BMJ Open

Identification of delayed diagnosis of paediatric appendicitis in administrative data: a multicentre retrospective validation study

Kenneth A Michelson [ID] ,[1] Richard G Bachur,[1] Arianna H Dart,[1]
Pradip P Chaudhari,[2] Andrea T Cruz,[3] Joseph A Grubenhoff,[4,5] Scott D Reeves,[6]
Michael C Monuteaux,[1] Jonathan A Finkelstein[7]

For numbered affiliations see
end of article.

Correspondence to
Dr Kenneth A Michelson;
kenneth.michelson@childrens.
harvard.edu

ABSTRACT

Objective To derive and validate a tool that retrospectively identifies delayed diagnosis of appendicitis in administrative data with high accuracy.

Design Cross-sectional study.

Setting Five paediatric emergency departments (EDs).

Participants 669 patients under 21 years old with possible delayed diagnosis of appendicitis, defined as two ED encounters within 7 days, the second with appendicitis.

Outcome Delayed diagnosis was defined as appendicitis being present but not diagnosed at the first ED encounter based on standardised record review. The cohort was split into derivation (2/3) and validation (1/3) groups. We derived a prediction rule using logistic regression, with covariates including variables obtainable only from administrative data. The resulting trigger tool was applied to the validation group to determine area under the curve (AUC). Test characteristics were determined at two predicted probability thresholds.

Results Delayed diagnosis occurred in 471 (70.4%) patients. The tool had an AUC of 0.892 (95% CI 0.858 to 0.925) in the derivation group and 0.859 (95% CI 0.806 to 0.912) in the validation group. The positive predictive value (PPV) for delay at a maximal accuracy threshold was 84.7% (95% CI 78.2% to 89.8%) and identified 87.3% of delayed cases. The PPV at a stricter threshold was 94.9% (95% CI 87.4% to 98.6%) and identified 46.8% of delayed cases.

Conclusions This tool accurately identified delayed diagnosis of appendicitis. It may be used to screen for potential missed diagnoses or to specifically identify a cohort of children with delayed diagnosis.

STRENGTHS AND LIMITATIONS OF THIS STUDY

⇒ The study establishes a method for specifically flagging cases with delayed diagnosis, allowing study in large datasets where medical records are not available.

⇒ The tool was derived and validated in separate cohorts.

⇒ Expert medical record review was conducted based on a previously-defined objective rubric.

⇒ All patients were evaluated in paediatric emergency departments, which may differ from non-paediatric emergency departments, affecting rule performance outside of paediatric settings.

formation, sepsis and rarely a need for bowel resection.[4] Timely diagnosis can prevent these complications. The emergency department (ED) environment accentuates factors that predispose patients to delayed diagnosis because of high cognitive load on clinicians, frequent high-stakes decisions and patients who are typically not previously known to clinicians.[5 6]

Systematic identification of diagnostic error is the first step in preventing clinical delays in diagnosis, but reporting of diagnostic errors is unreliable, challenging and typically relies on expert case review.[7–11] However, case review is labourious, expensive and difficult to scale. Automated approaches promise to screen and identify potential causes of error, but nevertheless require manual case review after screening.[12 13] Case review depends on access to records and resources to perform the review, biasing samples towards hospitals willing to participate.[14]

Despite these obstacles, tools to assess diagnostic accuracy across hospitals of all types are needed to improve delays in diagnosis of serious emergency conditions for children. Most childhood ED encounters occur in

INTRODUCTION

Appendicitis is the most common serious surgical emergency in children.[1] Appendicitis may be more difficult to diagnose in younger children, often exhibits fewer classic features of the disease, and the symptoms and signs can overlap with other, more common illnesses, such as gastroenteritis.[2 3] Delays are associated with complications including perforated appendicitis, abdominal abscess

community EDs not staffed by clinicians who primarily treat children, and one-third of EDs evaluate fewer than five children per day on average.[15 16] This may magnify the challenge of diagnosis in children, who are more likely to be developmentally unable to provide accurate historical information, and in whom early symptoms of disease are often non-specific.[17] Administrative data are the only current widespread means of assessing care at all types of hospitals and thus are the only currently realistic approach for understanding a broad cross-section of care.[14]

Approaches identifying delayed diagnosis in administrative data, if shown to be accurate, would have several advantages. First, they could be used in administrative data to illuminate hospital-level factors and rates of delayed diagnosis that would inform improvement efforts. Second, they could be used to identify high-performing hospitals or hospital systems that could serve as models or benchmarks to other institutions seeking to improve diagnosis. Third, they could save substantial effort in identifying cases for local review and feedback. Finally, they could be used to assess improvement efforts focused on diagnostic accuracy.

To address the challenges of efficiently identifying potential diagnostic error, we previously piloted a method for accurately identifying delayed diagnosis for conditions using the information contained within administrative data.[18] Here, we report on a multicentre investigation to validate that methodology for the identification of a delayed diagnosis of appendicitis in children and young adults aged less than 21 years old.

