 according to our SROC analysis was consistent with that reported by van Randen et al. Terasawa et al. reported summary estimates of 94% (95% CI 91 to 95%) for sensitivity, 95% (95% CI 93 to 96%) for specificity, and 0.09 for the negative likelihood ratio. Terasawa et al. observed that the test characteristics among the individual studies were similar despite variation in the use of contrast and CT technology but methodological limitations may have inflated estimates of diagnostic accuracy.

Our systematic review is unique in that it specifically focuses on the diagnostic accuracy of noncontrast CT in adult patients suspected of having acute appendicitis. Reviews by Terasawa et al. and van Randen et al. compare the use of CT to ultrasonography in diagnosis of acute appendicitis but include only 2 to 3 studies that assess the accuracy of noncontrast CT and also include pediatric populations. The systematic review by Anderson et al. focuses on whether oral contrast is necessary for the CT diagnosis of appendicitis in adults but includes rectal contrast in the noncontrast CT group.

As with any diagnostic test that does not have perfect sensitivity or specificity, CT cannot exclude appendicitis with 100% certainty and must be interpreted within the clinical context. Depending on the individual patient’s condition and circumstances, clinical judgment must be used when deciding to perform contrast-enhanced or unenhanced CT for suspected appendicitis. Our 7.3% summary estimate for the false-negative rate within the range of false-negative rates (3% to 17%) reported in a systematic review that included various CT contrast protocols. The ultimate goal of CT imaging in patients presenting with abdominal pain suspicious for appendicitis is to make a prompt diagnosis and decrease the rate of appendectomies performed on patients without appendicitis. Although some authors have reported a decrease in the rate of appendectomies performed on patients without appendicitis with the use of helical CT, others argue that there has been little change in the rate of surgical intervention or rate of perforation. Unfortunately, these observations were made using older technology and this type of outcome data has not been reported since the introduction of multislice CT. Despite the lack of data on long-term outcomes, our systematic review demonstrates high sensitivity and specificity for non-enhanced CT in the diagnosis of acute appendicitis and should be considered as an alternative to contrast CT, particularly in patients with contraindications to contrast such as those at risk for contrast-induced nephropathy.

The authors thank Patrick Bossuyt, PhD, for his recommendations about the bivariate method of analysis. The authors also thank Diana Delgado, AHIP, and Paul Albert, MLS, of the Weill Cornell Medical Library for their library and technical assistance. The authors acknowledge the following corresponding authors: Kashif Ashraf, MD, Shivanand Gamanagatti, MD, Marc D. Horton, MD, and Stefania Tamburrini, MD

**Supervising editor:** Allan B. Wolfson, MD

**Author contributions:** VH and JAD conceived the study and reviewed and assessed all relevant studies. VH, JAD, and MDB supervised the conduct of the systematic review and data collection. VH, JAD, and LF conducted the search of all relevant electronic databases, meeting abstracts, and bibliographies. ZJ and MDB provided statistical advice. VH, JAD, ZJ, and MDB analyzed the data. VH and JAD drafted the article, and all authors contributed substantially to its revision. VH takes responsibility for the paper as a whole.

**Funding and support:** By Annals policy, all authors are required to disclose any and all commercial, financial, and other relationships in any way related to the subject of this article that might create any potential conflict of interest. The authors have stated that no such relationships exist. See the Manuscript Submission Agreement in this issue for examples of specific conflicts covered by this statement.

**Publication dates:** Received for publication March 25, 2009. Revisions received May 20, 2009, and June 15, 2009. Accepted for publication June 24, 2009. Available online September 5, 2009.

Earn CME Credit: Continuing Medical Education is available for this article at: [http://www.ACEP.EMedHome.com](http://www.ACEP.EMedHome.com).

Reprints not available from the authors.

**Address for correspondence:** Veronica Hlibczuk, MD, Division of Emergency Medicine, Columbia University Medical Center, New York-Presbyterian Hospital, 622 West 168th St, PH1-137, New York, NY 10032; 212-305-2995, fax 212-305-6792; E-mail vhlibczuk@aol.com.

### REFERENCES


Noncontrast CT for Diagnosis of Adult Appendicitis

Hlibczuk et al


1. Appendicitis/
2. Appendectomy/
3. append$.tw.
4. or/1-3
5. exp Tomography, X-Ray Computed/
6. exp Contrast Media/
7. ra.fs.
8. computed tomography.tw.
9. cat scan$.tw.
10. ct.tw.
11. helical.tw.
12. enhanced.tw.
13. unenhanced.tw.
14. contrast$.tw.
15. or/5-15
16. 4 and 16
17. exp “Sensitivity and Specificity”/
18. sensitivity.tw.
19. specificity.tw.
20. (pre-test or pretest) adj probability).tw.
21. post-test probability.tw.
22. predictive value$.tw.
23. likelihood ratio$.tw.
24. or/18-24
25. 17 and 25

APPENDIX E2. Technical information.

The traditional SROC curve analysis was carried out with a regression model based on unweighted least-squares estimation, along with plots on the difference between the logit of the true-positive rate (TPR) and the logit of the false-positive rate (FPR) (D=logit TPR−logit FPR) on the y axis and the sum (S=logit TPR+logit FPR) on the x axis.1,2 The y axis (D) is equivalent to the log diagnostic odds ratio, and the x axis (S) is a measure of how the test characteristics vary with the test threshold. A regression equation (D=α+β×S) derived from the SROC curve analysis was used to assess for a threshold effect.4,5 A threshold effect was assumed to be absent if the β coefficient was near zero, implying a symmetric SROC curve.1,6 The sensitivity and specificity were pooled independently of each other. The logits of the sensitivity and its variance were calculated, and then the random-effects model of DerSimonian and Laird7 was used for pooling. The pooled logit sensitivity was then transformed back to the original representation. The pooled specificity was obtained similarly. If there was a value of zero in a 2×2 table, then a correction factor of one-half was added to each cell for the logit calculation to avoid numeric problems.1,7 The validity of using formal statistical tests for assessing heterogeneity in diagnostic meta-analysis has been recently questioned8; therefore, we did not calculate the I-squared statistic for the diagnostic odds ratio as originally described in our protocol.

REFERENCES