

Critical Review Form

Clinical Practice Guidelines

Selective Imaging Strategies for the Diagnosis of Appendicitis in Children, *Pediatrics*
2004: 13:24-28

Objective: “To define and test selective imaging guidelines based on risk stratification to increase diagnostic accuracy and reduce unnecessary testing for children with suspected appendicitis”. (p. 25)

Methods: Using a retrospectively identified cohort of children ages 3 – 21 years presenting to Children’s Hospital Boston with equivocal presentations for acute appendicitis between January 1996 – December 1999 the authors identified low – medium-, and high-risk appendicitis populations using recursive partitioning analysis. Appendicitis patients were identified by ICD-9 codes for appendicitis, perforated appendicitis, appendectomy, and abdominal pain. Using a standardized data abstraction tool a single reviewer obtained chart data on symptoms, clinical examination findings, radiographic studies and pathology reports.

Recursive partitioning analysis incorporated the following variables into risk stratification tools: age, gender, hours of abdominal pain, nausea or vomiting, diarrhea, anorexia, temperature $>38^{\circ}\text{C}$, RLQ tenderness, rebound, guarding, rectal tenderness, stool occult blood, WBC $>10,000/\text{mm}^3$, neutrophil count, and bands $>5\%$. Cross-validation was conducted statistically, but not on a new cohort. Missing a case of appendicitis was weighted as 10-times worse than diagnosing appendicitis in a child that did not have it for minimum cost-tree analysis. The following risk groups were defined by the above analysis:

Low-risk: neutrophils $< 67\%$, bands $< 5\%$, no guarding.

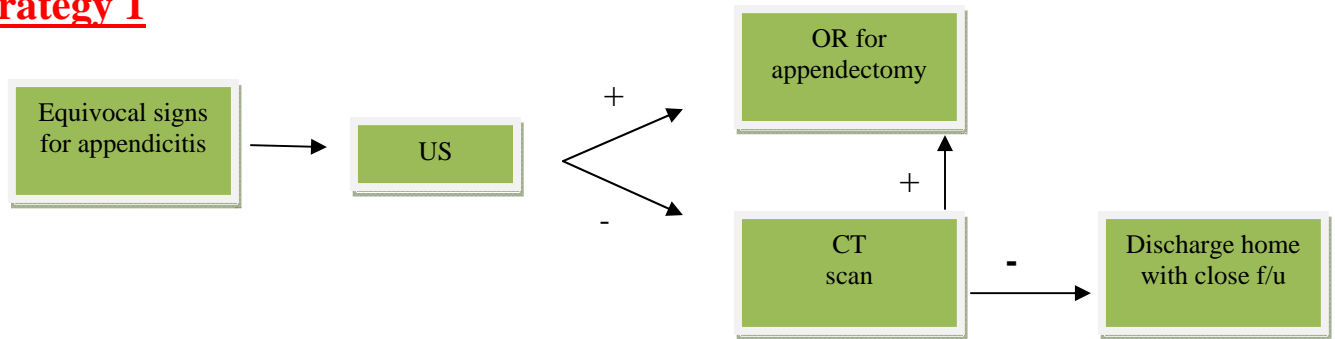
High-risk: neutrophils $> 67\%$, WBC $> 10,000$, guarding, abdominal pain > 13 hours.

Medium-risk: Neither low nor high risk.

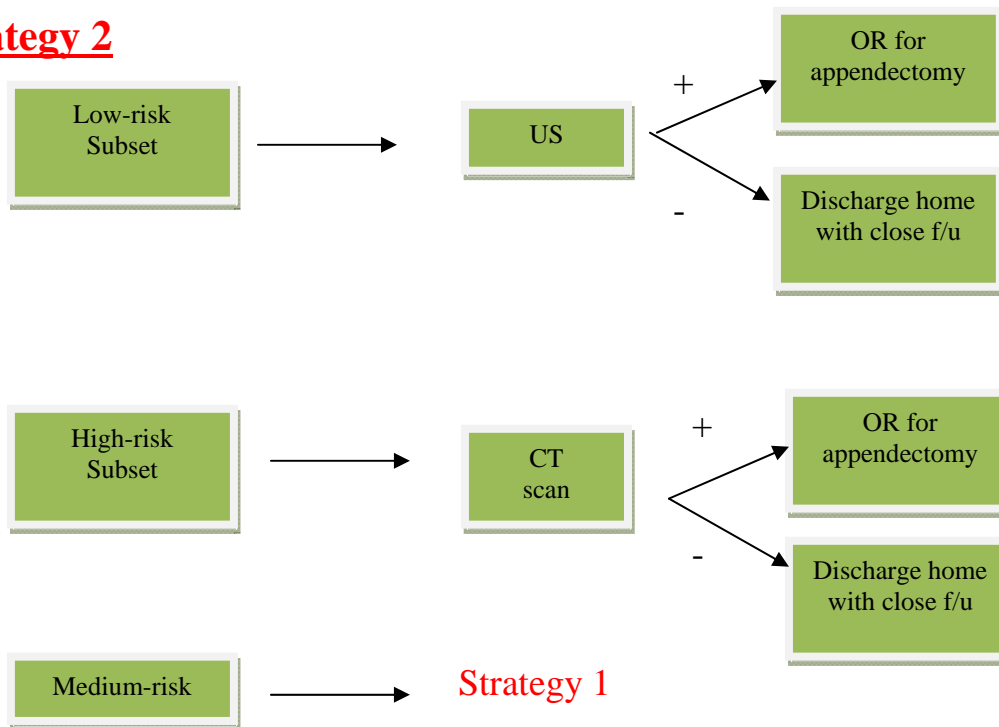
Based upon each patient’s above risk stratification, the authors analyzed three management guideline options:



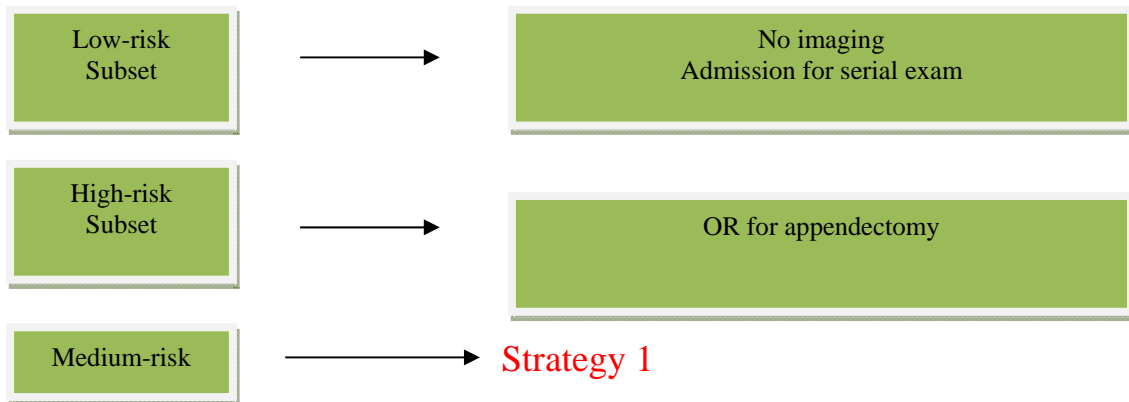
Strategy 1



Strategy 2



Strategy 3



By using the highest known sensitivity, specificities, positive and negative predictive values of US and CT the authors computed the number of negative appendectomies and missed/delayed appendicitis diagnoses each strategy would produce. They also calculated the number of US and CT needed to avoid one negative appendectomy or one missed/delayed appendicitis.

Guide		Comments
I.	Are the Recommendations Valid?	Answer questions IA-D below
A.	<p>Did the recommendations consider all relevant patient groups, management options, and possible outcomes?</p> <p><i>You should look for a clear description of how the panel was assembled along with the members' specialties and any organizations they are representing.</i></p>	No. "There are many more management strategies using selective imaging protocols that can be explored, and those chosen may not be the ideal strategies for managing suspected appendicitis in many institutions. However, to model > 3 strategies would be unwieldy in the scope of a single investigation." (p. 26)
B.	<p>If necessary, was an explicit, systematic, and reliable process used to tap expert opinion?</p> <p><i>You should look for a clear description of how the panel was assembled along with the members' specialties and any organizations they are representing.</i></p>	No. The authors did not reference any prior guidelines, diagnostic SRs, or decision making analyses to support their diagnostic pathways or model probabilities.
C.	<p>Is there an explicit, systematic specification of values or preferences?</p> <p><i>Panelists' ratings presumably reflect the risk-benefit trade-offs of specific interventions, but whether other physicians or patients themselves would make the same decisions remains uncertain. Whether given options are value or preference related should be clearly stated in the guideline.</i></p>	No The authors do not reference or contemplate patient, payer, societal or provider values or preferences. They assume minimizing imaging concurrent with minimizing missed appendicitis and negative appendectomy rates would be preferred by all parties.
D.	<p>If the quality of the evidence used in originally framing the criteria was weak, have the criteria themselves been correlated with patient outcomes?</p> <p><i>When the studies utilized to produce guidelines are less than randomized-controlled trials, conclusions can be strengthened by noting how outcomes can be correlated with adherence to the guidelines.</i></p>	No. The level of evidence supporting the diagnostic pathways is not rated, discussed, or quantified.



II.	Were the Criteria Applied Appropriately?	Answer questions II A-B below.
A.	Was the process of applying the criteria reliable, unbiased, and likely to yield robust conclusions?	<p>Uncertain – and this is fairly important for a chart review study. If findings were not documented (example guarding) was it coded absent or not documented? If coded “absent” was it truly not present or present but not assessed? In general, this is why prospective trial designs are superior and why chart review studies must reference and follow methodological guidelines (Gilbert 1996, Worster 2004).</p> <p>Because the risk-stratification tools were statistical models and not applied prospectively by clinicians to individual patients – and because experienced pediatric Radiologists were likely not blinded to the components of these stratification instruments, the external validity and reliability in other settings remain uncertain.</p>
B.	What is the impact of uncertainty associated with evidence and values on the criteria based ratings of process of care?	Unknown. The authors only use the highest reported sensitivity and specificity of US and CT for their analysis. A sensitivity analysis using the <u>lowest</u> value on 95% CIs would be illustrative.
III.	How Can I Apply the Criteria to Patient Care?	
A.	<p>Are the criteria relevant to your practice setting?</p> <p><i>Medical practice is shaped by an amalgam of evidence, values, and circumstances; clinicians should consider their local medical culture and practice circumstances before importing a particular set of audit criteria.</i></p>	Uncertain. These are pediatric patients presenting to a single Pediatric ED in the northeast United States with an established diagnostic protocol in place. Furthermore, risk stratification occurred retrospectively and the validity of these stratification models (not to mention their reliability, clinician/patient/parent acceptability) have not been assessed. Implementation of these strategies in diverse settings without 24/7 access to pediatric EM/surgery/radiology may produce extremely different conclusions. Finally, sensitivity analyses that incorporate the worst diagnostic test characteristics for CT/US would better inform the potential benefit/harm trade-offs of these strategies.



B.	Have the criteria been field-tested for feasibility of use in diverse settings, include settings similar to yours?	<ul style="list-style-type: none"> • From 1401 potential subjects 958 (63.4%) had complete data sets with mean age 11 years and 55% were male. • The prevalence of acute appendicitis was 61.4% (588/958). • The prevalence of appendicitis was 10.5%, 62.9%, and 90% in the low, medium, and high-risk groups. • The three pathways offered the following characteristics:
-----------	---	---

<u>Guideline</u>	<u>Missed or Delayed Dx (%)</u>	<u>Neg Appy (%)</u>	<u>Correct Dx (%)</u>	<u>Imaging to avoid one missed/delayed Dx</u>	<u>Imaging to avoid one negative appy</u>
1	6	5.9	94.0	19.2 CT + 27.4 US	30.6 CT + 43.5 US
2	6.1	6.2	93.8	17.7 CT + 20.4 US	27.7 CT + 31.9 US
3	6.3	9.7	92.4	11.1 CT + 15.9 US	11.4 CT + 16.4 US

Limitations

- 1) No chart review methods, details, or references.
- 2) Limited external validity at one academic pediatric hospital.
- 3) No prospective application of hypothetical guidelines # 2 and #3 in order to assess their internal validity, reliability, and acceptability.
- 4) No sensitivity analysis incorporating the lower edge of the 95% CI's.
- 5) No 95% CI's presented for proportional data.
- 6) No cost-effectiveness analyses.
- 7) Tabular presentation of results and graphical demonstration of treatment algorithms (as above) would have been far easier to quickly assess and analyze.
- 8) No description of how the diagnosis of appendicitis was ruled-in or ruled-out amongst those without surgery. What was the follow-up interval? Was it person-to-person (if so by whom?) or telephone follow-up? What was the lost to follow-up rate?



Bottom Line

Selective imaging guidelines for pediatric patients with suspected appendicitis may reduce the number of radiographic studies performed, but prospective validation of both the risk stratification algorithm and the proposed imaging protocols at less specialized institutions are needed before these results can be confidently applied. Currently these data represent a Level IV clinical decision rule at best – not yet ready for prime time.