Validation of the American Association for the Surgery of Trauma grading system for acute appendicitis severity

Charles A. Mouch, MD, Anne H. Cain-Nielsen, MS, Beckie L. Hoppe, RN, BSN, MSN, Maria P. Giudici, BSN, MSN, John R. Montgomery, MD, John W. Scott, MD, MPH, David A. Machado-Aranda, MD, *and* **Mark R. Hemmila, MD***, Ann Arbor, Michigan*

AAST Continuing Medical Education Article

Accreditation Statement

This activity has been planned and implemented in accordance with the Essential Areas and Policies of the Accreditation Council for Continuing Medical Education through the joint providership of the American College of Surgeons and the American Association for the Surgery of Trauma. The American College Surgeons is accredited by the ACCME to provide continuing medical education for physicians.

AMA PRA Category 1 Credits™

The American College of Surgeons designates this journal-based CME activity for a maximum of 1 AMA PRA Category 1 Credit™. Physicians should claim only the credit commensurate with the extent of their participation in the activity.

Of the AMA PRA Category 1 Credit™ listed above, a maximum of 1 credit meets the requirements for self-assessment.

Credits can only be claimed online

AMERICAN COLLEGE OF SURGEONS

Inspiring Quality: Highest Standards, Better Outcomes

100+*years*

Objectives

After reading the featured articles published in the *Journal of Trauma and Acute Care Surgery*, participants should be able to demonstrate increased understanding of the material specific to the article. Objectives for each article are featured at the beginning of each article and online. Test questions are at the end of the article, with a critique and specific location in the article referencing the question topic.

Claiming Credit

To claim credit, please visit the AAST website at http://www.aast.org/ and click on the "e-Learning/MOC" tab. You must read the article, successfully complete the post-test and evaluation. Your CME certificate will be available immediately upon receiving a passing score of 75% or higher on the post-test. Post-tests receiving a score of below 75% will require a retake of the test to receive credit.

Disclosure Information

In accordance with the ACCME Accreditation Criteria, the American College of Surgeons, as the accredited provider of this journal activity, must ensure that anyone in a position to control the content of J Trauma Acute Care Surg articles selected for CME credit has disclosed all relevant financial relationships with any commercial interest. Disclosure forms are completed by the editorial staff, associate editors, reviewers, and all authors. The ACCME defines a `commercial interest' as "any entity producing, marketing, re-selling, or distributing health care goods or services consumed by, or used on, patients." "Relevant" financial relationships are those (in any amount) that may create a conflict of interest and occur within the 12'months preceding and during the time that the individual is engaged in writing the article. All reported conflicts are thoroughly managed in order to ensure any potential bias within the content is eliminated. However, if you'perceive a bias within the article, please report the circumstances on the evaluation form.

Please note we have advised the authors that it is their responsibility to disclose within the article if they are describing the use of a device, product, or drug that is not FDA approved or the off-label use of an approved device, product, or drug or unapproved usage.

Disclosures of Significant Relationships with Relevant Commercial Companies/Organizations by the Editorial Staff

Ernest E. Moore, Editor: PI, research support and shared U.S. patents Haemonetics; PI, research support, Instrumentation Laboratory, Inc.; Co-founder, Thrombo Therapeutics. Associate Editors David Hoyt, Ronald V. Maier and Steven Shackford have nothing to disclose. Editorial staff and Angela Sauaia have nothing to disclose.

Author Disclosures

Anne Cain-Nelson: Research Grant, Blue Cross Blue Shield of Michigan; Blue Care Network, software; Beckie Hoppe: Research Grant, Blue Cross Blue Shield of Michigan and Blue Care Network, software; Mark Hemmila: Research Grant, Blue Cross Blue Shield of Michigan and Blue Care Network, software.

Reviewer Disclosures

The reviewers have nothing to disclose.

Cost

For AAST members and Journal of Trauma and Acute Care Surgery subscribers there is no charge to participate in this activity. For those who are not a member or subscriber, the cost for each credit is $25.

System Requirements

The system requirements are as follows: Adobe® Reader 7.0 or above installed; Internet Explorer® 7 and above; Firefox® 3.0 and above, Chrome® 8.0 and above, or Safari™ 4.0 and above.

Questions

If you have any questions, please contact AAST at 800-789-4006. Paper test and evaluations will not be accepted.

OBJECTIVE: The American Association for the Surgery of Trauma (AAST) developed an anatomic grading system to assess disease severity through increasing grades of inflammation. Severity grading can then be utilized in risk-adjustment and stratification of patient outcomes for clinical benchmarking. We sought to validate the AAST appendicitis grading system by examining the ability of AAST grade to predict clinical outcomes used for clinical benchmarking.

METHODS: Surgical quality program data were prospectively collected on all adult patients undergoing appendectomy for acute appendicitis at our institution between December 2013 and May 2018. The AAST acute appendicitis grade from 1 to 5 was assigned for all patients undergoing open or laparoscopic appendectomy. Primary outcomes were occurrence of major complications, any complications, and index hospitalization length of stay. Multivariable models were constructed for each outcome without and with inclusion of the AAST grade as an ordinal variable. We also developed models using International Classification of Diseases, 9th or 10th Rev.—Clinical Modification codes to determine presence of perforation for comparison.

RESULTS: A total of 734 patients underwent appendectomy for acute appendicitis. The AAST score distribution included 561 (76%) in grade 1, 49 (6.7%) in grade 2, 79 (10.8%) in grade 3, 33 (4.5%) in grade 4, and 12 (1.6%) in grade 5. The mean age was 35.3 ± 14.7 years, 47% were female, 20% were nonwhite, and 69% had private insurance. Major complications, any complications, and hospital length of stay were all positively associated with AAST grade ($p < 0.05$). Risk-adjustment model fit improved after including AAST grade in the major complications, any complications, and length of stay multivariable regression models. The AAST grade was a better predictor than perforation status derived from diagnosis codes for all primary outcomes studied.

CONCLUSION: Increasing AAST grade is associated with higher complication rates and longer length of stay in patients with acute appendicitis. The AAST grade can be prospectively collected and improves risk-adjusted modeling of appendicitis outcomes. (*J Trauma Acute Care Surg.* 2020;88: 839–846. Copyright © 2020 Wolters Kluwer Health, Inc. All rights reserved.)

LEVEL OF EVIDENCE: Prospective/Epidemiologic, Level III.

KEY WORDS: Emergency general surgery; AAST grade; grading scale; appendicitis; clinical outcomes.

The American Association for the Surgery of Trauma (AAST) created and published an anatomical disease severity grading system for emergency general surgery (EGS) in 2014.[1,2] Similar to the AAST Organ Injury Scale utilized to grade the severity of traumatic organ injury, the AAST grading system was developed to provide a uniform method to assess disease severity for a variety of EGS conditions.[2] Reflecting the broad nature of inflammatory conditions encountered in EGS practice, the AAST grading system includes specific grading criteria for the diseases of appendicitis, breast infection, cholecystitis, diverticulitis, esophageal perforation, hernia, infectious colitis, bowel obstruction, intestinal ischemia, pancreatitis, pelvic inflammatory disease, perforated peptic ulcer, perirectal abscess, pleural space infection, soft tissue infection, and surgical site infections.[3] Within each condition, the AAST grading system uses clinical, radiographic, operative, and pathologic criteria to assign a score of 1 to 5 (1, least severe; 5, most severe) based on the overall severity of the disease.

Surgeons intuitively recognize that the degree of inflammation due to an acute disease such as appendicitis can impact an operation or hospital course. The AAST grading system is intended to provide researchers with a way to uniformly quantify this difference using objective standards. As such, AAST grade has the potential to serve as a useful clinical benchmarking measure by allowing the comparison of patients according to objective measures of disease severity. Previous research has suggested that increasing AAST grades for acute appendicitis are associated with increasing cost, complications, operative duration, length of stay, and need for open surgical technique in a variety of populations.[4–9] However, previous studies exploring the relationship of AAST grade with clinical outcomes in appendicitis have not examined the feasibility of using AAST grade for risk adjustment in clinical benchmarking. It therefore remains to be seen whether AAST grade can be used as an effective metric for future risk-adjustment.

Our objective was to conduct a retrospective cross-sectional analysis of prospectively collected data to determine the relationship between AAST grade and clinical outcomes in acute appendicitis. We hypothesized that prospectively-collected AAST grades for appendicitis would be associated with clinical outcomes such as complications and hospital length of stay. By conducting this study, we provide support for use of the AAST grading scale as a valid and optimal measure for risk-adjustment in clinical benchmarking and outcomes research.

Submitted: August 20, 2019, Revised: February 11, 2020, Accepted: February 12, 2020, Published online: March 16, 2020.

From the Department of Surgery (C.A.M., J.R.M., J.W.S., D.A.M.-A., M.R.H.), and Center for Health Outcomes and Policy (A.H.C.-N., B.L.H., M.P.G., J.W.S., M.R.H.), University of Michigan, Ann Arbor, Michigan.

This study was presented at the American Association for the Surgery of Trauma 78th Meeting on September 18–21, 2019 in Dallas, TX as well as at the Coller Surgical Society Annual Meeting on September 27, 2019 in Ann Arbor, MI.

Supplemental digital content is available for this article. Direct URL citations appear in the printed text, and links to the digital files are provided in the HTML text of this article on the journal's Web site (www.jtrauma.com).

Address for reprints: Charles A. Mouch, MD, 2101 Taubman Center, 1500 E. Medical Center Dr., Ann Arbor, MI 48109; email: cmouch@med.umich.edu.

DOI: 10.1097/TA.0000000000002674

METHODS

Data

We utilized data from the Michigan Medicine Emergency General Surgery Database, an institutional quality improvement program that prospectively records clinical data from patients receiving both operative and nonoperative treatment by the Acute Care Surgery service at Michigan Medicine. All patients were treated at University Hospital, a 550-bed adult medical center located in Ann Arbor, MI. This hospital serves as a major tertiary referral center for Michigan and the northern Great Lakes region. The Michigan Medicine Acute Care Surgery service performs approximately 150 appendectomies per year at this location. All adult (age, ≥18 years) patients at our institution who underwent open (Current Procedural Terminology (CPT) code 44950, 44960) or laparoscopic (CPT 44970, 44979) appendectomy for acute appendicitis between December 2013 and May 2018 were eligible for inclusion in the study. Pediatric patients (age, <18 years) and patients with acute appendicitis who did not undergo operation were excluded from the analysis.

The data collection platform and methodology utilized by the Michigan Surgical Quality Collaborative (MSQC) was used for data capture and entry into a secure database.[10] Michigan Surgical Quality Collaborative is a collaborative quality initiative sponsored by Blue Cross Blue Shield of Michigan focused on improving the care of surgical patients in the state of Michigan. All of the data collected were abstracted from the electronic medical record at Michigan Medicine and only single-center data from Michigan Medicine were used in the analysis. Data on patient outcomes were collected from the electronic medical record from the time of surgery through 30 days postoperatively, which included ED visits, clinic appointments, and readmissions that occurred during this time. The data definitions followed for data abstraction were provided by MSQC in their annually published program manual. Prior to beginning data collection, the data abstractor completed the training program required by MSQC to be a Surgical Clinical Quality Reviewer. The same data abstractor collected all of data longitudinally over the 5-year study period and was blinded with regard to the outcomes to be included in this study.

Variables

Our independent variable was AAST grade, recorded as an ordinal variable from 1 to 5 by trained data abstraction staff based on review of clinical, imaging, operative, and pathology characteristics documented in the medical record. Table 1 illustrates the published criteria for each AAST acute appendicitis severity grade.[11] The hierarchy from highest to lowest source of information is as follows: pathology report, operative report, radiology studies. The abstractor resolved differences within this hierarchy using clinical judgment. For the few instances where there was confusion that the abstractor could not resolve, she sought consultation/review from the senior author regarding which AAST code to record. We also recorded International Classification of Diseases, 9th or 10th Rev.—Clinical Modification codes (ICD-9, ICD-10) related to acute appendicitis and used these codes to determine perforation status (i.e., perforated [540.0, 540.1, K35.21, K35.32, K35.33] or nonperforated appendicitis) of each patient. Our primary outcomes were major complications, any complications, and hospital length of stay.

Complication groupings were defined in a manner similar to previously published work and are consistent with MSQC benchmark reporting for major and minor complications.[12,13] We defined major complications as the occurrence of any of the following: anastomotic leak, deep incisional skin infection, organ space infection, severe sepsis, septic shock, deep venous thrombosis, pulmonary embolism, myocardial infarction, stroke, unplanned postoperative intubation, cardiac arrest, or death. We defined any complications as the occurrence of a major complication or any of the following: superficial incisional skin infection, urinary tract infection, *Clostridium difficile* infection, central line infection, pneumonia, sepsis, postoperative ileus, cardiac dysrhythmia, transfusion within 72 hours postoperatively, and acute renal insufficiency/failure. We measured hospital length of stay as the duration, in days, of the index hospitalization during which the patient underwent appendectomy for acute appendicitis. Other covariates included demographic information (age, sex, race), insurance type, patient comorbid conditions, and laparoscopic or open surgical technique. Consistent with recommendations from the International Committee of Medical Journal Editors, race was only included as a covariate due to prior data demonstrating an association between race/ethnicity and appendicitis presentation not accounted for by other variables in our models.[14,15]

Statistical Analysis

We performed univariate analyses (Fisher's exact tests for binary outcomes, Kruskall-Wallis tests for continuous outcomes)

TABLE 1. AAST Anatomic Severity of Disease Grading System for Acute Appendicitis

AAST Grade	Description	Clinical Criteria	Imaging Criteria	Operative Criteria	Pathologic Criteria
1	Intact but acutely inflamed appendix	Pain, leukocytosis, and RLQ tenderness	Inflammatory changes localized to the appendix with possible appendiceal dilation or contrast nonfilling	Intact but acutely inflamed appendix	Presence of neutrophils at the base of crypts, submucosa, and possibly in the muscular wall
2	Intact but gangrenous appendix	Pain, leukocytosis, and RLQ tenderness	Appendiceal wall necrosis with contrast nonenhancement and possible air in appendiceal wall	Intact but gangrenous appendix	Mucosa and muscular wall digestion but not identifiable on H&E stain
3	Perforated appendix with local contamination	Pain, leukocytosis, and RLQ tenderness	Any of above with local periappendiceal fluid and possible contrast extravasation	Any of above with evidence of local contamination	Gross perforation or focal dissolution of muscular wall
4	Perforated appendix with periappendiceal phlegmon or abscess	Pain, leukocytosis, and RLQ tenderness. May have palpable mass.	Regional soft tissue inflammatory changes, phlegmon, or abscess	Any of above with abscess or phlegmon in region of appendix	Gross perforation
5	Perforated appendix with generalized peritonitis	Generalized peritonitis	Diffuse abdominal or pelvic inflammatory changes with possible free intra-peritoneal air or fluid	Any of above with generalized purulent contamination distant from appendix	Gross perforation

Reference: http://www.aast.org/emergency-general-surgery-anatomic-grading-scales.
RLQ, right lower quadrant.

examining the association of major complications, any complications, and length of stay with AAST grade. We then created multivariable logistic regression models for the major and any complications outcomes. Models for length of stay, a right-skewed outcome, violated assumptions of normality so multivariable negative binomial regression models were used, which are more appropriate for skewed count data.[16,17] For all outcomes, we used a stepwise selection process to create a parsimonious multivariable model for each primary outcome while controlling for relevant confounders. We then added AAST grade to the multivariable model for each outcome. Our goal was to create a parsimonious model based on the current data and our group's previous modeling experience in appendicitis that could determine whether or not model fit could be improved by the addition of AAST grade to the model.[12,13,18] We measured goodness-of-fit before and after addition of AAST grade using the area under the receiver operating curve (concordance statistic or c-statistic) for the logistic regression analyses and McFadden's pseudo-R^2 for the negative binomial regression analysis.[19,20] The c-statistic describes how effectively a model predicts the outcome of interest. Its value ranges from 0.5 to 1, with 0.5 representing model prediction that is no better than chance and 1 representing perfect model prediction.[21] Comparing the c-statistic with and without AAST grade included in the model, therefore, allows us to evaluate how the inclusion of AAST grade influences model fit. We evaluated differences between c-statistics by comparing 95% percentile-based confidence intervals after bootstrapping, using 500 replications. We also assessed model calibration for logistic regression models by using calibration plots, which graphically depict observed versus predicted probabilities by decile of predicted outcome. Finally, we added AAST grade, appendiceal perforation status, and both AAST grade and appendiceal perforation status to the multivariable model for each outcome. We then performed likelihood ratio tests to assess model fit and compare the ability of each variable to predict the outcome of interest.

Results are expressed as odds ratios or incidence rate ratios along with 95% confidence intervals. A p value less than 0.05 was used as the threshold for statistical significance, and all reported p values were two-sided. All statistical analyses were performed in Stata 15.0 (StataCorp, College Station, TX). Approval for this study was obtained from the Michigan Medicine Institutional Review Board.

RESULTS

Participants

The final study population included 734 patients who underwent appendectomy for treatment of acute appendicitis. The mean patient age was 35.3 ± 14.7 years, 346 (47.1%) were female, 147 (20.0%) were of nonwhite race, and 506 (68.9%) had private insurance. None of the patients died. Laparoscopic appendectomy was performed in 94.7% of the cohort. Table 2 provides a detailed description of the study population demographics and medical comorbidities.

Six hundred fifty-four patients had documented postoperative follow-up after discharge. The mean length of follow-up was 18.3 ± 10.2 days and the median was 17 days (interquartile range, 13–29 days).

TABLE 2. Patient Characteristics

Characteristics	Patients (N = 734)
Mean age (SD), y	35.3 (14.7)
Age: n (%), y	
18–25	242 (33.0)
26–45	317 (43.2)
>45	175 (23.8)
Male, n (%)	388 (52.9)
Nonwhite race, n (%)	147 (20.0)
Private insurance, n (%)	506 (68.9)
ASA classification, n (%)	
1 to 2	650 (88.6)
3 to 5	84 (11.4)
Comorbid risk factors, n (%)	
Smoker within last year	99 (13.5)
Hypertension	82 (11.2)
Sleep apnea	76 (10.4)
Diabetes	36 (4.9)
Alcohol use (>2 drinks/day in last 2 wk)	17 (2.3)
Steroid use for chronic condition	15 (2.0)
Coronary artery disease	12 (1.6)
History of severe COPD	11 (1.5)
Ascites within 30 d	1 (0.1)
Currently on dialysis	1 (0.1)
Surgical technique, n (%)	
Laparoscopic	695 (94.7)
Laparoscopic converted to open	25 (3.4)
Open	14 (1.9)
Wound classification, n (%)	
Clean-contaminated	117 (15.9)
Contaminated	494 (67.3)
Dirty	123 (16.7)

SD, standard deviation; ASA, American Society of Anesthesiologists; COPD, chronic obstructive pulmonary disease.

Univariate Analyses

Results for any complications, major complications, and hospital length of stay stratified for each AAST grade are shown in Table 3. The AAST grade was significantly associated with both any complications ($p < 0.001$) and major complications ($p < 0.001$) in the univariate analysis. The frequency of each individual complication according to AAST grade for acute appendicitis is provided in a supplemental table (Supplemental Digital Content 1, Table, http://links.lww.com/TA/B605). Increasing AAST grade was also significantly associated with longer hospital length of stay ($p < 0.001$) in the univariate analysis.

Multivariable Analyses

Table 4 illustrates the final multivariable logistic regression models with and without inclusion of AAST grade for the any complications and major complications outcomes. Model discrimination as assessed by the c-statistic was nonsignificantly different with addition of the AAST grade in both the any complications and major complications models (Table 5), although point estimates increased. Model goodness-of-fit improved with addition of the AAST grade to the length of stay model (pseudo-R^2 = 0.10 before, 0.14 after). Likelihood ratio statistical tests

TABLE 3. Frequency of Events According to AAST Grade for Acute Appendicitis

	AAST Acute Appendicitis Severity Grade						
Outcome	**1 (n = 561)**	**2 (n = 49)**	**3 (n = 79)**	**4 (n = 33)**	**5 (n = 12)**	**Total (N = 734)**	***p* Value**
Any complications, n (%)	13 (2.3)	6 (12.2)	13 (16.5)	1 (3.0)	3 (25)	36 (4.9)	<0.001
Major complications, n (%)	1 (0.2)	2 (4.1)	5 (6.3)	0 (0)	1 (8.3)	9 (1.2)	<0.001
Length of stay: median (IQR), d	1 (1–1)	1 (1–2)	2 (1–4)	3 (1–5)	5 (2.5–8.5)	1 (1, 1)	<0.001

IQR, interquartile range.

showed a statistically significant improvement in each of the three outcome models with the addition of AAST grade as a covariate.

Table 6 shows a head-to-head comparison of risk-adjustment models without an assessment of disease severity, with inclusion of perforation status as determined by ICD coding, and lastly using the AAST severity grade in the models. Both of these disease description variables improved model discrimination for all of the studied outcomes; however, the addition of the AAST grade resulted in the highest c-statistic for major complications (c-statistic, 0.89; 95% confidence interval, 0.83–0.98) and any complications (c-statistic, 0.76; 95% confidence interval, 0.69–0.89). Addition of AAST grade and appendiceal perforation status resulted in a significant improvement in model fit for the length of stay over addition of perforation alone (as measured by the likelihood ratio test, $p < 0.001$). Including AAST grade alone versus AAST grade and perforation status for the length of stay outcome resulted in similar model fit (pseudo-R^2 for both models = 0.14).

Visual inspection of calibration plots, available in the supplemental figure (Supplemental Digital Content 2, Figure, http://links.lww.com/TA/B606), show that model calibration was improved by the inclusion of AAST grade for the outcome of major complications, but that model calibration was not improved by the inclusion of AAST grade for the outcome of any complications, except in the highest decile of risk.

DISCUSSION

The main finding of our study is that prospective collection of AAST severity grade is feasible, and inclusion of this variable in risk-adjustment models of acute appendicitis outcomes significantly improved model performance for prediction of complications and hospital length of stay following appendectomy. In addition, utilization of the AAST severity grade produced superior performance as a disease severity descriptor when compared with perforation status as determined by ICD coding. This finding represents a significant opportunity for improving the accuracy of risk-adjustment modeling in the future. We were able to demonstrate the feasibility of prospective data collection and use of AAST grade as a clinical benchmarking description of disease severity for acute appendicitis in a quality improvement database.

Our univariate results showing an association between increasing AAST grade and complications are in line with previous retrospective reports in the literature. For example, a retrospective single-center study found that AAST appendicitis severity grade was significantly associated with complication rate, length of stay, and likelihood of undergoing an open operation.[7] Another retrospective study found that determination of AAST appendicitis severity from operative reports was associated with both the occurrence and severity of complications at an urban South African

TABLE 4. Multivariate Logistic Regression Models of Outcomes Utilizing AAST Grade as a Covariate

Variable	Major Complication OR (95% CI)	*p* Value	Any Complication OR (95% CI)	*p* Value	Length of Stay IRR (95% CI)	*p* Value
Age, y						
18–25	Ref	—	Ref	—	Ref	—
26–45	0.08 (0.01–0.84)	0.04	0.94 (0.37–2.38)	0.9	1.01 (0.85–1.20)	0.9
>45	0.26 (0.04–1.56)	0.1	1.51 (0.57–4.05)	0.4	1.27 (1.03–1.55)	0.02
Male	0.32 (0.07–1.38)	0.1	0.80 (0.40–1.62)	0.5	0.83 (0.73–0.96)	0.01
Nonwhite race	1.05 (0.20–5.60)	0.9	1.21 (0.51–2.84)	0.7	1.07 (0.90–1.27)	0.5
ASA class 3–5	7.35 (1.48–36.4)	0.02	1.83 (0.76–4.43)	0.2	1.72 (1.41–2.11)	<0.001
Type of operation						
Laparoscopic	—	—	Ref	—	Ref	—
Laparoscopic to open	—		1.95 (0.60–6.37)	0.3	1.74 (1.30–2.32)	<0.001
Open	—		0.58 (0.06–5.53)	0.6	1.50 (1.01–2.21)	0.04
Alcohol use	—		—		2.28 (1.63–3.19)	<0.001
Private insurance	—		—		0.89 (0.77–1.04)	0.1
Sleep apnea	-		-		0.82 (0.65–1.04)	0.1
AAST grade	2.43 (1.44–4.09)	0.001	1.64 (1.22–2.23)	0.001	1.42 (1.33–1.52)	<0.001

OR, odds ratio; IRR, incidence rate ratio; CI, confidence interval; Ref, reference.

TABLE 5. Statistical Fit Parameters for Multivariable Models With and Without Inclusion of AAST Grade

Outcome	Model	AUC c-Statistic (95% CI)	Pseudo R^2	Likelihood Ratio Test (*p*-Value)
Major complications	With AAST grade	0.89 (0.83–0.98)	—	0.001
	Without AAST grade	0.81 (0.72–0.95)	—	
Any complications	With AAST grade	0.76 (0.69–0.89)	—	0.002
	Without AAST grade	0.68 (0.63–0.86)	—	
Length of stay	With AAST grade	—	0.14	<0.001
	Without AAST grade	—	0.10	

AUC, area under the curve.

hospital.[8] Finally, the same group also found a significant association between AAST appendicitis severity grade and complications in a pediatric population.[9] In addition, a post hoc analysis of the Eastern Association for the Surgery of Trauma's prospective multicenter MUSTANG observational study demonstrated that AAST grade was a valid predictor for outcomes, such as complications, length of stay, and need for a secondary intervention.[5] Taken together, these studies along with our results suggest that the AAST appendicitis severity grade is a useful tool that can be used to compare and stratify outcomes in patients with appendicitis based on the severity of their disease.

Performance of multivariate risk-adjustment modeling allowed us to demonstrate that inclusion of the AAST severity grade significantly improved the discrimination and fit of these models used to predict clinical outcomes. The AAST grading system represents the first uniform and stepwise measure of disease severity for appendicitis that uses a comprehensive constellation of clinical, imaging, operative, and pathology criteria. Patients with appendicitis can present with a wide range of clinical severity reflecting the variable impact of the acute process. Our results show that AAST grade can be used as an accurate indicator of disease severity and performs better than ICD-derived perforation status in the risk-adjusted prediction of complications and length of stay. We did not attempt to study AAST severity grading in nonoperative patients. However, this is an option for additional use of the scoring system when tracking outcomes, as Hernandez et al.[7] found no difference in AAST grade assigned by radiologic imaging as compared with operative characteristics.

The credibility of outcomes reporting is reliant upon high-quality inputted data and accurate risk-adjustment models to account for confounding factors that impact patient care and influence results. Public reporting of surgeon and hospital outcomes is increasing in prevalence (e.g., Leapfrog Hospital Safety Grade, ProPublica Surgeon Scorecard), but represents an area of considerable controversy with regard to methods of reporting and risk-adjustment. A published critique of the ProPublica Surgeon Scorecard highlighted significant deficiencies in the adequacy of case-mix adjustment with regard to important patient risk factors.[22] The prevalence of disease and spending on patients admitted to acute care hospitals is considerable.[23–25] The acute care surgeon is potentially responsible for the outcomes of 20% of hospital admissions when providing trauma, EGS, and surgical critical care.[26,32] Surgeon-level and hospital-level variations have both been found to influence outcomes in EGS.[27–29] Provision of credible and reliable benchmark reporting to acute care surgeons that takes into account disease severity in risk-adjustment and distinct practice measures found in EGS is likely to result in maximal optimization of outcomes through collaborative quality improvement.[30–32]

TABLE 6. Comparison of AAST Grade and Perforation Status as Risk-Adjustment Model Covariates

Outcome	Model	AUC c-Statistic (95% CI)	Pseudo-R^2	Model Comparison	Likelihood Ratio Test (*p* Value)
Major complications	Base[a]	0.81 (0.72–0.95)	—	—	—
	Base[a] + AAST grade	0.89 (0.83–0.98)	—	Base[a]	0.001
	Base[a] + perforation status	0.82 (0.67–0.96)	—	Base[a]	0.1
	Base[a] + AAST grade + perforation status	0.89 (0.83–0.99)	—	Base[a] + perforation status	0.005
Any complications	Base[b]	0.68 (0.63–0.86)	—	—	—
	Base[b] + AAST grade	0.76 (0.69–0.89)	—	Base[b]	0.002
	Base[b] + perforation status	0.70 (0.64–0.87)	—	Base[b]	0.1
	Base[b] + AAST grade + perforation status	0.76 (0.69–0.89)	—	Base[b] + perforation status	0.008
Length of stay	Base[c]	—	0.10	—	—
	Base[c] + AAST grade	—	0.14	Base[c]	<0.001
	Base[c] + perforation status	—	0.12	Base[c]	<0.001
	Base[c] + AAST grade + perforation status	—	0.14	Base[c] + perforation status	<0.001

Base model description: age, sex, and race were included in all base models. Additional variables in the base models were selected using stepwise selection.
Base[a] model included the following variables: age, sex, race, and ASA classification.
Base[b] model included the following variables: age, sex, race, ASA classification, and surgical technique.
Base[c] model included the following variables: age, sex, race, ASA classification, surgical technique, insurance, alcohol use, and sleep apnea.

Our findings must be interpreted in light of the study's limitations. First, the observational nature of the study limits ability to determine causal inference. Second, our data were drawn from a single tertiary referral center with patients who largely have private insurance, which may limit generalization to smaller practices or those with different patient populations. In addition, our cohort includes only patients who underwent surgery, so the applicability to patients treated nonoperatively is not yet known. Our cohort also represents predominantly younger, healthy patients without many medical comorbidities, which may limit generalizability to more comorbid populations. Postoperative follow-up was limited in that we did not have access to readmissions data at other hospitals and did not have documentation of follow-up in 80 (10.8%) of the 734 total patients. Despite these limitations, this is the first study to prospectively collect data used to evaluate the use of AAST grade as a risk-adjustment tool in EGS, and these findings may direct future research.

