Accuracy of computed tomography in predicting appendiceal perforation

Jason D. Fraser, Pablo Aguayo, Susan W. Sharp, Charles L. Snyder, Douglas C. Rivard, Brent E. Cully, Ronald J. Sharp, Daniel J. Ostlie, Shawn D. St. Peter*

Department of Surgery, The Children’s Mercy Hospital, Kansas City, MO 64108, USA

Received 5 October 2009; accepted 6 October 2009

Key words:
Computed tomography; Perforated appendicitis; Accuracy; Pediatrics

Abstract
Background: Some surgeons use nonoperative management with or without interval appendectomy for patients who present with perforated appendicitis. These strategies depend on accurately delineating perforation by computed tomography (CT). Since 2005, our institution has used an evidence-based definition for perforation as a hole in the appendix or fecalith in the abdomen. This has been shown to clearly separate those with a high risk of abscess from those without. To quantify the ability of CT to identify which patients would meet these criteria for perforation, we tested 6 surgeons and 2 radiologists who evaluated blinded CT scans.

Methods: A junior and senior surgical residents, 2 staff interventional radiologists, and 4 attending pediatric surgeons with 3 to 30 years of experience reviewed 200 CT scans of pediatric patients who had undergone a laparoscopic appendectomy. All CT scans were reviewed electronically, and the reviewers were blinded to the results, outcome, and intraoperative findings. None of the patients had a well-formed abscess on CT. The reviewers were asked to decide only on perforated or nonperforated appendicitis according to our intraoperative definition. Clinical admission data were reviewed and compared between groups.

Results: In total, the reviewers were correct 72% of the time with an overall sensitivity of 62% and a specificity of 81%. The overall positive predictive value was 67%, and the negative predictive value was 77%.

Conclusions: This study shows that in the absence of a well-formed abscess, the triage of patient care based on a preoperative diagnosis of perforation from CT may be imprudent and subject a portion of the population to an unnecessarily prolonged course of care.

© 2010 Elsevier Inc. All rights reserved.

Nonoperative management for patients who present with perforated appendicitis is an established treatment plan. Treating the abdominal disease with antibiotics can be done with or without interval appendectomy [1-6]. However, these strategies depend on accurately delineating which patients have a perforation with the information available at presentation. The most accurate tool available for precisely identifying appendiceal pathology is computed tomography (CT) [7-10]. There have been many studies investigating which patients are at risk for failure of nonoperative management in this population [4-6]. However, it remains uninvestigated what percentage of patients would be triaged to nonoperative management on the basis of presumptive
perforation, when in fact they have no perforation. This group of patients may represent a misguided utilization of resources. There are 2 difficulties in assessing this possibility. First is the fact that most patients triaged to nonoperative management will not undergo early operation to decipher if perforation is present. The second is that in centers where early operation is standard, surgeons do not agree on what constitutes perforation. Our institution is unique in that we perform early appendectomy on any patient without a well-formed abscess, and since 2005, our institution has used a rigid definition for perforation as a hole in the appendix or fecalith in the abdomen. This has been shown to clearly separate those with a high risk of abscess from those without [11]. Because we have a large cohort of patients now who were found to meet these criteria intraoperatively, we conducted this study to precisely quantify the ability of surgeons and radiologists to identify the presence of perforation from the preoperative CT scan.

### 1. Methods

A file of 200 patients who had a preoperative CT scan and subsequently underwent appendectomy was compiled by our clinical research coordinator. All patients were treated from the time when we began enrolling in appendicitis trials under the definition of perforation as a hole in the appendix or a fecalith in the abdomen. The list comprised 101 patients with perforation and 99 without. Blinded to the previous radiologic read and operative findings, each CT was evaluated by a total of 8 different caregivers. The caregivers were asked to decide from these images whether the patient was found to have a diagnosis of perforated or nonperforated appendicitis for each CT according to our intraoperative definition. No imaging criteria or suggestions were given to reviewers to make this decision, but a column was available for comment to justify the decision if felt compelled. The 8 participants included 2 surgical residents, 4 attending pediatric surgeons, and 2 pediatric interventional radiologists. The surgical residents include a junior (resident 1) and senior (resident 2). The 4 attending pediatric surgeons have 3, 6, 15, and 30 years’ experience (surgeons 1-4, respectively).

#### Accuracy numbers were based on using the read of perforation as being a positive test.

Admission data recorded included age, weight, body mass index, days of symptoms, maximum temperature, and white blood cell (WBC) count.

Interrater variability was assessed with 2-way intraclass correlation with single measures.

### 2. Results

The results of each reviewer’s interpretation of the CTs are listed below in Tables 1 and 2. In total, the reviewers were correct 72% of the time. The overall sensitivity of CT on predicting appendiceal perforation was 62% and the specificity was 81%. The overall positive predictive value was 67% and negative predictive value was 77%. Three reviewers (attending 1 and radiologists 1 and 2) commented on the presence of an appendiceal wall defect in a total of 55 cases, 41 of which were found to be perforated at operation for an accuracy of 75%.

On admission, patients with perforated appendicitis had significantly higher temperatures ($P < .001$), higher WBC counts ($P = .01$), and higher body mass indexes ($P = .01$) and were older ($P = .02$). There was no difference in length of symptoms ($P = .4$) However, among the CT scans where perforation was missed or was incorrectly called, there was no significant differences found among the admission variables. Range in WBC count ($10^3$ leukocytes) was 3.2 to 30.3 for those with nonperforated appendicitis and 5.2 to 31.2 for those with perforation. Range in maximum temperature was 36.1 to 40.9°C for nonperforated patients and 36.3 to 40.6°C for perforated patients.

Cronbach $\alpha$ was .867 for the reviewer’s responses showing a level of agreement among responders.

<table>
<thead>
<tr>
<th>Table 1</th>
<th>Accuracy of CT interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Number missed</td>
</tr>
<tr>
<td>Junior resident</td>
<td>59</td>
</tr>
<tr>
<td>Senior resident</td>
<td>53</td>
</tr>
<tr>
<td>Surgery attending 1</td>
<td>48</td>
</tr>
<tr>
<td>Surgery attending 2</td>
<td>61</td>
</tr>
<tr>
<td>Surgery attending 3</td>
<td>61</td>
</tr>
<tr>
<td>Surgery attending 4</td>
<td>68</td>
</tr>
<tr>
<td>Radiology attending 1</td>
<td>39</td>
</tr>
<tr>
<td>Radiology attending 2</td>
<td>64</td>
</tr>
<tr>
<td>Totals</td>
<td>453</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 2</th>
<th>Accuracy of CT interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Sensitivity (%)</td>
</tr>
<tr>
<td>Junior resident</td>
<td>52</td>
</tr>
<tr>
<td>Senior resident</td>
<td>66</td>
</tr>
<tr>
<td>Surgery attending 1</td>
<td>75</td>
</tr>
<tr>
<td>Surgery attending 2</td>
<td>63</td>
</tr>
<tr>
<td>Surgery attending 3</td>
<td>64</td>
</tr>
<tr>
<td>Surgery attending 4</td>
<td>53</td>
</tr>
<tr>
<td>Radiology attending 1</td>
<td>78</td>
</tr>
<tr>
<td>Radiology attending 2</td>
<td>45</td>
</tr>
<tr>
<td>Totals</td>
<td>62</td>
</tr>
</tbody>
</table>
3. Discussion

Nonoperative management of perforated appendicitis is predicated upon accurate and reliable interpretation of the available imaging. Computed tomography has been shown to be accurate in diagnosing acute appendicitis, with most reported sensitivities and specificities ranging around 90% [12-14]. There has also been a concordant decrease in the negative appendectomy rates from the previously accepted 15% to 20% to 2% to 12% with the use of CT as a diagnostic adjunct [12-18]. Although it is clear that CT is excellent at diagnosing nonperforated appendicitis, the demarcation of perforated from nonperforated appendicitis is not perfect.

The difficulty in distinguishing perforated from nonperforated appendicitis lies within the spectrum of manifestations of perforation as seen on CT. The presence of an appendicolith with intraluminal appendiceal air, extraluminal air, bowel wall thickening, ileal wall enhancement, extraluminal appendicolith, abscess, phlegmon, periappendiceal inflammatory stranding, and free fluid have all been demonstrated in varying degrees to be features consistent with CT evidence of perforation [19-24]. This range of radiographic signs makes accurate interpretation of perforation variable. Other reports have examined the use of CT scan for perforation, but the accuracy is based on the final pathologic diagnosis or operative findings without explicit and validated criteria [13,19-22,24,25,26]. Although such reports lay the framework for future studies, they are inherently flawed by a lack of consistent definition for perforation. Without a standardized definition of perforation, it would be predictable to demonstrate high concordance with a preoperative diagnosis of perforation from preoperative variables because patients with purulent peritoneal fluid without actual perforation can be arbitrarily termed perforated during the operation. We have used an evidence-based definition for perforation, which has separated those at risk for abscess from those without. Because we know the risk of abscess is extremely low (0.6 %) in the absence of identifying a hole in the appendix or a fecalith in the abdomen at the time of operation [11], we are concerned about patients being treated with interval appendectomy on the basis of imaging when they could go home postoperative day 1 without antibiotics. This study shows that typical CT scans obtained in everyday practice reviewed by surgeons or radiologists cannot perfectly delineate which patients have the appendiceal disease that will put them at risk of abscess and thus require a longer course of care.

We found that CT has a sensitivity of 62% with a specificity of 81% in predicting appendiceal perforation. The specificity varied on how many patients the examiners were willing to call perforated. As can be seen from the specificity, there was a substantial risk of calling a patient perforated when there was no perforation at operation. The percentage of patients who were not perforated, but appeared to be so on CT images, ranged from 11% to 38%. The providers who had the highest specificity did not have the highest accuracy. An example is radiologist 2, who was very strict on declaring an image perforated and called perforation in only 59 patients, despite the knowledge that they were intended to be evenly distributed. As a result, his accuracy was low, but the clinically important variable of specificity was quite good with only 12% false positive. To commit a patient to an extended treatment course of nonoperative management with a long antibiotic course and interval appendectomy based on CT evaluation alone, one must be sure that the result is not a false positive. Our data found that even a radiologist with a conservative standard for perforation was wrong approximately 10% of the time [26].

One retrospective study found that using a defect in the enhancing appendiceal wall as the sole CT finding to determine perforation increased the sensitivity, specificity, and accuracy to 95.0%, 96.8%, and 96.1%, respectively [27]. However, this was allowed by consistent, high-quality CT scans, and our data represent the full spectrum as many patients are transferred with outside CT scans. Furthermore, without a fixed definition of perforation, it is difficult to place meaning on studies evaluating prediction of perforation because the threshold to call the appendix perforated may be low resulting in high concordance with preoperative variables. Although in our study, we did not set out to look for mural defect, nor did we give the reviewers criteria for deciding on perforation, there was a column for comments. In this column, 3 reviewers commented on a mural defect. When a specific comment was made that the reviewer felt a mural defect was present, there was an accuracy of 75%. This was unchanged from their aggregate overall accuracy of 75%. This suggests that when there is heterogeneity in the quality of CT and the surgeons and radiologists specifically look for the mural defect under a fixed definition of perforation, then the appearance of a mural defect on imaging does not accurately predict that a tangible hole in the appendix will be found at operation.

There was a significant difference in admission characteristics between the patients who had perforated compared with nonperforated appendicitis. As one might expect, those with perforation as a population had higher temperatures and WBC counts. However, there were no significant differences when looking at those with a CT read incorrectly. Therefore, it seems that adjunctive admission data are less helpful in the tougher cases to delineate perforated from nonperforated preoperatively. In addition, the patients that have an obvious diagnosis of perforated or nonperforated on CT are likely to represent the extremes of data on admission, thus creating the separation in means seen in the whole population. This was emphasized by the lack of random error and dominant component of systematic error detected by the intraclass correlation coefficient, which shows a high rate of correlation between reviewers. This finding might be expected because the answer of perforation is determined during the operation, not on imaging, so there should be more systemic error. This simply shows that surgeons and
radiologist will often agree on the misrepresentation the CT manifests. The ranges in temperature and WBC and ABO are also identical between those who were found to have a perforation and those who were not; therefore, we are unable to recommend thresholds in these variable where a patient can be declared perforated.

This study shows that in the absence of a well-formed abscess, the triage of patient care based on a preoperative diagnosis of perforation from CT may be imprudent and subject a portion of the population to an unnecessarily prolonged course of care.

References


Discussion

Max R. Langham, MD, Memphis, TN: We have shared your group’s interest in appendicitis. Reagan Williams recently published a study in the Journal of the American College of Surgeons that was about the same size as yours but looked at a number of other variables and created a score, which had a much better ROC and performance than CT scan alone. CT was one of a number of variables, and it was pleasant to find that generalized peritonitis on examination was one of the most important variables, so the physical examination is an important part of this.

My question to you and your colleagues is how you think, given that many of us employ delayed therapy for appendicitis in patients who are perforated, but do not have an abscess, this information fits in what is really a multifaceted evaluation of a patient, and whether or not CT scan and in of itself is what we should be depending on.

Dr Jason D. Fraser, MD, Kansas City, MO: I think that is a very valid point. We actually did look at the admission characteristics of all of the patients. Of course patients that had perforation had higher WBC counts, temperatures, and age, but among the patients in whom the CT scans missed the diagnosis, there was not a difference between any of the admission characteristics between the patients. To me that suggests that in patients with an equivocal CT, those patients are the ones where not only CT but also the
admission characteristics of white blood cell count, temperature, etc, don’t help much with the diagnosis. I think CT scan is a part of making the diagnosis of perforation, but I think it needs to be taken in context with the entire clinical situation as well.

Doug Barnhardt, Salt Lake City, UT: I think some of us would make a distinction that all perforated appendicitis isn’t equal, and if you are asking clinicians where there is free fluid to tell you whether or not it is perforated, I think that is a different question than an abscess or a phlegmon, which would be the situation where many people would recommend nonoperative management. Did you subclassify these groups? If you take a group of patients where there is some free fluid in the pelvis and ask people to tell you whether or not that is perforated, that is probably a different clinical question than many of us would be sorting out in terms of whether are we in fact recognizing instances of well-formed abscesses and phlegmons versus just early perforated appendicitis.

Dr Fraser: Thank you for the question. In our series, of the 101 patients with perforated appendicitis, there was a wide gamut of patients. Some of them did have a well-formed abscess. Some of them just had free fluid. Some of them had a phlegmon, and I think that is where some of the variation comes from. Of course it is much more difficult at that time to diagnose a perforated appendicitis just on CT scan.