Development and Validation of a Natural Language Processing Tool to Identify Injuries in Infants Associated With Abuse

Gunjan Tiyyagura, MD, MHS; Andrea G. Asnes, MD, MSW; John M. Leventhal, MD; Eugene D. Shapiro, MD; Marc Auerbach, MD, MSc; Wei Teng, PhD; Emily Powers, MD, MHS; Amy Thomas, MS; Daniel M. Lindberg, MD; Justin McClelland, BS; Carol Kutryb, PhD; Thomas Polzin, PhD; Karen Daughtridge, BA; Virginia Sevin, MBA, BSN, CPC; Allen L. Hsiao, MD

From the Yale University School of Medicine (G Tiyyagura, AG Asnes, JM Leventhal, ED Shapiro, M Auerbach, W Teng, E Powers, A Thomas, AL Hsiao), New Haven, CT; University of Colorado Anschutz Medical Campus (DM Lindberg), Aurora Colo; and 3M | M*Modal Health Information Systems (J McClelland, C Kutryb, T Polzin, K Daughtridge, V Sevin), Pittsburg, Pa

Conflict of interest: Andrea Asnes reports that the Department of Pediatrics receives payment for her expert testimony in child abuse cases and that she receives grants from the State of Connecticut to support the Yale Child Abuse Programs. Dr Lindberg has served as a paid expert witness in cases with concern for child abuse. Virginia Sevin, Carol Kutryb, Thomas Polzin, Karen Daughtridge, and Justin McClelland are employed by 3M|MModal but the contents of this manuscript are solely the responsibility of the authors and are not influenced by the official view of 3M. The other authors have no relevant conflicts of interest.

Address correspondence to Gunjan Tiyyagura MD, MHS Yale University School of Medicine, 100 York St, Suite 1d, New Haven, CT 06511. (e-mail: gunjan.tiyyagura@yale.edu).

Received for publication July 20, 2021; accepted November 6, 2021.

ABSTRACT

OBJECTIVES: Medically minor but clinically important findings associated with physical child abuse, such as bruises in pre-mobile infants, may be identified by frontline clinicians yet the association of these injuries with child abuse is often not recognized, potentially allowing the abuse to continue and even to escalate. An accurate natural language processing (NLP) algorithm to identify high-risk injuries in electronic health record notes could improve detection and awareness of abuse. The objectives were to: 1) develop an NLP algorithm that accurately identifies injuries in infants associated with abuse and 2) determine the accuracy of this algorithm.

METHODS: An NLP algorithm was designed to identify ten specific injuries known to be associated with physical abuse in infants. Iterative cycles of review identified inaccurate triggers, and coding of the algorithm was adjusted. The optimized NLP algorithm was applied to emergency department (ED) providers' notes on 1344 consecutive sample of infants seen in 9 EDs over 3.5 months. Results were compared with review of the same notes conducted by a trained reviewer blind to the NLP results with discrepancies adjudicated by a child abuse expert.

RESULTS: Among the 1344 encounters, 41 (3.1%) had one of the high-risk injuries. The NLP algorithm had a sensitivity and specificity of 92.7% (95% confidence interval [CI]: 79.0%−98.1%) and 98.1% (95% CI: 97.1%−98.7%), respectively, and positive and negative predictive values were 60.3% and 99.8%, respectively, for identifying high-risk injuries.

CONCLUSIONS: An NLP algorithm to identify infants with high-risk injuries in EDs has good accuracy and may be useful to aid clinicians in the identification of infants with injuries associated with child abuse.

KEYWORDS: child abuse; emergency department; natural language processing; test characteristics

ACADEMIC PEDIATRICS 2022;22:981−988

WHAT'S NEW

A natural language processing algorithm was designed to identify ten specific injuries known to be associated with physical abuse in infants and had a sensitivity and specificity of 92.7% (95% CI: 79.0%−98.1%) and 98.1% (95% CI: 97.1%−98.7%).

