Critical Review Form

Clinical Decision Analysis

An Interdisciplinary Initiative to Reduce Radiation Exposure: Evaluation of Appendicitis in a Pediatric Emergency Department with Clinical Assessment Supported by a Staged Ultrasound and Computed Tomography Pathway

Acad Emerg Med 2009; 16:1258-1265

Objective: To retrospectively evaluate a multi-disciplinary pathway using a staged US and CT pathway to maintain the sensitivity and specificity of CT while reducing overall pediatric radiation exposures.

Methods: Single-center chart review for all ED patients presenting between 6PM – 6AM or on weekends/holidays between January 2003 and December 2008. Inclusion criterion was simply US-ordered as first-line imaging study for "rule out appendicitis". Exclusion criteria included patients brought directly to the OR, outside hospital CT, or inability to obtain CT (refusal, pregnancy).

Three-investigators conducted the chart review although specific methods are not well described (Gilbert 1996 or Worster 2004 – see Limitations). The chart review occurred during a period of time following the development of an interdisciplinary (EM, Surgery, and Radiology) pathway for the diagnosis of appendicitis (US first, CT only, equivocal US cases) although clinicians were not compelled to follow the algorithm. Additionally, no protocol was recommended for sonographers.

The following definitions were used:

Positive US = visualization of ≥ 6mm diameter non-compressible appendix with or without appendicolith, periappendiceal fluid, or increased appendix wall flow on Doppler.

Negative US = complete visualization of compressible appendix < 6mm or establishment of alternative diagnosis.

Equivocal US = Non-visualization of the entire appendix.

Positive appendicitis = pathology report of appendicolith or appendicitis.

The initial (not final) radiology report was used in this analysis. Two investigators conducted an inter-rater reliability of the radiology report interpretation (positive, negative, equivocal). Outcomes assessed included negative appendectomy rate, missed appendicitis rate, and number of CTs potentially reduced.

Guide		Comments
I.	**Are the results valid?**	
A.	**Were all important strategies and outcomes included?** *In other words, did the authors consider every potential course of action and possible outcome?*	No. Emergency physician performed US was not assessed.
B.	**Was an explicit and sensible process used to identify, select, and combine the evidence into probabilities?** *In other words, the authors should perform as comprehensive a literature review as is required for a meta-analysis. In addition, probabilities must be assigned to each branch emanating from a chance node, and for each chance node, the sum of probabilities must add to 1.0.*	No, the investigators did not conduct a systematic review of the US or CT diagnostic literature for appendicitis and did not report probabilities for each decision tree.
C.	**Were the utilities obtained in an explicit and sensible way from credible sources?** *Utility represents the value to the patient of remaining expected life. A utility threshold of 0.92 means that your patient feels he would be willing to sacrifice 8% of his/her remaining life to avoid that limb of the decision tree (going on dialysis, taking Coumadin, etc.).* *In other words, were the quantitative measurements of the value to the decision maker of the various outcomes provided by someone who understands the outcomes and the condition being rated? Whatever the measurement method, the authors should report the source of the ratings. In a decision analysis built for an individual patient, the most credible ratings are those measured directly from the patient.*	No utilities were developed. In fact, it is unclear which perspective would reflect benefit for this algorithmic approach to pediatric appendicitis. Perspectives one might consider would include • Patient • Provider • Payer • Society
D.	**Was the potential impact of any uncertainty in the evidence determined?** *Much of the uncertainty in clinical decision making arises from the lack of valid evidence in the literature. Even when present, published evidence is often imprecise with wide confidence intervals around estimates for important variables. Sensitivity analysis asks the question "Is the conclusion generated by the decision analysis affected by the uncertainties in our*	No. The investigators did not report precision (CI) or conduct a sensitivity analysis to test the assumptions of their algorithm. **Assumption #1**: CT radiation is dangerous

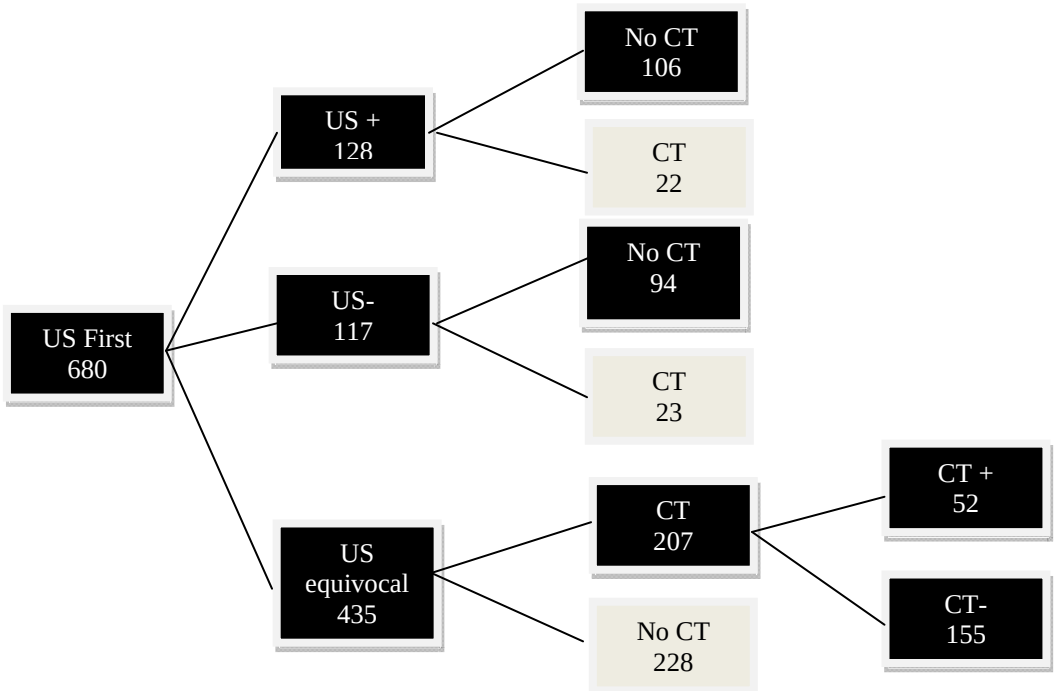
	estimates of the likelihood of the outcomes?" Satisfy yourself that all of the clinically important variables were included.	**Assumption #2**: US is available, accurate, and reliable.

II.	What are the results?	
A.	**In the baseline analysis, does one strategy result in a clinically important gain for patients? If not, is the result a "toss-up"?** *For a clinical decision analysis that compares two clinical strategies, there are three possible results: strategy 1 is better than strategy 2, strategy 2 is better than strategy 1, or both strategies are equally good or bad. A gain in life expectancy or quality-adjusted life expectancy of 2 or more months is considered an important gain.*	In this case we are trying to avoid two competing negative outcomes: missed appendicitis and CT radiation exposure. Theoretically, reducing CT radiation exposure (by reducing CT ordering) will increase missed appendicitis rates. These investigators demonstrate that an US-first protocol reduces CT ordering without increasing missed appendicitis rates.
B.	**How strong is the evidence used in the analysis?** *Ideally, every probability estimate at every node in the tree is supported by precise estimates from primary and integrative studies of high methodological quality. The fewer the probabilities that can be precisely estimated from high quality primary studies, the weaker the overall inference one can make from the results.*	Only one other prospective study of 139 children is presented to support this imaging algorithm.
C.	**Could the uncertainty in the evidence change the result?** *For any clinical variable the decision analyst can calculate the value or "threshold" above which the results favor one strategy and below which the results favor another strategy. If the result of the analysis would change by choosing different values for one of the variables, the result is said to be "sensitive" to that variable.*	Yes. If the lower limit of the 95% CIs are sufficiently low to suggest unacceptably high false-negative rates one would access further evidence. Furthermore, the investigators did not obtain a CT or operate on the majority of these subjects, and they relied upon self-return to the same hospital captured by their unspecified chart review strategy to identify all cases of "true-negatives". These are a lot of assumptions and a sensitivity analysis testing how robust their algorithm would be if they misidentified patients would be useful.

