



BRIEF REPORT

Pediatrics



Check for updates

Automated Calculation of the Pediatric Emergency Care Applied Research Network (PECARN) Intra-Abdominal Injury Rule from Unstructured Emergency Department Notes Using a Large Language Model

Leisha Devisetti¹, Belinda Chen MS², Newton Addo BS², Kush Narang BS², Chris Alvarez BS², John Kornak PhD³, Chandan Singh PhD⁴, Aaron E. Kornblith MD, MS^{2,5}

¹Department of Electrical Engineering and Computer Science; University of California, Berkeley; Berkeley, California, USA

²Department of Emergency Medicine; University of California, San Francisco; San Francisco, California, USA

³Department of Epidemiology and Biostatistics; University of California, San Francisco; San Francisco, California, USA

⁴Voyage AI, Palo Alto, California, USA

⁵Bakar Computational Health Sciences Institute; University of California, San Francisco; San Francisco, California, USA

Correspondence

Aaron Kornblith, MD, MS, Bakar
Computational Health Sciences Institute,
University of California, 550 16th Street, Box
0649, San Francisco, CA 94143, USA. Email:
Aaron.kornblith@ucsf.edu

Received: July 7, 2026

Revised: September 1, 2026

Accepted: September 10, 2026

<https://doi.org/10.1016/j.jacepo.2026.100525>

Abstract

Objective: The Pediatric Emergency Care Applied Research Network (PECARN) intra-abdominal injury clinical prediction rule identifies children at very low risk for clinically significant intra-abdominal injury after blunt abdominal trauma. However, retrospective evaluation of the rule's use in practice requires resource-intensive chart review. We investigated whether large language models (LLMs) can automatically calculate PECARN intra-abdominal injury risk scores from real-world unstructured emergency department (ED) notes.

Methods: We conducted a retrospective secondary analysis of children after blunt abdominal trauma presenting to 2 urban pediatric EDs. The LLM, OpenAI o1, was compared with clinician-derived standards for PECARN risk scores and predictor variable extraction from unstructured physician ED notes. Performance was assessed using standard diagnostic test characteristics.

Results: Among 196 children with complete ED notes, the LLM calculated binary PECARN risk scores with 95.4% accuracy (95% confidence interval [CI] 92.5 to 98.3), 94.4% sensitivity (95% CI 90.6 to 98.2), 98.2% specificity (95% CI

abstract continues

Abstract (continued)

94.6 to 100), and an F1 score of 0.97 (95% CI 0.90 to 0.98). Agreement by weighted Cohen's kappa between multicategorical LLM-derived risk scores and clinician-derived risk scores was high, 0.96 (95% CI 0.92 to 1.00). Predictor variable extraction demonstrated strong performance across all PECARN variables, with accuracy greater than 94%.

Conclusion: The LLM had comparable performance to clinicians in calculating binary and multicategorical PECARN intra-abdominal injury risk scores and extracting predictor variables from unstructured ED notes. LLMs may enable automated calculation of clinical prediction rules from clinical documentation.

Keywords: *clinical decision rules, large language models, abdominal injuries, pediatric emergency medicine*

1 INTRODUCTION

1.1 Background

The Pediatric Emergency Care Applied Research Network (PECARN) intra-abdominal injury clinical prediction rule identifies children at very low risk for clinically significant intra-abdominal injury after blunt abdominal trauma.¹ It supports reduced use of computed tomography scans, which expose children to ionizing radiation and increased lifetime malignancy risk.² More broadly, clinical prediction rules support consistent, evidence-based clinical predictions in emergency medicine.^{1,3,4}

1.2 Importance

Although clinicians can apply the PECARN intra-abdominal injury rule at the bedside, retrospective evaluation of how the rule is used in practice remains challenging because predictor variables are documented inconsistently or require subjective interpretation during chart review, introducing inter-rater variability.^{5,6} Many PECARN predictor variables are documented within unstructured clinical notes, making retrospective calculation of PECARN risk scores labor-intensive and difficult to scale.⁷ Reliable methods to automatically extract predictor variables and calculate PECARN risk scores could support quality monitoring, research validation, and future clinical prediction support applications. LLMs offer a potential approach to automatically calculate clinical prediction rules from unstructured clinical notes and generate risk scores without manual data entry or review.^{8,9} However, it remains unclear whether LLMs can accurately calculate the PECARN intra-abdominal injury risk scores from real-world emergency department (ED) notes.

1.3 Goals of This Investigation

We evaluated whether an LLM can accurately calculate PECARN intra-abdominal injury risk scores from real-world unstructured ED notes compared with clinician-derived scores from chart review. We also assessed the LLM's ability to extract the individual predictor variables used to generate those scores.

2 METHODS

2.1 Study Design and Setting

We conducted a retrospective secondary analysis on a random sample of 198 electronic health records (EHRs) of children 18 years old or younger, who presented after blunt abdominal trauma and received a Focused Assessment with Sonography for Trauma (FAST). The 2 study sites were University of California, San Francisco (UCSF)-affiliated academic pediatric EDs: an urban tertiary/quaternary referral center (~25,000 annual visits) and a Level 1 pediatric trauma center (~45,000 annual visits). Both share an EHR instance, semistructured documentation templates, and some pediatric emergency medicine physicians. We reviewed physician ED notes only because the treating ED physician applies the PECARN intra-abdominal infection (IAI) rule. This study was approved as exempt by the UCSF Institutional Review Board and reported according to the TRIPOD-LLM guidelines and the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement: guidelines for reporting observational studies ([Supplemental Methods](#)).^{10,11}

2.2 Selection of Participants

We included children who met local trauma activation criteria in the period from March 2018 to January 2021, and excluded children with penetrating injuries and any patients missing a physician ED note or final diagnosis code ([Fig S1](#), [Fig S2](#)). The study window spans the period when EHR documentation and archived FAST data were harmonized.

2.3 Measurements and Outcomes

Our primary outcomes were measures of concordance between the LLM and clinicians across 3 levels of the PECARN intra-abdominal injury clinical prediction rule: binary risk scores (very low risk vs. not very low risk), multicategorical risk scores (very low, moderate, and high risk), and individual predictor variable extraction.¹

The Bottom Line

Can a large language model calculate the Pediatric Emergency Care Applied Research Network intra-abdominal injury rule from emergency physician notes?

In this retrospective study of 196 children with blunt abdominal trauma at 2 academic pediatric emergency departments, the model calculated rule-based risk scores with 95.4% accuracy compared with clinicians. Sensitivity was 94.4% and negative predictive value was 86.9% for identifying children at very low risk of injury. These results suggest that automated rule calculation from clinical notes is feasible for research and quality review, but larger external validation is needed before use to guide bedside imaging decisions.

2.3.1 Clinician predictor variable extraction and risk score calculation

To create the reference standard, 2 clinicians (LD; CA) independently reviewed physician ED notes to identify whether each PECARN intra-abdominal injury predictor variable and/or a dangerous mechanism of injury was positive, negative, or missing per prespecified criteria (Table 1). Disagreements were resolved by a physician-investigator (AK). To enable binary classification, missing predictor variables were classified as “negative” because missing documentation was judged more likely to be a negative than a positive finding, consistent with prior approaches.^{12,13} Using clinician-derived predictor variables, we calculated binary and multicategorical PECARN risk scores per the original decision tree.¹

2.3.2 Intra-abdominal injury reference standard development

For secondary outcomes, we reviewed charts to determine whether the child had intra-abdominal injury or intra-abdominal injury requiring acute intervention using hospital final diagnoses and International Classification of Diseases (ICD)-10 diagnosis codes. We reviewed charts to identify acute interventions using previously defined PECARN criteria.¹

2.4 Model Selection

We evaluated GPT-4o-mini (gpt-4o-mini-2024-07-18) and GPT-4.1 (gpt-4.1-2025-04-14) using default sampling parameters (temperature = 1.0, no seed set), and OpenAI o1 (o1-2024-12-17; medium reasoning effort). One inference run was performed per clinical note. OpenAI o1 was selected

for final analyses because it demonstrated the highest accuracy at the predictor variable level (Table S1).

2.5 Model Evaluation

2.5.1 LLM predictor variable extraction and risk score calculation

The LLM was prompted to extract the 7 PECARN predictor variables and dangerous mechanisms individually from unstructured physician ED notes using zero-shot prompting with function calls for structured outputs. Median note length was 845 words with no empty notes, and all note sections were provided. The prompt provided definitions for each predictor variable and instructed the model to classify each as positive or negative from the text alone. LLM prompts were developed over 2 rounds and are recorded in the Supplemental Methods. Two clinicians (LD; AK) assessed discordant LLM–clinician predictor variable extractions to identify misclassification patterns and refine prompts with clearer criteria (Table 1). For risk score calculation, we used the LLM’s extracted predictor variables to generate both a binary and a multicategorical risk score based on the original PECARN decision tree.¹

2.5.2 LLM-derived binary risk scores vs. intra-abdominal injury reference standards

We evaluated LLM-derived binary risk scores in discriminating patients with intra-abdominal injury from those without. Performance was assessed separately against the clinician-derived intra-abdominal injury and intra-abdominal injury requiring acute intervention reference standards.

2.6 Data Analysis

LLM-derived binary risk score calculation and predictor variable extraction were compared with clinician-derived calculation and extraction using standard diagnostic test characteristics (accuracy, sensitivity, specificity, positive predictive value, negative predictive value, and F1 score). For the multicategorical risk scores, we generated a confusion matrix and assessed inter-rater agreement between the LLM and clinicians using weighted Cohen’s kappa. All model inference and data processing were conducted within the secure UCSF Research Analysis Environment and Versa secure computing environments, and all analyses were performed in Python 3.11 (Python Software Foundation).

3 RESULTS

3.1 Characteristics of Study Subjects

Two of the 198 children were excluded because of missing ED notes, yielding a final study cohort of 196 patients (112 boys, 84 girls) with a mean age of 7.8 years (SD 5.3, range 0 to 18 years). Four children (2.0%) had intra-abdominal injury, of whom one required acute intervention (0.5%).

TABLE 1. Operational definitions of the PECARN intra-abdominal injury predictor variables.

Glasgow Coma Scale (GCS) < 14	Patients with a lowest recorded GCS score < 14 at any point en route to the hospital or during the hospital encounter
Evidence of abdominal wall trauma or seat belt sign	- Mentions of seatbelt sign in the abdominal or flank region - Abrasions, ecchymoses, lacerations, or any other external signs of trauma in the abdominal or flank region
Abdominal Tenderness	Abdominal tenderness mentioned
Chest wall trauma	- Mentions of seatbelt sign in the anterior chest region - Abrasions, ecchymoses, lacerations, or any other external signs of trauma in the chest region
Vomiting	Mentions of vomiting or emesis
Decreased Breath Sounds	Reduction in the intensity of respiratory noises heard on auscultation on one or both sides of the chest
Abdominal Pain	Measured by patient complain of abdominal pain
Dangerous Mechanism	- Motor vehicle crashes with ejection, rollover, or death in the same crash - Motor vehicle crashes with speed greater than 20 miles per hour and patient unrestrained - Falls that are documented as or appear to be greater than 10 feet (eg, falling from the second story of a building) - Pedestrians or bicyclists struck by vehicles moving greater than 20 miles per hour - Bicycle collision with handlebars striking the abdomen Motor vehicle crashes with airbag deployment

Bold text indicates refinements made during iterative review.

3.2 Main Results

The LLM binary risk scores achieved a sensitivity of 94.4% (95% CI 90.6 to 98.2) and a negative predictive value of 86.9% (95% CI 78.4 to 95.4) compared with the clinician-

derived binary risk scores. Overall accuracy was 95.4% (95% CI 92.5 to 98.3), specificity was 98.2% (95% CI 94.6 to 100), positive predictive value was 99.3% (95% CI 97.8 to 100), and F1 score was 0.97 (95% CI 0.90 to 0.98). The

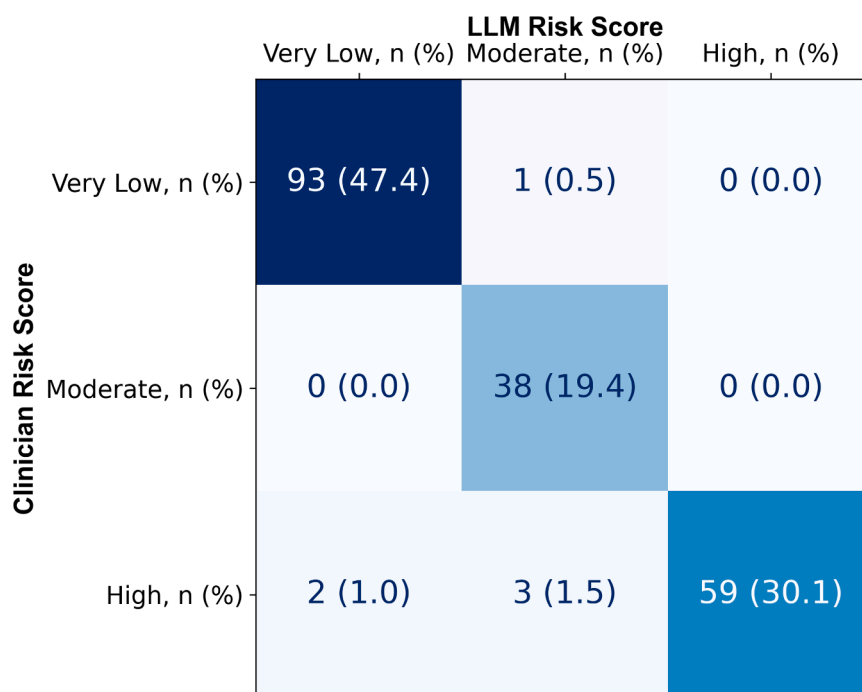


FIGURE 1. Confusion matrix for Large Language Model (LLM)-derived and clinician-derived multi-categorical PECARN risk scores.

TABLE 2. Test characteristics of Large Language Model extraction of PECARN intra-abdominal injury clinical prediction rule predictor variables (N=196).

	GCS < 14	Abdominal Wall Trauma	Abdominal Pain	Abdominal Tenderness	Chest Wall Trauma	Decreased breath sounds	Vomiting since injury	Dangerous Mechanism
Positive, n (%)	7 (3.6)	24 (12.2)	40 (20.4)	39 (19.9)	21 (10.7)	3 (1.5)	31 (15.8)	64 (32.7)
Accuracy	100% [196/196]	98.0% [192/196]	99.5% [195/196]	99.0% [194/196]	98.4% [193/196]	99.0% [194/196]	100% [196/196]	94.4% [185/196]
Sensitivity	100% [7/7]	83.3% [20/24]	100% [40/40]	94.9% [37/39]	95.0% [20/21]	33.3% [1/3]	100% [31/31]	82.8% [53/64]
Specificity	100% [189/189]	100% [172/172]	99.4% [155/156]	100% [157/157]	98.8% [173/175]	100% [193/193]	100% [165/165]	100% [132/132]
PPV	100% [7/7]	100% [20/20]	97.6% [40/41]	100% [37/37]	90.5% [20/22]	100% [1/1]	100% [31/31]	100% [53/53]
NPV	100% [184/184]	97.7% [172/176]	100% [155/155]	98.7% [157/159]	99.4% [173/174]	99.0% [193/195]	100% [165/165]	92.3% [132/143]
F1 score	1.00	0.91	0.99	0.97	0.93	0.50	1.00	0.91

GCS, Glasgow Coma Scale score; PPV, positive predictive value; NPV, negative predictive value.

LLM required 2 hours to process all 196 encounters. Agreement between the LLM-derived and clinician-derived multicategorical risk scores by weighted Cohen's kappa was 0.96 (95% CI 0.92 to 1.00). [Figure 1](#) presents the confusion matrix comparing the multicategorical risk scores from the LLM and the clinician. In both cases where the clinician calculated a high-risk score and the LLM calculated a very low risk score, the LLM missed abdominal wall trauma; neither child had IAI.

Across all 8 variables evaluated, the LLM generated 1568 labels, of which 23 (1.47%) were discordant with the clinician-derived labels. Most discordance involved *dangerous mechanism of injury* (11/23, 47.8%), *abdominal wall trauma* (4/23, 17.4%), *chest wall trauma* (3/23, 13.0%), *decreased breath sounds* (2/23, 8.7%), *abdominal tenderness* (2/23, 8.7%), and *abdominal pain* (1/23, 4.3%). [Table 2](#) reports LLM performance for the extraction of individual PECARN intra-abdominal injury predictor variables.

Qualitative review of discordant labels revealed variable-specific error patterns, including missing quantitative details for dangerous mechanisms, failure to recognize anatomical descriptors for *abdominal wall trauma*, classification of *chest wall trauma* due to blunt trauma without evidence of external injury, nonstandard phrasing for *decreased breath sounds*, pain attributed to nontraumatic causes, and conflicting documentation for *abdominal tenderness* ([Figure S3](#)).

As a secondary outcome, LLM-derived risk scores demonstrated 100% sensitivity for intra-abdominal injury and intra-abdominal injury requiring acute intervention, although confidence intervals were wide, because of the low prevalence of these outcomes ([Table 3](#)).

4 LIMITATIONS

Our study has several limitations. First, clinicians establishing the reference standard were not blinded to LLM outputs, which may have inflated agreement. However, consensus adjudication is an accepted method for improving reference standard quality involving clinically interpreted variables.^{14–17} Fully blinded adjudication would strengthen future work. Second, the low prevalence of intra-abdominal injury and several PECARN predictor variables yielded binary risk scores with lower negative predictive value and wide confidence intervals, restricting utility beyond quality improvement and research applications. Because the primary objective was to evaluate LLM risk score calculation rather than clinical outcome prediction, larger cohorts enriched for clinical outcomes are needed. Third, developing operational definitions for predictor variables revealed inconsistencies in the reference standard that required adjudication, such as defining the extent of abrasion coverage in abdominal wall trauma. As a result, ambiguity in these definitions may have limited clinician and LLM performance. Fourth, iterative prompt refinement may have overfit to institution-specific documentation patterns. External validation and site-specific prompt tuning may be required, particularly for dangerous mechanism of injury. Fifth, classifying missing documentation as “negative” may have

TABLE 3. Test characteristics for Large Language Model calculation of binary PECARN intra-abdominal injury risk scores identifying intra-abdominal injury (N=196).

Diagnostic Test Characteristics	Intra-abdominal Injury (n=4)	Intra-abdominal Injury Requiring Acute Intervention (n=1)
Accuracy	50.5% [99/196] (95% CI 43.6% to 57.4%)	49.0% [96/196] (95% CI 42.1% to 55.9%)
Sensitivity	100% [4/4] (95% CI 51.0% to 100%)	100% [1/1] (95% CI 20.7% to 100%)
Specificity	49.5% [95/192] (95% CI 42.5% to 56.5%)	48.7% [95/195] (95% CI 41.8% to 55.7%)
PPV	4.0% [4/101] (95% CI 1.6% to 9.7%)	1.0% [1/101] (95% CI 0.2% to 5.4%)
NPV	100% [95/95] (95% CI 96.1% to 100%)	100% [95/95] (95% CI 96.1% to 100%)

PPV - positive predictive value; NPV - negative predictive value.

affected performance estimates. Although this reflects the binary structure of the PECARN IAI rule, some undocumented findings may have been positive, especially chest wall and abdominal wall trauma, which lack structured entries. Sixth, we analyzed only physician ED notes. Future work could investigate alternative missing-data strategies, including missingness indicators and additional documentation sources, and how objective definitions within clinical prediction rules impact clinician baseline extraction and downstream LLM performance.

5 DISCUSSION

In this retrospective single-system study, the LLM calculated the PECARN intra-abdominal injury risk scores from unstructured physician ED notes with performance comparable to clinician reviewers. LLM performance varied across PECARN predictor variables. Accuracy was highest for objective findings, such as Glasgow Coma Scale score <14, abdominal tenderness, and decreased breath sounds, and lowest for dangerous mechanism of injury, whereas all other variables demonstrated similarly high performance. The observed differences in performance highlighted the importance of clear operational definitions for individual predictor variables.

Developing explicit operational definitions emerged as a key determinant of both clinician consistency and LLM performance. Standardized definitions improved clinician consistency among reviewers when establishing the reference standard. For example, defining dangerous mechanisms required iterative review by the investigators because several mechanisms encountered during chart abstraction were not explicitly addressed in the original PECARN study definitions. LLM performance depended on the clarity and consistency of the clinical definitions used to operationalize predictor variables. Discordant cases frequently reflected ambiguity in the underlying definitions rather than model errors, suggesting that LLMs may help identify and refine imprecise clinical variable definitions, supporting rule development and bedside interpretability. Notably, the LLM applied definitions strictly and did not consider a mechanism of injury as dangerous if the necessary quantitative information, such as fall height or motor vehicle crash speed, was not documented.

Previous efforts to automate clinical prediction rules have relied on structured EHR data, traditional natural language processing, or manual chart abstraction. Although effective, each approach faces important scalability limitations.

Structured EHR-based approaches depend on documentation in designated fields, which is typically inconsistent in clinical practice.¹⁸ Conventional natural language processing methods can extract information from narrative text but typically require task-specific development.⁸ Manual chart abstraction remains the standard approach, but is labor-intensive and difficult to sustain at scale.¹⁹ Recent studies have demonstrated that LLMs can perform complex chart abstraction tasks from clinical notes, overcoming these issues.^{9,20,21} Our findings extend this work by demonstrating LLM calculation of a validated clinical prediction rule from routine documentation without structured data entry or task-specific development. Future work should evaluate whether this approach generalizes across institutions, EHR systems, and documentation cultures.

In conclusion, LLM-derived intra-abdominal injury risk scores may be useful for quality improvement and research applications, including auditing adherence to clinical prediction rules, standardizing predictor definitions, and evaluating prediction rule performance across sites without resource-intensive chart review. Future work should validate this LLM-based approach prospectively with greater outcome prevalence across institutions before implementing clinician-facing applications.

ETHICS DECLARATIONS

The authors have NOT obtained informed consent from participants or their legal representatives. This study was approved by the University of California, San Francisco Institutional Review Board (25-43740), which granted a waiver of informed consent and parental permission. The study was a retrospective analysis of existing emergency department records, involving no intervention and no contact with patients or families, and posing no more than minimal risk. Obtaining consent from every family in a cohort of this size, for encounters that had already occurred, was not practicable, and the waiver did not adversely affect the rights or welfare of the participants. All data were de-identified before analysis and stored on secure institutional servers. The study was conducted in accordance with the Declaration of Helsinki.

This study was performed in compliance with relevant laws, regulatory frameworks and guidelines where the research took place. This study was approved by the University of California, San Francisco. (Approval No. 25-43740)

FUNDING AND SUPPORT

By *JACEP Open* policy, all authors are required to disclose any and all commercial, financial, and other relationships in any way related to the subject of this article as per ICMJE conflict of interest guidelines (see www.icmje.org).

All authors attest to meeting the four [ICMJE.org](http://www.icmje.org) authorship criteria: (1) Substantial contributions to the conception or design of the work; or the acquisition, analysis, or interpretation of data for the work; AND (2) Drafting the work or revising it critically for important intellectual content; AND (3) Final approval of the version to be published; AND (4) Agreement to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. Research reported in this publication was supported by Eunice Kennedy Shriver National Institute of Child Health and Human Development of the National Institutes of Health under award number K23HD110716. This information or content and conclusions are those of the author and should not be construed as the official position or policy of, nor should any endorsements be inferred by HRSA, HHS or the U.S. Government.

CONFLICT OF INTEREST

AEK and NA are co-founders and consultants to CaptureDx. This entity did not have any role in the design, planning, or execution of the study, or interpretation of the findings. All other authors have affirmed they have no conflicts of interest to declare.

ORCID

Aaron E. Kornblith MD, MS  <https://orcid.org/0000-0002-1344-575X>

REFERENCES

- Holmes JF, Lillis K, Monroe D, et al. Identifying children at very low risk of clinically important blunt abdominal injuries. *Ann Emerg Med*. 2013;62(2):107-116.e2. <http://doi.org/10.1016/j.annemergmed.2012.11.009>
- Smith-Bindman R, Chu PW, Azman Firdaus H, et al. Projected lifetime cancer risks from current computed tomography imaging. *JAMA Intern Med*. 2025;185(6):710. <http://doi.org/10.1001/jamainternmed.2025.0505>
- Kuppermann N, Holmes JF, Dayan PS, et al. Identification of children at very low risk of clinically-important brain injuries after head trauma: a prospective cohort study. *Lancet*. 2009;374(9696):1160-1170. [http://doi.org/10.1016/S0140-6736\(09\)61558-0](http://doi.org/10.1016/S0140-6736(09)61558-0)
- Leonard JC, Harding M, Cook LJ, et al. PECARN prediction rule for cervical spine imaging of children presenting to the emergency department with blunt trauma: a multicentre prospective observational study. *Lancet Child Adolesc Health*. 2024;8(7):482-490. [http://doi.org/10.1016/S2352-4642\(24\)00104-4](http://doi.org/10.1016/S2352-4642(24)00104-4)
- Yen K, Kuppermann N, Lillis K, et al. Interobserver agreement in the clinical assessment of children with blunt abdominal trauma. *Acad Emerg Med*. 2013;20(5):426-432. <http://doi.org/10.1111/ajem.12132>
- Ahmad FA, Browne LR, Glomb NW, et al. Interrater reliability between surgeons and pediatric emergency providers in the cervical spine assessment of injured children. *J Trauma Acute Care Surg*. 2025;99(6):876-883. <http://doi.org/10.1097/TA.0000000000004695>
- Wu H, Toti G, Morley KI, et al. SemEHR: a general-purpose semantic search system to surface semantic data from clinical notes for tailored care, trial recruitment, and clinical research. *J Am Med Inform Assoc*. 2018;25(5):530-537. <http://doi.org/10.1093/jamia/ocx160>
- Paulraj NJ. Natural language processing on clinical notes: advanced techniques for risk prediction and summarization. *J Comp Sci Technol Stud*. 2025;7(3):494-502. <http://doi.org/10.32996/jcsts.2025.7.3.56>
- Wright DS, Socrates V, Huang T, et al. Automated computation of the HEART score with the GPT-4 large language model. *Am J Emerg Med*. 2025;93:120-125. <http://doi.org/10.1016/j.ajem.2025.03.065>
- Gallifant J, Afshar M, Ameen S, et al. The TRIPOD-LLM reporting guideline for studies using large language models. *Nat Med*. 2025;31(1):60-69. <http://doi.org/10.1038/s41591-024-03425-5>
- Strengthening the reporting of observational studies in epidemiology (STROBE) statement: guidelines for reporting observational studies. *BMJ*. 2007;335(7626):bmj/335/7626/0.2.atom. <http://doi.org/10.1136/bmj.39386.490150.94>
- Tsvetanova A, Sperrin M, Peek N, Buchan I, Hyland S, Martin GP. Missing data was handled inconsistently in UK prediction models: a review of method used. *J Clin Epidemiol*. 2021;140:149-158. <http://doi.org/10.1016/j.jclinepi.2021.09.008>
- Simper MA, Ahmad FA, Abraham J, Maddox TM, Flotken J, Rudloff JR. Safety and diagnostic accuracy of large-language model application of PECARN head injury algorithm. *Int J Med Inform*. 2026;216:106466. <http://doi.org/10.1016/j.ijmedinf.2026.106466>
- Feng J, Kothari A, Vossler P, et al. Human-AI co-design for clinical prediction models. *NPJ Digit Med*. 2026. Online ahead of print, <https://doi.org/10.1038/s41746-026-02838-5>
- Kothari A, Vossler P, Digitale J, et al. When the domain expert has no time and the LLM developer has no clinical expertise: real-world lessons from LLM co-design in a safety-net hospital. *AAAI*. 2026;40(45):38754-38762. <http://doi.org/10.1609/aaai.v40i45.41219>
- Agatstein K. Chart review is dead; long live chart review: how artificial intelligence will make human review of medical records obsolete, one day. *Popul Health Manag*. 2023;26(6):438-440. <http://doi.org/10.1089/pop.2023.0227>
- Vaccaro M, Almaatouq A, Malone T. When combinations of humans and AI are useful: a systematic review and meta-analysis. *Nat Hum Behav*. 2024;8(12):2293-2303. <http://doi.org/10.1038/s41562-024-02024-1>
- Perry WM, Hossain R, Taylor RA. Assessment of the feasibility of automated, real-time clinical decision support in the emergency department using electronic health record data. *BMC Emerg Med*. 2018;18(1):19. <http://doi.org/10.1186/s12873-018-0170-9>
- Panacek EA. Performing chart review studies. *Air Med J*. 2007;26(5):206-210. <http://doi.org/10.1016/j.amj.2007.06.007>
- Boussina A, Krishnamoorthy R, Quintero K, et al. Large language models for more efficient reporting of hospital quality measures. *NEJM AI*. 2024;1(11). <http://doi.org/10.1056/aics2400420>; doi:10.1056/AIcs2400420.
- Schipper A, Belgers P, O'Connor RD, et al. Large language model automated extraction of clinical signs and symptoms from emergency department reports for machine learning prediction models: development and validation study. *JMIR Med Inform*. 2026;14:e81500. <http://doi.org/10.2196/81500>

SUPPLEMENTARY MATERIAL

Supplementary material associated with this article can be found in the online version at <https://doi.org/10.1016/j.acepjo.2026.100525>.

How to cite this article: Devisetti L, Chen B, Addo N, et al. Automated Calculation of the Pediatric Emergency Care Applied Research Network (PECARN) Intra-Abdominal Injury Rule from Unstructured Emergency Department Notes Using a Large Language Model. *JACEP Open*. 2026;7:100525.

<https://doi.org/10.1016/j.acepjo.2026.100525>