



Published in final edited form as:

Hosp Pediatr. 2023 July 01; 13(7): e170–e174. doi:10.1542/hpeds.2023-007204.

Community Validation of an Approach to Detect Delayed Diagnosis of Appendicitis in Big Databases

Kenneth A Michelson, MD, MPH^a, Finn L E McGarghan, BS^a, Mark L Waltzman, MD^a, Margaret E Samuels-Kalow, MD MPhil MSHP^b, Richard G Bachur, MD^a

^aDivision of Emergency Medicine, Boston Children's Hospital, Boston, MA

^bDepartment of Emergency Medicine, Massachusetts General Hospital, Boston, MA

Abstract

Background: Detection of delayed diagnosis using administrative databases may illuminate the healthcare settings at highest risk. A method for detection of delays in claims has been validated in children's hospitals. We sought to further validate the method in community emergency departments (EDs).

Methods: We studied patients <21 years old diagnosed with appendicitis from 2008-2019 in eight eastern Massachusetts EDs. Eligible patients had 2 ED encounters within 7 days, the second with an appendicitis diagnosis. Delayed diagnosis was evaluated in medical records by trained reviewers. A previously validated trigger tool was applied to participants' electronic medical record data. The tool used data elements included in administrative data, including initial encounter diagnoses, time between encounters, presence of medical complexity, and ultimate length of stay. The tool assigned a probability of delayed diagnosis for each patient. Test characteristics at 4 confidence thresholds were determined, and the area under the receiver operating curve (AUC) was calculated.

Results: We analyzed 68 children with two encounters leading to a diagnosis of appendicitis (i.e. possible delay). When assigning a delayed diagnosis prediction to patients at 4 thresholds of confidence (>0%, >50%, >75%, and >90% confident), the positive predictive values (PPVs) were respectively 74%, 89%, 92%, and 89%; the negative predictive values were respectively 100%, 57%, 50%, and 33%. The AUC was 0.837 (95% confidence interval 0.719-0.954).

Conclusion: A trigger tool that identifies delays in diagnosis using only administrative data in community EDs has a high PPV for true delay. The tool may be applied in community EDs.

Address correspondence to: Kenneth A. Michelson; Boston Children's Hospital; Division of Emergency Medicine; 300 Longwood Avenue, BCH 3066; Boston, MA 02115; [kenneth.michelson@childrens.harvard.edu]; Tel: +1.617.355.6624.

Contributor statements:

Dr. Michelson conceptualized and designed the study; performed statistical analysis; drafted and approved the final manuscript.

Mr. McGarghan collected data; designed the data collection instruments; and revised and approved the final manuscript.

Drs. Waltzman and Samuels-Kalow directed regulatory and data collection efforts at study sites; and revised and approved the final manuscript.

Dr. Bachur supervised the design and analysis of the study; and revised and approved the final manuscript.

Conflict of Interest: The authors have no conflicts of interest relevant to this article to disclose

Keywords

Quality; measurement; diagnostic error; appendicitis

Introduction

Appendicitis is common in children but can be difficult to diagnose.¹⁻³ Timely diagnosis can prevent complications including perforated appendicitis, sepsis, and rarely a need for bowel resection.⁴ Approximately 5-10% of children with appendicitis had related visits preceding diagnosis, and complications are more likely after delayed treatment.⁵⁻⁸ Understanding systems factors that predict delays in diagnosis would be useful toward reducing the rate of delayed diagnosis. Identifying delayed diagnosis is challenging because case review is time consuming and requires significant expertise. Thus, a useful approach to delayed diagnosis identification would have high accuracy and would not require manual case review.

We recently developed and validated a method to identify delayed diagnosis in large administrative databases, which are databases that contain patient demographics and healthcare claim (i.e. billing) information generated in the course of patient care.⁹ The advantage of a method that uses only administrative data is that it allows for the study of children who visit community hospitals, which represent most childhood ED visits.¹⁰ Such hospitals do not ordinarily share data for research, so using administrative data are the only feasible source of information for large-scale studies of care. The method consists of a trigger tool, which assigns a probability that delayed diagnosis occurred for each child with appendicitis in the database. This method was created and validated using children's hospital data. Thus, it is unclear how generalizable it is to community hospitals, where case mix, acuity, reasons for delayed diagnosis, and diagnosis coding differ. The ability to use the tool with community hospital data would allow for broad study of rates and consequences of diagnostic delays across all hospital types.

Our objective was to externally validate an approach for retrospectively detecting delayed diagnosis of appendicitis in administrative data from general hospitals, extending our prior work in children's hospitals. Successful validation would indicate that the approach could be used in all types of hospitals.