III.	**Will the results help me in caring for my patients?**	
A.	**Do the probability estimates fit my patients' clinical features?** *If the analysis was intended for patients different from yours, review the results of the sensitivity analyses. If the clinical characteristics of the intended patients are different from yours, you should discard the results. If a clinical decision analysis shows that the preferred strategy is sensitive to a given variable, you will need to gauge where your patient fits on the scale of that variable.*	Investigators do not report any sensitivity analysis so uncertain how this staged pathway would work with less qualified ultrasonographers or more obese patients. The results of this research may be unique to their institution. Without assessing potential confounding variables readers cannot be confident that these results would extrapolate to their clinical setting (external validity).
B.	**Do the utilities reflect how my patients would value the outcomes of the decision?** *You must consider whether your patient's values are similar to those used in the decision analysis. If you were to ask your patient to rate the outcome states using the rating instrument in the article, you would know exactly what utility values to use.*	Patients' values are not incorporated into this analysis.
C.	**Can I use the results in the management of patients in my practice?**	• Investigators identified 680 subjects with a complete data set for analysis managed as illustrated in Fig 1 below. • Excluding the 435 equivocal US cases gives the following results for US. Appy + - sen 99 (95 – 99) US spec 86 (82 – 87) + 91 15 LR+ 7 (5 – 8) - 1 93 LR - 0.013 (0.002 –0.061) • Looking at the overall pathway gives the following diagnostic performance (p 1263) Appy + - sen 99 (96 – 100) Pathway + 134 24 spec 85 (82 – 85) LR+ 11 (9 – 12) - 1 248 LR- 0.008 (0.001-0.04)

		• 52% (228/435) of patients with an equivocal US did not receive further CT imaging: 4% went to OR, 53% improved and were discharged with the diagnosis of abdominal pain not otherwise specified and 43% were given alternative diagnoses. • 60% of patients had imaging according to the pathway (407/680) and had a prevalence of appendicitis 33% with 0.5% (1/407) missed appendicitis rate and a 7% negative appendectomy rate. • Among the 407 patients managed by the pathway 200 had definitive US (106 positive, 94 negative) so 49% of CT's could have been avoided (200/407). • Among the 45 who had CT after definitive US, 41/45 were concordant with US of which US had been inaccurate in two and CT was inaccurate in the other two so CT did not alter management in 43/45 cases.

Figure 1. Staged ultrasound (US) and computed tomography (CT) pathway: imaging results of patients undergoing US followed by CT scans. Patients who followed the pathway are in black. The 228 patients in gray did not receive a CT scan following an equivocal US



Limitations: - Significant

1) **Chart reviews with no methods. Many unanswered questions which could impact the validity of results. Were data abstractors blinded to study hypothesis? How did abstraction QA occur? How were discrepancies resolved? Was a standardized data abstraction form used?**

2) **No discussion of Radiologist experience (performing or interpreting US). Years in residency, # prior appendix US, etc. Also no mention of how many different Radiologists performed US.**

3) **No assessment of subject weight or BMI.**

4) **No assessment of time-to-diagnosis or treatment delays.**

5) **No assessment of patient or clinician satisfaction with the algorithm or the two imaging modalities.**

6) **No reporting of LR's or 95% CI.**

7) **No follow-up of the 263 (39% of the cohort!) who did not have operative management and were assumed "true negatives" because they did not return to the same hospital. How many hospitals are in that area? How were re-visits ascertained and what was the average period of time between ED evaluation and end of the follow-up interval?**

8) **Single-center with exceptional cooperation between EM, surgery and Radiology. Could this process be replicated at other institutions with less congenial relationships (external validity)?**

Bottom Line:

This significantly flawed study suggests that a multi-disciplinary, staged imaging protocol for pediatric appendicitis reduces CT rates without increasing perforation or missed appendicitis rates or increasing negative laparotomy rates. Prospective evaluation at heterogeneous ED settings incorporating patient BMI and co-morbidity, radiology delays and expertise, staff acceptability, and time-to-diagnosis as well as patient satisfaction will be needed before this protocol can be confidently advocated. Future studies will need a valid surrogate outcome for the criterion standard when pathologic specimens are not available.