Accuracy of automated identification of delayed diagnosis of pediatric appendicitis and sepsis in the ED

Kenneth A Michelson ,[1] Lillian C Buchhalter,[2] Richard G Bachur,[1] Prashant Mahajan ,[3] Michael C Monuteaux,[1] Jonathan A Finkelstein[2]

[1]Division of Emergency Medicine, Boston Children's Hospital, Boston, Massachusetts, USA
[2]Division of General Pediatrics, Boston Children's Hospital, Boston, Massachusetts, USA
[3]Department of Emergency Medicine, University of Michigan, Ann Arbor, Michigan, USA

Correspondence to
Dr Kenneth A Michelson, Division of Emergency Medicine, Boston Children's Hospital, Boston, MA 02115, USA; kenneth.michelson@childrens.harvard.edu

Received 17 June 2019
Revised 5 September 2019
Accepted 28 September 2019

To cite: Michelson KA, Buchhalter LC, Bachur RG, *et al*. *Emerg Med J* Epub ahead of print: [*please include* Day Month Year]. doi:10.1136/emermed-2019-208841

ABSTRACT

Background Delayed diagnoses of serious emergency conditions can lead to morbidity in children, but are challenging to identify and measure. We developed and piloted an automated tool for identifying delayed diagnosis of two serious conditions commonly seen in the ED using administrative data.

Methods We identified cases with a final diagnosis of appendicitis or sepsis in a freestanding children's hospital from 2008 to 2018, with any hospital ED encounter within the preceding 7 days. Two investigators reviewed a subset of these cases using the electronic health records (EHR) to determine if there was a delayed diagnosis and interrater reliability was assessed using the intraclass correlation coefficient (ICC). An automated tool was applied to the same cases to assess its positive predictive value (PPV) to identify those with a delayed diagnosis, using the manual chart review as the gold standard. The tool used number of days since visit, presence of a related diagnosis on the initial visit, and whether or not the patient was discharged.

Results Previous ED encounters preceded 91/3703 (2.5%) appendicitis cases and 159/1754 (9.1%) sepsis cases; 78 cases of each were sampled for review. In manual review, 73.4% and 22.8% were thought to have delayed diagnoses; reviewer agreement was excellent (appendicitis ICC 0.77, 95% CI 0.62 to 0.86 and sepsis ICC 0.77, 95% CI 0.43 to 0.89). The PPVs of the automated tool for determination of delayed diagnosis for appendicitis within 1, 3 or 7 days were 96.2%, 95.1% and 93.6%, respectively. For sepsis, the PPVs were 71.4%, 63.6% and 41.2% within 1, 3 or 7 days, respectively.

Conclusions This automated tool performed well compared with expert EHR review. Performance was stronger for appendicitis. Further tool refinement could improve performance.

Key messages

What is already known on this subject

- Delayed diagnosis is an important cause of morbidity in EDs, and children are at high risk. There is no validated, automated method of identifying patients who have experienced delayed diagnosis. Such a method would allow for tracking and identification of delays in diagnosis to continuously learn from cases of delays in diagnosis.

What this study adds

- We created and piloted a tool to identify delayed diagnosis of sepsis and appendicitis using administrative data from a Children's Hospital ED. Positive predictive value (PPV) for a delayed diagnosis of appendicitis ranged from 96.2% to 93.6% depending on whether the prior visits was 1, 3 or 7 days earlier. PPVs for sepsis were not as robust. Our study suggests it is possible to automate this type of quality assurance effort for large-scale efforts to detect, predict and prevent delayed diagnosis.

INTRODUCTION

Delayed diagnosis contributes to at least 73 000 deaths per year and is often preventable.[1–7] Delayed diagnosis in the ED is particularly consequential because of the severity and time sensitivity of emergency conditions. High cognitive load, acuity and decision density in the ED are among the factors that have been implicated in diagnostic errors.[8–13] Even among ED patients, children may be at increased risk of delayed diagnosis because they are developmentally unable to articulate symptoms with specificity, early symptoms of serious disease are frequently non-specific, and because EDs disproportionately care for children with poor access to other forms of care.[14–16] Patients with poor access may also be less likely to receive the follow-up care that could act as a safeguard in cases of delayed diagnoses.