BACKGROUND

PHYSICAL ABUSE AFFECTS approximately 115,100 children in the United States (US) each year, with infants (children <12 months old) at highest risk.[1] More than 30% of children who die or suffer serious injuries from abuse have been previously evaluated, often in emergency departments (EDs), for signs, such as bruising or symptoms such as vomiting or irritability that were not recognized as being secondary to abusive injuries.[2−6] This problem is especially important in EDs that treat both adults and children (general EDs), which have relatively fewer resources to recognize and to manage child abuse, even though it is where the majority of US children present for emergency care.[6−9] To improve identification and evaluation of injuries that are associated with high rates of

Copyright © 2021 by Academic Pediatric Association

physical abuse (eg, "sentinel injuries"), child abuse pediatricians have promoted the use of clinical pathways and electronic-health record-based clinical decision support in all EDs.[10–14] Despite these efforts, identification of occult injuries and reporting of injuries to child protective services (CPS) when patients present with high-risk injuries are highly variable due to a provider's clinical judgment and are often impacted by socioeconomic, racial, and ethnic biases.[8,11,15,16]

Existing clinical pathways for child abuse vary in both content and process, and in the US have led to increased testing for occult injuries, increased reporting to CPS, decreased variability of care, and increased compliance with national guidelines about the evaluation of abuse.[17–23] However, to date, pathway implementation has not demonstrated a decline in cases of missed abuse. Additionally, while pathway implementation may provide guidance about which children to evaluate for suspected abuse, it does not obviate the need for frontline providers to recognize that an injury is high-risk for being due to abuse; a pathway may be ineffective if a high-risk physical finding is overlooked by a frontline clinician.

Researchers have also examined the use of electronic-health record-based clinical decision support (CDS) systems to improve evaluation of suspected abuse. Berger et al reported on the development and validation of a system of triggers from discrete fields in the electronic health record (EHR) to assist providers in identifying children with physical abuse. Two main limitations were the possibility of an overestimation of the sensitivity of the trigger system and the inability of free text to generate alerts.[24] Another study by Berger's team examined the impact of using a combination of a child abuse screen, physician orders, discharge diagnoses and natural language processing (NLP) within very focused fields of the note (ie, the "chief complaint" and the "focused assessment" fields) to identify injuries concerning for child abuse.[22] The majority of the over triggers in this study were due to the NLP even though potentially important information sources such as imaging were not included. Furthermore, due to the study design in which false-negative cases could not be ascertained, the test characteristics of this combination of tools remained unknown.

IMPORTANCE

NLP is a method of automated data extraction that converts unstructured free text to structured, encoded data by applying rule-based algorithms to combinations of written terms.[25] Advanced NLP, such as the system we use in this study, uses a mixture of both machine learning of deep neural networks and rule based reasoning. Advanced NLP interfaced with the EHR, when optimized, may be a more effective and acceptable tool for clinicians compared with some of the examples noted above as it does not require changes in workflow or extra efforts to initiate. It has been used to accurately identify various disease processes and to automate the estimation of the severity of illness.[26-28] NLP may provide a unique opportunity to identify high-risk injuries associated with child physical abuse documented in free text but not recognized as high-risk by the clinician.[5] An accurate NLP tool, combined with CDS, may be used to decrease missed cases of abuse in young children who present for emergency care and may decrease bias in both the evaluation of and the reporting of suspected child abuse.

GOALS OF THE INVESTIGATION

Studies to date have not described the development of or test characteristics of an optimized NLP algorithm to identify high-risk injuries associated with child physical abuse. Therefore, the objectives of this study were to: 1) develop an NLP algorithm that accurately identifies injuries potentially associated with abuse in children <12 months old and 2) determine the accuracy of the optimized NLP algorithm.

METHODS

DEVELOPMENT OF THE NLP ALGORITHM

The NLP system used in this work is 3M's M*Modal Natural Language Understanding (NLU, 3M M*Modal, Pittsburgh, Pa) platform that provides the foundation for speech recognition, NLP indexing, and transcription in the Yale New Haven Health (YNHH) System. The NLP system processes text using a sequence of dedicated engines that are grouped into 3 main modules. The initial "Core" module establishes the document structure from sections down to words. The "Core" module identifies paragraphs, tables, and lists and assigns standardized codes to these sections using industry-accepted standard ontologies, such as the Logical Observation Identifiers Names and Codes (LOINC) code system. Once the higher-level structure is established, the text is broken up into sentences and words. The other main task of the "Core" module is to define the scoping of phraseology affecting temporality, certainty (including negation), and subjects within sentences. The second "Clinical" module builds up the clinical content of the text. Using a combination of text matching, parsing, and rule-based techniques, clinical content, such as body parts, disorders, procedures, devices, lab results, vital signs, allergies, and medications are identified and structured in information models. These are assigned codes from SNOMED-CT (Systematized Nomenclature of Medicine − Clinical Terms), used to encode EHR data and capture clinical information, LOINC, a universal standard for identifying medical laboratory observations, and Rx-Norm, a standardized drug nomenclature that links systems that use different vocabularies for use in the following module. The final "Reasoning" module uses customizable rules to reason over or draw conclusions based on the given information about the document on the basis of the previously generated coded information models to trigger alerts, aggregate clinical information, or produce summaries (Figure).

For this study, we developed rules at the Reasoning module level to examine nursing, provider, and social

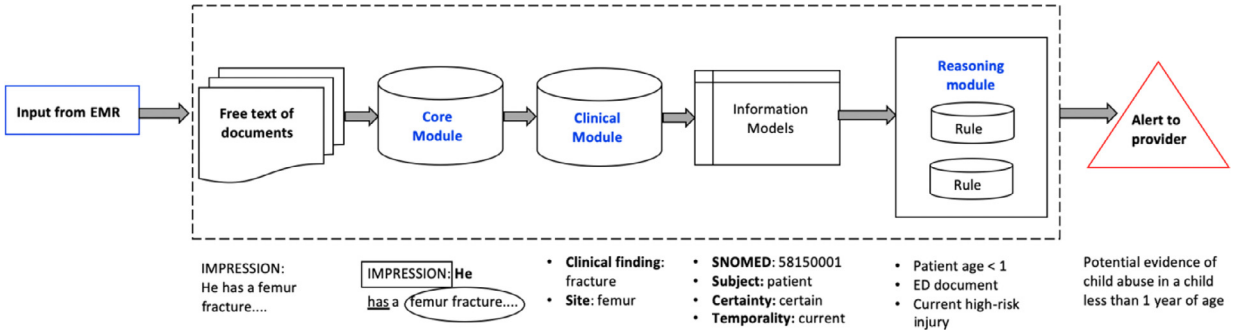

Figure. *Example of the NLP system. NLP indicates natural language processing.*

work notes in the ED to identify high-risk injuries concerning for abuse in children <1 year of age (Table 1).[29–33] Notes could be typed, transcribed, created with speech recognition, or combination of the above. Variations of the injury terms were also included. For example, contusion, ecchymoses, and hematoma (other than umbilical) were included as synonyms for the term bruising whereas the terms mark or swelling were excluded. Intracranial hemorrhage, subarachnoid intracranial hemorrhage, subdural intracranial hemorrhage, cerebral hemorrhage, cerebral edema and traumatic hemorrhage were considered synonymous terms for intracranial injury. For solid organ injury, aspartate transaminase or alanine aminotransferase >80 would trigger as positive as long as there was no mention of biliary atresia, short gut syndrome, or liver failure in the text. The rules were tested and refined using data from the Yale New Haven Children's Hospital (YNHCH) Pediatric ED and 8 general EDs in the YNHH health system. Epic (Epic Systems Corporation, Verona, Wis) is the EHR used in all the EDs of the Health System. Approximately 4500 documents created in the Epic ASAP (Epic's emergency department information system application) module by ED clinicians (physicians, advanced practitioners, nurses, and social workers) from visits of children <1 year of age evaluated between January1, 2018 and October 23, 2019 were included in the initial dataset. The documents were processed through the Core, Clinical and Reasoning modules as above, which resulted in 200 positive alerts. A trained research assistant and multiple members of the clinical NLP development team performed 6 rounds of review. Over multiple cycles, inaccurate triggers (both false-positive and false-negative triggers) were identified, and coding was adjusted. To minimize alert fatigue in clinical practice,[34,35] the cycles of review emphasized specificity over sensitivity for detecting the 10 high-risk injuries to permit eventual reduction in the number of unnecessary alerts for a clinical decision support system.

Adjustments were made to the Core, Clinical, and Reasoning modules to optimize accuracy of the alerts. For example, Core engines were modified to better identify and exclude injuries that were written as part of a differential diagnosis but that were not present at the ED visit, so they would not trigger an alert. Similarly, adjustments were made to the Subject engine to better account for the language in pediatric documents in which the reporter (eg, parent/guardian) is different from the patient (eg, child). The Clinical module was adjusted by adding new terminology, such as synonyms and acronyms and by enhancing rule-based techniques to improve recognition of the injuries of interest. For example, improvements were made to identify the anatomical site of various injury types; we then were able to use this information at the Reasoning level to exclude injuries in locations on the body that were less concerning for abusive etiologies (eg, hematoma of frontal scalp in infants ≥5 months old). Finally, the rules themselves were refined to exclude high risk injuries with an obvious known cause other than child abuse (eg, due to birth trauma or motor vehicle accident) and to remove terms, such as erythema or redness, which were originally included as variations for bruising but were also present frequently in the setting of infant visits for rash, allergy, infections, and vomiting.

A final round of review to further refine the performance of the NLP algorithm was completed on a separate data set from the same hospital sites that consisted of

Table 1. High-Risk Injuries Concerning for Physical Abuse in Infants

Younger than 12 Months	Younger than 5 Months
• Fracture (long bone, skull, rib) • Intracranial injury • Abdominal injury (lab or imaging) • Burns • Bruising: - TEN (torso, ears, neck) - FACES (frenum, angle of jaw, cheeks, eyelids, subconjunctiva) - P (patterned)	• Any bruising • Any oral injury

2000 documents with 50 positive alerts from October 24, 2019 to January 31, 2020.

VALIDATION

The optimized NLP algorithm was tested on a validation data set consisting of 1344 patient encounters of all children <1-year-old evaluated in the same hospital sites in the YNHH system from September 1, 2020 to December 15, 2020. Data on whether the NLP algorithm detected a high-risk injury and which high-risk injury it detected was documented. To determine whether or not a high-risk injury was actually noted in the child's encounter, a blinded review of the same documents was also completed by a trained research assistant who recorded the presence/absence of a high-risk injury and which high-risk injury was present. Discordance between the NLP algorithm and the manual review was managed by a review of the entire document by one of the investigators (G.T.) who is a Pediatric Emergency Medicine provider with interest and training in child abuse.

This study received exemption from the Yale University Human Research Protection Program.

ANALYSIS

Descriptive statistics were used to characterize the infants in the validation sample. Sensitivity, specificity, and positive and negative predictive values as well as the 95% confidence intervals (CIs) were calculated using IBM SPSS Statistics version 26. We also qualitatively analyzed false positive and false negative triggers. These classification errors were then grouped and quantified in different categories.

RESULTS

In the 3.5-month validation period, 1344 infants presented for care, and 41 (3.1%) had at least one high-risk injury.

Sensitivity and specificity of the NLP algorithm for the detection of high-risk injuries were 92.7% (95% CI: 79.0 −98.1) and 98.1% (95% CI: 97.1−98.7), respectively (Table 2). Positive and negative predictive values were 60.3% and 99.8%, respectively. Overall, 25 of 1303 encounters (1.9%) were interpreted as false-positive by the NLP, and 3 of 41 (7.3%) encounters were false-negatives.

There were 28 errors for the NLP system, 25 false positives and 3 false negatives. Descriptions, examples, and counts of each type are shown in Table 3.

Of the 41 high-risk injuries identified in the study, manual chart review revealed that 2 were due to birth injury and 2 had elevated transaminases (a criterion for abdominal injury for the NLP) in the presence of fever, making infection a more likely cause of the abnormality. In the remaining 37 encounters of infants presenting with high-risk injuries, ED providers did not document consideration of physical abuse in 9 charts (24.3%), and in these encounters (3 for fractures, 5 for bruising or hematoma in infants <5 months of age and 1 for frenum injury), no further evaluation for suspected abuse was completed. This percentage may be relatively low compared to other time periods as active quality improvement work to improve recognition of abuse was being done in the pediatric ED and in 5 of the 8 community EDs from which patient encounters for this study were reviewed.

DISCUSSION

To our knowledge, this is one of the first studies to describe the development and validation of an NLP algorithm to identify high-risk injuries concerning for physical abuse in children <1-year-old. We demonstrated that the algorithm has excellent test characteristics. If we had implemented this algorithm in real-time, this tool would have alerted positive in 63 encounters over a 3.5-month period, (fewer than 1 encounter per day), with the majority (38, 60.3%) accurately raising concerns about injuries associated with physical abuse, while in 25 encounters (39.7%), the alert would have been a false positive. Three encounters, or 7.3% of the true positives, would have been missed.

In the 28 encounters in which the NLP system erred, identification and linguistic context errors typically occurred due to complex and/or idiosyncratic language or document formatting. The relatively high frequency of these errors may be explained by the fact that the NLP analyzed text from an entire encounter, including provider, nursing and social worker notes about a patient's history, physical exam, and ED course, as well as each clinician's decision making. As a result, the NLP tool captured cases when providers were examining possibilities and discussing relevant events from the past medical history. It is possible that by limiting the scope of the NLP (ie, to only the final diagnostic impression in a provider's

Table 2. Accuracy of the NLP Algorithm in Detecting High-Risk Injuries Associated With Abuse Among 1344 Infants Presenting for Care to 9 Emergency Departments Over a 3.5-Month Period

	High-Risk Injury Present	High-Risk Injury Absent	Total	Predictive Values
NLP +	38	25	63	PPV * = 60.3%
NLP −	3	1278	1281	NPV * = 99.8%
Total	41	1303	1344	
	Sensitivity=92.70%	Specificity=98.10%		

*PPV indicates positive predictive value; NPV, negative predictive value.
NLP, natural language processing; ED, Emergency Department.

text), we could avoid some of these false positives; however, this could mean possibly missing a larger number of true cases as a result. Among the ambiguity errors, there were some patterns that repeated more than once. For example, NLP interpreted the word parietal multiple times to mean parietal lobe of the brain rather than parietal area of the scalp. This distinction is especially important in this case, because intracranial hematomas are concerning for any infant, but scalp hematomas are considered high-risk injuries only for younger infants. Similarly, a history of "IVH" (intraventricular hemorrhage) was described in present tense language in several cases (eg, "*Patient is a 4-month-old male with IVH"*). Other information in the document can be used to make the meaning of the terms clearer, and it is possible that with additional data, the NLP algorithm could be modified to better detect such distinctions.

Previous work to identify children with injuries considered high-risk of being due to abuse have focused on trigger systems that mainly use structured or discrete choices of documentation. While Berger et al's physical abuse trigger system, using triggers in discrete fields of the EHR, resulted in a high sensitivity and specificity for identifying children with injuries associated with physical abuse, limitations included a potential for an overestimate of the sensitivity of the system (eg, if the patient's findings were not captured within the discrete fields or the patient was not evaluated by the child protection team) and a high rate of over-triggering with a reported positive predictive value of 26.5%.[24] To improve their trigger system, Rosenthal and Berger then used the combination of a child abuse screen, physician orders, discharge diagnoses and NLP within focused sections of a provider's note to identify injuries concerning for child abuse. The authors demonstrated a four-fold increase in the identification of high-risk injuries associated with abuse in children <2 years old in 2 Eds: 10/86 (11.6%) preintervention versus 44 of 156 (28.2%) postintervention, $P = .02$) without changes in compliance with national recommendations about the evaluation of physical abuse. Furthermore, 70% of the over-triggers in their system were due to the NLP algorithm.[22] Future studies examining the effectiveness of advanced NLP-triggered CDS (compared with the current practice of pathways and CDS triggered by traditional mechanisms) to decrease cases of missed abuse are needed.

Among the unintended outcomes of integrating trigger systems with computerized CDS systems is the alert fatigue that occurs when providers feel there are too many positive triggers or alerts. Alert fatigue has been found to be a common and pervasive problem, which often results in the clinicians' desensitization to alerts with clinicians overlooking, developing workarounds, or submitting default responses when alerted by CDS in their clinical workflow.[34,36] To attempt to mitigate the impact of alert fatigue, we focused on decreasing the false-positive rate of injury identification (eg, by removing terms such as erythema or redness as synonymous concepts for bruising) and improving the specificity of our NLP algorithm. Future iterations of the NLP tool may find a work-around to include these words when not associated with terms like dermatitis or viral exanthem. In addition, we continued to make adjustments in the rule-based engines to address precision. For example, in the two encounters with elevated transaminases (a criterion for abdominal injury for the NLP) in the presence of fever, infection was a more likely cause of the abnormality than injury. Based on these findings, the criterion for abdominal injury was further refined to exclude abdominal injury when text had documentation of fever or the presence of vital signs with a temperature greater than or equal to 100.4 Fahrenheit. Considering the balance between sensitivity and specificity and aiming for higher specificity during development may lead to improved compliance of the clinician with the eventual CDS system.

There are 4 main strengths of our study: 1) we focused on infants who are at highest risk of being physically abused, and whose entry into the emergency care system may serve as an opportunity to recognize abuse and to intervene to prevent escalating injury; 2) our NLP algorithm was developed using the entirety of nursing, provider, and social worker notes of all children <1 year of age presenting for various reasons for emergency care in multiple EDs, both pediatric and general; 3) the algorithm was developed to minimize false-positive rates of injury identification in an effort to minimize the risk of alert fatigue; and 4) even if this NLP algorithm that identifies high-risk injuries is under-utilized by front line clinicians, it may serve as a surveillance tool, which, based on institutional ethical standards, may aid in the identification of high-risk injuries that may be retroactively reviewed to detect possible missed cases of abuse.

LIMITATIONS

There are at least 4 limitations of this study. First, development and validation of the NLP algorithm was performed using encounters that presented to one health system in which quality improvement work to improve recognition of and evaluation for suspected abuse was underway and could have influenced what providers included in their notes.[19, 23] The positive predictive value of the algorithm may be decreased in EDs in which subtle high-risk injuries are not documented in unstructured text. However, our study did include children who were evaluated at 3 general EDs in which improvement efforts had not yet started. Additionally, Epic, the shared EHR in the health system, is the leading EHR used by approximately 70% of clinicians across the US, making this work more generalizable. Further development and testing of the algorithm in a different health system and with a different EHR may improve generalizability.

Second, our algorithm was used only to identify high-risk injuries, not to predict abuse. Thus, its utility in identifying abuse needs to be further evaluated. Furthermore, our tool was not developed to identify nonspecific chief complaints, such as irritability or vomiting, that may be concerning for abuse. In the future, NLP may be used to

Table 3. Examples of NLP Errors

Error Category	Error Type	Examples of Text in the ED Document.	Rationale	Count
Identification – NLP made an error in identifying the injury itself, including the wrong type of injury or site of injury.	False Negative	*ALT > 80*	Lab value (liver function test) missed due to novel surrounding wording.	1
	False Positive	*She accidentally caught his toe on a metal piece of the wall, where it is broken.*	NLP flagged this as a broken toe.	5
Linguistic Context – NLP correctly identified the injury, but incorrectly interpreted the surrounding words or formatting, resulting in an injury being flagged even though it was uncertain or in the past.	False Negative	*N/A*		0
	False Positive	*Per mom "I think his thumb is fractured." Patient did briefly have a small burn on her hand, but this is no longer visible.*	Flagged as a thumb fracture. NLP identified this as a current burn injury.	10
Ambiguity – NLP missed or raised a flag based on a technically correct interpretation of ambiguous text, but that interpretation was less favorable or unlikely given the full document or clinical context.	False Negative	*. . .concerning for ecchymosis to... chest wall. . . Clinical Impressions... Ecchymosis*	First notation of ecchymosis contains a site reference but is marked as only a concern. Second notation is more certain, but does not mention the site. NLP could not be certain these two notations in separate areas of the document referred to the same injury and did not raise a flag.	2
	False Positive	*Left parietal hematoma.*	NLP interpreted 'parietal' as area of the brain, rather than area of scalp.	7
Clinical Interpretation – NLP correctly identified an injury according to the algorithm, but the expert reviewer did not consider it to be a high-risk injury.	False Negative	*N/A*		0
	False Positive	*Traumatic ulcer of oral mucosa.*	NLP correctly identified traumatic injury of mouth, but it was not considered a high-risk injury.	3

NLP indicates natural language processing; ED, Emergency Department.

detect additional patterns in the clinical, past medical and psychosocial histories to predict abuse even within single encounters. Third, while the positive predictive value of our NLP algorithm was only moderate, the prevalence of high-risk injuries in infants in EDs in our study was quite low, which would ultimately result in a very low number of absolute false-positive triggers per individual provider. Finally, the validity of our NLP tool may be diminished if the ED provider's documentation in the EHR is inaccurate or subject to observational error (ie, the full skin exam is not done) or biased judgment (ie, the full skin exam is only done in selected patient populations).

In summary, an NLP algorithm to identify children <1-year-old with injuries associated with a high risk of abuse has good accuracy and could be used to aid frontline ED providers in the identification of suspected child abuse. It may also provide the technological foundation for surveillance of high-risk injuries to retrospectively identify children at high-risk of being abused.

ACKNOWLEDGMENTS

Financial statement: This work was supported by grants to Dr. Tiyyagura from the Doris Duke Fund to Retain Clinical Scientists, ID 2015216 (GT) and UL1TR001863 (EDS) from the National Center for Advancing Translational Science (NCATS), a component of the National Institutes of Health (NIH). The contents of this manuscript are solely the responsibility of the authors and do not necessarily represent the official view of NIH.

Authorship statement: Dr. Tiyyagura conceptualized and designed the study, interpreted the analyses, drafted the initial manuscript, and critically revised the manuscript. Drs. Asnes, Hsiao and Leventhal conceptualized and designed the study and reviewed and revised the manuscript critically for important intellectual content. Ms. Kutryb, Sevin, and Woodly and Mr. McClelland and Gibson helped design the study, carried out data analysis and reviewed and revised the manuscript critically for important intellectual content. Dr. Teng carried out data analysis and participated in review and interpretation of the data and revised the manuscript critically for important intellectual content. Drs. Shapiro, Powers, Auerbach, Polzin and Lindberg and Ms. Shah helped design the study, interpret the data and reviewed and revised the manuscript critically for important intellectual content. All authors approved the final manuscript as submitted and agree to be accountable for all aspects of the work.

Previous presentations: 2021 Platform Presentation at Virtual Helfer Child Abuse Society, Poster Presentation at Virtual Pediatric Academic Societies.

REFERENCES

1. Child Maltreatment 2019. In: Administration on Children, Children's Bureau (2021), ed: US Department of Health and Human Services; 2019. https://www.acf.hhs.gov/cb/data-research/child-maltreatment.
2. King WK, Kiesel EL, Simon HK. Child abuse fatalities: are we missing opportunities for intervention? *Pediatr Emerg Care*. 2006;22:211–214.
3. Jenny C, Hymel KP, Ritzen A, et al. Analysis of missed cases of abusive head trauma. *JAMA*. 1999;281:621–626.
4. Thorpe EL, Zuckerbraun NS, Wolford JE, et al. Missed opportunities to diagnose child physical abuse. *Pediatr Emerg Care*. 2014;30:771–776.
5. Sheets LK, Leach ME, Koszewski IJ, et al. Sentinel injuries in infants evaluated for child physical abuse. *Pediatrics*. 2013;131:701–707.
6. Ravichandiran N, Schuh S, Bejuk M, et al. Delayed identification of pediatric abuse-related fractures. *Pediatrics*. 2010;125:60–66.
7. Gausche-Hill M, Ely M, Schmuhl P, et al. A national assessment of pediatric readiness of emergency departments. *JAMA Pediatr*. 2015;169:527–534.
8. Trokel M, Waddimba A, Griffith J, et al. Variation in the diagnosis of child abuse in severely injured infants. *Pediatrics*. 2006;117:722–728.
9. Tiyyagura G, Leventhal JM, Auerbach M, et al. Improving the care of abused children presenting to community emergency departments: the evolving landscape. *Acad Pediatr*. 2021;21:221–222.
10. Christian CW, Committee On Child A, Neglect. The evaluation of suspected child physical abuse. *Pediatrics*. 2015;135:e1337–e1354.
11. Lindberg DM, Beaty B, Juarez-Colunga E, et al. Testing for abuse in children with sentinel injuries. *Pediatrics*. 2015;136:831–838.
12. Wood JN, Fakeye O, Feudtner C, et al. Development of guidelines for skeletal survey in young children with fractures. *Pediatrics*. 2014;134:45–53.
13. Wood JN, Fakeye O, Mondestin V, et al. Development of hospital-based guidelines for skeletal survey in young children with bruises. *Pediatrics*. 2015;135:e312–e320.
14. Louwers EC, Korfage IJ, Affourtit MJ, et al. Accuracy of a screening instrument to identify potential child abuse in emergency departments. *Child Abuse Negl*. 2014;38:1275–1281.
15. Hymel KP, Laskey AL, Crowell KR, et al. Racial and ethnic disparities and bias in the evaluation and reporting of abusive head trauma. *J Pediatr*. 2018;198:137–143.
16. Wood JN, Feudtner C, Medina SP, et al. Variation in occult injury screening for children with suspected abuse in selected US children's hospitals. *Pediatrics*. 2012;130:853–860.
17. Rumball-Smith J, Fromkin J, Rosenthal B, et al. Implementation of routine electronic health record-based child abuse screening in general Emergency Departments. *Child Abuse Negl*. 2018;85:58–67.
18. Louwers EC, Korfage IJ, Affourtit MJ, et al. Effects of systematic screening and detection of child abuse in emergency departments. *Pediatrics*. 2012;130:457–464.
19. Powers E, Tiyyagura G, Asnes AG, et al. Early involvement of the Child Protection Team in the care of injured infants in a Pediatric Emergency Department. *J Emerg Med*. 2019.
20. Rangel E, Cook B, Bennett B, et al. Eliminating disparity in evaluation for abuse in infants with head injury: use of a screening guideline. *J Pediatr Surg*. 2009;44:1229–1235.
21. Stavas N, Paine C, Song L, et al. Impact of child abuse clinical pathways on skeletal survey performance in high risk infants. *Acad Pediatr*. 2020;20:39–45.
22. Rosenthal B, Skrbin J, Fromkin J, et al. Integration of physical abuse clinical decision support at 2 general emergency departments. *J Am Med Inform Assoc*. 2019;26:1020–1029.
23. Tiyyagura GK EB, Gaither JR, Bechtel K, et al. Child Protection Team Consultation for injuries potentially due to child abuse in Community Emergency Departments. *Acad Emerg Med*. 2020.
24. Berger RP, Saladino RA, Fromkin J, et al. Development of an electronic medical record-based child physical abuse alert system. *J Am Med Inform Assoc*. 2018;25:142–149.
25. Lee S, Mohr NM, Street WN, et al. Machine learning in relation to emergency medicine clinical and operational scenarios: an overview. *West J Emerg Med*. 2019;20:219–227.
26. Doan S, Maehara CK, Chaparro JD, et al. Building a natural language processing tool to identify patients with high clinical suspicion for Kawasaki disease from Emergency Department Notes. *Acad Emerg Med*. 2016;23:628–636.
27. Kaur H, Sohn S, Wi CI, et al. Automated chart review utilizing natural language processing algorithm for asthma predictive index. *BMC Pulm Med*. 2018;18:34.
28. Castro VM, Dligach D, Finan S, et al. Large-scale identification of patients with cerebral aneurysms using natural language processing. *Neurology*. 2017;88:164–168.
29. Woolf SM, Leventhal JM, Gaither JR, et al. Oral injuries in children less than 24 months of age in a pediatric emergency department. *Child Abuse Negl*. 2019;89:70–77.
30. Kemp AM, Hollén L, Emond AM, et al. Raising suspicion of maltreatment from burns: derivation and validation of the BuRN-Tool. *Burns*. 2018;44:335–343.

31. Flaherty EG, Perez-Rossello JM, Levine MA, et al. Evaluating children with fractures for child physical abuse. *Pediatrics*. 2014;133: e477–e489.
32. Christian CW, Committee on Child Abuse and Neglect AAoP. The evaluation of suspected child physical abuse. *Pediatrics*. 2015;135: e1337–e1354.
33. Pierce MC, Kaczor K, Lorenz DJ, et al. Validation of a clinical decision rule to predict abuse in young children based on bruising characteristics. *JAMA Netw Open*. 2021;4: e215832.
34. Ash JS, Sittig DF, Campbell EM, et al. Some unintended consequences of clinical decision support systems. *AMIA Annu Symp Proc*. 2007:26–30.
35. Press A, Khan S, McCullagh L, et al. Avoiding alert fatigue in pulmonary embolism decision support: a new method to examine 'trigger rates'. *Evid Based Med*. 2016;21:203–207.
36. Ash JS, Sittig DF, Poon EG, et al. The extent and importance of unintended consequences related to computerized provider order entry. *J Am Med Inform Assoc*. 2007;14:415–423.