Methods

We performed a retrospective, cross-sectional study to test a trigger tool that incorporates only variables typically included in administrative data. The tool is used to predict the presence of delayed diagnosis of appendicitis, which we compared to the criterion standard of detailed electronic health record (EHR) review (see Trigger Tool below). Participants were under 21 years old and visited 1 of 8 general EDs in eastern Massachusetts (children with appendicitis per year range 3-127), had a first-time diagnosis of appendicitis, and had an ED visit in the preceding 7 days at any of the sites. EHRs became available at different sites in different years, ranging from 2008-2017. The data were originally collected as part of a study on diagnostic error rates across several diseases.¹¹

The ED encounter associated with the appendicitis diagnosis was designated as the “diagnosis encounter,” and the preceding encounter was designated as the “initial encounter.” For patients with more than one previous encounter, the most recent was designated the initial encounter. Cases were identified for inclusion using diagnosis codes (International Classification of Diseases, 9th Edition, Clinical Modification [ICD-9-CM] 540.x, 541, 542 and ICD-10-CM K35.x-K37.x). Patients were excluded if insufficient medical records existed to determine whether a delayed diagnosis occurred, if no record of a prior encounter existed, if the patient left the ED without being seen, or the patient was transferred at the conclusion of the initial ED visit (which made determination of a delayed diagnosis impossible). This study was approved by the facilities’ institutional review boards under a waiver of informed consent.

All data were drawn from the hospitals’ EHRs, including the administrative components (diagnosis and procedure codes, timestamps, demographics) and clinical components (operative and clinical notes, medication administration records, test results).

Outcome

The reference standard primary outcome was delayed diagnosis as determined by manual case review of the EHR. Delay was defined as appendicitis being present at the initial encounter. Reviewers rated the likelihood that appendicitis was present as “near-definitely not,” “probably not,” “possibly,” “probably,” or “near-definitely” using the same definitions as in the prior validation study.⁹ The definitions were originally developed by multispecialty consensus panel.¹² The reviewer assessment of delayed diagnosis was dichotomized as delayed diagnosis (probably or near-definitely delay) or not delayed diagnosis (possibly, probably not, or near-definitely not delay). A subset of cases (34%) was evaluated by a second reviewer.

Trigger Tool

The goal of the trigger tool was to assign a probability that a patient’s administrative data (i.e. their billing records) represented a real clinical delay in diagnosis. It was originally developed using administrative data from children’s hospitals and validated through chart review.⁹ The tool is a logistic regression model that takes a patient’s administrative data and outputs the probability that a real delay in diagnosis occurred. The inputs to the tool are age, sex, history of a complex chronic condition,¹³ revisit interval (days between initial and diagnosis encounters), diagnosis code for perforated appendicitis (ICD-9-CM 540.0-1, ICD-10-CM K35.2x, K35.32-33), length of stay of the diagnosis encounter (0-1, 2-3, 4-7, or >7 days), and individual presence or absence of specific diagnoses at the initial encounter including abdominal pain, constipation, dehydration, fever, gastroenteritis, genitourinary condition, head/ear/eye/nose/throat condition, leukocytosis, urinary tract infection, viral infection, or none of the above. The trigger tool was not modified from its original form. Thus, this study represents an external cohort validation.

Analysis

The prevalence of delayed diagnosis was determined in the whole cohort. We constructed a receiver operating characteristic (ROC) curve to illustrate the tradeoff of sensitivity

versus specificity of the trigger tool in correctly classifying delayed diagnosis. The areas under the receiver operating curve (AUC) were computed. Sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and accuracy were determined at several thresholds of delayed diagnosis likelihood: >0%, >50%, >75%, and >90%. Test characteristics were reported as percentages with 95% binomial exact confidence intervals. We determined interrater reliability using Cohen's kappa.

We conducted a sensitivity analysis by varying the threshold for determination of the outcome by recategorizing "possible delayed diagnosis" cases as true delays in diagnosis.

Results

There were 2777 children with appendicitis. Among them, we included 72 (2.6%) children with possible delayed diagnosis of appendicitis based on having at least two ED encounters leading to an appendicitis diagnosis. Four were excluded: 1 for insufficient records to perform the case review, 2 for leaving without being seen, and 1 for being transferred out after the initial encounter. We analyzed 68 (94%) cases arising from the 8 hospitals (Table 1).

The prevalence of true delayed diagnosis was 50/68 (74%), of whom 10 were classified as probable delay and 40 were classified as near-definite delay. The ROC curve for the trigger tool prediction of delayed diagnosis is shown in the Figure. The AUC was 0.84 (95% confidence interval [CI] 0.72-0.96). Test characteristics are shown in Table 2 at varying thresholds of confidence in the trigger tool prediction of delayed diagnosis.

Twenty-three (34%) cases underwent determination of interrater reliability. Overall agreement occurred in 21/23 (91%) cases, and Cohen's kappa was 0.78, representing moderate agreement.¹⁴

The sensitivity analysis involved recategorizing patients judged on review to have a possible delayed diagnosis. After reassigning such cases to be considered as having a delayed diagnosis, the proportion with delay increased to 54/68 (79%). The AUC improved to 0.93 (95% CI 0.87-0.99). The positive predictive value of the trigger tool at a threshold of >75% was 97% (95% CI 87-100). Cohen's kappa was 1.0, representing perfect agreement.

Discussion

In a cohort of 68 children with possible delayed diagnosis of appendicitis, a previously validated trigger tool accurately distinguished between children with and without true delays. At a trigger tool confidence threshold of >75%, the positive predictive value was 92%, indicating that cases flagged as having delayed diagnosis nearly always do. Trigger tool sensitivity was reasonable (72%). Taken together, these findings suggest that this trigger tool can produce reasonably accurate counts of children with delayed diagnosis.

The goal of the trigger tool is to allow population research on rates and systems risk factors for delayed diagnosis. Additionally, the tool would be useful to quality/safety managers. They could use it to monitor and identify cases of delayed diagnosis within health

systems, which would allow common cause analysis or root cause analysis and feedback to clinicians. Such tools make manageable the number of case reviews needed to perform quality assurance work.¹⁵

The trigger tool uses only information available in claims data. The advantage of this approach is that no human review is required to determine whether a delayed diagnosis occurred.¹⁵ Because the trigger tool has been validated in both pediatric and general EDs, it can be used on large claims datasets to evaluate rates and predictors of delayed diagnosis. In the future, the trigger tool is intended to be used at a prespecified threshold of 75%. However, the tool is flexible: a user may use a lower threshold if greater sensitivity is desired, or a higher threshold of perfect specificity is needed.

In the sensitivity analysis, children reviewed as having a possible delayed diagnosis were categorized as having a true delay. This recategorization improved model performance significantly, with a nearly perfect positive predictive value of 97% and perfect interrater reliability. This indicates that false positive results from the trigger tool are largely due to children with possible delayed diagnosis, some of whom are likely to have experienced delay. It also highlights the challenges of human review of delayed diagnosis. Assigning a level of confidence to the determination of delayed diagnosis is inherently subjective, particularly for the cases that exist in a grey area (i.e. cases with a “possible” delayed diagnosis).

Study limitations include the restricted geography (eastern Massachusetts only) and the use of EHR administrative data (rather than true claims).

In conclusion, a trigger tool that identifies delays in diagnosis using only health claims in community EDs has a high positive predictive value for true delayed diagnosis. The tool may be applied in community EDs to evaluate diagnostic quality.

Funding:

Dr. Michelson received funding from CRICO and through award K08HS026503 from the Agency for Healthcare Research and Quality.

Role of Funder:

The funder/sponsor did not participate in the work.

Data statement:

The data are available upon reasonable request.

Abbreviations:

AUC	area under the receiver operating curve
CI	confidence interval
ED	emergency department

EHR	electronic health record
ICD-9/10	International Classification of Diseases, 9th/10th Editions Clinical Modification
NPV	negative predictive value
PPV	positive predictive value

References

1. Somme S, Bronsert M, Morrato E, Ziegler M. Frequency and Variety of Inpatient Pediatric Surgical Procedures in the United States. *Pediatrics*. 2013;132(6):e1466–e1472. doi:10.1542/peds.2013-1243 [PubMed: 24276846]
2. Colvin JM, Bachur R, Kharbanda A. The Presentation of Appendicitis in Preadolescent Children. *Pediatr Emerg Care*. 2007;23(12):849–855. doi:10.1097/pec.0b013e31815c9d7f [PubMed: 18091591]
3. Staab S, Black T, Leonard J, Bruny J, Bajaj L, Grubenhoff JA. Diagnostic Accuracy of Suspected Appendicitis. *Pediatr Emerg Care*. 2022;38(2):e690–e696. doi:10.1097/PEC.0000000000002323 [PubMed: 34170096]
4. Sawin RS. Principles and Practice of Pediatric Surgery: Chapter 80 - Appendix and Meckel's Diverticulum. In: Oldham Colombani, Foglia Skinner, eds. Lippincott Williams & Wilkins; 2005:1271–1282.
5. Papandria D, Goldstein SD, Rhee D, et al. Risk of perforation increases with delay in recognition and surgery for acute appendicitis. *Journal of Surgical Research*. 2013;184(2):723–729. doi:10.1016/j.jss.2012.12.008 [PubMed: 23290595]
6. Flum D, Morris A, Koepsell T, Dellinger E. Has misdiagnosis of appendicitis decreased over time? A population-based analysis. *JAMA : the journal of the American Medical Association*. 2001;286(14):1748–1753. doi:10.1016/S0196-0644(03)00613-9 [PubMed: 11594900]
7. Naiditch JA, Lautz TB, Daley S, Pierce MC, Reynolds M. The implications of missed opportunities to diagnose appendicitis in children. *Academic Emergency Medicine*. 2013;20(6):592–596. doi:10.1111/acem.12144 [PubMed: 23758306]
8. Goyal MK, Chamberlain JM, Webb M, et al. Racial and ethnic disparities in the delayed diagnosis of appendicitis among children. Zonfrillo MR, ed. *Academic Emergency Medicine*. 2021;28(9):949–956. doi:10.1111/acem.14142 [PubMed: 32991770]
9. Michelson KA, Bachur RG, Dart AH, et al. Identification of delayed diagnosis of paediatric appendicitis in administrative data: a multicentre retrospective validation study. *BMJ Open*. 2023;13(2):e064852. doi:10.1136/bmjopen-2022-064852
10. Michelson KA, Hudgins JD, Lyons TW, Monuteaux MC, Bachur RG, Finkelstein JA. Trends in Capability of Hospitals to Provide Definitive Acute Care for Children: 2008 to 2016. *Pediatrics*. 2020;145(1):e20192203. doi:10.1542/peds.2019-2203 [PubMed: 31882440]
11. Michelson KA, McGarghan FLE, Patterson EE, Samuels-Kalow ME, Waltzman ML, Greco KF. Delayed diagnosis of serious paediatric conditions in 13 regional emergency departments. *BMJ Qual Saf*. Published online September 30, 2022. doi:10.1136/bmjqs-2022-015314
12. Michelson KA, Williams DN, Dart AH, et al. Development of a rubric for assessing delayed diagnosis of appendicitis, diabetic ketoacidosis and sepsis. *Diagnosis*. 2021;8(2):219–225. doi:10.1515/dx-2020-0035 [PubMed: 32589599]
13. Feudtner C, Feinstein JA, Zhong W, Hall M, Dai D. Pediatric complex chronic conditions classification system version 2: updated for ICD-10 and complex medical technology dependence and transplantation. *BMC Pediatr*. 2014;14(1):199. doi:10.1186/1471-2431-14-199 [PubMed: 25102958]
14. McHugh ML. Interrater reliability: the kappa statistic. *Biochem Med (Zagreb)*. 2012;22(3):276–282. [PubMed: 23092060]

15. Stockwell DC, Sharek P. Diagnosing diagnostic errors: it's time to evolve the patient safety research paradigm. *BMJ Qual Saf.* 2022;31(10):701–703. doi:10.1136/bmjqs-2021-014517

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

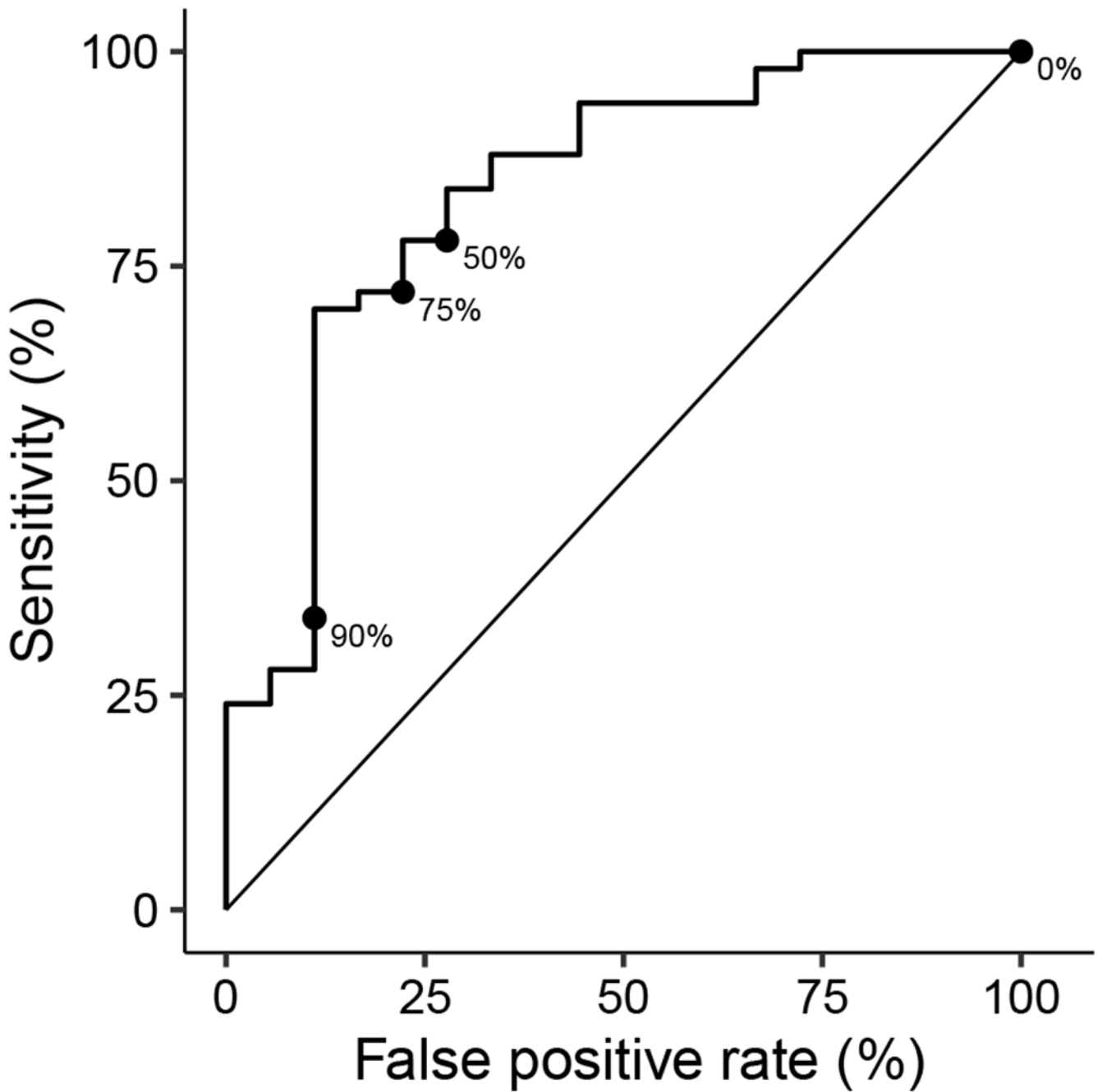


Figure: Receiver operating characteristic curve for trigger tool prediction of delayed diagnosis of appendicitis at varying prediction thresholds. Prespecified cutoffs for evaluating test characteristics included a delayed diagnosis likelihood of >0%, >50%, >75%, and >90%.

Table 1:

Demographic characteristics of the 68 analyzed patients

	All patients N=68 n (%)	Patients with probable or near-definite delayed diagnosis upon manual review N=50 n (%)
Age, median (IQR)	14.5 (10.5, 17.2)	14.2 (9.8, 16.6)
Female	33 (48.5)	25 (50.0)
Race		
Asian	3 (10.3)	2 (9.1)
Hispanic	8 (27.6)	8 (36.4)
Non-Hispanic Black	3 (10.3)	1 (4.5)
Non-Hispanic White	15 (51.7)	11 (50.0)
Primarily English speaking	60 (88.2)	42 (84.0)
Primary insurance		
Private	42 (61.8)	33 (66.0)
Public	16 (23.5)	11 (22.0)
Other	10 (14.7)	6 (12.0)
Complex chronic condition	11 (16.2)	7 (14.0)
Perforated appendicitis at time of diagnosis	17 (28.3)	17 (34.0)

Numbers do not add up to 100% due to missing data

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 2:

Test characteristics of the trigger tool’s prediction of delayed diagnosis, using the criterion standard of electronic health record review.

Trigger tool predicted delay threshold	Sensitivity % (95% CI)	Specificity % (95% CI)	PPV % (95% CI)	NPV % (95% CI)
>0%	100 (93-100)	0 (0-19)	74 (61-83)	NA
>50%	78 (64-88)	72 (47-90)	89 (75-96)	54 (33-74)
>75%	72 (58-84)	83 (59-96)	92 (79-98)	52 (33-71)
>90%	34 (21-49)	89 (65-99)	89 (67-99)	33 (20-48)

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript