Risk assessment for intra-abdominal injury following blunt trauma in children: Derivation and validation of a machine learning model

Christopher Pennell, MD, Conner Polet, MD, L. Grier Arthur, MD, Harsh Grewal, MD, *and* **Stephen Aronoff, MD**, *Philadelphia, Pennsylvania*

BACKGROUND: Computed tomography is the criterion standard for diagnosing intra-abdominal injury (IAI) but is expensive and risks radiation exposure. The Pediatric Emergency Care Applied Research Network (PECARN) model identifies children at low risk of IAI requiring intervention (IAI-I) in whom computed tomography may be omitted but does not provide an individualized risk assessment to positively predict IAI-I. We sought to apply machine learning algorithms to the PECARN blunt abdominal trauma (BAT) data set experimentally to create models for predicting both the presence and absence of IAI-I for pediatric BAT victims.

METHODS: Using the PECARN data set, we derived and validated predictive models for IAI-I. The data set was divided into derivation (n = 7,940) and validation (n = 4,089) subsets. Six algorithms were tested to create 2 models using 19 clinical variables including emesis, dyspnea, Glasgow Coma Scale score of <15, visible thoracic or abdominal trauma, seatbelt sign, abdominal distension, tenderness or rectal bleeding, peritoneal signs, absent bowel sounds, flank pain, pelvic pain or instability, sex, age, heart rate, and respiratory rate (RR). Five algorithms were fitted to predict the absence (low-risk model) or presence (high-risk model) of IAI-I. Models were validated using the test subset.

RESULTS: For the low-risk model, four algorithms were significantly better than the baseline rate (2.28%) when validated using the test set. The random forest model identified 73% of children as low risk, having a predicted IAI-I rate of 0.54%. For the high-risk model, all six algorithms had added predictive power compared with the baseline rate with the highest reportable risk being 39.0%. By incorporating both models into a web application, child-specific risks of IAI-I can be estimated ranging from 0.28% to 39.0%

CONCLUSION: We developed a tool that provides a child-specific risk estimate for IAI-I after BAT. This publically available model provides a powerful tool for clinicians triaging pediatric victims of blunt abdominal trauma. (*J Trauma Acute Care Surg.* 2020;89: 153–159. Copyright © 2020 Wolters Kluwer Health, Inc. All rights reserved.)

LEVEL OF EVIDENCE: Prognostic, Level II.

KEY WORDS: Pediatric trauma; blunt abdominal trauma; PECARN; intra-abdominal injury; machine learning.

With more than 7 million visits annually in the United States, trauma is the second most common reason that children seek medical care in an emergency department.[1,2] Approximately 9,000 children die as a result of traumatic injuries every year, three times more than any other cause of childhood death beyond infancy.[2] Blunt mechanisms such as falls, motor vehicle collisions, and trauma by falling objects account for 70% of childhood injuries.[3,4] Among pediatric victims of blunt trauma, about 6% suffer intra-abdominal injuries (IAIs) with about one third of these injuries requiring acute intervention (IAI requiring intervention [IAI-I]).[5]

Submitted: November 23, 2019, Revised: March 9, 2020, Accepted: March 26, 2020, Published online: April 8, 2020.

From the Department of Pediatric General, Thoracic, and Minimally Invasive Surgery (C. Pennell, L.G.A., H.G.), St. Christopher's Hospital for Children, Philadelphia, Pennsylvania; Department of Psychiatry (C. Polet), SUNY Downstate Medical Center, Brooklyn, New York; Department of Pediatrics (S.A.), Lewis Katz School of Medicine at Temple University; Section of Pediatric Infectious Diseases (S.A.), St. Christopher's Hospital for Children; and College of Medicine (L.G.A., H.G.), Drexel University, Philadelphia, Pennsylvania.

This study was presented at the 33rd Annual Meeting of the Eastern Association for the Surgery of Trauma, January 14–18, in Orlando, Florida.

Address for reprints: Christopher Pennell, MD, Department of Pediatric General, Thoracic, and Minimally Invasive Surgery, St. Christopher's Hospital for Children, 160 E Erie Ave, Philadelphia, PA 19134; email: Christopher.Pennell@towerhealth.org.

DOI: 10.1097/TA.0000000000002717

Computed tomography is the criterion standard for diagnosis of blunt abdominal injuries, and its use has been increasing.[6] Because the vast majority of children do not have IAIs that require intervention, clinicians are faced with the challenge of identifying those likely to have an injury without unnecessarily irradiating children at lowest risk, a significant issue given exponentially increasing health care costs and the risk of radiation-induced malignancy.[7] The Pediatric Emergency Care Applied Research Network (PECARN) has established criteria for identifying children at low risk for IAI-I in whom imaging can be safely omitted.[5] This prediction rule is helpful but has important limitations. First, it was derived using a single machine learning algorithm, a significant limitation given the hundreds of available algorithms for modeling. Second, the model was not validated on an independent set of patients, raising concerns about the accuracy of the model because of overfitting. Third, the model was not tested against the *a priori* or naive rate of IAI-I to demonstrate added value and improved predictive power. Finally, only a negative predictive model for identifying low-risk children was developed, which does not stratify children at increased risk for IAI-I.

The purpose of this study was to apply machine learning algorithms to the PECARN blunt abdominal trauma data set experimentally to derive and validate new models for predicting IAI-I in children with blunt abdominal trauma. Our goal was to create both a low-risk model that identifies a cohort of children in whom CT imaging could be avoided and a high-risk model to accurately risk stratify those children at increased risk for IAI-I. We sought to provide clinicians with a powerful tool by combining these two models in a web-based application that provides real-time, patient-specific estimates of risk for IAI-I.

PATIENTS AND METHODS

Data Sources and Participants

A publicly available, prospectively collected data set from PECARN was used for this study. The data set included children presenting to 20 participating PECARN emergency departments between May 2007 and January 2010 after suffering blunt abdominal trauma. Detailed inclusion criteria and demographics have previously been published.[5] This study was exempt from institutional review board approval because the data in use are publicly available and deidentified. Results are reported according to the Transparent Reporting of multivariable prediction model for Individual Prognosis or Diagnosis guidelines.[8]

Predictors

To produce a useful tool to guide diagnostic evaluation, all variables used in the prediction model were clinical variables readily assessed at the time of initial physical examination. No laboratory or imaging variables were used as predictors. A summary of the predictors is listed in Table 1.

Outcomes

The primary outcome was IAI requiring acute intervention. Intra-abdominal injury was defined as any injury to the spleen, liver, genitourinary tract, gastrointestinal tract, pancreas, gallbladder, adrenal gland, intra-abdominal vascular structure, or a traumatic abdominal wall hernia identified radiographically or at the time of surgery. The definition of IAI-I was adopted from the original PECARN study and included injuries that resulted in death, required therapeutic angiography or laparotomy, blood transfusion, or admission to the hospital for two or more nights to receive intravenous fluids deemed necessary as a result of the IAI-I.

Database Preparation and Model Development

All of the relational data sets were downloaded from the PECARN website in .csv format. All data analyses and database manipulations were performed in the R environment.

TABLE 1. Predictor Variables Included in the Models

Predictors	
Age	Abdominal distension
Sex	Absent bowel sounds
Heart rate	Abdominal tenderness
Respiratory rate	Peritoneal signs
Glasgow Coma Scale <15	Visible thoracic trauma
Dyspnea	Flank pain
Emesis	Pelvic pain
Visible abdominal trauma	Unstable pelvis
Seatbelt sign	Occult rectal blood
Abdominal pain	

Following review of the codebook, individual data files were selected and variables were formatted using the dplyr package.[9] Variables were segregated into two groups: predictors and outcomes. Using the outcome criteria outlined previously, a single binary variable of IAI-I (present or absent) was developed. The single outcome variable was merged with the predictors to generate a final data set. Only complete cases were used for the study. The rate of IAI-I in the final data set was 2.28%.

The partitioning function from the caret package randomly created the training and test subsets in a 2:1 ratio from the data set of complete cases.[10] The Mann-Whitney *U* test and the χ^2 test were used to compare the outcomes and predictors in the training and test sets in a univariate manner; *p* values of <0.05 were considered significant. Since the outcome was severely unbalanced in the complete case data set because of the overall low rate of IAI-I, synthetic minority oversampling (Synthetic Minority Oversampling Technique [SMOTE] package) was used to balance the positive outcome rate in the training set.[11] Briefly, this technique creates synthetic cases of IAI-I by oversampling the minority outcome (IAI-I present) and undersampling the majority outcome (IAI-I absent). This improves classification when receiver operator curves and its components are used as the loss function for the predictive models. The balanced training set was used to develop the machine learning models. The test set was not altered and was used for experimental testing among the final optimized algorithms to determine which of them had the greatest predictive power.

Model Creation

Modeling was performed using the caret package as a wrapper. Each algorithm was optimized using fivefold cross validation. The loss function was the true-positive rate (sensitivity) from the receiver operating curve (pROC package).[12] Two models were developed: (1) a low-risk model optimized for negative prediction (positivity defined as the absence of IAI-I) and (2) a high-risk model optimized for positive prediction (positivity defined as the presence of IAI-I). Six machine learning algorithms were studied for each model. (1) Generalized linear modeling[13] uses classical multivariate logistic regression with least weighted squares. (2) Linear discriminant analysis[14] produces a linear decision boundary similar to that produced by logistic regression but assumes a normal distribution of predictors. It optimizes differences between classes. (3) Recursive partition analysis[15] is a single tree-based decision model that uses a greedy algorithm to develop multiple decision points (nodes) and endpoints (leaves) based on nodal purity. (4) Random forest (RF)[16] is an ensemble model that uses multiple decision trees with nodes generated by the random selection of subgroups of predictors using a greedy algorithm and. (5) Support vector machines[17] are parametric-based classifiers that develop dimensional decision boundaries that may be linear or nonlinear in shape; two different boundary shapes were used: support vector machines—quadratic (SVM) and support vector machines—radial (SVMR).

The training procedure consisted of recursive cycles of parameter fitting and tuning. For each algorithm, training terminated when successive sensitivity rates were comparable (convergence). Algorithms were trained separately for the low-risk and high-risk models.

Model Assessment (Experimentation)

For each model, the fitted version of each algorithm was assessed experimentally to determine which had the greatest predictive power. Two methods were used to compare the fitted algorithms.[18] First, the area under the receiver operator curve (ROC) and the sensitivity of each fitted algorithm were compared for the low-risk and high-risk models separately via resampling of the training set. Second, outcome predictions were made by each fitted algorithm using the previously unseen test set for both the low-risk and high-risk models. Since the output of these algorithms is the probability of either the presence or absence of IAI-I, depending on the model, 0.5 was used as the standardized cutoff for each algorithm. From the predictions, the number of correctly and incorrectly classified subjects was determined, and statistical differences among the algorithms were assessed using the χ^2 test. The test was applied sequentially to determine statistical differences among the algorithms' predicted rates and the baseline rate of IAI-I in the test data set (2.28%) with p values of <0.05 considered significant. Algorithms generating predicted rates of IAI-I that were significantly different than the baseline rate in the study population (2.28%) were considered to have added predictive value.

RESULTS

Database

The raw data set had 13,360 subjects. After 1,331 patients with missing data were removed, a final data set of 12,029 complete cases remained. The mean age of this group was 10.0 years (interquartile range, 6–15 years). The most common mechanisms of injury were motor vehicle collisions (32%) and pedestrians or bicyclists struck by a moving vehicle (19%). The training subset consisted of 7,940 patients; the test subset contained 4,089 patients. The rate of IAI-I was 2.28% in each group. Sex was the only predictor that differed in distribution between the training and the test subsets (males, 63% vs. 60%, respectively; $p = 0.011$).

In the training group, 181 subjects had IAI-I, while 7,759 did not. Because data sets with unequal numbers of the outcome of interest perform poorly in machine learning algorithms, we

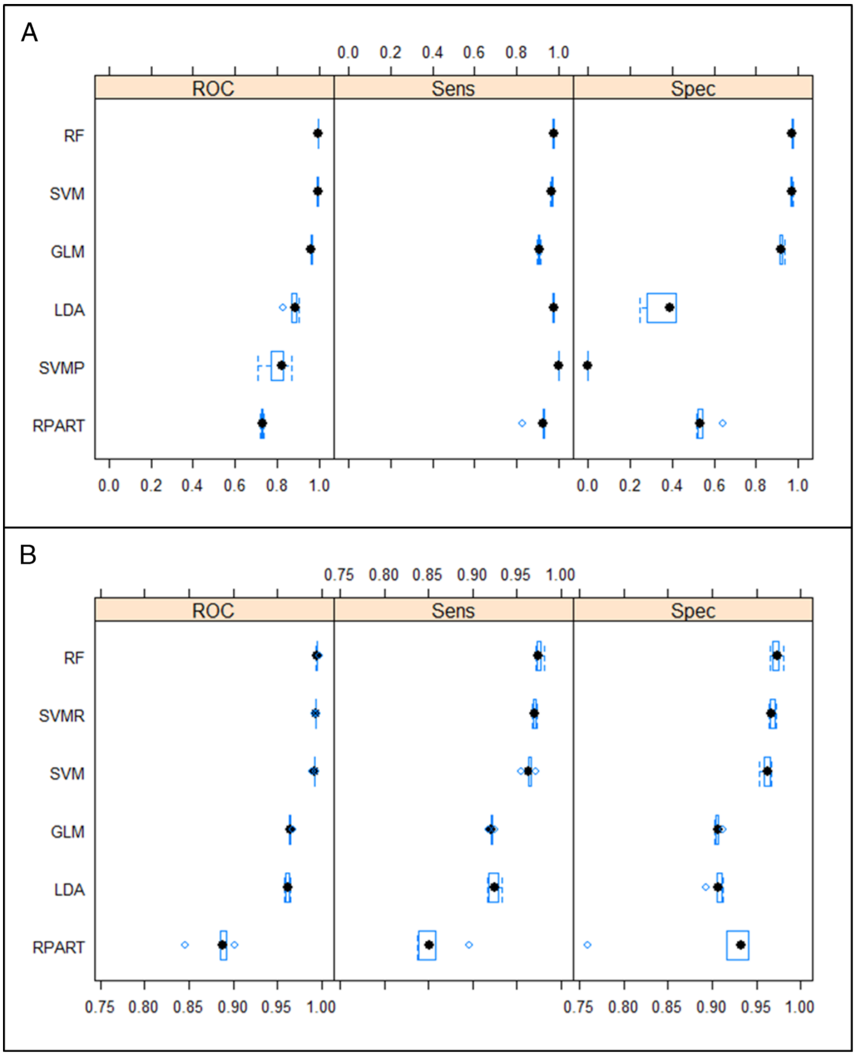

Figure 1. Comparison of six fitted algorithms to predict risk for IAI-I. A: Low-risk model, B: High-risk model.

TABLE 2. Performance of Six Fitted Algorithms to Predict Low Risk for IAI-I

Model	n	True Negative	False Negative	False-Negative Rate	*p* Vs. Naive	*p* Vs. SVMR
Naive	4,089	3,996	93	0.023	Ref	—
SVM	4,089	3,996	93	0.023	1	—
SVMR	3,860	3,799	61	0.016	0.031	Ref
LDA	3,956	3,895	61	0.015	0.021	0.96
RF	3,836	3,779	57	0.015	0.013	0.81
GLM	3,631	3,592	39	0.011	0.000071	0.071
RPART	3,631	3,592	39	0.011	0.000071	0.071

LDA, linear discriminant analysis; GLM, generalized linear modeling; RPART, recursive partition analysis.

TABLE 3. Performance of Six Fitted Algorithms to Predict High Risk for IAI-I

Model	n	True Positive	False Positive	True-Positive Rate	*p* Vs. Naive	*p* Vs. RPART	*p* Vs. GLM
RF	275	47	228	0.171	<0.00001	0.0002	0.057
SVM	203	31	172	0.153	<0.00001	0.004	0.27
SVMR	236	35	201	0.148	<0.00001	0.004	0.31
GLM	458	54	404	0.118	<0.00001	0.039	Ref
LDA	457	52	405	0.114	<0.00001	0.058	—
RPART	337	24	313	0.071	<0.00001	Ref	—
Naive	4,089	93	3,996	0.023	Ref	—	—

LDA, linear discriminant analysis; GLM, generalized linear modeling; RPART, recursive partition analysis.

created hypothetical patients using the SMOTE package. This produced a balanced data set with more equal numbers of cases (7,421 with and 7,240 without IAI-I, respectively). This balanced data set was used to train both the low-risk and high-risk models.

Low-Risk Model

The predictive performance of the six fitted algorithms was compared by resampling the balanced training set (Fig. 1A). Using the ROC as the metric of interest, the fitted RF, SVM, and generalized linear modeling algorithms were the best predictors of the absence of IAI-I (mean ROC, 0.99, 0.99, and 0.96, respectively). The sensitivity of the linear discriminant analysis model for identifying children without IAI-I was highest at 0.98. The RF and SVM algorithms were also highly sensitive with sensitivities of 0.97 each. When the final fitted algorithms were assessed using the validation data set, all models except SVM had superior predictive power for identifying low-risk children compared with the naive rate (Table 2). With the removal of the naive and SVM algorithms, the predictive results from the remaining five algorithms were not significantly different.

A calibration curve for the low-risk model (the relationship between the probability that a child has no injury and the observed rate of injury) using the fitted RF algorithm is shown in Figure 2. Probabilities of 0.75 or greater were associated with a risk of less than 1%, and probabilities of 0.90 or greater were associated with risks less than 0.7%. Defining low-risk children as those with a ≥0.95 probability that no injury is present classifies 73% (2,994 of 4,089 patients) of the patients as low risk. The corresponding rate of IAI-I in this group is 0.54%.

High-Risk Model

The performance of the six fitted algorithms for predicting the presence of IAI-I by resampling the derivation set is shown in Figure 1B. By both ROC and sensitivity, the fitted RF, SVMR, and SVM algorithms performed the best. Using the validation data set, Table 3 compares the performance of the six algorithms fitted to predict the presence of IAI-I using a cutoff probability of 0.5. All of the fitted algorithms outperformed the naive

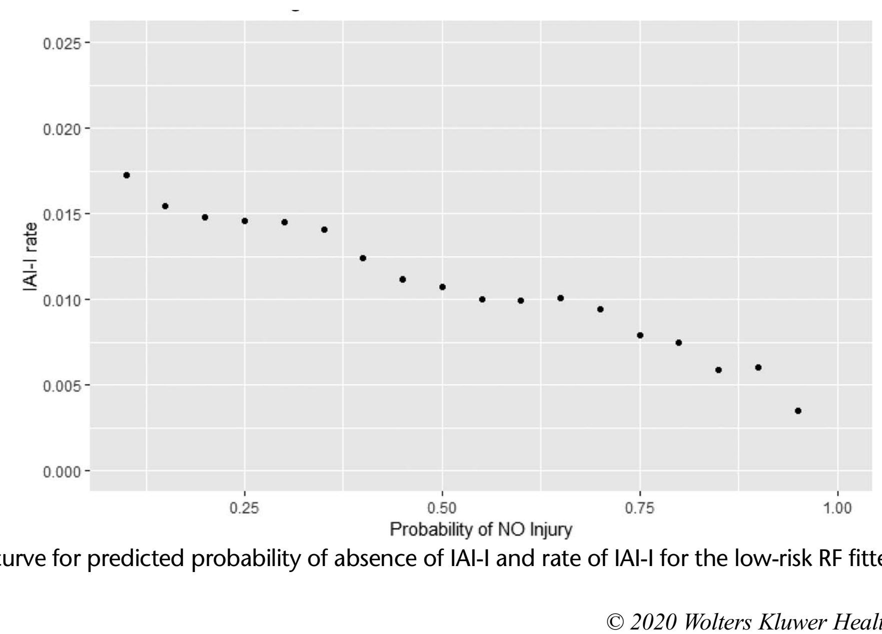

Figure 2. Calibration curve for predicted probability of absence of IAI-I and rate of IAI-I for the low-risk RF fitted algorithm.

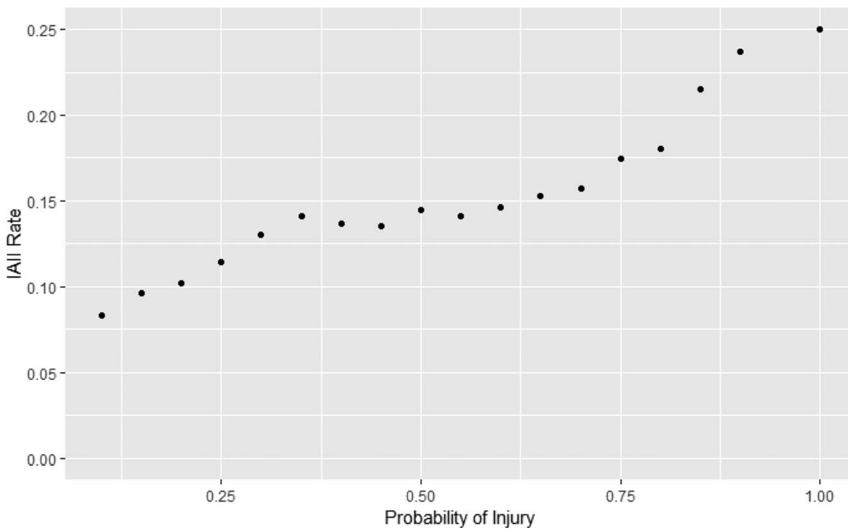

Figure 3. Calibration curve for predicted probability of presence of IAI-I and rate of IAI-I for the low-risk RF fitted algorithm.

model. Three algorithms (RF, SVM, and SVMR) had significantly higher rates of IAI-I than the others. The relationship between the predicted probability of IAI-I and the observed rate of IAI-I for the high-risk fitted RF algorithm is shown in Figure 3. Probabilities of ≥0.85 were associated with IAI-I rates greater than 16%, and probabilities greater than 0.95 were associated with IAI-I rates as high as 20%. The highest rate of IAI-I that can be predicted by the high-risk model is 39.0%.

Web Application

To meaningfully apply these models in clinical practice, a prototype web application was developed (https://www.stchristophershospital.com/SitePages/Our-Services/Trauma-Center.aspx). Patient-specific variables (Table 1) are entered into the application at the time of initial assessment, and the child's risk of IAI-I is estimated using both the low- and high-risk models (Figs. 2 and 3). Because of the low rate of IAI-I in the data set, two models, optimized using different metrics, were needed. The low-risk model accurately identifies children without injuries and reports risks as low as 0.28%, while the high-risk model accurately identifies children at increased risk of injury and reports risks as high as 39.0%. The application has a built in toggle switch; the rate estimated by the low-risk model is reported if it is predicted to be <1%. If it is ≥1%, the rate estimated by the high-risk model is reported. As a result, the web application can report a patient-specific estimated risk of IAI-I ranging from 0.28% to 39.0%.

DISCUSSION

Using a large patient data set and machine learning technology, we trained and validated a model that predicts both the presence and absence of IAI-I in children after blunt abdominal trauma. Incorporation of these models into a publically available web application provides clinicians with a patient-specific estimate of risk that can inform diagnostic and management decisions. Compared with existing prediction tools, this risk assessment tool identifies nearly three fourths of children as low risk for injury and provides patient-specific estimates of risk for all patients, including those not meeting existing low-risk criteria.

Clinical decision rules have been developed to limit unnecessary CT scanning in children with a low likelihood of clinically important injuries for both head and abdominal trauma.[5,19,20] The creation of these rules followed a dramatic increase in the use of CT imaging along with greater recognition of the risks of radiation in children.[6] For abdominal trauma, the current PECARN rule uses seven physical examination findings to identify children at low risk for IAI-I.[5] Similarly, the Pediatric Surgery Research Consortium (PedSRC) model uses five physical, radiographic, and laboratory findings to define a low-risk population.[19] Authors of both models recommend no imaging be performed in these low-risk children. The publication of these prediction rules is an important step toward responsible use of abdominal CT in pediatric trauma; however, both rules have important limitations to their clinical application as well as their derivation and validation.

The major limitation to the clinical application of the PECARN and PedSRC models is that they are binary, classifying children as low-risk or not low-risk for IAI-I. They are designed to identify children in whom the risk of injury is sufficiently low that CT imaging is not necessary; this population includes only 34% to 42% of children experiencing blunt abdominal trauma.[5,19] By comparison, defining low-risk children as those with a probability of no injury ≥0.95, our model classified 73% of children as low risk with a predicted rate of IAI-I of 0.54%.

An additional limitation of the PECARN and PedSRC models is that they do not provide granular estimates of risk for children who do not meet low-risk criteria. Rather, they provide grouped estimates based on the number of risk factors present. In the PECARN model, children with one risk factor have a predicted IAI-I rate of 1.4%, while those with three risk factors have a risk of 4.5%. The problem with these grouped estimates is that they assume each variable portends an equal risk of injury, and they do not account for interactions between variables. For example, a child with vomiting and thoracic wall trauma may not have the same risk as a child with abdominal tenderness and a seatbelt sign. Both, however, screen positive for two variables and would therefore be assigned a risk of 1.8% by the PECARN rule. Using our model, an 8-year-old boy with vomiting and thoracic wall trauma would have a risk of only 0.29%, while the risk for the same child with abdominal tenderness and a seatbelt sign would be 0.91%. The former patient meets our low-risk criteria, while the latter is not considered low risk. Furthermore, in both cases, the predicted risk is considerably different than the PECARN models, which significantly changes the risk-benefit discussion that should be had with parents.

Modern machine learning is based on a strict experimental methodology consisting of a three-step process.[18] Initially, a number of different computer algorithms are selected based on an exploratory analysis of data. Next, the metric of interest is determined, each algorithm is optimally fitted to the training data set, and error is measured. Finally, the model with the greatest predictive power is selected by comparing all the fitted models to the baseline rate of IAI-I in the test set using standard statistical methods. The algorithm with the most statistically superior outcomes is selected as the model with the best predictive power.

The methodology by which PECARN was developed is limited in two ways. First, only a single computer algorithm was explored (single decision tree), despite the fact that hundreds of potential algorithms exist. Second, because it was only validated using the training data set, it is at risk of being overfit, meaning it can accurately predict outcomes in the training data set, but not on new, unseen data.[21] Models validated in this way are at risk of overfitting because conceptually they are being asked to make predictions about data they have already learned. We used a conventional methodology to avoid overfitting by validating our models on new and unseen data, which had been held out as a test subset from the complete data set before training the algorithms. In contrast, the only published validation study for the PECARN prediction rule is a retrospective, single-center review of children with IAI-I.[22]

Our study has important limitations that warrant discussion. The number of predictive variables in our model is greater than those required by existing decision rules. Moreover, the algorithm is complex and not summarized in a simple flowchart, so it is not possible to memorize and make a bedside decision after considering a small number of easily recalled variables. The included variables, however, are all routinely documented by a scribe nurse as a part of a standard trauma evaluation, so transcribing them into the web application adds only minimally to the current workflow.

Attempts to reduce the number of predictors during the derivation process resulted in models with significantly lower predictive power. Using all 19 variables provides optimal accuracy, but the web application does allow clinicians to record “unknown” for any of the dichotomous variables. On average, predictors were documented as unknown in 9.5% of cases in the PECARN data set, ranging from 0% for presence of thoracic wall trauma to 52% for the presence of occult rectal blood. Therefore, our model incorporates this uncertainty, and its accuracy holds up in spite of a significant number of unknown values.

Unlike other models that state children should not undergo CT imaging if they meet low-risk criteria, our prediction tool is not prescriptive with regard to how clinicians should use the information. Rather, it is designed to provide an individualized probability of injury for use when counseling parents and making diagnostic and therapeutic decisions for children after blunt abdominal trauma. Whether or not a clinician feels it is safe to forgo CT imaging will depend on many factors including hospital location, availability of expert consultants and imaging resources, ability to serially monitor patients, radiation dosing for abdominal CTs at the treating hospital, and individual tolerance of risk.

CONCLUSIONS

We have created and validated a novel clinical decision tool using modern machine learning techniques to estimate the risk of IAI-I in children. Our publicly available web application fills a gap in the existing literature by providing risk estimates for all children, not just those deemed low risk by traditional decision rules. It is a powerful tool that has the potential to facilitate more responsible use of CT for children with blunt abdominal trauma.

AUTHORSHIP

C. Pennell, C. Polet, and S.A. contributed in the study concept and design. C. Polet and S.A. contributed in the data acquisition. C. Pennell, C. Polet, L.G.A., H.G., and S.A. contributed in the analysis and data interpretation. C. Pennell and S.A. contributed in the drafting of the article. C. Pennell, C. Polet, L.G.A., H.G., and S.A. contributed in the critical revision.

ACKNOWLEDGMENTS

We thank the PECARN for allowing and supporting our use of the Identifying Children at Very Low Risk of Clinically Important Blunt Abdominal Injuries data set.
This article was prepared using the Identifying Children at Very Low Risk of Clinically Important Blunt Abdominal Injuries data set obtained from the Data Coordinating Center, University of Utah School of Medicine, and does not necessarily reflect the opinions or views of the study investigators or the Health Resources Services Administration/Maternal Child Health Bureau/Emergency Medical Services for Children. The PECARN was funded by the Health Resources Services Administration/Maternal Child Health Bureau/Emergency Medical Services for Children.

DISCLOSURE

The authors declare no conflicts of interest.

REFERENCES

1. McDermott KW, Stocks C, Freeman WJ. Overview of Pediatric Emergency Department Visits, 2015. HCUP Statistical Brief #242. August 2018. Agency for Healthcare Research and Quality, Rockville, MD. Available at: www.hcup-us.ahrq.gov/reports/statsbriefs/sb242-Pediatric-ED-Visits-2015.pdf. Accessed November 5, 2019.
2. Centers for Disease Control and Prevention. WISQeARS (Web-based Injury Statistics Query and Reporting System)|Injury Center|CDC. Published 2019. Available at: www.cdc.gov/injury/wisqars. Accessed November 5, 2019.
3. Villaveces A, Mutter R, Owens PL, Barrett ML. Causes of Injuries Treated in the Emergency Department, 2010. HCUP Statistical Brief #156. May 2013 *Health Cost Util Proj Stat Briefs*. Agency for Healthcare Research and Quality, Rockville, MD. Available at: www.hcup-us.ahrq.gov/reports/statsbriefs/sb156.pdf. Accessed November 5, 2019.
4. Chang MC, Stewart RM, Rotondo MF, Nathens AB. National Trauma Data Bank Pediatric Report 2016. *Am Coll Surg*. 2016;128. Available at: www.facs.org/~/media/files/quality programs/trauma/ntdb/ntdb pediatric annual report 2016.ashx. Accessed November 5, 2019.
5. 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.

6. Miglioretti DL, Johnson E, Williams A, et al. The use of computed tomography in pediatrics and the associated radiation exposure and estimated cancer risk. *JAMA Pediatr*. 2013;167(8):700–707.
7. Brenner DJ, Hall EJ. Computed tomography — an increasing source of radiation exposure. *N Engl J Med*. 2007;357(22):2277–2284.
8. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *Br J Surg*. 2015;102(3):148–158.
9. CRAN. Package Dplyr: A Grammar of Data Manipulation. Available at: https://cran.r-project.org/web/packages/dplyr/index.html. Accessed November 5, 2019.
10. Kuhn M. Building predictive models in R using the caret package. *J Stat Softw*. 2008;28(5):1–26.
11. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: synthetic minority over-sampling technique. *J Artif Intell Res*. 2002;16:321–357.
12. Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez JC, Müller M. pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics*. 2011;12:77.
13. Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. *J Stat Softw*. 2010;33(1):1–22.
14. Hastie T, Tibshirani R, Friedman J. Linear methods for classification. In: *The Elements of Statistical Learning*. New York, NY: Springer; 2009:101–135.
15. Therneau TM, Atkinson EJ. Technical Report Series No. 61: An Introduction to Recursive Partitioning Using the RPART Routines. Department of Health Sciences Research, Mayo Clinic, Rochester, Minn; 1997.
16. Breiman L. Random forests. *Mach Learn*. 2001;45(1):5–32.
17. Karatzoglou A, Meyer D, Hornik K. Support vector machines in R. *J Stat Softw*. 2006;15(9):1–28.
18. Hothorn T, Leisch F, Zeileis A, Hornick K. The design and analysis of benchmark experiments. *J Comput Graph Stat*. 2005;14(3):675–699.
19. Streck CJ, Vogel AM, Zhang J, et al. Identifying children at very low risk for blunt intra-abdominal injury in whom CT of the abdomen can be avoided safely. *J Am Coll Surg*. 2017;224(4):449–458.e3.
20. 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.
21. Wu H, Shapiro JL. Does overfitting affect performance in estimation of distribution algorithms. In: *Proceedings of the 8th Annual Conference on Genetic and Evolutionary Computation*. Seattle, WA; July 8–12, 2006:433–434.
22. Springer E, Frazier SB, Arnold DH, Vukovic AA. External validation of a clinical prediction rule for very low risk pediatric blunt abdominal trauma. *Am J Emerg Med*. 2019;37(9):1643–1648.