The AAST disease severity grading system provides the first uniform mechanism to account for disease burden when performing comparative effectiveness research in EGS. It can be collected in a prospective manner, is easy to measure, and improves model prediction for complications and length of stay. This study, therefore, provides validation for the use of the AAST appendicitis severity grade in future clinical benchmarking and outcomes research. Quality improvement databases capturing AAST disease severity in EGS patient will allow us to better understand optimal methods for managing patients with regard to their clinical outcomes.

AUTHORSHIP

C.A.M. participated in the literature search, data analysis, data interpretation, writing, editing. A.H.C.-N. participated in the study design, analysis, writing, editing. B.L.H. participated in the study design, data collection, data interpretation, editing. M.P.G. participated in the study design, data collection, data interpretation, editing. J.R.M. participated in the data analysis, data interpretation, editing. J.W.S. participated in the literature search, study design, data interpretation, editing. D.A.M.-A. participated in the study design, data collection, data interpretation, editing. M.R.H. participated in the study design, data collection, data interpretation, writing, editing.

ACKNOWLEDGMENTS

Funding: This study was supported by a Blue Cross Blue Shield of Michigan and Blue Care Network Collaborative Quality Initiatives grant awarded to Mark R. Hemmila for administration of the Michigan Trauma Quality Improvement Program.
Role of the Funder/Sponsor: The funding organizations had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the article; and decision to submit the article for publication.

DISCLOSURE

Anne H. Cain-Nielsen, Beckie L. Hoppe, Maria P. Giudici, and Mark R. Hemmila receive salary support from Blue Cross Blue Shield of Michigan and Blue Care Network (a nonprofit mutual company) through grant funding of the Michigan Trauma Quality Improvement Program. The remaining authors declare no conflicts of interest.

REFERENCES

1. Crandall ML, Agarwal S, Muskat P, Ross S, Savage S, Schuster K, Tominaga GT, Shafi S. Application of a uniform anatomic grading system to measure disease severity in eight emergency general surgical illnesses. *J Trauma Acute Care Surg*. 2014;77(5):705–708.
2. Shafi S, Aboutanos M, Brown CV, Ciesla D, Cohen MJ, Crandall ML, Inaba K, Miller PR, Mowery NT. Measuring anatomic severity of disease in emergency general surgery. *J Trauma Acute Care Surg*. 2014;76(3): 884–887.
3. Tominaga GT, Staudenmayer KL, Shafi S, et al. The American Association for the Surgery of Trauma grading scale for 16 emergency general surgery conditions: disease-specific criteria characterizing anatomic severity grading. *J Trauma Acute Care Surg*. 2016;81(3):593–602.
4. Collins CM, Davenport DL, Talley CL, Bernard AC. Appendicitis grade, operative duration, and hospital cost. *J Am Coll Surg*. 2018;226(4):578–583.
5. Vasileiou G, Ray-Zack M, Zielinski M, Qian S, Yeh DD, Crandall M. Validation of the American Association for the Surgery of Trauma emergency general surgery score for acute appendicitis—an EAST multicenter study. *J Trauma Acute Care Surg*. 2019;87(1):134–139.
6. Finnesgard EJ, Hernandez MC, Aho JM, Zielinski MD. The American Association for the Surgery of Trauma emergency general surgery anatomic severity scoring system as a predictor of cost in appendicitis. *Surg Endosc*. 2018;32(12):4798–4804.
7. Hernandez MC, Aho JM, Habermann EB, Choudhry AJ, Morris DS, Zielinski MD. Increased anatomic severity predicts outcomes: validation of the American Association for the Surgery of Trauma's emergency general surgery score in appendicitis. *J Trauma Acute Care Surg*. 2017;82(1):73–79.
8. Hernandez MC, Kong VY, Aho JM, Bruce JL, Polites SF, Laing GL, Zielinski MD, Clarke DL. Increased anatomic severity in appendicitis is associated with outcomes in a South African population. *J Trauma Acute Care Surg*. 2017;83(1):175–181.
9. Hernandez MC, Polites SF, Aho JM, Haddad NN, Kong VY, Saleem H, Bruce JL, Laing GL, Clarke DL, Zielinski MD. Measuring anatomic severity in pediatric appendicitis: validation of the American Association for the Surgery of Trauma appendicitis severity grade. *J Pediatr*. 2018;192:229–233.
10. Michigan Surgical Quality Collaborative Web site. https://msqc.org/. Accessed January 22, 2020.
11. The American Association for the Surgery of Trauma. Data dictionaries for AAST grading system for EGS conditions web site. http://www.aast.org/emergency-general-surgery-anatomic-grading-scales. Accessed January 22, 2020.
12. Hemmila MR, Birkmeyer NJ, Arbabi S, Osborne NH, Wahl WL, Dimick JB. Introduction to propensity scores: a case study on the comparative effectiveness of laparoscopic vs. open appendectomy. *Arch Surg*. 2010;145(10): 939–945.
13. Jenkins PC, Oerline MK, Mullard AJ, Englesbe MJ, Campbell DAJr., Hemmila MR. Hospital variation in outcomes following appendectomy in a regional quality improvement program. *Am J Surg*. 2016;212(5):857–862.
14. International Committee of Medical Journal Editors. Preparing a manuscript for submission to a medical journal web site. http://www.icmje.org/recommendations/browse/manuscript-preparation/preparing-for-submission.html. Updated July 2013. Accessed January 22, 2020.
15. Agency for Healthcare Research and Quality—healthcare cost and utilization project web site. Statistical Brief #139. Barrett ML, Hines, AL, Andrews RM. Trends in rates of perforated appendix 2001-2010. https://www.hcup-us.ahrq.gov/reports/statbriefs/sb159.pdf. Accessed January 22, 2020.
16. Deb P, Norton EC. Modeling health care expenditures and use. *Annu Rev Public Health*. 2018;39:489–505.
17. Mihaylova B, Briggs A, O'Hagan A, Thompson SG. Review of statistical methods for analysing healthcare resources and costs. *Health Econ*. 2011; 20(8):897–916.
18. To KB, Kamdar NS, Patil P, Collins SD, Seese E, Krapohl GL, Campbell DS Jr., Englesbe MJ, Hemmila MR, Napolitano LM, Michigan Surgical Quality Collaborative (MSQC) Emergency General Surgery Study Group and the MSQC Research Advisory Group. Acute care surgery model and outcomes in emergency general surgery. *J Am Coll Surg*. 2019;228(1):21–28.e7.
19. UCLA. Institute for Digital Research & Education: Statistical Consulting Web site. Negative binomial regression—Stata annotated output. https://stats.idre.ucla.edu/stata/output/negative-binomial-regression/. Accessed January 22, 2020.
20. McFadden D. *Conditional logit analysis of qualitative choice behavior*. In: Zarembka P, ed. *Frontiers in Econometrics*. New York, NY: Wiley; 1973.
21. Caetano SJ, Sonpavde G, Pond GR. C-statistic: a brief explanation of its construction, interpretation and limitations. *Eur J Cancer*. 2018; 90:130–132.

22. Friedberg MW, Pronovost PJ, Shahian DM, Safran DG, Bilimoria KY, Elliott MN, Damberg CL, Dimick JB, Zaslavsky AM. A methodological critique of the ProPublica *surgeon* scorecard. *Rand Health Q*. 2016;5(4):1.
23. Montgomery JR, Cain-Nielsen AH, Jenkins PC, Regenbogen SE, Hemmila MR. Prevalence and payments for traumatic injury compared with common acute diseases by episode of care in medicare beneficiaries, 2008–2014. *JAMA*. 2019;321(21):2129–2131.
24. Bernard A, Staudenmayer K, Minei JP, Doucet J, Haider A, Scherer T, Davis KA. Macroeconomic trends and practice models impacting acute care surgery. *Trauma Surg Acute Care Open*. 2019;4(1):e000295.
25. Scott JW, Olufajo OA, Brat GA, Rose JA, Zogg CK, Haider AH, Salim A, Havens JM. Use of national burden to define operative emergency general surgery. *JAMA Surg*. 2016;151(6):e160480.
26. Knowlton LM, Minei J, Tennakoon L, Davis KA, Doucet J, Bernard A, Haider A, Scherer LRT3rd, Spain DA, Staudenmayer KL. The economic footprint of acute care surgery in the United States: implications for systems development. *J Trauma Acute Care Surg*. 2019; 86(4):609–616.
27. Scott JW, Tsai TC, Neiman PU, Jurkovich GJ, Utter GH, Haider AH, Salim A, Havens JM. Lower emergency general surgery (EGS) mortality among hospitals with higher-quality trauma care. *J Trauma Acute Care Surg*. 2018;84(3):433–440.
28. Feeney T, Castillo-Angeles M, Scott JW, Nitzschke SL, Salim A, Haider AH, Havens JM. The independent effect of emergency general surgery on outcomes varies depending on case type: a NSQIP outcomes study. *Am J Surg*. 2018;216(5):856–862.
29. Udyavar R, Cornwell EE, Havens JM, Hashmi ZG, Scott JW, Sturgeon D, Uribe-Leitz T, Lipsitz SR, Salim A, Haider AH. Surgeon driven variability in emergency general surgery outcomes: does it matter who is on call? *Surgery*. 2018;164(5):1109–1116.
30. Havens JM, Peetz AB, Do WS, Cooper Z, Kelly E, Askari R, Reznor G, Salim A. The excess morbidity and mortality of emergency general surgery. *J Trauma Acute Care Surg*. 2015;78(2):306–311.
31. Ingraham AM, Cohen ME, Bilimoria KY, Feinglass JM, Richards KE, Hall BL, Ko CY. Comparison of hospital performance in nonemergency versus emergency colorectal operations at 142 hospitals. *J Am Coll Surg*. 2010; 210(2):155–165.
32. Hemmila MR, Jakubus JL, Cain-Nielsen AH, Kepros JP, Vander Kolk WE, Wahl WL, Mikhail JN. The Michigan Trauma Quality Improvement Program: results from a collaborative quality initiative. *J Trauma Acute Care Surg*. 2017; 82(5):867–876.