METHODS
Design, setting and participants
We performed a retrospective cross-sectional study to develop and test a decision rule using variables only available in administrative data to predict delayed diagnosis of appendicitis, as determined by expert case review. The study was designed in accordance with Standards for Reporting of Diagnostic Accuracy Studies guidelines for studies on diagnostic accuracy.[19] Participants were children and young adults age <21 years who visited one of five paediatric EDs across the country from 2010 to 2019, had a first-time diagnosis of appendicitis, and had an ED visit in the preceding 7 days. The ED encounter associated with the appendicitis diagnosis was designated as the 'diagnosis encounter,' and the preceding encounter was designated as 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).

Data sources
The source of administrative data was the Pediatric Health Information System (PHIS). The PHIS database contains clinical and billing data from 44 not-for-profit, tertiary care children's hospitals. The data collection, validation and safeguarding procedures are assured through a joint effort between the Children's Hospital Association (Lenexa, Kansas, USA) and participating hospitals, and have previously been described.[20–22] Data are deidentified at the time of data submission, and data are subjected to a number of reliability and validity checks before being included in the database. For this study, data from five hospitals were included. Cases from PHIS were reidentified locally and linked to the electronic health record (EHR) at each participating site for manual review.

Outcome
The reference standard primary outcome was delayed diagnosis as determined by manual expert case review of the EHR. It 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' (definitions provided in online supplemental table 1).

Case reviewers were all board-certified paediatric emergency medicine faculty. Reviewers were trained on the assessment of delay using study reading material, and then were tested and retested grading 40 standard appendicitis cases. Real-time feedback was given after each response. The correct answers and feedback were determined by a multispecialty expert consensus panel.[23] The reviewer assessment of delayed diagnosis was dichotomised as delayed diagnosis (probably or near-definitely delay) or not delayed diagnosis (possibly, probably not or near-definitely not delay). This approach to case review was previously shown to have high inter-rater reliability in a very similar cohort.[18] After training, reviewers evaluated study cases. Reviewers were blinded to the decision rule assessment of delayed diagnosis.

Development of the decision rule
The decision rule evaluated the likelihood of delayed diagnosis of appendicitis using variables contained in administrative data and based on investigators' clinical expertise. These included age (<3 years, 3–10 years or ≥11 years), sex, history of a complex chronic condition,[24] 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, leucocytosis, urinary tract infection, viral infection or none of the above (diagnosis codes in online supplemental table 2).

Table 1 Demographics and outcomes of the derivation and validation study cohorts

Characteristic	Derivation cohort n=444 (66.4%) n (%)	Validation cohort n=225 (33.6%) n (%)	P value
Age, years			0.62
<3	33 (7.4)	16 (7.1)	
3–10	243 (54.7)	132 (58.7)	
11–21	168 (37.8)	77 (34.2)	
Male	224 (50.5)	111 (49.3)	0.81
Race			0.66
White	236 (53.2)	116 (51.6)	
Black	34 (7.7)	23 (10.2)	
Asian	4 (0.9)	2 (0.9)	
American Indian	0 (0.0)	1 (0.4)	
Pacific Islander	2 (0.5)	1 (0.4)	
Other	168 (37.8)	82 (36.4)	
Ethnicity			0.45
Hispanic or Latino	249 (56.1)	117 (52.0)	
Not Hispanic or Latino	189 (42.6)	103 (45.8)	
Unknown	6 (1.4)	5 (2.2)	
Primary payer			0.05
Public	276 (62.2)	126 (56.0)	
Private	151 (34.0)	92 (40.9)	
Self-pay	16 (3.6)	4 (1.8)	
Free care	0 (0.0)	0 (0.0)	
Unknown	1 (0.2)	3 (1.3)	
Complex chronic condition	52 (11.7)	25 (11.1)	0.90
Perforated appendicitis	257 (63.3)	128 (61.5)	0.72
Revisit interval, days			0.09
0	73 (16.4)	17 (7.6)	
1	149 (33.6)	83 (36.9)	
2	94 (21.2)	51 (22.7)	
3	41 (9.2)	20 (8.9)	
4	36 (8.1)	20 (8.9)	
5	21 (4.7)	18 (8.0)	
6	17 (3.8)	9 (4.0)	
7	13 (2.9)	7 (3.1)	
Diagnosis encounter LOS, days, median (IQR)	4 (1–6)	3 (1–6)	0.85
Diagnosis code from initial encounter			
Abdominal pain	187 (42.1)	92 (40.9)	0.80
Constipation	73 (16.4)	46 (20.4)	0.20
Fever	47 (10.6)	23 (10.2)	1
Gastroenteritis	200 (45.0)	98 (43.6)	0.74
Urinary tract infection	34 (7.7)	13 (5.8)	0.43
Genitourinary problem	10 (2.3)	5 (2.2)	1
Viral syndrome	34 (7.7)	14 (6.2)	0.53
Dehydration	21 (4.7)	14 (6.2)	0.46

Continued

Table 1 Continued

Characteristic	Derivation cohort n=444 (66.4%) n (%)	Validation cohort n=225 (33.6%) n (%)	P value
Head, ear, eyes, nose or throat problem	30 (6.8)	13 (5.8)	0.74
None of the above	26 (5.9)	20 (8.9)	0.15
Outcomes			
Delayed diagnosis			0.04
Near-definite delay	222 (54.4)	96 (46.2)	
Probable delay	91 (22.3)	62 (29.8)	
Not delayed diagnosis			
Possible delay	44 (10.8)	26 (12.5)	
Probable non-delay	21 (5.1)	4 (1.9)	
Near-definite non-delay	30 (7.4)	20 (9.6)	
LOS, length of stay.			

