Risk Stratification to Decrease Unnecessary Diagnostic Imaging for Acute Appendicitis

Holly Depinet, MD, MPH, Daniel von Allmen, MD, Alex Towbin, MD, Richard Hornung, PHD, Mona Ho, MS, Evaline Alessandrini, MD, MSCE

BACKGROUND: There has been an increase in the use of imaging modalities to diagnose appendicitis despite evidence that can help identify children at especially high or low risk of appendicitis who may not benefit. We hypothesized that the passive diffusion of a standardized care pathway (including diagnostic imaging recommendations) would improve the diagnostic workup of appendicitis by safely decreasing the use of unnecessary imaging when compared with historical controls and that an electronic, real-time decision support tool would decrease unnecessary imaging.

METHODS: We used an interrupted time series trial to compare proportions of patients who underwent diagnostic imaging (computed tomography [CT] and ultrasound) between 3 time periods: baseline historical controls, after passive diffusion of a diagnostic workup clinical pathway, and after introduction of an electronic medical record–embedded clinical decision support tool that provides point-of-care imaging recommendations (active intervention).

RESULTS: The moderate- and high-risk groups showed lower proportions of CT in the passive and active intervention time periods compared with the historical control group. Proportions of patients undergoing ultrasound in all 3 risk groups showed an increase from the historical baseline. Time series analysis confirmed that time trends within any individual time period were not significant; thus, incidental secular trends over time did not appear to explain the decreased use of CT.

CONCLUSIONS: Passive and active decision support tools minimized unnecessary CT imaging; long-term effects remain an important area of study.
Clinical scoring systems have been developed to identify patients at especially high or low risk of appendicitis based on clinical features, including the 10-point Pediatric Appendicitis Score (PAS). Initially, a single scoring cutoff was used, but subsequent studies found better validity when scores delineated 3 risk strata. Specifically, high-risk scores (≥7) have a specificity of 95% to 98%, and low-risk scores (≤3) have a negative predictive value of 98%, identifying cases where diagnostic imaging could be avoided.

Evidence-based clinical pathways often standardize care and use clinical decision support (CDS) tools or systems to achieve their goals. Studies have shown that CDS tools that are presented to the clinician through active interventions showed improved outcomes, and early evidence indicates that computerized CDS systems that present recommendations at the point of care also increase adherence to specific clinical recommendations.

We hypothesized that the passive diffusion of a standardized care pathway (including diagnostic workup recommendations) would safely decrease the use of unnecessary imaging when compared with baseline historical controls and that an active intervention at the point of care would decrease the use of unnecessary imaging (Fig 1).

**METHODS**

**Setting**

This study was conducted at an urban tertiary care hospital with a pediatric ED with 90,000 annual visits and ~500 cases of appendicitis per year. There is 24-hour availability of surgical subspecialty consultation, anesthesia, and radiology (including ultrasound and CT). The ED used a well-established electronic medical record (EMR) system with decision support capabilities throughout the study period.

**Study Design**

This was a prospective, interrupted time series trial comparing imaging use during 3 time periods: a historical...
baseline, after traditional passive diffusion of a clinical pathway, and after implementation of an active CDS intervention that used the same clinical pathway (Fig 2).

**Interventions**

**Historical Controls (January–December 2010)**

We used a cohort with clinical characteristics that approximated those of our study sample, to determine baseline rates of outcomes before any interventions were made.\(^{18}\)

**Passive Intervention (January 2012–October 2012)**

We developed a clinical pathway for the diagnostic evaluation of suspected acute appendicitis that was based on the well-validated PAS and incorporated our standardized ultrasound report templates and test characteristics.\(^{19}\) The pathway recommended minimizing imaging in the high- and low-risk groups and encouraged ultrasound over CT use in the moderate risk category. The passive intervention (typical evidence dissemination) included conventional passive diffusion of the clinical pathway via review at divisional meetings and educational conferences, posted copies of the pathway in clinical areas and on our divisional Web site, and e-mail reminders.

**Active Intervention (October 2012–June 2013)**

We implemented a real-time, computerized CDS tool that integrated the above appendicitis clinical pathway at the point of care in the EMR. The tool was technically simple, and physicians had used similar tools in our setting previously. Workflow analysis and focus groups with key stakeholders informed the development of a CDS tool with 3 interacting components, which were extensively tested:

- **EMR-based data collection template:** An alert (which fired in response to chief complaints: vomiting and abdominal pain) triggered clinicians to enter elements of the PAS in a timely fashion without interrupting their clinical workflow; laboratory and fever data were automatically incorporated (Fig 3). The tool was iterative; if elements were changed by the user or a more senior physician, new data were incorporated.