A clear understanding of the population variation, causes and outcomes of delayed diagnosis of serious conditions in children in EDs is needed and would lead to interventions to reduce morbidity and mortality among this vulnerable population.

Most children visit non-academic EDs, in which there is most often no infrastructure to measure, monitor or report diagnostic error.[17 18] Measurement of diagnostic delays using administrative data has the potential to overcome this limitation in monitoring and reporting, and to provide an opportunity to compare different types of hospitals. However, it is not known whether apparent delays in diagnosis using administrative data can accurately measure actual delays in diagnosis. As appendicitis and sepsis are conditions frequently associated with delays in diagnosis, they represent two candidate conditions for which an automated approach to identifying delayed diagnosis would be most useful.

Our objective was to develop and pilot a screening tool called DelayDx to identify delayed diagnoses of appendicitis and sepsis in an administrative database, testing it the standard of expert manual health record review.

METHODS

Study design

We conducted a retrospective diagnostic accuracy study. The diagnostic test was DelayDx, a tool to identify ED visits for appendicitis and sepsis with delayed diagnosis within administrative data. The reference standard was the expert opinion of two emergency physicians after manual review of electronic records to assess the presence of a true delayed diagnosis. We adhered to the STARD 2015[19] guidelines for reporting studies on diagnostic accuracy.[19] The local Institutional Review Board approved this study.

Participants

We included patients under 18 years old presenting between 1 July 2008 and 30 June 2018 to Boston Children's Hospital (BCH) with a first-time diagnosis of appendicitis or sepsis, with a preceding ED visit within 7 days. Cases were identified using International Classification of Diseases-9 and -10 (ICD-9 and -10) final admission codes from the Pediatric Health Information System (PHIS) for appendicitis[3] (ICD-9 540.x, ICD-10 K35.x) and sepsis[20 21] (ICD-9 785.52, 995.91, 995.92; ICD-10 A40.x, A41.x, R65.2x).

Data sources

We used two databases. First, the PHIS was used to identify patients. 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 and participating hospitals, and have previously been described.[22–24] We used PHIS data for only BCH, as we intended to link it to local review of our electronic health record (EHR). The BCH EHR was then used to review the cases that were identified in the PHIS.

Reference standard: EHR review

We performed a manual record review on a random sample of the potentially delayed cases identified by our search of the PHIS database. The most recent encounter prior to the visit where the diagnosis was made was called the 'prior encounter'. The encounter in which a diagnosis was made defined the 'diagnosis encounter'. We excluded cases from the manual record review if they did not have sufficient information in the medical record to make an outcome determination (for instance, because of incomplete or missing documentation).

We assigned all cases for separate review to two authors (KAM and LCB). The lead author first developed a draft of the case report form for manual review. Both reviewers piloted the form on 10 random potentially delayed cases with the target diagnoses that were not included in the study sample. The reviewers adapted the report form based on issues identified in the 10-case review, after which the case report form was deemed final and ready for use.

The assigned reviewers reviewed all available documentation from the prior encounter through the end of the diagnosis encounter to determine whether appendicitis or sepsis was present at the prior encounter. Raters considered the following factors: (1) symptoms, signs and test results from prior and diagnosis encounters; (2) the clinician's impression in both encounters; (3) the patient's comorbidities; (4) the typical course of illness; (5) variability in presentation of the illness; and (6) any context for the encounters, such as documentation from before the prior encounter. Reviewers were not asked to assess harm associated with the potential delay, in order to reduce the possibility of hindsight bias.[25]

Reviewers each rated the percentage likelihood the diagnosis was present at the prior encounter, informed by a list of example cases with preset delay likelihoods. A patient's likelihood of delay was assigned as the mean likelihood between reviewers. In this study, we defined potential delay as a patient with a study diagnosis encounter preceded within 7 days by any hospital encounter. Based on the consensus of the authors, we prospectively defined delayed diagnosis as a delay likelihood ≥40%.

To determine interrater reliability of the determination of delayed diagnosis, we used the two reviewers' percentage likelihoods of delayed diagnosis. We calculated a two-way mixed effects intraclass coefficient (ICC) with 95% CI for agreement between the two reviewers.

Diagnostic test: DelayDx, an automated tool to identify possible delayed diagnosis in administrative data

DelayDx is based on a set of primary care trigger tools described by Singh and colleagues.[25] The tool identifies a care episode as a potentially delayed diagnosis based on three parameters: (1) the number of days between the prior encounter and diagnosis encounter, (2) the disposition of the prior encounter (discharged or not discharged) and (3) whether the prior encounter diagnosis was related to the ultimate diagnosis. The parameters are fixed before the tool is applied; it does not learn or adapt to identify delays in diagnosis. Once parameters are chosen, the tool is automated in that it rapidly flags individual encounters as a potential delayed diagnosis or not in large administrative datasets.

In order to develop the tool for any particular diagnosis, it must have a database of potentially related diagnoses for the prior visit. The investigators created this list prior to application of the tool or review of any cases. First, the study authors reviewed the diagnoses, ordered by frequency, that accounted for 90% of the prior visits among all patients with a study condition. Next, each author judged each prior diagnosis as related or not (binary categories). For instance, a diagnosis of 'right lower quadrant pain' on a prior visit was considered related to appendicitis on the diagnosis visit, while a diagnosis of 'tibia fracture' would be considered unrelated. Disagreements in the judgments of the authors were resolved by consensus discussion. Separate lists were generated for appendicitis and sepsis (online supplementary table 1). In the future, additional lists would need to be created if additional conditions beyond appendicitis or sepsis were studied. However, once a list for a specific diagnosis is created, it can be used repeatedly in future audits.

Analysis

We applied the DelayDx tool to all the encounters for which we had performed a manual case review. The opinion of the reviewers as to whether the diagnosis was delayed was the standard against which the DelayDx tool was measured.

The main outcome was the positive predictive value (PPV) of the DelayDx tool, as this takes into account the prevalence of delayed diagnosis. The PPV was the number of delayed diagnoses as determined by the expert reviewers based on EHR review, and were also flagged by DelayDx as having a delay, divided by all cases flagged by the tool as having a delayed diagnosis. This included all true positive delayed cases of the condition (appendicitis or sepsis), patients judged on EHR review never to have had the condition (ie, incorrectly labelled as having the condition by DelayDx) and cases in which the diagnosis was actually made at the prior encounter (incorrectly flagged by DelayDx as having a delay when in fact there was no previous encounter for the condition).

We assessed the tool with multiple fixed, predetermined parameter combinations, including (1) testing windows of time of 1, 3 or 7 days between the prior encounter and diagnosis encounter, (2) testing the requirement that the prior ED encounter resulted in discharge and (3) testing the requirement that prior encounter diagnosis and ultimate diagnosis be related, using the lists of related prior diagnoses generated by the process above. We reported PPVs for each of these combinations. Additionally, we reported sensitivity, specificity and negative predictive value. We reported outcomes of the DelayDx tool using proportions and exact 95% CIs.

Data were collected using Research Electronic Data Capture. Data were analysed using R V.3.5.0.

Power

For this exploratory analysis, we predetermined that 41 cases of each condition in the final tool would provide us with sufficient precision to have a CI with lower and upper bounds no >15% away from a PPV estimate of 80%. Since we preplanned to vary the parameters of the tool, we doubled the number of reviewed cases to 82 per condition, assuming that other reasonable sets of parameters we might wish to test would be present in at least half of the reviewed cases.

RESULTS

We identified 3703 cases of appendicitis and 1754 cases of sepsis from BCH in PHIS (flow diagram, online supplementary figure l). Cases were preceded within 7 days by hospital encounters in 91 (2.5%) appendicitis cases and 159 (9.1%) sepsis cases. We randomly sampled 82 cases for manual review for each diagnosis based on our prespecified power cut-off. Three cases each of appendicitis and sepsis were excluded due to insufficient records available for review, leaving 79 cases for analysis each (table 1).

Interrater reliability was excellent for determination of delayed diagnosis on manual review for both appendicitis (ICC=0.77) and sepsis (0.77) (table 2).

Using the tool with the least restrictive criteria for identifying delayed diagnoses (all revisits within 7 days) correctly identified true delayed diagnoses in 58/79 (PPV 73%) of appendicitis cases and 18/79 (PPV 23%) of sepsis cases (online supplementary table 2). PPV improved with shorter revisit intervals and when requiring a related diagnosis on the prior visit (table 2). The DelayDx parameters with the highest-performing predictions for appendicitis (revisits within 1 day following any ED discharge) yielded a PPV of 96%; for sepsis the best-performing parameters (revisits within 1 day following related ED discharge) had a PPV of 71%. Sensitivity declined with more restrictive tool parameters (table 2). A complete report of all test characteristics for all permutations of tool parameters is shown in online supplementary table 2.

DISCUSSION

We developed an automated tool for identification of delayed diagnoses of appendicitis and sepsis. Among children with appendicitis who had a prior related ED discharge within 1, 3 or 7 days, the tool correctly identified a delayed diagnosis in 96.2%, 95.1% and 93.6% of cases, respectively. The tool was less predictive in sepsis. Among children with sepsis, the tool had a PPV of 71.4%, 63.6% and 41.2% with revisit intervals of 1, 3 or 7 days, respectively.

The difference in performance between appendicitis and sepsis is partly attributable to the significantly different prevalence of delayed diagnoses; 73% of appendicitis and 23% of sepsis revisits represented delayed diagnosis. This prevalence difference is likely due to differences in the types of patients who are at risk for appendicitis versus sepsis. Patients with appendicitis often have no comorbidities and are infrequent visitors to the hospital, so multiple encounters in a short period of time are likely to be related. In contrast, most patients with sepsis have predisposing risk factors such as cancer, indwelling central lines or immune deficiencies; they visit the hospital frequently, so any given encounter is less likely to be directly related to the next. Moreover, the progression of appendicitis is typically over days, whereas progression of sepsis is most frequently over hours. This means that applying longer revisit intervals is more likely to pick up delayed diagnoses for appendicitis than sepsis.

The DelayDx tool was not designed to identify preventable delayed diagnoses. Preventability is a more subjective determination and may not be possible using administrative data.[25 26] Furthermore, preventability may change as diagnostic approaches and tests improve. We believe all delayed diagnoses represent opportunities to improve diagnosis, and provide a better basis for comparison and monitoring between health systems.

ED revisits have been proposed as a quality measure, as revisits may be a marker for patient dissatisfaction or uncertainty.[25 26] However, most revisits are not related to deficiencies in care and revisits are not associated with increased costs or illness severity.[27 28] Despite that, the concept of using a subset of revisits

Table 1 Demographics

	Appendicitis n=79 n (%)	Sepsis n=79 n (%)
Age, years, mean±SD	9.8±4.3	7.5±6.1
Female	37 (46.8)	30 (38.0)
Race		
White	29 (36.7)	38 (48.1)
Black	4 (17.7)	8 (10.1)
Asian	3 (3.8)	4 (5.1)
Pacific Islander	1 (1.3)	0 (0.0)
Other*	32 (40.5)	29 (36.7)
Ethnicity		
Hispanic	24 (30.4)	13 (16.5)
Not Hispanic	45 (57.0)	59 (74.7)
Insurance		
Public	30 (38.0)	37 (46.8)
Private	44 (55.7)	42 (53.2)
Other	5 (6.3)	0 (0.0)
Days between prior encounter and diagnosis encounter, median (25th-75th percentile)	2 (1–4)	3 (1–5)
Admission on prior encounter	20 (25.3)	43 (54.4)
Delayed diagnosis on manual EHR review	58 (73.4)	18 (22.8)

*Among 61 patients with race listed as 'other', 30 were Hispanic, 18 had 'other race' listed and 13 had missing race.
EHR, electronic health record.

Table 2 Performance of DelayDx algorithm for identifying delayed diagnoses depending on the revisit interval (time between prior visit and diagnosis). The algorithm flagged only cases preceded by a related ED discharge. 2×2 table counts are given. Positive predictive value indicates the proportion of identified cases with a true delayed diagnosis. Sensitivity indicates the proportion of cases with truly delayed diagnosis detected by the DelayDx algorithm.

Diagnosis	Revisit interval (days)	TP n	TN n	FP n	FN n	Sensitivity % (95% CI)	Positive predictive value % (95% CI)
Appendicitis	1	25	20	1	33	43.1 (30.2 to 56.8)	96.2 (80.4 to 99.9)
	3	39	19	2	19	67.2 (53.7 to 79.0)	95.1 (83.5 to 99.4)
	7	44	18	3	14	75.9 (62.8 to 86.1)	93.6 (82.5 to 98.7)
Sepsis	1	5	59	2	13	27.8 (9.7 to 53.5)	71.4 (29.0 to 96.3)
	3	7	57	4	11	38.9 (17.3 to 64.3)	63.6 (30.8 to 89.1)
	7	7	51	10	11	38.9 (17.3 to 64.3)	41.2 (18.4 to 67.1)

Positive predictive value: TP/(TP + FP).
Sensitivity, TP/(TP + FN).
FN, false negative; FP, false positive;TN, true negative; TP, true positive.

to identify quality problems such as delayed diagnosis has been employed successfully, including as a trigger for manual review.[2]

Our results are promising in several ways. First, this study suggests that automated revisit analysis using large administrative datasets can identify instances of delayed diagnosis of serious conditions. Our method could be used as part of quality assurance efforts by identifying cases of appendicitis or sepsis in which a delayed diagnosis may have occurred, for manual review. Second, our work suggests the types of diagnoses amenable to this approach: those that progress over days rather than hours, and perhaps those that apply to all children rather than mostly those with serious comorbidities.

This study has several limitations. First, this research is exploratory, using a small number of cases; consequently CIs around estimates of performance are wide. Our approach to assessment of relatedness of diagnoses could be further refined with larger samples, non-binary measurement of relatedness and multicentre consensus assignment of likelihood of relatedness. Fourth, the study assumed that delayed diagnoses occurred in the ED or hospital; however, delayed diagnoses can also occur in other healthcare settings. We chose to focus on ED-attributable delayed diagnoses in an effort to improve diagnosis in EDs, but additional work could include primary care or other urgent care settings. Finally, this single-centre study had a limited number of clinician reviewers, and a tertiary referral paediatric hospital's cases may not represent the larger population of hospitals. This may limit the generalisability of our list of related conditions, and the high tool performance should be interpreted with caution. This only reinforces the need for approaches that can be applied to all hospitals. Taken together, the promising results of this study and limitations justify a larger, multicentre effort with improved ability to assess diagnostic relatedness.

CONCLUSION

An automated tool can be applied to administrative data to identify delayed diagnoses of appendicitis and sepsis with good PPV for true delayed diagnoses. Further study across multiple centres, reviewers and conditions is warranted.

Contributors KAM contributed to study planning, data collection, data analysis and drafted the manuscript. RGB, LCB and JAF contributed to study planning, data analysis and substantially revised the manuscript. PM contributed to study design and substantially revised the manuscript. MCM substantially contributed to data collection and statistical planning.

Funding KAM was supported by an institutional grant from the Boston Children's Hospital Medical Staff Office.

Competing interests None declared.

Patient consent for publication Not required.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available upon reasonable request.

ORCID iDs
Kenneth A Michelson http://orcid.org/0000-0003-1763-7262
Prashant Mahajan http://orcid.org/0000-0003-4278-3560

REFERENCES

1 Balogh EP, Miller BT, Ball JR. *Improving diagnosis in health care*. Washington, DC: National Academies Press, 2015.
2 Singh H, Meyer AN, Thomas EJ. The frequency of diagnostic errors in outpatient care: estimations from three large observational studies involving us adult populations. *BMJ Qual Saf* 2014:1–5.
3 Papandria D, Goldstein SD, Rhee D, *et al*. Risk of perforation increases with delay in recognition and surgery for acute appendicitis. *J Surg Res* 2013;184:723–9.
4 Proulx N, Frechette D, Toye B, *et al*. Delays in the administration of antibiotics are associated with mortality from adult acute bacterial meningitis. *QJM* 2005;98:291–8.
5 Anders JF, Powell EC. Urgency of evaluation and outcome of acute ovarian torsion in pediatric patients. *Arch Pediatr Adolesc Med* 2005;159:532–5.
6 Muszynski JA, Knatz NL, Sargel CL, *et al*. Timing of correct parenteral antibiotic initiation and outcomes from severe bacterial community-acquired pneumonia in children. *Pediatr Infect Dis J* 2011;30:295–301.
7 Kline JA, Hernandez-Nino J, Jones AE, *et al*. Prospective study of the clinical features and outcomes of emergency department patients with delayed diagnosis of pulmonary embolism. *Acad Emerg Med* 2007;14:592–8.
8 Croskerry P, Sinclair D. Emergency medicine: a practice prone to error? *CJEM* 2001;3:271–6.
9 Graber ML, Franklin N, Gordon R. Diagnostic error in internal medicine. *Arch Intern Med* 2005;165:1493–9.
10 Warrick C, Patel P, Hyer W, *et al*. Diagnostic error in children presenting with acute medical illness to a community hospital. *Int J Qual Heal Care* 2014;26:538–46.
11 Remick K, Kaji AH, Olson L, *et al*. Pediatric readiness and facility verification. *Ann Emerg Med* 2016;67:320–8.
12 Croskerry P. Ed cognition: any decision by anyone at any time. *CJEM* 2014;16:13–19.
13 Hudgins JD, Monuteaux MC, Bourgeois FT, *et al*. Complexity and severity of pediatric patients treated at United States emergency departments. *J Pediatr*.
14 Nath JB, Costigan S, Lin F, *et al*. Federally qualified health center access and emergency department use among children. *Pediatrics* 2016;138:e20160479.
15 Committee on the Future of Emergency Care in the United States Health System. *Emergency care for children: growing pains*. Washington, DC: National Academies Press, 2007.
16 Guttmann A, Shipman SA, Lam K, *et al*. Primary care physician supply and children's health care use, access, and outcomes: findings from Canada. *Pediatrics* 2010;125:1119–26.
17 Heisey-Grove DM. Variation in rural health information technology adoption and use. *Health Aff* 2016;35:365–70.
18 DesRoches CM, Worzala C, Joshi MS, *et al*. Small, Nonteaching, and rural hospitals continue to be slow in adopting electronic health record systems. *Health Aff* 2012;31:1092–9.
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 Jolley RJ, Quan H, Jetté N, *et al*. Validation and optimisation of an ICD-10-coded case definition for sepsis using administrative health data. *BMJ Open* 2015;5:e009487–10.
21 Balamuth F, Weiss SL, Hall M, *et al*. Identifying pediatric severe sepsis and septic shock: accuracy of diagnosis codes. *J Pediatr* 2015;167:1295–300.
22 Mongelluzzo J, Mohamad Z, Ten Have TR, *et al*. Corticosteroids and mortality in children with bacterial meningitis. *JAMA* 2008;299:2048–55.
23 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.
24 Weiss AK, Hall M, Lee GE, *et al*. Adjunct corticosteroids in children hospitalized with community-acquired pneumonia. *Pediatrics* 2011;127:e255–63.
25 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.
26 Rising KL, Padrez KA, O'Brien M, *et al*. Return visits to the emergency department: the patient perspective. *Ann Emerg Med* 2015;65:377–86.
27 Jiménez-Puente A, Del Río-Mata J, Arjona-Huertas JL, *et al*. Which unscheduled return visits indicate a quality-of-care issue? *Emerg Med J* 2017;34:145–50.
28 Sills MR, Macy ML, Kocher KE, *et al*. Return visit admissions may not indicate quality of emergency department care for children. *Acad Emerg Med* 2017:1–10.