The full cohort was randomly divided into derivation (2/3) and validation (1/3) sets, stratified on the outcome. The decision rule was trained using only the derivation set. Variables were selected for inclusion in the decision rule using univariable logistic regressions. All variables associated with the outcome with p<0.20 were included in the decision rule. The final model underlying the decision rule was created using multivariable logistic regression within the derivation set using delayed diagnosis (determined by expert case review) as the outcome and all screened-in variables as predictors. The decision rule classified cases as delayed or not delayed using two thresholds: (1) a maximal accuracy threshold, based on the model predicted probability being greater than or equal to the value that maximises the proportion of correct classifications[25] and (2) a near-definite delay threshold if the predicted probability of delay was ≥90%.

Analysis

The prevalence of delayed diagnosis was determined in the whole cohort and then separately by site. We constructed receiver operating characteristic (ROC) curves in the derivation and validation sets to illustrate the trade-off of sensitivity vs specificity of the decision rule in correctly classifying delayed diagnosis. Areas under the ROC curve (AUC) were computed. Sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV) and accuracy were determined for the rule in both derivation and validation sets at the two thresholds (maximal accuracy and near-definite delay). We determined binomial exact 95% CIs for each test characteristic.

Calibration of the rule was determined separately in the derivation and validation sets. We first used the predicted probability of delayed diagnosis to categorise patients as 0% to <20%, 20% to <40%, 40% to <60%, 60% to <80% or 80% to 100% likely to have delayed diagnosis. We then computed the actual proportion of patients who had a

delayed diagnosis and its 95% binomial exact CI within each of these subgroups. One-sample binomial proportion tests were computed for the validation set comparing expected frequencies of delay (10%, 30%, 50%, 70% and 90%, respectively) with actual proportions.

Sensitivity analyses were performed, recreating the rule derivation to predict the outcome of (1) possible, probable or near-definite delay (a permissive rule) or (2) near-definite delay only (a strict rule).

Statistical significance was defined as p<0.05. A prestudy power analysis suggested that we would need 193 patients in the validation cohort to have 80% power to estimate the rule PPV within 10 percentage points based on a binomial exact CI around 0.9 (the expected PPV based on pilot work).[18]

Patient and public involvement
Patients and the public were not involved in the development of this research, as the topic of the research was focused on informatics.

RESULTS
Among 801 patients included in the study because of a revisit within 7 days leading to appendicitis diagnosis, we excluded 14 (1.7%) for having insufficient records, 5 (0.6%) for no record of an initial encounter, 32 (4.0%) for leaving without being seen and 81 (10.1%) for being transferred at the initial encounter. We analysed 669 (83.5%) patients. Demographics of the cohort are shown in table 1, and there were no significant differences between characteristics of children in the derivation or validation sets. Delayed diagnosis of appendicitis occurred in 471 (70.4%) of patients.

Derivation of the decision rule
Among all possible variables screened for inclusion, all except age were associated with the outcome with p<0.20. The final logistic regression model used for the decision rule is shown in table 2. A risk calculator is available as online supplemental file 2. The variables most associated with delayed diagnosis were perforated appendicitis and the interval between the initial and diagnosis encounters. The maximum accuracy threshold of predicted probability of delayed diagnosis was 0.568. Therefore, based on the decision rule determined from the derivation set, predictions of delayed diagnosis were most accurate when a case had a predicted probability of delay >56.8%.

Validation of the decision rule
ROC curves depicting the trade-off of sensitivity and specificity at differing thresholds of predicted probability of delay are shown in figure 1. The AUC for the derivation set was 0.892 (95% CI 0.858 to 0.925) and for the validation set was 0.859 (95% CI 0.806 to 0.912). We applied the decision rule using the maximal accuracy threshold of 56.8% and separately of >90% in both derivation and validation sets. The validation set PPV of the prediction

Table 2 Final model predicting delayed diagnosis of appendicitis based on administrative data in the derivation cohort only (model pseudo-R^2=0.42)

Variable	Adjusted OR (95% CI)	P value
Complex chronic condition	0.71 (0.30 to 1.69)	0.44
Revisit interval, days (compared with 0)		
1	1.05 (0.40 to 2.76)	0.92
2	0.21 (0.08 to 0.54)	0.001
3	0.14 (0.04 to 0.46)	0.001
4	0.03 (0.01 to 0.10)	<0.001
5	0.01 (0.00 to 0.06)	<0.001
6	0.01 (0.00 to 0.05)	<0.001
7	0.03 (0.01 to 0.21)	<0.001
Initial encounter diagnosis		
Abdominal pain	0.98 (0.50 to 1.95)	0.97
Constipation	1.16 (0.47 to 2.91)	0.75
Dehydration	1.23 (0.30 to 5.03)	0.77
Fever	0.60 (0.23 to 1.57)	0.30
Gastroenteritis or vomiting	2.03 (1.00 to 4.11)	0.05
Genitourinary problem	0.31 (0.05 to 2.03)	0.22
Head, ear, eyes, nose or throat problem	0.48 (0.16 to 1.44)	0.19
UTI or dysuria	8.01 (2.09 to 30.65)	0.002
Viral syndrome	0.61 (0.20 to 1.81)	0.37
None of the above	0.04 (0.01 to 0.26)	<0.001
Final diagnosis of perforated appendicitis during diagnosis encounter	5.67 (2.56 to 12.56)	<0.001
Diagnosis encounter LOS, days (compared with 0–1)		
2–3	1.32 (0.59 to 2.95)	0.51
4–7	2.45 (0.96 to 6.23)	0.06
>7	1.50 (0.52 to 4.33)	0.46

LOS, length of stay; UTI, urinary tract infection.

rule was 84.7% (95% CI 78.2% to 89.8%) and NPV was 67.7% (95% CI 54.7% to 79.1%). Using a stricter cut-off predicted probability of ≥90% yielded a PPV of 94.9% (95% CI 87.4% to 98.6%) and NPV of 42.9% (95% CI 34.7% to 51.3%). Test characteristics of the decision rule are shown in table 3. The calibration of the model was excellent. In the validation cohort, predictions of the probability of delayed diagnosis were not significantly different than actual probabilities of delayed diagnosis, except for children with a predicted probability of 20% to <40%, in whom delayed diagnosis was underestimated (table 4).

Sensitivity analyses
Permissive and strict decision rules had similar performance to the main decision rule (Details of the rules and test characteristics are shown in online supplemental tables 3 and 4). The validation AUCs were 0.865 (95% CI
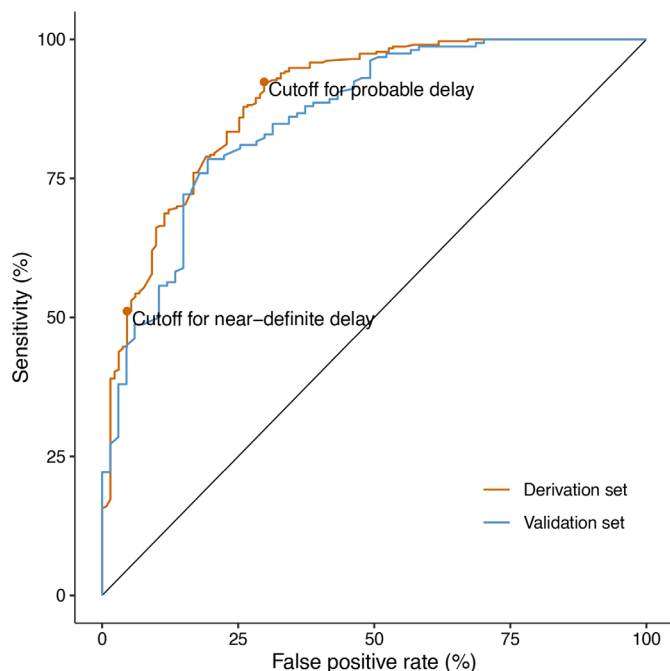
Figure 1 Receiver operating characteristic curves depict the trade-off between sensitivity and false positive rate (1-specificity) in predicting delayed diagnosis. The AUC for the derivation set was 0.892 (95% CI 0.858 to 0.925) and for the validation set was 0.859 (95% CI 0.806 to 0.912). AUC, area under the curve.

Table 4 Calibration of the predicted probability of delayed diagnosis at different predicted probabilities

Predicted probability of delayed diagnosis	n/N	Proportion with delayed diagnosis % (95% CI)	P value for difference in observed vs expected frequency of delayed diagnosis
Derivation cohort			
0% to <20%	4/63	6.3 (1.8 to 15.5)	0.53
20% to <40%	7/19	36.8 (16.3 to 61.6)	0.62
40% to <60%	22/45	48.9 (33.7 to 64.2)	1.0
60% to <80%	51/66	77.3 (65.3 to 86.7)	0.23
80% to 100%	229/251	91.2 (87.0 to 94.4)	0.60
Validation cohort			
0% to <20%	4/34	11.8 (3.3 to 27.5)	0.77
20% to <40%	10/16	62.5 (35.4 to 84.8)	0.01
40% to <60%	10/19	52.6 (28.9 to 75.6)	1.0
60% to <80%	29/41	70.7 (54.5 to 83.9)	1.0
80% to 100%	105/115	91.3 (84.6 to 95.8)	0.76

0.800 to 0.930) for the permissive rule and 0.803 (95% CI 0.747 to 0.859) for the strict rule. PPV was 92.6% (95% CI 87.7% to 96.0%) for the permissive rule and 63.1% (95% CI 53.9% to 71.7%) for the strict rule.

DISCUSSION

We successfully derived and validated an accurate decision rule for retrospectively identifying cases of delayed diagnosis of appendicitis in administrative data, with a PPV of 84.7%. Importantly, the model underlying the decision rule is well calibrated, provides accurate estimates of delay likelihood, and can identify a subcohort of patients who almost certainly experienced a delayed diagnosis of appendicitis: a stricter model threshold had a PPV of 94.9%. The rule relies only on information contained with administrative databases, including patient demographics, encounter length of stay and diagnosis codes.

The model is therefore amenable to assessment of care, research and improvement efforts both locally and at the state and national levels.

We believe the decision rule will be useful for several different applications by varying the threshold for detection of delay. Since the rule is well calibrated, using a lower threshold of predicted probability to detect delay (eg, 0.2) will provide sensitive detection but would require further review to confirm delay, and using a higher threshold of detection (eg, 0.8) will provide specific detection but will miss cases with delay. At higher thresholds, the decision rule is specific enough to estimate rates of delayed diagnosis of appendicitis in populations drawn from large administrative databases, without the need for subsequent case review. These features of the rule are crucial, because they allow for a direct assessment of diagnostic performance in hospitals without considerable quality measurement infrastructure or investments in research. Using a sensitive threshold, hospitals could track their diagnostic performance and screen for potential cases of delay, aiding quality assurance efforts by balancing good

Table 3 Test characteristics of the delayed diagnosis prediction model, applied to the derivation and validation cohorts

Test characteristic	Derivation cohort % (95% CI)	Validation cohort, probable delay % (95% CI)	Validation cohort, near-definite delay % (95% CI)
Sensitivity	92.3 (88.8 to 95.0)	87.3 (81.1 to 92.1)	46.8 (38.9 to 54.9)
Specificity	70.2 (61.6 to 77.9)	62.7 (50.0 to 74.2)	94.0 (85.4 to 98.3)
Positive predictive value	88.1 (84.1 to 91.4)	84.7 (78.2 to 89.8)	94.9 (87.4 to 98.6)
Negative predictive value	79.3 (70.8 to 86.3)	67.7 (54.7 to 79.1)	42.9 (34.7 to 51.3)

Probable delay was defined by a predicted probability of delayed diagnosis of 0.568 (the value that maximises rule accuracy). Near-definite delay was defined by a predicted probability of 0.9.

case capture with the feasibility of many case reviews. A tiered approach would be to screen only cases above the sensitive threshold but assume that those above the higher threshold constitute delays.

The final model mirrors the clinical factors known to predispose to delayed diagnosis of appendicitis. A shorter period between initial and diagnosis encounters increases the likelihood that the initial one was related, mirroring evidence suggesting that the relatedness of two ED visits decreases with time.[26] Perforation at diagnosis is associated with the likelihood that a delayed diagnosis occurred, probably because it increases the likely duration of disease that existed before diagnosis. Conditions commonly misdiagnosed before appendicitis were associated with a higher likelihood of delay and included gastroenteritis and urinary tract infection. In contrast, an absence of an apparently related diagnosis made delay less likely.

The approach used to generate this rule is generalisable to other emergency conditions. First, we convened a panel of experts to define the standards for grading a delayed diagnosis. Second, expert reviewers were trained to evaluate case records. Finally, reviewers analysed hundreds of records to generate enough data to develop a reliable decision rule. The variables could be repurposed for other conditions, but the rule itself is unique to paediatric appendicitis. We believe duplicating this approach for other conditions would be useful, because once developed, a rule is applicable for ongoing quality monitoring and research. Once expanded to multiple conditions, it would provide a realistic view of a hospital's overall diagnostic performance, which has proved elusive to date.

A major reason that we developed this decision rule is that identifying and thus preventing diagnostic errors is challenging in general, as self-report is unreliable and labourious case review is needed.[7 27] It is specifically challenging in children because most paediatric care happens outside of paediatric hospitals, where research is most commonly conducted and EHRs may not be available.[14 28] Although trigger tools exist to identify diagnostic errors in abdominal cases, they are too non-specific to forego the review step, which requires access to records.[29] A key advantage of our approach is that, with a high predicted probability threshold of 90%, delay can be specifically identified.

This study has several strengths, including reliance on a multidisciplinary consensus definition of delayed diagnosis, the validation of the model on a cohort distinct from that used to train it, the large sample size, use of data from multiple centres and face validity of the factors predicting delay. Limitations include the use of data from only paediatric hospitals (suggesting the value of a future independent validation in general hospitals) and the complex nature of the decision rule model. Additionally, we did not perform tests of inter-rater reliability, though we previously showed in pilot work that inter-rater reliability for this approach is excellent.[18] Finally,

the development of the decision rule using a random split cohort can result in optimistic predictions; thus, we intend to further validate this rule in external populations in the future.

In conclusion, we developed and validated a model that can accurately identify delayed diagnoses of paediatric appendicitis in administrative data, without the need for manual record review for confirmation. This model may be applied to hospital data sources in which policymakers and researchers do not have access to patients' records, allowing for accurate study of diagnostic error in most hospitals. The model may also be used by hospital systems to identify errors and improve care.