- **CDS engine:** The CDS engine above the data and provided the individualized risk score, strata, and recommendations.

- **EMR feedback interface:** A feedback interface provided the results and recommendations.
to clinician users in a clear, visually simple manner at a logical point in the ED workflow; recommendations were associated with automatic orders (based on age, gender, and PAS risk strata) for the appropriate testing or treatment, which the physician could accept or modify (Fig 4).

**Study of the Intervention**

**Inclusion and Exclusion Criteria and Population Identification**

Our study included 3 time periods: historical baseline, passive intervention, and active intervention. During the active and passive intervention phases, we included a convenience sample of patients age 3 to 21, with symptoms for <72 hours, where the ED attending physician had a clinical suspicion for appendicitis or the patient had right lower quadrant pain. We excluded patients with significant abdominal trauma, outside institution imaging, or an underlying medical condition that can confound the diagnosis of appendicitis.

A paper data collection tool was used to collect prospective data for analysis (PASs determined by a combination of attending physician impression of clinical variables at the time of their first physical examination and laboratory and vital sign data) during both intervention phases for 2 reasons. First, this step ensured that the data in both time periods were measured similarly; additionally, it allowed us to study the active intervention (the CDS tool) itself by comparing scores obtained from the CDS tool with the prospective data. To minimize the influence of this data collection tool (ie, limit the possibility that it would serve as an intervention itself), we included extraneous variables and had a run-in period before the passive intervention started. Because our clinical pathway did not firmly recommend a complete blood cell count (CBC) for low-risk patients, in cases where a CBC was not ordered, we analyzed the PAS without complete data.

To find historical controls that represented the true baseline before the clinical pathway was developed (the pathway was developed during the year before the interventions), thus before any clinicians had knowledge of its recommendations, we sought a population cared for >1 year before our study began. Identifying such a population is challenging, because patients with issues of diagnostic accuracy such as “potential appendicitis” are difficult to identify retrospectively.
To identify a historical population that would accurately approximate our study sample, we used a database of ED patients with abdominal pain, from a 1-year period that preceded our study by 13 months, who had PAS assigned and validated by chart review by pediatric emergency medicine providers as part of another study. To select the patients in this database who were as similar as possible to our study population, we additionally analyzed charts to determine whether patients would have met study inclusion and exclusion criteria. Specifically, ED physician notes were analyzed via a natural language processing program to identify any mention of concern for appendicitis; this is a well-validated technique to extract relevant clinical information from free text notes, and in a previous study that used this database, it performed comparably to physician chart reviewers in identifying clinical elements of the PAS. Additionally, we reviewed imaging orders and indications. This database fully sampled patients with a diagnosis of appendicitis and also included a random sample of other patients with abdominal pain. Therefore, the moderate- and high-risk patients in the historical control group approximate the same groups in our study sample well. The low-risk group was proportionally larger in the historical sample, but it includes patients with a low risk of appendicitis based on our criteria.

Measurement

Outcome Variables

Our primary outcome of interest was the proportion of patients who underwent CT as part of their evaluation for suspected appendicitis. Outcomes were calculated overall and stratified by high, moderate, or low risk for appendicitis. Secondary outcomes were proportion undergoing ultrasound and ED length of stay (LOS).

Additional process measures included use of the CDS tool, agreement between the CDS tool and prospectively collected data elements, and presence of a CBC. Balancing measures include negative appendectomy rate and missed diagnosis rate (based on 1-month follow-up review).

Covariates included age, gender, race, shift of arrival (day, evening, or overnight), resident physician involvement (yes or no), and referral status (referred or not referred).

Analysis

Outcomes were examined via statistical process control analyses, logistic regression, and a time series analysis to assess for trends within each time period. Specifically, we used p-charts, which are a type of Shewhart chart (commonly called a control chart) used to graphically display trends and explore variation in binomial data (eg, CT no CT). We ran separate multivariable logistic regression models for the outcomes of CT and ultrasound by using separate 3-level dummy variables for the variable representing PAS risk group (high, moderate, or low) and intervention time interval (historical, passive diffusion, or active intervention) across time intervals and stratified by risk group, controlled for all covariates.

Time series analyses assessed for secular trends in major outcomes between study time periods (via comparison of means) and within each study time period (via comparison of slopes). This was done to expand on the results of the regression analyses and determine whether our outcomes were stable within each time period or changing incrementally over time and thus help determine whether the changes we saw between time periods were probably attributable to our intervention or to secular trends already under way.

We determined frequency of CDS use and compared PAS risk strata generated by the CDS tool with risk strata determined by the prospective data collection. ED LOS was reported for each time period, stratified by PAS risk group.

RESULTS

Our final sample included 809 patients in the historical group, 588 in the passive intervention group, and 489 in the active intervention group (Fig 2). Groups were similar with respect to age, gender, and race; however, patients in the active and passive groups were more likely to have had a resident physician involved in their care, to have been referred to the ED, and to have presented to the ED earlier in the day compared with historical controls. Additionally, the historical group had a higher proportion of patients in the low-risk strata (Table 1).

The CDS tool was used 61% of time in the active study time period; the final PAS strata assigned by the CDS tool (including patients where all clinical PAS elements were completed by the physician user) were in agreement with the prospectively assigned PAS 84.5% of the time; however, in 16% of cases the CDS tool had ≥1 missing or unknown clinical variable (Table 1).

P-charts showed a decrease in the proportion of patients with CT from the historical baseline to the passive and active intervention phases, in the moderate-risk (from 22.3% to 10.2% and 12.2%) and high-risk (from 25.2% to 15.7% and 14.4%) strata (Figs 5 and 6). There were too few patients in the low-risk strata to create control charts. Ultrasound increased in all groups combined (we combined this graphical display for all 3 groups because similar statistically significant increases were seen in all groups over
from 59.5% in the historical group to 83.7% in the passive group and 88.8% in the active group (Fig 7).

In the regression analysis, moderate- and high-risk patients during both the active and passive phases had significantly lower proportions of CTs than those in the historical time period (but neither differed from passive to active phases); all risk groups during both the active and passive phases had significantly higher proportions of ultrasounds than those in the historical time period (Table 2).

Time series analyses found that a comparison of the slopes of the proportion of patients with imaging within each time period was nonsignificant, so changes seen were not thought to be attributable to secular trends occurring incidentally. In the time series comparison of the

| TABLE 1 Characteristics of the Study Population (With 95% Confidence Intervals) |
|---------------------------------|-----------------|-----------------|-----------------|
|                                | Historical, n = 809 | Passive, n = 588 | Active, n = 489 |
| Age, y (median)                | 12              | 12              | 12              |
| Gender, % male                 | 48.2% (44.8–51.7) | 43.5% (39.5–47.5) | 46.6% (42.2–51.1) |
| Race, % white                  | 76.1% (75.3–81.0) | 76.6% (73.1–78.9) | 77.9% (74.2–81.6) |
| Resident involved, % yes       | 29.5% (26.4–32.7) | 70.6% (66.9–74.3) | 77.5% (73.8–81.2) |
| Referral, % yes                | 15.6% (13.1–18.1) | 56.8% (52.8–60.8) | 58.7% (54.3–63.1) |
| PAS distribution               |                 |                 |                 |
| Low risk                       | 18.9% (16.2–21.6) | 10.4% (7.9–12.8) | 6.8% (4.5–9.0) |
| Moderate risk                  | 38.8% (35.5–42.2) | 51.9% (47.8–55.9) | 60.5% (56.2–64.9) |
| High risk                      | 42.3% (38.9–45.7) | 37.8% (33.8–41.7) | 32.7% (28.6–38.9) |
| Time of day of arrival         |                 |                 |                 |
| Day                            | 42.4% (39.0–45.8) | 60.4% (56.4–64.3) | 58.5% (54.1–62.9) |
| Evening                        | 41.9% (38.5–45.3) | 34.9% (31.0–38.7) | 34.4% (30.1–38.6) |
| Overnight                      | 15.7% (13.2–18.2) | 4.8% (3.0–6.5)   | 7.2% (4.9–9.4)   |
| % of time CDS tool used        | n/a             | n/a             | 61.6% (57.2–65.9) |
| % with CBC                     | 62.8% (59.5–66.1) | 71.9% (68.3–75.6) | 76.9% (73.2–80.1) |
| Negative appendectomy ratea    | 17.2% (10.7–23.7) | 8.6% (4.6–12.6)  | 16.8% (12.9–20.7) |
| Missed appendicitis rateb      | n/a             | 0               | 0               |

n/a, not applicable.
a Based on operative histology.
b Based on 1-mo follow-up.
FIGURE 6
P-chart for proportion of patients with CT imaging over time in moderate-risk patients (subgroups of 20).

FIGURE 7
P-chart for the proportion of patients with ultrasound over time in all risk groups (subgroups of 20).
Analysis of ED LOS showed no appreciable change over the 3 time periods (Table 3).

**DISCUSSION**

We found lower rates of CT imaging in the moderate- and high-risk strata, in the passive diffusion and active intervention groups compared with historical controls, and an unintended consequence of an increase in the use of ultrasound.

The additional ultrasounds could be explained by several factors:

Some patients in the final high-risk category did not have laboratory results ready at the time of ordering an ultrasound, and so they appeared to be in the moderate-risk category until their CBC was completed (the CDS tool was iterative and would have given preliminary recommendations for imaging even without final data); also, trainees may have scored the patients higher than attending physicians.

Finally, some high-risk patients may have appropriately had ultrasound ordered after surgical consultation.

The CDS tool was used 61% of the time in the active intervention period, and the CDS-assigned risk strata correlated with prospective data collected 84.4% of the time. Although the CDS tool offered patient-tailored recommendations at the point of care, challenges of real-world implementation may have led to less than ideal practice (eg, missing or unknown variables entered into the CDS tool). For instance, clinicians may have believed that the PAS strata were simple enough to memorize, and therefore they did not need decision support, or the many rotating trainees from various institutions may have been less familiar with CDS tools or not reliably trained to use them. Also, the firing of the CDS tool each time there was a chief complaint related to appendicitis may have led to alert fatigue. Overall, more work is needed to introduce a culture of standardized care in which such a decision support tool could work optimally.

Although these results showed improved care in the direction we thought was most important (minimizing diagnostic imaging associated with ionizing radiation), we did not see more improvement in imaging rates from the passive to the active intervention time periods, as we hypothesized we would. Active interventions, especially those embedded in the EMR at the point of decision making, have shown improved outcomes over guidelines alone.15, 22–24 It is possible that the initial positive effect of traditional passive diffusion of knowledge extinguishes over time25; because our passive pathway was present for only 9 months before the introduction of the active pathway, it was still fairly novel to clinicians, and its effects did not have time to extinguish. Additionally, it is possible that all the preliminary work that occurred in the year between the historical control period and the passive and active intervention periods, which enabled development of the clinical pathway for implementation (eg, obtaining consensus among specialty services about imaging goals, changes in workflow and processes for multiple divisions to support the pathway, education about the utility of the PAS and improved test characteristics with standardized ultrasound interpretations), changed our practice in a way that is more sustainable than an individual pathway. We did not see a significant difference in

---

**TABLE 2 Multivariable Regression for the Outcomes of CT and Ultrasound**

<table>
<thead>
<tr>
<th></th>
<th>Historical Controls, n = 809</th>
<th>Passive Diffusion, n = 588</th>
<th>Active Intervention, n = 489</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>CT</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low riskb</td>
<td>1.0</td>
<td>3.27 (0.85–12.6)</td>
<td>3.71 (0.77–17.5)</td>
</tr>
<tr>
<td>Moderate risk</td>
<td>1.0</td>
<td>0.38 (0.24–0.61)</td>
<td>0.48 (0.31–0.75)</td>
</tr>
<tr>
<td>High risk</td>
<td>1.0</td>
<td>0.55 (0.36–0.85)</td>
<td>0.49 (0.3–0.82)</td>
</tr>
<tr>
<td><strong>Ultrasound</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low risk</td>
<td>1.0</td>
<td>10.5 (5.0–22.0)</td>
<td>10.8 (4.2–27.5)</td>
</tr>
<tr>
<td>Moderate risk</td>
<td>1.0</td>
<td>8.4 (4.8–14.7)</td>
<td>17.8 (9.5–33.3)</td>
</tr>
<tr>
<td>High risk</td>
<td>1.0</td>
<td>14.9 (7.4–29.8)</td>
<td>18.3 (8.6–39)</td>
</tr>
</tbody>
</table>

CI, confidence interval; OR, odds ratio.

a Adjusted for age, gender, race, shift of arrival (day, evening, or overnight), resident physician involvement (yes or no), and referral status (referred or not referred).

b All risk strata based on our institutional pathway.

**TABLE 3 ED LOS for Patients in All Risk Strata Across Time Intervals**

<table>
<thead>
<tr>
<th></th>
<th>Historical Controls, n = 809</th>
<th>Passive Diffusion, n = 588</th>
<th>Active Intervention, n = 489</th>
</tr>
</thead>
<tbody>
<tr>
<td>LOS, min, high risk</td>
<td>355.5 (339.4–371.6)</td>
<td>367.8 (350.4–385.3)</td>
<td>372.9 (351.4–394.4)</td>
</tr>
<tr>
<td>LOS, min, moderate risk</td>
<td>354.3 (337.9–370.6)</td>
<td>356.5 (321.8–351.2)</td>
<td>371.3 (355.7–386.9)</td>
</tr>
<tr>
<td>LOS, min, low risk</td>
<td>251.7 (234.6–268.7)</td>
<td>325.4 (284.5–366.3)</td>
<td>288.8 (244.0–335.1)</td>
</tr>
</tbody>
</table>

CI, confidence interval.
ED LOS between the time intervals; however, this factor is affected by many variables other than the targets of our interventions (eg, hospital boarding, changes in available rooms due to construction, other initiatives).

One challenge in studies of diagnostic accuracy is the difficulty with population identification. Unlike in studies of children with known diagnoses such as asthma or diabetes, there is no simple way to identify patients with “possible diagnoses,” such as appendicitis, intussusception, or sepsis. Our method of population identification for the active and passive samples (asking the attending physician whether he or she is considering appendicitis) has been used successfully in previous studies of pediatric appendicitis and other types of studies. However, there is a risk that this method may introduce bias (toward the null) by prompting the ED attending use the pathway.

Additionally, we believe that the low-risk population, in which the clinician has some thought of appendicitis but has ruled it out clinically, may be underrepresented in the passive and active diffusion phases because of the hesitation to identify a patient as having some risk of appendicitis (however low) despite not planning to do any more evaluation. Thus, the historical sample, with its population identified by less subjective methods, may represent a truer low-risk sample than the prospectively identified samples.

Challenges with population identification also led to the need to approximate the historical control group differently than the intervention groups, which may introduce bias. For instance, the historical group included more patients seen overnight, fewer patients cared for by resident physicians, and patients from all areas of the ED (whereas the active and passive study populations were not drawn from urgent care areas). These factors could have affected imaging outcomes in the historical group, although this effect is minimized by close supervision of learners, introduction of the pathway to residents, and constant availability of ultrasound. Additionally, some of these factors would theoretically have introduced bias in differing directions (ie, some would bias toward and some away from the null), but we believe that our stringent selection criteria for historical controls minimized this possibility and that the moderate- and high-risk groups were similar across study periods.

Additionally, although a decrease in CT was temporally correlated with our interventions, it is possible that other more global factors, such as nationally disseminated knowledge about the risks of CT, were partially responsible for the decrease in CT that we observed.

Finally, generalizability may be limited to institutions with the specific capabilities needed for the intervention: 24-hour availability of surgeons and radiologists, acceptable test characteristics of ultrasound, and EMRs with decision support capabilities and the technical resources to program them.

CONCLUSIONS

Overall, we saw an improvement in the use of CT from our historical baseline, although CT use did not differ between active and passive intervention groups, and we saw an unintended consequence of increased use of ultrasound. Additional work is needed to determine whether this effect will diminish over time. Additionally, this study demonstrates the value of standardized care pathways in reducing variation in imaging in cases of diagnostic accuracy.

ABBREVIATIONS

CBC: complete blood cell count
CDS: clinical decision support
CT: computed tomography
ED: emergency department
EMR: electronic medical record
LOS: length of stay
PAS: Pediatric Appendicitis Score

REFERENCES


Risk Stratification to Decrease Unnecessary Diagnostic Imaging for Acute Appendicitis
Holly Depinet, Daniel von Allmen, Alex Towbin, Richard Hornung, Mona Ho and Evaline Alessandrini

*Pediatrics* 2016;138;
DOI: 10.1542/peds.2015-4031 originally published online August 23, 2016;

The online version of this article, along with updated information and services, is located on the World Wide Web at:

http://pediatrics.aappublications.org/content/138/3/e20154031

---

Pediatrics is the official journal of the American Academy of Pediatrics. A monthly publication, it has been published continuously since . Pediatrics is owned, published, and trademarked by the American Academy of Pediatrics, 141 Northwest Point Boulevard, Elk Grove Village, Illinois, 60007. Copyright © 2016 by the American Academy of Pediatrics. All rights reserved. Print ISSN: