Progress in the Diagnosis of Appendicitis

A Report From Washington State's Surgical Care and Outcomes Assessment Program

The SCOAP Collaborative, Frederick Thurston Drake, MD, Michael G. Florence, MD,† Morris G. Johnson, MD,‡ Gregory J. Jurkovich, MD,§ Steve Kwon, MD, MPH,* Zeila Schmidt, MLIS,|| Richard C. Thirlby, MD,¶ and David R. Flum, MD, MPH*||*

Background and Objectives: Studies suggest that computed tomography and ultrasonography can effectively diagnose and rule out appendicitis, safely reducing negative appendectomies (NAs); however, some within the surgical community remain reluctant to add imaging to clinical evaluation of patients with suspected appendicitis. The Surgical Care and Outcomes Assessment Program (SCOAP) is a physician-led quality initiative that monitors performance by benchmarking processes of care and outcomes. Since 2006, accurate diagnosis of appendicitis has been a priority for SCOAP. The objective of this study was to evaluate the association between imaging and NA in the general community.

Methods: Data were collected prospectively for consecutive appendectomy patients (age > 15 years) at nearly 60 hospitals. SCOAP data are obtained directly from clinical records, including radiological, operative, and pathological reports. Multivariate logistic regression models were used to examine the association between imaging and NA. Tests for trends over time were also conducted.

Results: Among 19,327 patients (47.9% female) who underwent appendectomy, 5.4% had NA. Among patients who were imaged, frequency of NA was 4.5%, whereas among those who were not imaged, it was 15.4% ($P < 0.001$). This association was consistent for men (3% vs 10%, $P < 0.001$) and for women of reproductive age (6.9% vs 24.7%, $P < 0.001$). In a multivariate model adjusted for age, sex, and white blood cell count, odds of NA for patients not imaged were 3.7 times the odds for those who received imaging (95% CI: 3.0–4.4). Among SCOAP hospitals, use of imaging increased and NA decreased significantly over time; frequency of perforation was unchanged.

Conclusions: Patients who were not imaged during workup for suspected appendicitis had more than 3 times the odds of NA as those who were imaged. Routine imaging in the evaluation of patients suspected to have appendicitis can safely reduce unnecessary operations. Programs such as SCOAP improve care through peer-led, benchmarked practice change.

Keywords: appendicitis, computed tomography, negative appendectomy, quality improvement, ultrasonography, surgery

(*Ann Surg* 2012;256: 586–594)

Surgical convention suggests that clinical assessment is usually sufficient to make the diagnosis of acute appendicitis. Under this view, a certain frequency of so-called negative appendectomies (NAs)—in which a noninflamed appendix is removed from patients mistakenly suspected to have appendicitis—is acceptable to prevent underdiagnosis, delay in definitive therapy, and an attendant increase in the risk of appendiceal perforation. However, recent studies have found that the addition of advanced diagnostic imaging to the clinical evaluation of suspected appendicitis is associated with a reduction in the frequency of NA without an associated increase in the frequency of perforation.[1–16] Surgeons and emergency medicine physicians now commonly employ imaging in the workup of appendicitis, and many of the most recent studies are devoted to evaluating diagnostic protocols, such as sequenced ultrasonographic (US) and computed tomographic (CT) pathways designed to limit exposure to ionizing radiation.[17–23] Despite growing acceptance of imaging and widely replicated performance results in tertiary centers, the accuracy of diagnostic imaging in some community settings has not achieved the level reported by clinical studies, the utility of imaging across diverse community settings has not been established as a safe means of reducing unnecessary operations, and many surgeons feel that CT is not necessary and overused.[24–27]

The Surgical Care Outcomes and Assessment Program (SCOAP) is a physician-led quality surveillance program that began in 2006 and has subsequently enrolled essentially all hospitals in Washington State. Data are collected prospectively by trained abstractors, and statewide reports are issued (individual institutions are de-identified). Many aspects of surgical care are reported including specific processes of care and clinical outcomes. Performance benchmarks are established by high-achieving hospitals for both processes and outcomes. Although SCOAP data is collected primarily as a quality improvement endeavor, it is a source of data for observational research studies. Unlike administrative data sets in which *International Classification of Diseases, Ninth Revision*, codes are used to obtain information about diagnosis and treatment, SCOAP relies on review of clinical records for consecutive patients undergoing specific procedures, including those who have appendectomies.

In 2008, we reported results from 15 SCOAP hospitals on the frequency of NA and use of imaging. That report noted substantial variation in both the use of imaging and in NA between hospitals and that NA correlated most closely with diagnostic accuracy (concordance between radiology reports and pathology reports).[24] The current report describes results from 55 hospitals over the last 5 years to (1) investigate the association between the use of imaging and NA in the general community, with a focus on patients at high-risk for misdiagnosis; (2) estimate performance characteristics of imaging modalities within a broad clinical environment; and (3) evaluate whether progress in safely reducing NA has continued as SCOAP expanded.

From the *University of Washington Medical Center, Seattle; †Swedish Medical Center, Seattle; ‡Skagit Valley Hospital, Mt Vernon; §Harborview Medical Center, Seattle; ||University of Washington Surgical Outcomes Research Center; and ¶Virginia Mason Medical Center, Seattle, WA.

Disclosure: The authors declare no conflicts of interest.

Presented at the 132nd Annual Meeting at the American Surgical Association, April 27, 2012, San Francisco, CA.

Reprints: Frederick Thurston Drake, MD, Department of Surgery, University of Washington, 1859 NE Pacific Street, Box 356410, Seattle, WA 98195. E-mail: ftdrake@u.washington.edu.

ISSN: 0003-4932/12/25604-0586

DOI: 10.1097/SLA.0b013e31826a9602

METHODS

Study Population and Setting

Although 55 Washington hospitals currently participate in SCOAP, hospital enrollment has been a gradual process as hospitals have joined each quarter over the past 6 years. Data are collected at the 2 pediatric hospitals in the state, and SCOAP also gathers information on children having operations at general hospitals; however, the current study population has been restricted to patients aged 15 years or older who underwent appendectomy in a nonpediatric SCOAP hospital between January 1, 2006, and December 31, 2011. Participating hospitals submit data for all appendectomies performed within the institution for each enrollment year.

Data Characteristics and Primary Outcome

Demographic information, clinical characteristics, radiology interpretations, operative indications, operative findings, and pathology results are abstracted from the clinical record using standardized definitions. Abstracted data are audited for quality control and to verify that charts are being evaluated in a similar way among participating sites. The data for appendectomy represent consecutive nonelective appendectomies performed at participating sites. A comorbidity index score, modeled on the Charlson comorbidity index, is calculated on the basis of documentation in the clinical record of the following comorbid conditions: coronary artery disease, asthma, diabetes, human immunodeficiency virus/AIDS, diabetes, and/or elevated serum creatinine. White blood cell (WBC) count is based on the result obtained most proximal to the appendectomy. Body mass index (BMI) is calculated from recorded height and weight. Review of the patient's pathology report determines whether the appendix was diseased at the time of operation. Positive pathology results include confirmed or consistent with appendicitis or appendiceal tumor. Perforation of the appendix is based on pathologic diagnosis, and frequency of perforation was calculated excluding patients with NA (ie, percent perforation = patients with perforation/all with positive appendiceal pathology). Imaging results are based on the final radiologist interpretation and are reported as consistent with appendicitis, not consistent with appendicitis, or indeterminate. An appendectomy is characterized as an NA in the absence of appendicitis or tumor/mass. The imaging report and pathology report are considered concordant if the imaging results are consistent with appendicitis and the pathology is positive or if imaging results are not consistent with appendicitis and pathology does not show evidence of disease. Indeterminate radiographic findings are considered discordant, regardless of pathologic findings. The primary outcome of interest was NA. Research projects utilizing SCOAP data are approved by the Washington State Department of Health Institutional Review Board.

Analytic Methods

Patients with appendiceal pathology were compared with those without appendiceal pathology to identify distinguishing characteristics between the 2 groups. Categorical variable comparisons were evaluated for significance using Pearson χ^2 test (significance set at $\alpha = 0.05$). Student *t* test was used to compare continuous variables ($\alpha = 0.05$). Odds ratios (ORs) [and 95% confidence intervals (CIs)] for variables predictive of misdiagnosis were calculated on the basis of a priori hypotheses. A one-way analysis of variance model (multiple linear regression on a binary variable) was used to evaluate whether the proportion of NA differed significantly among comorbidity categories. Tests of trend over time were calculated using the Cochran-Armitage test for trends in the odds. After the unadjusted analysis, we evaluated the association between imaging and NA for the presence of confounding by other covariates; variables potentially available to be included in this logistic regression model were those patient characteristics listed in Table 1. Covariates were included in this explanatory logistic regression model if they were known from the surgical literature or from clinical experience to be associated with misdiagnosis *and* if a differential association was detected in univariate analysis between the exposures of interest (ie, imaging vs. no imaging) and the potential covariate. Using these criteria, a parsimonious, logistic regression model was developed that included age, sex, and WBC count as covariates in the relationship between imaging use and NA. Using a generalized estimating equation, the model was also adjusted for clustering of patients by institution. Women of reproductive age were previously identified as a group of patients at high risk for misdiagnosis; therefore, we separately considered a subcohort of women ages 15 to 50 years. STATA version 12 was used for all analyses (STATA Corp, College Station, TX).

We estimated sensitivity and positive predictive value (PPV) for CT and US. In addition, we compared frequency of NA among patients imaged by the 2 most common modalities, US and CT. An overall comparison was performed, and, because some institutions

TABLE 1. All Appendectomy, Appendicitis, and NA

	All Appendectomy (n = 19,327)	Appendicitis (n = 18,193)	NA (n = 1042)	*P*
No. female patients	47.86%	46.9%	64.0%	$P < 0.001$*
Mean age (SD)	39.4 (16.6)	39.5 (16.6)	37.4 (15.9)	$P < 0.001$†
Preoperative Imaging	91.3%	92.2%	75.3%	$P < 0.001$*
NA	5.4%	–	–	
Laparoscopy	84.8%	84.9%	83.5%	$P = 0.24$*
Perforation	15.0%	15.8%	–	
Mean WBC count (SD)	13.2 (4.3)	13.3 (4.3)	10.4 (4.0)	$P < 0.001$†
Mean BMI (SD)	27.7 (6.0)	27.8 (6.2)	27 (6.3)	$P < 0.001$†
Comorbidity Index:				NS
0	86.5%	86.6%	85.5%	
1	10.9%	10.8%	12.6%	
2	1.7%	1.7%	1%	
≥3	0.9%	0.9%	0.5%	

P values reflect results of statistical comparison between patients with appendicitis and patients with NA.
*Pearson χ^2 test.
†Independent Student *t* test (2-sided).
SD indicates standard deviation.

have imaging protocols based on age, we also made comparisons within 3 age groups (15–30, 31–65, and >65 years).

RESULTS

Cohort Characteristics

A total of 19,327 adolescent and adult patients underwent appendectomy (47.9% female, mean age 39.4 years, standard deviation 16.6). Ninety-one percent of patients underwent some form of preoperative imaging (CT, US, and/or magnetic resonance imaging). Among all patients with appendectomy, 1042 (5.4%) had NA. Overall frequency of perforation, as a percentage of patients with appendicitis, was 15.8%. Patients with NA were more often female, younger, and with a lower WBC count. BMI and comorbidity score were similar between patients with NA and those with appendicitis. Equal proportions of patients underwent a laparoscopic procedure (Table 1).

Imaging Versus No Imaging

Among patients with NA, a significantly smaller proportion received preoperative imaging compared with those with appendicitis (75.3% vs 92.6%, $P < 0.001$). For patients who had preoperative imaging, the frequency of NA was 4.5%, significantly lower than the frequency of NA (15.4%) among those who did not have preoperative imaging (OR = 3.90, 95% CI: 3.34–4.55, $P < 0.001$). After adjusting for sex, age, WBC count, and clustering by site, the odds of NA among those who did not undergo preoperative imaging were 3.7 times the odds of NA for those who did undergo preoperative imaging (95% CI: 3.01–4.42, see Table 2). Adjusted for imaging, the OR for NA among women compared to men was 2.10 (95% CI: 1.76–2.51, $P < 0.001$). Although women were twice as likely as men to undergo NA, imaging among male patients was also associated with a significantly lower frequency of NA (3% vs 10%, $P < 0.001$). Frequency of perforation was the same between patients who were and were not imaged: among those who were imaged (and who had appendicitis), perforation was 15.8% and among those who were not, perforation was 15.6% ($P =$ 0.16).

Women of Reproductive Age

There were 6632 women ages 15 to 50 years who underwent appendectomy, representing 34.4% of all appendectomies. Almost 95% underwent some form of diagnostic imaging. Among women of reproductive age, frequency of NA was 8.1%. Nine percent of these patients were perforated compared with 15.8% in the entire cohort. Among women of reproductive age who received any form of preoperative imaging, the frequency of NA was 6.9%, whereas it was 24.7% among women of reproductive age who received no imaging (crude OR = 4.48, 95% CI: 3.49–5.64, $P < 0.001$). In the multivariate model adjusted for age, WBC count, and clustering by hospital, the odds of NA were 3.46 times the odds for those who did undergo preoperative imaging (95% CI: 2.43–4.94, see Table 3). Frequency of perforation was the same between those who had imaging and those who did not (9.9% vs 9.7% respectively, $P = 0.48$).

TABLE 2. Multiple Logistic Regression on Odds of NA (Full Cohort)

	OR	95% CI
Imaging vs no imaging (unadjusted)	3.90	3.34–4.55
Imaging vs no imaging (adjusted)	3.65	3.01–4.42
Sex (female vs male)	2.10	1.76–2.51
Age	0.99	0.98–0.99
WBC count	0.86	0.84–0.89

TABLE 3. Multiple Logistic Regression on Odds of NA (Women of Reproductive Age)

	OR	95% CI
Imaging vs no imaging (unadjusted)	4.48	3.49–5.64
Imaging vs no imaging (adjusted)	3.46	2.42–4.94
Age	0.99	0.99–1.0
WBC count	0.86	0.84–0.88

Outcome of interest is NA; primary exposure is the use of imaging versus no imaging (imaging could be CT scan or US). Potential confounders (sex, age, and WBC count) were identified a priori and assessed in a univariate fashion (see Table 1) before inclusion in the regression model. WBC count and age are included as continuous variables. A generalized estimating equation function was utilized to adjust the model for clustering by hospital.

Performance Characteristics of CT and US

Among all ages, 4.1% of patients who had CT underwent NA compared with 10.4% of patients who had US ($P < 0.001$). In both, the adolescent/young adult and middle-age categories, NA was significantly less common when CT was used compared with US (4.6% vs 12% and 3.8% vs 8.6%, respectively, $P < 0.001$ for both). Only 29 elderly patients underwent US, so a comparison with CT was not considered robust in this age group. Among elderly patients who underwent CT, frequency of NA was 3.6%. In patients who were not imaged, percent NA ranged from 14.1% to 16.3% depending on age group (Fig. 1). The sensitivity of CT scan for appendicitis was estimated to be 93.2%, and for US, sensitivity was estimated to be 47.8%. PPV of CT scan was 97.6% and of US was 94.3%.

Temporal Trends

We evaluated trends in imaging use and percent NA over the duration of SCOAP. The proportion of patients who received imaging in the workup of suspected appendicitis has been consistently rising (Fig. 2). This is seen among SCOAP hospitals overall ($P < 0.001$), and also within hospital groups stratified by the year in which they joined SCOAP (though with more year-to-year variability). Concomitantly, in SCOAP overall, there has been a significant decline in the annual rate of NA ($P < 0.001$), though, again, there is year-to-year variability

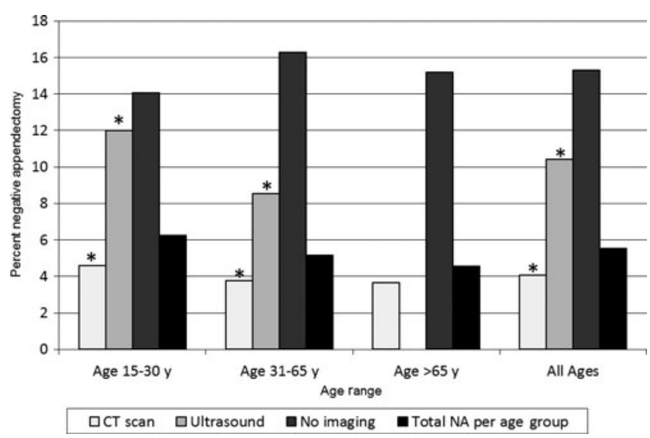

FIGURE 1. Percent NA by imaging modality, stratified by age range. Asterisks indicate statistically significant comparison within each age category. No US results are reported for the elderly age group; only 29 elderly patients underwent preoperative US, which was felt too small of a number to make a comparison valid.

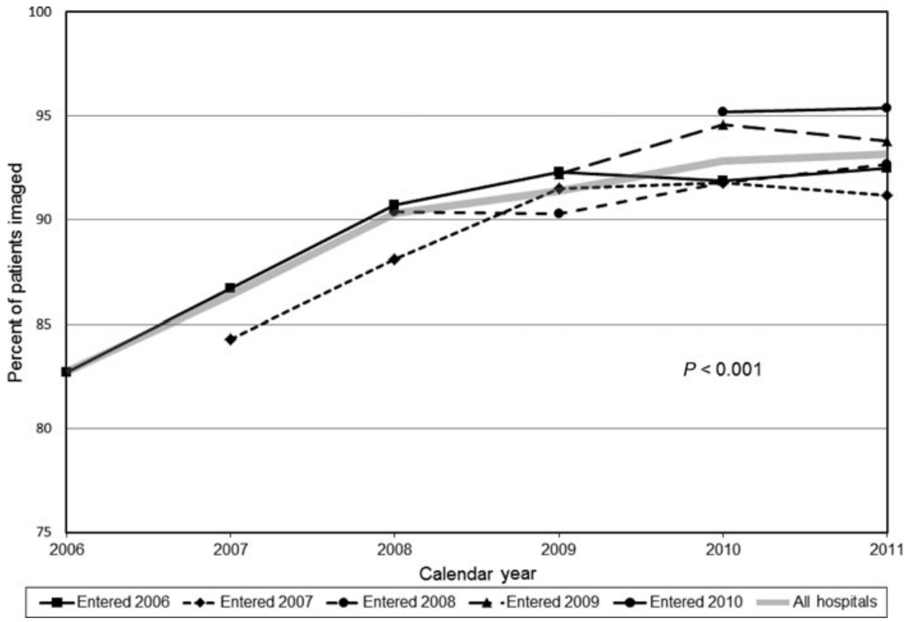

FIGURE 2. Appendectomy patients who received preoperative imaging. Hospitals are stratified by the year in which they joined SCOAP (represented by black lines that begin with the year in which that group of hospitals first submitted cases to SCOAP). The solid gray line indicates the number of patients who underwent preoperative imaging as a percentage of all appendectomy patients from all hospitals participating in SCOAP in any particular year. Test for trend for the overall percentage of preoperative imaging was significant ($P < 0.001$).

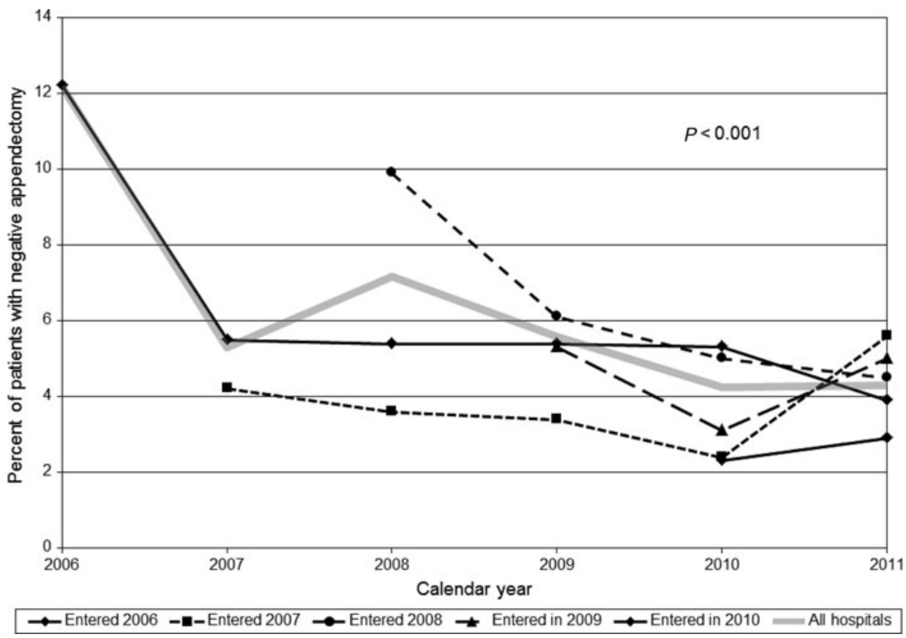

FIGURE 3. Appendectomy patients with NA. Hospitals are stratified by the year in which they joined SCOAP (represented by black lines that begin with the year in which that group of hospitals first submitted cases to SCOAP). The solid gray line indicates the number of NAs as a percentage of all appendectomy patients from all hospitals participating in SCOAP in any particular year. Test for trend for the overall percentage of NAs was significant ($P < 0.001$).

within subgroups of hospitals (Fig. 3). Over this same time period, the percent of appendicitis patients who were perforated has not changed (Fig. 4). Cumulative frequency of perforation ranged from 14.9% to 16.8%, but there was no temporal trend ($P = 0.63$).

DISCUSSION

In this cohort of older-adolescent and adult patients cared for in SCOAP hospitals over a 6-year period, the use of advanced diagnostic imaging increased and the frequency of NA decreased. Among patients who received preoperative imaging, NA was substantially less frequent than among patients who did not receive preoperative imaging. When this relationship was adjusted for other predictors of NA, failure to obtain imaging was associated with a 3.7-fold increase in odds of NA. Among women of reproductive age, the relationship with imaging was especially pronounced (25% NA vs 7% NA). However, the age- and sex-adjusted regression suggests that, even among men, there is a strong association between preoperative imaging and decreased odds of NA. As a group, SCOAP hospitals have prioritized the use of diagnostic imaging in the evaluation of suspected appendicitis as part of a commitment to safely reducing unnecessary operations. Although yearly variation is evident, data over the last 6 years suggest that these goals are being met by SCOAP hospitals.

The sensitivity of CT scan in this population (93.2%) was lower than some of the studies of CT in the highly structured environment of studies in academic centers; however, this is within the range reported in the literature. Cumulative sensitivity of studies related to ultrasonography in SCOAP hospitals was disturbingly low at 47.8%. Close inspection of this data revealed that a large number of patients

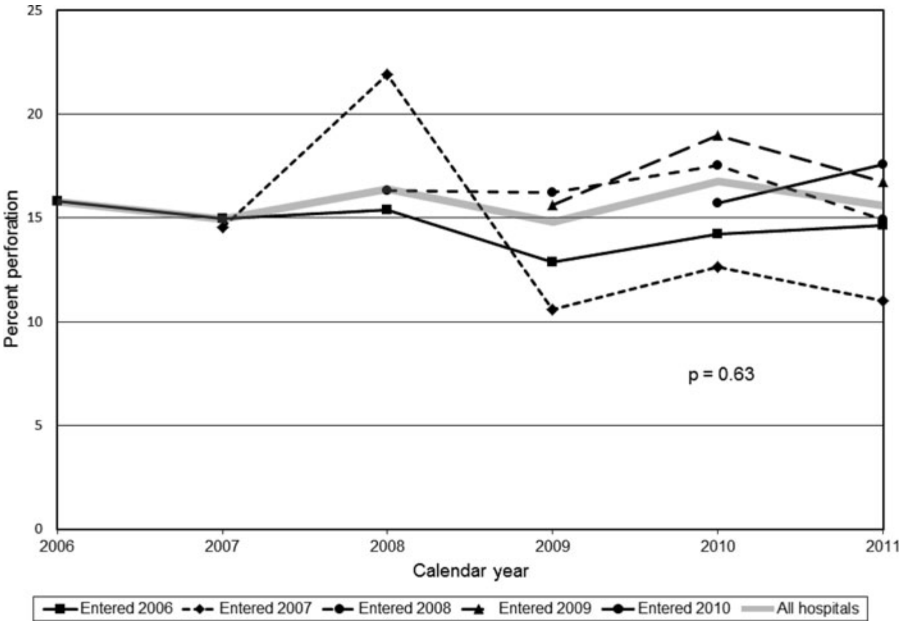

FIGURE 4. Appendicitis patients with perforation. Hospitals are stratified by the year in which they joined SCOAP (represented by black lines that begin with the year in which that group of hospitals first submitted cases to SCOAP). The solid gray line indicates the number of patients with perforation as a percentage of all patients with appendicitis from all hospitals participating in SCOAP in any particular year. There was no temporal trend for the overall frequency of perforation ($P = 0.63$).

with indeterminate results on ultrasonography were ultimately found to have appendicitis at appendectomy, which substantially reduced the modality's sensitivity. PPV for both studies was high (94% and 98%), however, suggesting that positive results on either CT scan or US are useful findings in the evaluation of a patient with suspected appendicitis.

One of the aims of this study was to evaluate diagnostic performance of CT and US in the general community as it compares to that published in the literature. A rigorous meta-analysis of CT and US published by Doria et al[28] in 2006, included 57 studies (both retrospective and prospective) and more than 13,000 patients. Studies were included only if absolute numbers of true-positives, true-negatives, false-positives, and false-negatives were available and adults and children were considered separately. In adults, sensitivity and specificity of CT were both 94%, and those of US were 83% and 93%, respectively. More recently, a large single-center study prospectively evaluated CT performance in 2871 consecutive adults imaged for suspected appendicitis and obtained thorough clinical follow-up of operative and nonoperative patients; sensitivity was 98.5%, specificity was 98%, NPV was 99.5%, and PPV was 93.9%.[29] Other recent studies have shown similar high performance for CT,[30,31] including one study that evaluated low-dose radiation CT.[32] Regarding US, Rettenbacher et al[1] prospectively followed 350 patients evaluated by US for suspected appendicitis and determined a sensitivity of 98%, a specificity of 98%, a PPV of 96%, and an NPV of 99%.[1] The sensitivity we estimated for CT among SCOAP hospitals is within the 95% CI reported by Doria (92%–95%), but as a group, SCOAP hospitals have not achieved the high bar set by studies performed in academic centers. Furthermore, although US had a substantial PPV, the frequency of equivocal results limited its performance in terms of sensitivity. Certainly, for surgeons to include imaging results in their clinical decision making, they have to have confidence in the results, and SCOAP has made imaging accuracy a priority. Performance measures and statewide benchmarks for CT and US accuracy are provided to participating hospitals, and SCOAP is collaborating with radiology colleagues to address mechanisms for improvements.

Accurate imaging can provide 3 important functions in the evaluation of suspected appendicitis: provide evidence for a diagnosis of appendicitis, provide evidence against a diagnosis of appendicitis, and suggest alternative diagnoses. All are important, but the current study focused primarily on the second function. Reducing unnecessary operations is good for patients and for health care systems; previous studies have shown substantial increases in both length of stay and hospital charges for patients with NA compared with patients with appendicitis.[33] The current data from SCOAP hospitals suggest that the use of imaging is associated with a reduction in NA in the general community. Two other recent studies have assessed this association prospectively in patients with suspected appendicitis, and in both, preoperative imaging changed management decisions, reducing NA.[2,3] In one of these studies, 152 patients were randomized to mandatory CT or selective CT on the basis of clinical examination. In the mandatory CT group, the frequency of NA was 2.6% compare with 13.9% in the selective CT group with no difference in perforation.[2] In addition to these prospective studies, numerous observational, retrospective analyses of appendectomy patients have shown an association between increased use of imaging and a decrease in the frequency of NA.[4–14] This association was found for pediatric patients in some studies[4] but not all.[5,34]

The current study has several limitations. How patients were allocated to imaging or no-imaging is not captured by our data set, and although the logistic regression models control for confounding by age, sex, and WBC count (a marker of clinical severity), unmeasured confounding by indication may still be present. It is possible that this would lead to a conservative bias if complex or clinically uncertain patients were more likely to undergo imaging. The potential influence of laparoscopy on the measured frequency of NA is also uncertain. Administrative database analysis from the 1990s in Washington state[35] suggested that patients undergoing laparoscopy were more likely to have NA than those undergoing open appendectomy; however, a later analysis of SCOAP hospitals showed that there was no trend between a hospital's use of laparoscopy and frequency of NA.[14] This latest analysis of SCOAP data is consistent with the latter finding in that patients undergoing laparoscopy were no more likely to have NA than those undergoing open appendectomy. Because SCOAP does not collect data for patients who undergo laparoscopy but do not have appendectomy if no appendicitis is found, the contribution of exploratory laparoscopy to decreasing NA cannot be judged from this data set. Finally, in an earlier SCOAP study, in which hospitals were the unit of analysis, the correlation between accuracy of imaging (defined as pathology and radiology concordance) and institutional rates of NA was stronger than the correlation between NA and frequency of use of imaging.[24] In the current study, which treats patients as the unit of analysis, the impact of accuracy was not assessed. Although this would not be expected to change the association between the use of imaging and frequency of NA (because accuracy of imaging does not impact patients who are not imaged), it could confound the inference that increased imaging among SCOAP hospitals has led to less NA among SCOAP hospitals. If accuracy is also improving, institutional rates of NA could decrease both from improved accuracy and from increased use of imaging. This is the topic of ongoing analyses. There may also be other variables beyond indication for imaging, use of laparoscopy, and accuracy of imaging that are unmeasured confounders. One solution to such confounding would be a statewide trial that randomized patients to mandatory imaging or selective use of imaging, but this may not be feasible.

A further potential limitation is the possibility for sampling bias because the SCOAP cohort does not represent a truly random sample of the state's total appendectomy volume; however, for this to substantially alter the study results, those hospitals not participating would have to be outliers in terms of appendicitis care. By the end of 2011, 55 of the 75 hospitals in the state that perform appendectomies were actively contributing data to SCOAP. Contributors include both pediatric hospitals in the state and the state's active-duty military hospital, but it does not include the Veterans Affairs Medical Center in Seattle. The 20 hospitals that do not contribute to SCOAP are diverse in size, geographic location, and ownership; of those not enrolled, median appendectomy volume for 2010 was 22 cases, and only 4 nonenrolled hospitals performed greater than 100 cases. Utilizing the Washington State Comprehensive Hospital Abstract Reporting System, which collects information on all discharges from nonfederal hospitals in the state, we estimated Washington's total 2010 volume of nonincidental, nonelective appendectomy (for patients age $\geq$18 years only) to be 6124 cases. SCOAP collected 5005 such cases for the same year, representing 82% of the state's total appendectomy volume. For 2011, data of Comprehensive Hospital Abstract Reporting System was not available, but with the addition of 6 new hospitals to SCOAP between the end of 2010 and the end of 2011, we expect that the proportion of the state's appendectomies captured by SCOAP has continued to increase as it has every year since 2006, SCOAP's first year. In 2006, data were captured from 14 hospitals, representing approximately 20% of the state's appendectomy volume.

Finally, our estimates of diagnostic performance (sensitivity and PPV) also involve limitations. Because this patient cohort is generated by patients who undergo appendectomy, data on most "true negatives" are not available. Patients who were correctly determined by CT or US not to have appendicitis were not included in this data set unless the study was overruled, the patient was operated on, and had an NA. This makes a determination of specificity [true-negatives / (true-negatives + false-positives)] and negative predictive value [true-negatives / (true-negatives + false-negatives)] impossible. However, if it is assumed that very few patients with acute appendicitis do not undergo appendectomy, estimations of sensitivity [true-positives / (true-positives + false-negatives)] and PPV [true-positives / (true-positives + false-positives)] are possible. There may be some loss of "true-positives" if the scan was overruled by the physicians, the patient was discharged, and ultimately had an appendectomy at another hospital (if the patient returned to the same hospital, the original CT information would be captured by SCOAP). There may be some loss of "false-positives" if the CT scan was correctly overruled and the patient did not proceed to surgery. There may be some loss of "false-negatives," if the patient was discharged and ultimately had an appendectomy at another hospital. Loss of "true-positives" tend to reduce the observed performance of the imaging modality; loss of "false-negatives" and "false-positives" tend to increase the observed performance.

Many of the overlapping issues that arise in a consideration of imaging in suspected appendicitis are areas of active investigation and collaboration among SCOAP-affiliated surgeons and an increasingly broad coalition of academic and community radiologist partners. Given the previously detected association between accuracy of imaging and reductions in NA, we are currently developing a standardized CT report for imaging in suspected appendicitis that will soon undergo piloting and validation testing. Attention to CT radiation dose, a variable newly captured by SCOAP, has revealed substantial variation in levels of radiation delivered during CT scan for appendicitis. Standardization of dose levels may be one way of reducing unnecessarily high radiation exposure, and the potential benefit to patient safety is being investigated by the SCOAP community. There is an ongoing effort to compare accuracy of CT scans in which intravenous (IV) and enteral contrast are both used to CT scans in which only IV contrast is used; given the time and cost savings that accrue from not using oral contrast, plus the advantage of avoiding oral intake among patients who typically feel very poorly, abandoning oral contrast has potential for significant improvements in the CT evaluation of appendicitis. For care of patients with suspected appendicitis, these developments represent some of the latest efforts within this physician-led system of continuous quality improvement.

The current investigation evaluated the association between imaging and NA across a large population served by diverse institutions. The data suggest that including preoperative imaging in the workup of suspected appendicitis can lead to a reduction in unnecessary operations, especially among women of reproductive age; these modalities may also uncover alternative diagnoses (eg, inflammatory bowel disease or gynecologic pathology), some of which (eg, Crohn disease) are better managed nonoperatively. Data from SCOAP further suggest that CT is more effective than ultrasonography at accurately detecting acute appendicitis. In populations for which ionizing radiation is a concern, however, sequenced algorithms of US followed by CT scan for inconclusive US results may be appropriate. This latest report from SCOAP demonstrates the value of programs that facilitate collaborative, peer-driven quality improvement based on benchmarks for processes of care and for outcomes.

REFERENCES

1. Rettenbacher T, Hollerweger A, Gritzmann N, et al. Appendicitis: should diagnostic imaging be performed if the clinical presentation is highly suggestive of the disease? *Gastroenterology.* 2002;123:992–998.
2. Lee CC, Golub R, Singer AJ, et al. Routine versus selective abdominal computed tomography scan in the evaluation of right lower quadrant pain: a randomized controlled trial. *Acad Emerg Med.* 2007;14:117–122.
3. Santos DA, Manunga J, Jr, Hohman D, et al. How often does computed tomography change the management of acute appendicitis? *Am Surg.* 2009;75:918–921.
4. Applegate KE, Sivit CJ, Salvator AE, et al. Effect of cross-sectional imaging on negative appendectomy and perforation rates in children. *Radiology.* 2001;220:103–107.
5. Bendeck SE, Nino-Murcia M, Berry GJ, et al. Imaging for suspected appendicitis: negative appendectomy and perforation rates. *Radiology.* 2002;225:131–136.
6. Coursey CA, Nelson RC, Patel MB, et al. Making the diagnosis of acute appendicitis: do more preoperative CT scans mean fewer negative appendectomies? A 10-year study. *Radiology.* 2010;254:460–468.
7. Guller U, Rosella L, McCall J, et al. Negative appendicectomy and perforation rates in patients undergoing laparoscopic surgery for suspected appendicitis. *Br J Surg.* 2011;98:589–595.
8. Jones K, Pena AA, Dunn EL, et al. Are negative appendectomies still acceptable? *Am J Surg.* 2004;188:748–754.

9. Piper HG, Rusnak C, Orrom W, et al. Current management of appendicitis at a community center—how can we improve? *Am J Surg.* 2008;195:585–588; discussion 588–589.
10. Rao PM, Rhea JT, Rattner DW, et al. Introduction of appendiceal CT: impact on negative appendectomy and appendiceal perforation rates. *Ann Surg.* 1999;229:344–349.
11. Rao PM, Feltmate CM, Rhea JT, et al. Helical computed tomography in differentiating appendicitis and acute gynecologic conditions. *Obstet Gynecol.* 1999;93:417–421.
12. Wagner PL, Eachempati SR, Soe K, et al. Defining the current negative appendectomy rate: for whom is preoperative computed tomography making an impact? *Surgery.* 2008;144:276—282.
13. Webb EM, Nguyen A, Wang ZJ, et al. The negative appendectomy rate: who benefits from preoperative CT? *AJR Am J Roentgenol*. 2011;197:861–866.
14. Weyant MJ, Eachempati SR, Maluccio MA, et al. Interpretation of computed tomography does not correlate with laboratory or pathologic findings in surgically confirmed acute appendicitis. *Surgery.* 2000;128:145–152.
15. Horton MD, Counter SF, Florence MG, et al. A prospective trial of computed tomography and ultrasonography for diagnosing appendicitis in the atypical patient. *Am J Surg.* 2000;179:379–381.
16. Walker S, Haun W, Clark J, et al. The value of limited computed tomography with rectal contrast in the diagnosis of acute appendicitis. *Am J Surg.* 2000;180:450–454; discussion 454–455.
17. Adibe OO, Amin SR, Hansen EN, et al. An evidence-based clinical protocol for diagnosis of acute appendicitis decreased the use of computed tomography in children. *J Pediatr Surg.* 2011;46:192–196.
18. Gaitini D, Beck-Razi N, Mor-Yosef D, et al. Diagnosing acute appendicitis in adults: accuracy of color Doppler sonography and MDCT compared with surgery and clinical follow-up. *AJR Am J Roentgenol.* 2008;190:1300–1306.
19. Garcia Pena BM, Mandl KD, Kraus SJ, et al. Ultrasonography and limited computed tomography in the diagnosis and management of appendicitis in children. *JAMA.* 1999;282:1041–1046.
20. Krishnamoorthi R, Ramarajan N, Wang NE, et al. Effectiveness of a staged US and CT protocol for the diagnosis of pediatric appendicitis: reducing radiation exposure in the age of ALARA. *Radiology.* 2011;259:231–239.
21. Poletti PA, Platon A, De Perrot T, et al. Acute appendicitis: prospective evaluation of a diagnostic algorithm integrating ultrasound and low-dose CT to reduce the need of standard CT. *Eur Radiol.* 2011;21:2558–2566.
22. Poortman P, Oostvogel HJ, Bosma E, et al. Improving diagnosis of acute appendicitis: results of a diagnostic pathway with standard use of ultrasonography followed by selective use of CT. *J Am Coll Surg.* 2009;208:434–441.
23. Ramarajan N, Krishnamoorthi R, Barth R, et al. 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.
24. Cuschieri J, Florence M, Flum DR, et al. Negative appendectomy and imaging accuracy in the Washington State Surgical Care and Outcomes Assessment Program. *Ann Surg.* 2008;248:557–563.
25. Huynh V, Lalezarzadeh F, Lawandy S, et al. Abdominal computed tomography in the evaluation of acute and perforated appendicitis in the community setting. *Am Surg.* 2007;73:1002–1005.
26. Frei SP, Bond WF, Bazuro RK, et al. Appendicitis outcomes with increasing computed tomographic scanning. *Am J Emerg Med.* 2008;26:39–44.
27. Sarkaria IS, Eachempati SR, Weyant MJ, et al. Current surgical opinion of computed tomography for acute appendicitis. *Surg Infect (Larchmt).* 2004;5:243–252.
28. 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.
29. Pickhardt PJ, Lawrence EM, Pooler BD, et al. Diagnostic performance of multidetector computed tomography for suspected acute appendicitis. *Ann Intern Med.* 2011;154:789–796, W-291.
30. Dearing DD, Recabaren JA, Alexander M. Can computed tomography scan be performed effectively in the diagnosis of acute appendicitis without the added morbidity of rectal contrast? *Am Surg.* 2008;74:917–920.
31. Kim K, Rhee JE, Lee CC, et al. Impact of helical computed tomography in clinically evident appendicitis. *Emerg Med J.* 2008;25:477–481
32. Platon A, Jlassi H, Rutschmann OT, et al. Evaluation of a low-dose CT protocol with oral contrast for assessment of acute appendicitis. *Eur Radiol.* 2009;19:446–454.
33. Flum DR, Koepsell T. The clinical and economic correlates of misdiagnosed appendicitis: nationwide analysis. *Arch Surg.* 2002;137:799–804; discussion 804.
34. Weyant MJ, Eachempati SR, Maluccio MA, et al. The use of computed tomography for the diagnosis of acute appendicitis in children does not influence the overall rate of negative appendectomy or perforation. *Surg Infect (Larchmt).* 2001;2:19–23.
35. Flum DR, Morris A, Koepsell T, et al. Has misdiagnosis of appendicitis decreased over time? A population-based analysis. *JAMA.* 2001;286:1748–1753.

DISCUSSANT

DR. PHILIP BARIE (New York, NY): It is sobering to reflect on our nearly 15–year quest to improve appendicitis diagnostics. The journey and the debate, as is often the case for a transformative idea, began with the finding by Row, et al, in 1998 that helical-computed tomography improved the accuracy of diagnosis. At the time, it was a marvel that clinicians would consider changing practice based on a single noncomparative study that contained a mere 52 patients with the disease under study, but such is the nature of disruptive innovation.

The report described CT imaging of the pelvis after a water-soluble contrast enema alone. Neither oral nor intravenous contrast was administered. The reported accuracy was 98%.

Concurrently, we reported the accuracy of helical CT to be only 88% for the diagnosis of appendicitis, with much depending on who was, and at what time we were, interpreting the studies. Although we concurred at the time that increased CT utilization was associated with a decreased negative appendectomy rate.

As experience has accrued, it appears that CT is 92% to 97% accurate for the diagnosis of appendicitis. Moreover, oral contrast, now standard, may not be necessary. These types of changes over time are but a few of the many potential confounders in attempting to do these longitudinal analyses.

In 2004, in New York State, for example, only 32% of appendectomies were performed laparoscopically and only 8% of hospitals in the state performed more than 75% of appendectomies by laparoscopy.

The practicing surgeon was arguably slow on the uptake, as Dr. Drake mentioned. In 2004, we found that 70% of general surgeons believed that the accuracy of appendiceal CT was lower than reported, that 62% believed CT to be overutilized as a diagnostic test, and that fewer than 40% of surgeons ordered CT for more than one half of their patients. That has become moot as it is now a rarity to be called as a surgeon to see a possible case of appendicitis that has not already been imaged by the primary physician or an emergency physician.

Ironically, the single randomized prospective trial of clinical acumen versus imaging for the diagnosis of appendicitis from the University of Miami showed no benefit in imaging, even in reproductive age females, the subacute group at punitive high risk for misdiagnosis.

By contrast, we revisited the question of who benefits from preoperative imaging in the 2008 report in the context of having increased our rate of imaging to 95% and our negative appendectomy rate to 5%, and found that only reproductive age females benefited from imaging insofar as the negative appendectomy rate is concerned, and that males and children did not. This is in contradistinction to the findings of the present study.

I have four questions for the authors.

First, is there any incentive for Washington hospitals to participate in SCOAP, other than altruism?

Two, according to Wikipedia, that bastion of knowledge, there are 114 hospitals in the state of Washington. Not all of them are acute-care general hospitals, but, arguably, SCOAP has the participation of about one half of the hospitals in the state of Washington. What proportion of inpatient care do the SCOAP hospitals provide and, specifically, what proportion of appendectomies?

Obviously, not all hospitals are created equal. And what is the potential for error to be introduced from studying what is in essence a convenient sample, although a highly convenient one, as opposed to a population-based sample, understanding the limits that you have written about in using administrative data in the case of the latter?

Third, in the essayist's experience, overriding a false CT scan in favor of a brief period of observation is commonplace, and an important management consideration for the patient with a normal white blood cell count, an uncaptured error in your study that would result in an overestimate of positive predictive value.

The problem with the false-positive CT scan leading to a negative appendectomy seems to me to be a larger problem than negative studies being overridden by a decision to operate. Among the negative appendectomies you evaluated, what was the proportion of incorrect imaging versus incorrect judgment?

Finally, current data indicate that 95% of patients diagnosed with appendicitis undergo appendectomy. As Dr. Lucas alluded moments ago, given the emerging hypothesis that some cases are manageable nonoperatively, could imaging also be driving an expensive overutilization of surgical services, another category of unnecessary surgery not accounted for by your model? I wonder if you could comment on that.

CLOSING DISCUSSANT

DR. F. THURSTON DRAKE: One of the major incentives for Washington State hospitals to participate in SCOAP is the opportunity for surgeons to be in the driver's seat of a process that's going on across the country, which is an increasing emphasis on and attention to quality coming from policy makers and payers. Rather than having priorities imposed upon them, surgeons and hospitals in Washington are themselves proactively establishing priorities and driving improvements. We think one of the additional strengths of SCOAP is that it is led by surgeons but is also a collaborative effort among multiple disciplines, for instance, radiologists. So that is one of the biggest incentives, which is surgeons in each of their individual hospitals are really driving the ways in which they approach improvements in quality and how they measure the outcomes of their efforts.

Your second question noted that there are as many as 114 hospitals in the state of Washington and yet we report that about 60 are currently participating in SCOAP. Although I don't have an exact percentage, a substantial majority of appendectomies in the state take place in a SCOAP hospital. For a separate project, I have been looking at a list of all the hospitals in the state published by the state department of health, and there are a fair amount of institutions registered as hospitals that are, for instance, long-term acute-care hospitals, psychiatric hospitals, and a fair number of small hospitals in the state that do not provide surgery at all or only have the capacity for elective surgeries.

As to the point about convenience sampling, it's necessary to note that, although we currently capture more than 85% of the state's appendectomies, in 2006 it was much less. And so this certainly has the potential to introduce bias, but this has lessened considerably over time, and, at this point, we have a very close to a population-based sample.

Dr. Barie, you also mentioned the issue, which we discuss extensively in the paper, of the potential for error in the estimation of positive predictive value. Our estimate is several percentage points higher than the 94% reported just last year in *Annals of Internal Medicine* by Pickhardt – a study I consider one of the most thorough prospective studies of CT scan in appendicitis. You are right to point out that this is likely related to the fact that we do not capture those false positives that are correctly overruled by the surgeon, because those patients are either admitted or discharged, and they never go to the operating room. I think it is an accurate criticism of our positive predictive value calculation, one that we also noted in the paper and can only take it in that context.

As for incorrect surgical judgment versus incorrect CT scan, leading to negative appendectomy, we did not differentiate negative appendectomies in this manner. Our calculated positive predictive value would suggest that, of those patients taken to the OR with a positive CT scan, between 2 and 3 percent had a false positive scan.

Finally, you mentioned the data that has emerged from Europe, and primarily the Scandinavian countries, though British researchers have just published a meta-analysis on the same topic, on treating appendicitis with antibiotics instead of surgery. They have a good success rate. I believe around 60% of the patients in the longest series have not needed to come back for an operation.

This is not a practice model that we have undertaken in the United States, so I don't think we can comment on how that would fare in SCOAP. To your specific point about unnecessary surgery, several researchers, primarily Roland Andersson, who is writing papers from Sweden, make exactly your point, which is that CT scan may be driving the diagnosis of appendicitis that might otherwise have resolved and never been taken to the operating room. I think that this is an important consideration, especially in this era of managing limited healthcare resources.

DISCUSSANT

DR. JOSEPH COFER (Chattanooga, TN): Greater than 25% negative surgical rate in nonimaged women alarmed me. I have two questions.

First, have we as surgeons lost the ability to diagnose appendicitis with history and physical, including a pelvic?

Second, did you look at your negative appendectomies and try to correlate them with either years away from training or age of the surgeon as a surrogate?

CLOSING DISCUSSANT

DR. F. THURSTON DRAKE: We were not able to evaluate how experienced was the surgeon seeing the patients in the emergency department. As to your first question about whether surgeons have lost their ability to diagnose appendicitis, I think that the diagnostic acumen of surgeons, based on clinical examination, laboratory findings, and history-taking, may very well be the same, and still, in most cases, very good; the message we take from our data is that advanced imaging is simply a tool one should add to these other tools in order to augment and improve clinical judgment.

DISCUSSANT

DR. FRANK LEWIS, JR. (Philadelphia, PA): From the large classical studies of appendicitis in the '70s and the '80s, where gender differences could be ascertained, the false-negative rate in women was mostly reported in the mid 30s, around 35 or so percent. So a 25% native rate is actually a fair improvement over what used to be reported. But bringing it down to single digits is certainly a further major improvement from that.

I would echo Dr. Barie's question, which I did not hear the clear answer to. Given the clinical diagnosis in males, most people think that an experienced clinician can see, in males, a false-negative rate of 7%, 8%, 9%. So the question is what is the cost-benefit ratio exactly from using CT scanning? Have you thought about that, assuming that the surgeon had the opportunity to decide whether to order the CT scan or not?

Secondly, who reads the CT scan? Does the surgeon read them, or do you rely entirely on radiologic diagnosis?

Thirdly, have you found that the issue is a significant problem? When you have a CT positive, if clinically you do not really see that it is a classical presentation, how do you override that?

Conversely, if you see what turns out to be a false negative that actually is appendicitis, have you had significant delays in deciding to go to the OR and recognized that? Because, certainly, in inexperienced hands, once an image is present, and there is a black-and-white reading from a radiologist, it is very hard to countermand that with the clinical data. I wonder if you would comment on that phenomenon?

CLOSING DISCUSSANT

DR. F. THURSTON DRAKE: Regarding the evaluation of male patients with suspected appendicitis, within the SCOAP cohort of adolescent and adult males, we found that 10% of those who do not have preoperative imaging had a negative appendectomy, and 3% of those who were imaged had a negative appendectomy. So even among males, our data have suggested a benefit for imaging. There are several conditions in males that present as right lower quadrant pain and may masquerade as appendicitis: Crohn's disease, mesenteric adenitis, and so forth. And both of those are instances in which they are best managed nonoperatively. So I think imaging can be of benefit in male patients as well.

Our analysis is based on the final radiology report in the patients' charts, so this may or may not be what is actually used for decision-making by the surgeon in the middle of the night. We do mention this as a limitation in our manuscript and, as Dr. Barie's group has observed in some of their work, the interpretive skill of the person reading the imaging has a big impact on the diagnostic performance of the tests.

As for costs, although a formal cost effectiveness model has not been produced comparing costs of routine US and/or CT imaging in suspected appendicitis to the costs of negative appendectomy at higher rates, previous work form our group has demonstrated that costs associated with negative appendectomy are substantially elevated over costs associated with positive appendectomy. That doesn't directly answer your question, but I mention it to illustrate the fact that negative appendectomy is a relatively costly procedure.

DISCUSSANT

DR. MARCO PATTI (Chicago, IL): When you compare the results of the physical examination with the CT and the findings in operated on, how often do you feel that you really needed a CT scan? This would suggest a selective approach rather than doing CTs in everybody.

My second question, in 2005, UCSF was one of the first programs to introduce a surgical hospitalist system. Within 30 minutes, an attending surgeon had to see a patient in the emergency room. Magically, the number of CTs decreased, but, also, the number of negative appendectomies decreased. Do you have any information about that?

CLOSING DISCUSSANT

DR. F. THURSTON DRAKE: Regarding your first question about how physicians are choosing whom to image, we thought about that a great deal, because one of the limitations of our data set is that we don't know how surgeons and emergency physicians in SCOAP hospitals are making this choice.

One of the considerations we had, however, was that it is likely that if surgeons and emergency department physicians are making a choice – and many are – they are more likely to image confusing patients more often than straightforward patients. And so we actually thought that this may be a conservative bias for our finding, in that more confusing patients are being imaged, and yet, in the patients who are imaged, there is a lower negative appendectomy rate. To directly answer your question, if the less confusing patients are the ones who are not imaged, our data would suggest that the selective approach to imaging still risks taking patients to the OR for operations that they do not need.

Regarding surgical hospitalists, that is a phenomenon that is occurring in the State of Washington as well, both in the academic institutions and private institutions. And I think there is data to suggest that this acute care model of surgery provides more timely care for the surgical patients, and that may be improving outcomes in these patients, including appendicitis patients.

DISCUSSANT

DR. SHERRY WREN (Stanford, CA): Do you know processes of care? It is sort of like Marco's question. Because if they are clinically using something like an Alvarado score and then only sending high-risk people based on that, in an effort to diminish scans, et cetera, that could absolutely change your positive predictive values.

Secondly, I disagree with your supposition that the ERs are only sending confusing patients. My observation is that they send every patient, until they implement Alvarado scoring.

CLOSING DISCUSSANT

DR. F. THURSTON DRAKE: Your point about processes of care is extremely important and is one of the main areas in which SCOAP focuses on improving care for patients. Not just looking at outcomes, but backing up and looking at how patients got to those outcomes and what changes could be made along the way.

This particular research study only looked at certain outcomes in terms of negative appendectomy and specific points in the diagnostic pathway that patients took to get there, but as a group of hospitals and surgeons, SCOAP is very much focused on processes and making improvements.

I certainly agree with you that most patients now receive CT scan as a relatively routine aspect of an emergency department workup. However, specifically in terms of a factor in our inferential analysis, our supposition was that complexity would potentially be a conservative bias.