Author affiliations
[1]Division of Emergency Medicine, Boston Children's Hospital, Boston, MA, USA
[2]Division of Emergency and Transport Medicine, Children's Hospital Los Angeles, Los Angeles, CA, USA
[3]Department of Pediatrics, Baylor College of Medicine, Houston, TX, USA
[4]Section of Pediatric Emergency Medicine, University of Colorado School of Medicine, Aurora, CO, USA
[5]Children's Hospital Colorado, Aurora, CO, USA
[6]Division of Pediatric Emergency Medicine, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, USA
[7]Kaiser Permanente Bernard J. Tyson School of Medicine, Pasadena, CA, USA

ORCID iD
Kenneth A Michelson http://orcid.org/0000-0003-1763-7262

REFERENCES

1 Somme S, Bronsert M, Morrato E, *et al*. Frequency and variety of inpatient pediatric surgical procedures in the United States. *Pediatrics* 2013;132:e1466–72.

2 Colvin JM, Bachur R, Kharbanda A. The presentation of appendicitis in preadolescent children. *Pediatr Emerg Care* 2007;23:849–55.

3 Staab S, Black T, Leonard J, *et al*. Diagnostic accuracy of suspected appendicitis. *Pediatr Emerg Care* 2022;38:e690–6.

4 Sawin RS. Chapter 80 - appendix and meckel's diverticulum. In: *Oldham, Colombani, Foglia, eds. Principles and Practice of Pediatric Surgery*. Philadelphia, PA: Lippincott Williams & Wilkins, 2005: 1271–82.

5 Croskerry P, Sinclair D. Emergency medicine: a practice prone to error? *CJEM* 2001;3:271–6.

6 Croskerry P. ED cognition: any decision by anyone at any time. *CJEM* 2014;16:13–9.

7 Sevdalis N, Jacklin R, Arora S, *et al*. Diagnostic error in a national incident reporting system in the UK. *J Eval Clin Pract* 2010;16:1276–81.

8 Schiff GD, Hasan O, Kim S, *et al*. Diagnostic error in medicine: analysis of 583 physician-reported errors. *Arch Intern Med* 2009;169:1881–7.

9 Singh H, Khanna A, Spitzmueller C, *et al*. Recommendations for using the revised safer dx instrument to help measure and improve diagnostic safety. *Diagnosis* 2019;6:315–23.

10 Graber ML, Franklin N, Gordon R. Diagnostic error in internal medicine. *Arch Intern Med* 2005;165:1493–9.

11 Balogh EP, Miller BT. *Improving diagnosis in health care*. Washington, D.C: National Academies Press, 2015.

12 Singh H, Giardina TD, Forjuoh SN, *et al*. Electronic health record-based surveillance of diagnostic errors in primary care. *BMJ Qual Saf* 2012;21:93–100.

13 Mahajan P, Pai C-W, Cosby KS, *et al*. Identifying trigger concepts to screen emergency department visits for diagnostic errors. *Diagnosis (Berl)* 2021;8:340–6.

14 Michelson KA, Bachur RG. The high value of blurry data in improving pediatric emergency care. *Hosp Pediatr* 2019;9:1007–9.

15 Gausche-Hill M, Ely M, Schmuhl P, *et al*. A national assessment of pediatric readiness of emergency departments. *JAMA Pediatr* 2015;169:527–34.

16 Michelson KA, Hudgins JD, Lyons TW, *et al*. Trends in capability of hospitals to provide definitive acute care for children: 2008 to 2016. *Pediatrics* 2020;145:e20192203.

17 Committee on the Future of Emergency Care in the United States Health System. *Emergency care for children: growing pains*. Washington, D.C: National Academies Press, 2007.

18 Michelson KA, Buchhalter LC, Bachur RG, *et al*. Accuracy of automated identification of delayed diagnosis of pediatric appendicitis and sepsis in the ED. *Emerg Med J* 2019;36:736–40.

19 Cohen JF, Korevaar DA, Altman DG, *et al*. STARD 2015 guidelines for reporting diagnostic accuracy studies: explanation and elaboration. *BMJ Open* 2016;6:e012799.

20 Mongelluzzo J, Mohamad Z, Ten Have TR, *et al*. Corticosteroids and mortality in children with bacterial meningitis. *JAMA* 2008;299:2048–55.

21 DeCourcey DD, Steil GM, Wypij D, *et al*. Increasing use of hypertonic saline over mannitol in the treatment of symptomatic cerebral edema in pediatric diabetic ketoacidosis: an 11-year retrospective analysis of mortality. *Pediatr Crit Care Med* 2013;14:694–700.

22 Weiss AK, Hall M, Lee GE, *et al*. Adjunct corticosteroids in children hospitalized with community-acquired pneumonia. *Pediatrics* 2011;127:e255–63.

23 Michelson KA, Williams DN, Dart AH, *et al*. Development of a rubric for assessing delayed diagnosis of appendicitis, diabetic ketoacidosis and sepsis. *Diagnosis* 2021;8:219–25.

24 Feudtner C, Feinstein JA, Zhong W, *et al*. Pediatric complex chronic conditions classification system version 2: updated for ICD-10 and complex medical technology dependence and transplantation. *BMC Pediatr* 2014;14:199.

25 Perkins NJ, Schisterman EF. The inconsistency of "optimal" cutpoints obtained using two criteria based on the receiver operating characteristic curve. *Am J Epidemiol* 2006;163:670–5.

26 Michelson KA, Lyons TW, Bachur RG, *et al*. Timing and location of emergency department revisits. *Pediatrics* 2018;141:e20174087.

27 Singh H, Meyer AND, Thomas EJ. The frequency of diagnostic errors in outpatient care: estimations from three large observational studies involving US adult populations. *BMJ Qual Saf* 2014;23:727–31.

28 Heisey-Grove DM. Variation in rural health information technology adoption and use. *Health Aff* 2016;35:365–70.

29 Perry MF, Melvin JE, Kasick RT, *et al*. The diagnostic error index: a quality improvement initiative to identify and measure diagnostic errors. *J Pediatr* 2021;232:257–63.

Supplemental material

Supplemental Table 1: Definitions of delayed diagnosis.
Supplemental Table 2: Diagnosis codes used to identify specific conditions presents at the initial encounter.
Supplemental Table 3: Permissive model predicting delayed diagnosis
Supplemental Table 4: Strict model predicting delayed diagnosis

Supplemental material

BMJ Publishing Group Limited (BMJ) disclaims all liability and responsibility arising from any reliance placed on this supplemental material which has been supplied by the author(s)

BMJ Open

Supplemental Table 1: Definitions of delayed diagnosis.

Likelihood of delay	Definition
Near-definitely not	(1) there were no signs of the condition on the prior encounter; (2) an alternative explanation for the prior encounter symptoms is definite or almost definite; or (3) the time course makes the condition virtually certain not to have been present.
Probably not	It is very unlikely that the condition was present on the prior encounter, and symptoms, signs, and other data at that time mostly pointed away from the condition.
Possibly	(1) it is possible the condition was present, but there are factors for and against that theory; (2) determination is confusing; (3) there is limited detail from which to decide; or (4) there are some alternative explanations for the case's features, but they have similar likelihood to the condition.
Probably	More likely than not, the patient had the condition on the prior encounter. Evidence pointed toward the condition, or few alternative explanations existed.
Near-definitely	The patient almost definitely or definitely had the condition on the prior encounter. Generally, these cases have clear evidence that the condition was present on the prior encounter. Alternatively, there is no other plausible explanation for the symptoms/signs on the prior encounter.

Supplemental Table 2: Diagnosis codes used to identify specific conditions presents at the initial encounter.

Diagnosis	Code System	Diagnosis Codes
Abdominal pain	ICD-9	289.2, 787.3, 789.00, 789.01, 789.02, 789.03, 789.04, 789.05, 789.06, 789.07, 789.09
Abdominal pain	ICD-10	I88.0, I88.9, N83.202, R10.0, R10.11, R10.12, R10.13, R10.2, R10.30, R10.31, R10.32, R10.33, R10.84, R10.9
Constipation	ICD-9	564.00, 564.09
Constipation	ICD-10	K59.00, K59.09
Dehydration	ICD-9	276.51
Dehydration	ICD-10	E86.0
Fever	ICD-9	780.6
Fever	ICD-10	R50.9
Gastroenteritis	ICD-9	008.8, 009.0, 535.00, 535.50, 558.9, 787.01, 787.02, 787.03, 787.91
Gastroenteritis	ICD-10	A08.4, A09, K29.00, K29.70, K52.9, R11.0, R11.10, R11.2, R19.7
Genitourinary condition	ICD-9	603.9, 604.90, 620.8
Genitourinary condition	ICD-10	N44.03, N45.1, N47.6, N50.819, N70.91, N76.0, N83.20, N83.9, N89.8, N93.8, N93.9, N94.6
Head, ears, eyes, nose, or throat condition	ICD-9	034.0, 381.01, 382.00, 382.9, 388.70, 462, 463, 464.4, 465.9
Head, ears, eyes, nose, or throat condition	ICD-10	H66.43, H66.91, J02.0, J02.8, J02.9, J05.0, J06.9, J11.1, J34.89
Urinary tract infection	ICD-9	590.80, 595.0, 595.9, 599.0, 788.1, 791.9
Urinary tract infection	ICD-10	N30.00, N39.0, R30.0
Viral syndrome	ICD-9	057.9, 079.99
Viral syndrome	ICD-10	B34.9, B97.89

Supplemental Table 3. Permissive model predicting delayed diagnosis, derived from the derivation cohort only. The permissive model predicted delay with a definition of delay that included "possible," "probable", and "near-definite" delayed diagnosis cases.

Variable	Adjusted odds ratio (95% confidence interval)	p
Complex chronic condition	1.30 (0.52, 3.27)	0.58
Revisit interval, days (compared with 0)		
1	0.91 (0.31, 2.70)	0.86
2	0.33 (0.11, 0.99)	0.05
3	0.26 (0.06, 1.18)	0.08
4	0.08 (0.02, 0.34)	<0.001
5	0.04 (0.01, 0.18)	<0.001
6	0.05 (0.01, 0.26)	<0.001
7	0.04 (0.01, 0.24)	<0.001
Initial encounter diagnosis		
Abdominal pain	0.64 (0.27, 1.50)	0.30
Constipation	2.85 (0.86, 9.46)	0.09
Dehydration	NA	0.98
Fever	0.75 (0.24, 2.31)	0.61
Gastroenteritis or vomiting	3.90 (1.63, 9.30)	<0.001
Genitourinary problem	1.04 (0.19, 5.69)	0.96
Head, ear, eyes, nose or throat problem	0.33 (0.10, 1.08)	0.07
UTI or dysuria	2.35 (0.56, 9.83)	0.24
Viral syndrome	0.95 (0.26, 3.47)	0.94
None of the above	0.07 (0.02, 0.34)	<0.001
Final diagnosis of perforated appendicitis during diagnosis encounter	8.29 (3.20, 21.52)	<0.001
Diagnosis encounter LOS, days (compared with 0-1)		
2-3	0.93 (0.40, 2.17)	0.86
4-7	3.98 (1.31, 12.07)	0.01
>7	0.61 (0.20, 1.87)	0.39

LOS; length of stay; NA: could not be estimated; UTI: urinary tract infection

Supplemental Table 4. Strict model predicting delayed diagnosis, derived from the derivation cohort only. The strict model predicted delay with a definition of delay that included "near-definite" delayed diagnosis cases.

Variable	Adjusted odds ratio (95% confidence interval)	p
Complex chronic condition	0.79 (0.37, 1.69)	0.54
Revisit interval, days (compared with 0)		
1	0.76 (0.39, 1.49)	0.43
2	0.16 (0.08, 0.35)	<0.001
3	0.10 (0.04, 0.27)	<0.001
4	0.07 (0.03, 0.21)	<0.001
5	0.03 (0.01, 0.15)	<0.001
6	0.01 (0.00, 0.11)	<0.001
7	0.03 (0.00, 0.25)	<0.001
Initial encounter diagnosis		
Abdominal pain	0.95 (0.56, 1.61)	0.85
Constipation	0.90 (0.44, 1.83)	0.76
Dehydration	1.57 (0.55, 4.49)	0.40
Fever	0.64 (0.30, 1.37)	0.25
Gastroenteritis or vomiting	1.01 (0.59, 1.74)	0.96
Genitourinary problem	0.87 (0.16, 4.59)	0.87
Head, ear, eyes, nose or throat problem	0.84 (0.31, 2.23)	0.72
UTI or dysuria	0.12 (0.02, 0.66)	0.01
Viral syndrome	3.47 (1.38, 8.72)	0.01
None of the above	0.38 (0.15, 0.95)	0.04
Final diagnosis of perforated appendicitis during diagnosis encounter	2.90 (1.58, 5.30)	<0.001
Diagnosis encounter LOS, days (compared with 0-1)		
2-3	1.11 (0.56, 2.22)	0.76
4-7	1.29 (0.62, 2.69)	0.49
>7	1.34 (0.59, 3.07)	0.49

LOS; length of stay; UTI: urinary tract infection

Supplemental material

BMJ Publishing Group Limited (BMJ) disclaims all liability and responsibility arising from any reliance
placed on this supplemental material which has been supplied by the author(s)

BMJ Open

Instructions:

Apply this calculator only to appendicitis patients with two ED encounters

within 7 days, with the second resulting in an appendicitis diagnosis

Note: This calculator should only be applied retrospectively to children who

have already had two encounters and cannot be used to predict the likelihood

of appendicitis during or after a single encounter.

Information from the first encounter

Complex chronic condition	No
Abdominal pain diagnosis	No
Constipation diagnosis	No
Dehydration diagnosis	No
Fever diagnosis	No
Gastroenteritis diagnosis	No
Genitourinary problem diagnosis	No
HEENT problem diagnosis	No
Urinary tract infection diagnosis	No
Viral syndrome diagnosis	No

Information from the second encounter

Days between encounters	3
Hospital days	0
Perforated appendicitis at diagnosis	No

Predicted probability of delayed diagnosis
1.7%

Predictor	Beta coefficient
Intercept	1.183116261
Perforated appy	1.73496391
CCC	-0.337679678
LOS 2-3	0.274417066
LOS 4-7	0.894800027
LOS > 7	0.402697272
Revisit 1d	0.051710402
Revisit 2d	-1.572028563
Revisit 3d	-1.992213603
Revisit 4d	-3.530247947
Revisit 5d	-4.380602173
Revisit 6d	-4.890563329
Revisit 7d	-3.406296157
Abdominal pain diagnosis	-0.015330246
Constipation diagnosis	0.150484262
Dehydration diagnosis	0.210586845
Fever diagnosis	-0.508220633
Gastroenteritis diagnosis	0.708369608
Genitourinary problem diagnosis	-1.173802692
HEENT problem diagnosis	-0.744005787
Urinary tract infection diagnosis	2.081011214
Viral syndrome diagnosis	-0.497649125
None of the above diagnoses	-3.246578991
Patient's total be	
Beta to pred prol	

Patient's betas
1.183116261
0
0
0
0
0
0
0
-1.992213603
0
0
0
0
0
0
0
0
0
0
0
-3.246578991
-4.055676333
0.017028757