Assessing the Accuracy of Diagnostic Imaging for Pediatric Appendicitis During the Course of Illness

Gregory P. Conners, MD, MPH, Lisa L. Schroeder, MD

From Children's Mercy Hospitals and Clinics and the Departments of Pediatrics (Conners, Schroeder) and Emergency Medicine (Conners), University of Missouri, Kansas City School of Medicine, Kansas City, MO.

0196-0644/$-see front matter

http://dx.doi.org/10.1016/j.annemergmed.2012.07.001

A **podcast** for this article is available at www.annemergmed.com.

SEE RELATED ARTICLE, P. 582.

[Ann Emerg Med. 2012;60:603-604.]

The goal of Bachur et al[1] in their article is to determine "the test performance characteristics of CT [computed tomography] and ultrasonography according to the duration of abdominal pain in children being assessed for appendicitis." The authors wondered whether advanced imaging is less accurate in children presenting early in the course of their illness.

But wait! There are many articles already reporting the sensitivity, specificity, and so forth of CT and ultrasonography for pediatric appendicitis. Certainly, results like this will vary over different studies and we need to study large numbers of children to get reliable values. However, as Bachur et al[1] cite, in 2006 Doria et al[2] published a meta-analysis of 26 of these studies and found that the pooled sensitivity and specificity of CT scan for appendicitis were 0.94 (95% confidence interval [CI] 0.92 to 0.97) and 0.95 (95% CI 0.94 to 0.97), respectively, whereas for ultrasonography the pooled sensitivity and specificity were 0.88 (95% CI 0.86 to 0.90) and 0.94 (95% CI 0.92 to 0.95), respectively. Their analysis reflected experience with thousands of children. With numbers like that, how could the current study add anything new?

Readers may recall coming away from their epidemiology course with an impression like this: although positive predictive value and negative predictive value change according to the prevalence of the disease in the population being studied, sensitivity and specificity stay constant. This impression is easy to confirm; for example, in their recent article, Lalkhen and McCluskey[3] told us that sensitivity and specificity are "independent of the population of interest subjected to the test," whereas, "unlike sensitivity and specificity, the positive predictive value and negative predictive value are dependent on the population being tested and are influenced by the prevalence of the disease." With this understanding, we might conclude that any variation that Bachur et al[1] found in sensitivity or specificity at various points in time was just the result of random variation and should wash out if we studied large enough numbers of patients. The truth about these test characteristics, though, is somewhat more subtle; even in standard textbooks, only footnotes or half-sentences suggest otherwise.[4-6] In fact, although their constancy is often a fairly accurate approximation, as Gallagher[7] explained in *Annals* in 1998, "sensitivity and specificity . . . may become unstable if disease severity in the study population shifts concurrently with a change in prevalence." In other words, although positive predictive value and negative predictive value almost certainly will vary with changes in prevalence and other study population characteristics, it is possible that sensitivity and specificity may vary as well.[8,9] Thus, if patients with longer duration of abdominal pain have increasingly "worse" appendicitis or are likelier to have appendicitis, both of which could be true, not only positive predictive value and negative predictive value but also sensitivity and specificity of imaging could potentially improve during this time.

As the authors have shown, the sensitivity and specificity of CT scanning for appendicitis in children were well above 90% even early in the course of abdominal pain and remained constant even into the third day of abdominal pain.[1] Similarly, when equivocal cases were not included, the specificity of ultrasonography remained 95% or better throughout the duration of abdominal pain. Although at least theoretically they could have increased, they did not. This makes sense; given how well these tests performed even early in the course of illness, there was simply not much room left for improvement, although the probability of an equivocal CT scan result did diminish with increased pain duration.

However, the sensitivity of ultrasonography for appendicitis increased significantly as the length of abdominal pain extended. This finding is important in 2 distinct ways. First, it instructs clinicians caring for children with abdominal pain that negative ultrasonography study results are less reliable for detecting appendicitis early in the course of illness. This is unfortunate because most patients (62% in this study[1]) present in the first 24 hours after developing abdominal pain and ultrasonographic study results of these children will usually be negative. As Bachur et al[1] told us, too many of these results will be false negatives. Thus, imaging strategies for many children must include CT scanning, serial ultrasonographic studies, forgoing imaging altogether, or some other form of imaging. Second, this finding reminds us that

studies reporting just 1 value for a test's sensitivity or specificity may be incomplete; patient and disease factors vary during the course of disease, and so may these important test performance characteristics. This is likely fairly common, largely underappreciated, and often ignored in study designs. The authors of this study are to be congratulated for analyzing the data from their large-scale, multicenter study to systematically look for and ultimately find important changes in the test performance characteristics of diagnostic imaging for pediatric appendicitis.

Supervising editor: Steven M. Green, MD

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 as per ICMJE conflict of interest guidelines (see www.icmje.org). The authors have stated that no such relationships exist.

Publication date: Available online July 27, 2012.

Address for correspondence: Gregory P. Conners, MD, MPH, E-mail: gpconners@cmh.edu.

REFERENCES

1. Bachur R, Dayan PS, Bajaj L, et al. The effect of abdominal pain duration on the accuracy of diagnostic imaging for pediatric appendicitis. *Ann Emerg Med*. 2012;60:582-590.
2. Doria AS, Moineddin R, Kellenberger CJ, et al. US or CT for diagnosis of appendicitis in children and adults? a meta-analysis. *Radiology*. 2006;241:83-94.
3. Lalkhen AG, McCluskey A. Clinical tests: sensitivity and specificity. *Continuing Educ Anesth Crit Care Pain*. 2008;8:221-223.
4. Riegelman RK. Studying a Study & Testing a Test. 5th ed. Philadelphia, PA: Lippincott Williams & Wilkins; 2005.
5. Sackett DL, Haynes RB, Guyatt GH, et al. Clinical Epidemiology: A Basic Science for Clinical Medicine. 2nd ed. Boston, MA: Little Brown; 1991.
6. Hulley SB, Cummings SR, Browner WS, et al. Designing Clinical Research. 2nd ed. Philadelphia, PA: Lippincott Williams & Wilkins; 2001.
7. Gallagher EJ. Clinical utility of likelihood ratios. *Ann Emerg Med*. 1998;31:391-397.
8. Ransohoff DF, Feinstein AR. Problems of spectrum and bias in evaluating the efficacy of diagnostic tests. *N Engl J Med*. 1978; 299:926-930.
9. Brenner H, Gefeller O. Variation of sensitivity, specificity, likelihood ratios and predictive values with disease prevalence. *Stat Med*. 1997;16:981-991.