Critical Review Form
Meta-analysis

US or CT for Diagnosis of Appendicitis in Children and Adults? A Meta-Analysis
Radiology 2006; 241:83-94

Objective: **"To perform a meta-analysis to evaluate the diagnostic performance of US and CT for the diagnosis of appendicitis in pediatric and adult populations". (p. 84)**

Methods: **Two investigators conducted independent literature searches of MEDLINE. EMBASE, CINAHL, Cochrane, and ACP Journal Club using the following search terms "appendicitis", "appendix", "sonography", "ultrasonography", "computed tomography", and "computed tomography scan". They also conducted a manual search of selected articles reference lists.**

Criteria for individual study inclusion were:

a) **Prospective or retrospective designs that used either histopathology or follow-up criterion standards for the diagnosis of abdominal CT or US to evaluate for appendicitis;**
b) **Availability of data to reproduce diagnostic 2x2 tables;**
c) **Segmentation of results into adults and children;**
d) **Explicit imaging criteria to define appendicitis (i.e. appendix diameter, compressibility, appendicolith, stranding, etc;**
e) **Inclusion of males and females;**
f) **Description of CT technique used;**
g) **Appendicitis prevalence 15% - 75%.**

No language restrictions were applied. Exclusion criteria included unsystematic observation (case reports, letters), pregnant women, cost-effectiveness or decision-analysis designs, or performance of serial imaging exams (report US or CT).

When data reported was insufficient to reproduce 2x2 tables, the original investigators were contacted. If they provided the data within a pre-specified time frame, their results were incorporated into this meta-analysis. If meta-analysis author computed sensitivity/specificity differed from the published values than the published values were used.

These investigators independently rated the evidence quality using previously described checklists (Irwig 1995, Mulrow 1989). They assessed rating reliability using the intraclass correlation coefficient. They also planned *a priori* to evaluate summary diagnostic performance based upon three recognized study design flaws: criterion standard (surgery alone or surgery vs. f/u); performance of both US & CT or only one study or the other; how equivocal cases were interpreted (equivocal = time – positive being best-case scenario vs. equivocal = false-positive or false-negative being worst case scenario).

After ensuring sufficiently small inter-study variance by the -2 techniques, the investigators derived pooled diagnostic statistics weighted by quality scores using the Mantel-Haenszel fixed-effects model (Higgins 2002). They conducted one-way sensitivity analyses removing outliers and also reported point estimates to compute diagnostic odds ratios and summary receiver operating characteristic (ROC) curves. Meta-regression was used to assess the impact of three confounding variables: study design (retrospective vs. prospective), year of publication (pre-1999 vs. post-1999), continent of origin (N. America vs. non N. America). Funnel plots were visually inspected to assess publication bias.

Guide	Question	Comments
I	***Are the results valid?***	
1.	Did the review explicitly address a sensible question?	Yes – what does the summation of published evidence conclude about the diagnostic preference of CT and US for the diagnosis of pediatric and adult appendicitis?
2.	Was the search for relevant studies details and exhaustive?	Yes. The investigators could have conducted a search of the gray literature, contacted industry and content experts for further (unpublished) research, but their search strategy was well-described and reasonably thorough.
3.	Were the primary studies of high methodological quality?	No. "The median score for the studies of children was 34.4% (maximum score, 32 points or 100%), and the median score for the studies of adults was 42.2%". (p. 86)
4.	Were the assessments of the included studies reproducible?	Yes. " The interrater agreement between the one unblinded reviewer and the two blinded reviewers (intraclass correlation coefficient, 0.78; 95% CI: 0.64, 0.87) and between the two blinded reviewers (intraclass correlation coefficient, 0.70; 95% CI: 0.46, 0.84) was good". (p. 86)

II.	*What are the results?*	
1.	What are the overall results of the study?	• 229 articles were screened and 57 met inclusion criteria (26 pediatric, 31 adult studies) published between 1988 – 2004 • The mean sample prevalence of appendicitis was 31% for pediatric studies (US and CT) versus 48% for adult US studies and 40% for adult CT studies. • Among the studies that reported perforation rates in children (n =10 studies) and adults (n =3 studies) the weighted perforation rates were 26.5% in kids and 18.5% in adults. • All but two studies used 3rd generation helical CT scanners. • The pooled sensitivity, specificity and DOR for US, CT, or US and CT for pediatric or adult populations were computed when outliers were removed.

	Sen (95% CI)	Spec(95% CI)	DOR (95% CI)
Pediatric			
US	88(86-90)	94(92-95)	202 (159-358)
CT	94(92-97)	95 (94-97)	239 (118-487)
US & CT			46 (32-67)
Adult			
US	83(78-87)	93(92-96)	15 (10-21)
CT	94(92-95)	94(94-96)	118 (85-165)
US & CT			100 (57-167)

• The meta-regression model indicated that CT had significantly better diagnostic sensitivity in adults (OR 3.1, p<0.001) and children (OR 2.5, p=0.02) but no difference in specificity between CT and US in either population.
• Study design, year and continent of publication did not significantly change results.
• No heterogeneity or publication bias was recognized.
• Assuming an appendicitis prevalence of 15% than the number of cases of missed appendicitis by using US rather than CT would be 10 cases per 1000 children imaged and 18 cases per 1000 adult imaged.
• Conversely, if the prevalence of appendicitis is 75% than the number of cases of missed appendicitis by using US either than CT would be 48 per 1000 children imaged and 83 per 1000 adult imaged. (p. 86)

2.	How precise are the results?	Precise enough. See 95% CI above that do not cross one (the line of indifference)
3.	Were the results similar from study to study?	"Meta–regression analysis failed to show any evidence of heterogeneity related to the potential confounders on comparison of studies by means of dichotomization of the covariates into prospective versus retrospective (study design), before 1999 versus 1999 or later (year of publication), and North America versus non–North America (continent of study origin)". (p. 87)
III.	***Will the results help me in caring for my patients?***	
1.	How can I best interpret the results to apply them to the care of my patients?	" Assuming a hypothetical sample prevalence of 0.31 for diagnosis of appendicitis in children, for every 10,000 children 11 years of age scanned with US rather than with CT, 280 would have a missed diagnosis of appendicitis and 13 could be prevented from developing cancer in the future. On the other hand, if we considered 10 000 adults 35 years of age scanned with US rather than with CT in a center with a sample prevalence of 0.40, the diagnosis of appendicitis would be missed in 480 patients, but only two patients could be prevented from developing cancer in the future.". (p. 90) Therefore, two general considerations. First, **CT should be increasingly preferred with higher appendicitis prevalence (i.e. in those with a higher pre-test probability)**. Second, **future cancer risks are much smaller in adults so the CT-test threshold probably ought to be set lower than in children**. But how much lower and where should that threshold be set?
2.	Were all patient important outcomes considered?	No assessments of patient references, priorities or satisfaction with different imaging strategies are reported.
3.	Are the benefits worth the costs and potential risks?	Yes. Although the specificity of US for appendicitis in adults and children does not differ significantly from CT, the sensitivity of CT is significantly higher in both populations. The difference between false-negative rates between CT and US is accentuated by higher prevalence rates so institutional standards may elect to use CT-first strategies among higher risk subsets.

Limitations

1) **Potential verification and selection bias (Sackett 1979, Lijmer 1999) since most surgeons were aware of imaging results while making management decisions. These biases can even over-estimate sensitivity and under-estimate specificity.**
2) **Lack of detail about illness severity or pre-test probability by which to judge spectrum bias.**
3) **No detail about BMI or obesity rates by which to extrapolate US results to one's institutional population.**
4) **No details about training or experience of CT or US personnel by which to gauge external validity.**
5) **Lack of randomized diagnostic trials (Lord 2006)**
6) **No reference to QUADAS guidelines for diagnostic meta-analyses.**

Bottom Line

The specificity of US for appendicitis does not significantly differ from CT in children or adults. However, the sensitivity of CT is significantly higher than US in both populations and the number of false-negative studies increase as appendicitis prevalence increases. Future research will need to better define the disease probability threshold whereby CT is the preferred first line imaging study in adults and children to simultaneously minimize false-negatives and future cancer rates in the emergency diagnosis of appendicitis.