APPRAISE-HRI: AN ARTIFICIAL INTELLIGENCE ALGORITHM FOR TRIAGE OF HEMORRHAGE CASUALTIES

Jonathan D. Stallings,* Srinivas Laxminarayan,†‡ Chenggang Yu,†‡ Adam Kapela,†‡ Andrew Frock,†‡ Andrew P. Cap,* Andrew T. Reisner,§ and Jaques Reifman†

*US Army Institute of Surgical Research, Fort Sam Houston, Texas; †Department of Defense Biotechnology High Performance Computing Software Applications Institute, Telemedicine and Advanced Technology Research Center, US Army Medical Research and Development Command, Fort Detrick, Maryland; ‡The Henry M. Jackson Foundation for the Advancement of Military Medicine, Inc., Bethesda, Maryland; and §Department of Emergency Medicine, Massachusetts General Hospital, Boston, Massachusetts

ABSTRACT—**Background:** Hemorrhage remains the leading cause of death on the battlefield. This study aims to assess the ability of an artificial intelligence triage algorithm to automatically analyze vital-sign data and stratify hemorrhage risk in trauma patients. **Methods:** Here, we developed the APPRAISE–Hemorrhage Risk Index (HRI) algorithm, which uses three routinely measured vital signs (heart rate and diastolic and systolic blood pressures) to identify trauma patients at greatest risk of hemorrhage. The algorithm preprocesses the vital signs to discard unreliable data, analyzes reliable data using an artificial intelligence–based linear regression model, and stratifies hemorrhage risk into low (HRI:I), average (HRI:II), and high (HRI:III). **Results:** To train and test the algorithm, we used 540 h of continuous vital-sign data collected from 1,659 trauma patients in prehospital and hospital (i.e., emergency department) settings. We defined hemorrhage cases (n = 198) as those patients who received ≥1 unit of packed red blood cells within 24 h of hospital admission *and* had documented hemorrhagic injuries. The APPRAISE-HRI stratification yielded a hemorrhage likelihood ratio (95% confidence interval) of 0.28 (0.13–0.43) for HRI:I, 1.00 (0.85–1.15) for HRI:II, and 5.75 (3.57–7.93) for HRI:III, suggesting that patients categorized in the low-risk (high-risk) category were at least 3-fold less (more) likely to have hemorrhage than those in the average trauma population. We obtained similar results in a cross-validation analysis. **Conclusions:** The APPRAISE-HRI algorithm provides a new capability to evaluate routine vital signs and alert medics to specific casualties who have the highest risk of hemorrhage, to optimize decision-making for triage, treatment, and evacuation.

KEYWORDS—Artificial intelligence; hemorrhage; trauma; triage algorithm

INTRODUCTION

Hemorrhage is recognized as the leading cause of preventable death in both civilian and military trauma (1–5). On the battlefield, 91% of potentially survivable deaths are attributed to hemorrhage (4), including 67% truncal, 19% junctional, and 14% extremity. Recent civilian trials demonstrated that the median time to death is 2 to 3 h after presentation of exsanguination (6–10). When hemorrhage is visible, providers initiate damage-control resuscitation to rapidly control bleeding, systematically correct major physiological derangements (e.g., hypoperfusion, shock, and coagulopathy), and notify trauma teams in advance of hospital arrival (6,11). When hemorrhage is less obvious, for example, because of internal bleeding, decisions related to triage, treatment, evacuation, and advanced activation of a trauma team may be delayed, resulting in suboptimal outcomes.

Data from the recent conflicts in Iraq and Afghanistan suggest that perhaps 6% to 24% of trauma casualties with hemorrhage who died before reaching a military medical treatment facility could have potentially survived if they had been quickly identified and treated (4,12). In the current Russian invasion of Ukraine as well as future conflicts in which air superiority, freedom of movement, and communication may be degraded or compromised, military medical capabilities are expected to be overwhelmed by high casualty numbers, and deaths from hemorrhage are expected to rise. Therefore, medical personnel need triage capabilities that support quick casualty assessment and identification of those most in need of immediate treatment, to optimize resource utilization (e.g., blood products) and adjust evacuation decisions based on available medical assets and capacity in both the prehospital and hospital settings.

Holcomb *et al.* previously demonstrated the value of manually assessed vital signs to predict the need for life-saving interventions in trauma patients (13). However, providing medics with advanced decision-support capabilities to detect and quantify intravascular volume loss and end points of resuscitation remains desirable (14) for trauma patient care in both civilian and military settings. For example, an automated computer algorithm that processes available data through artificial intelligence (AI) algorithms could augment the capability of medics in far-forward environments by providing accurate, data-driven information through predictive analytics

about the current state of the casualty and emerging risks. The ideal AI decision-support algorithm would quickly reveal nuanced gradations in casualty state through real-time fusion of routine vital-sign data, which are readily available throughout the echelons of care. In this scenario, decision support could optimize triage, evacuation decisions, and life-saving interventions, leading to improved clinical outcomes.

Over the years, our US Army team investigated methods to improve the use of routine vital signs for triage of trauma casualties by using novel AI and pattern-recognition algorithms. We examined adult trauma patients (age ≥18 years) in an emergency department (ED) setting and during helicopter transport to level I trauma centers and assessed physiological data sets consisting of physiological waveforms, such as electrocardiogram, and discrete vital signs, such as heart rate (HR), systolic blood pressure (SBP), and diastolic blood pressure (DBP), as well as demographic information and clinical outcomes (15–26). Our earlier studies focused on developing methods to identify and quantify reliable vital signs (15,16), demonstrating their ability to improve automated diagnosis of trauma patients (17–19) and determining the most predictive vital signs to diagnose uncontrolled bleeding (20,21). We also examined different approaches for analysis of continuous vital signs through time by investigating whether temporal trends in the data would improve hemorrhage diagnosis (22) and whether accumulation of evidence over time through simple vital-sign averaging would be beneficial compared with more sophisticated approaches, such as the sequential probability ratio test (23). Our analysis of new sensor modalities beyond routine vital signs, such as muscle oxygen saturation (24) and heart rate variability (25), indicated that by themselves, these modalities are not more predictive of hemorrhage than HR and blood pressures (BPs).

Previously, in terms of end point measures, we predicted hypovolemia (19,20) and the need for blood transfusion, for example, one or more units of packed red blood cells (PRBCs) and massive blood transfusion (26). In these investigations, we performed both univariate and multivariable regression analysis (15–26) and developed a platform for real-time collection of vital-sign data in the prehospital environment (27). In prior efforts to develop the Automated Processing of the Physiological Registry for Assessment of Injury Severity (APPRAISE) algorithm, we used an ensemble classifier, consisting of a set of 25 univariate and multivariable regression models whose numerical outputs are averaged to yield one binary output, to indicate whether trauma patients received one or more units of PRBCs within 24 h of hospital admission (26). This prior study, which involved analysis of both prospectively and retrospectively collected vital-sign data during helicopter transport to level I trauma centers, demonstrated that the APPRAISE algorithm could provide notification for patients at risk for hemorrhage within the first 10 min of transport and more than 20 min before arrival at a trauma center (26).

Here, we report our subsequent progress and the development of the APPRAISE-Hemorrhage Risk Index (HRI) algorithm, which is a streamlined version of our previous AI algorithms geared toward low echelons of care (e.g., roles 1 and 2). Role 1 is the first medical care military personnel receive, with major emphasis on return to duty or measures to stabilize the casualty to allow for evacuation to the next appropriate role of care, while role 2 provides advanced trauma management and emergency medical treatment, including

damage control surgery (28). The new algorithm only requires HR and BP measurements, maps these inputs to clinical outcomes (control or hemorrhage cases) using one multivariable linear regression model, and stratifies hemorrhage risk (the updated end point of the algorithm) by providing an index (I, II, or III) reflecting the likelihood of hemorrhage from low (I) to high (III). We hypothesize that the APPRAISE-HRI algorithm can stratify low and high hemorrhage risk with a minimum 2-fold change from the average risk in a trauma population.

METHODS

Study population

We combined data from three independent studies comprising more than 1,900 h of vital-sign data from 2,688 trauma patients to retrospectively train (i.e., develop) and test (i.e., challenge) the APPRAISE-HRI algorithm (Table 1). Briefly, we collected demographic, vital-sign, and clinical data from a convenience sample of adult (≥18 years) trauma patients who were either transported by air emergency medical service to participating level I trauma centers at (1) Memorial Hermann Life Flight (MHLF; protocol # 02101004, Houston, TX), previously described by Cooke *et al.* (29) and Holcomb *et al.* (13), and (2) Boston MedFlight (BMF; protocol # 2008P002042, Bedford, MA), previously described by Liu *et al.* (26), or who (3) presented to the ED of the Massachusetts General Hospital (MGH; protocol # 2011P002735, Boston, MA), previously described by Reisner *et al.* (24). For all data sets, we analyzed patients with at least two recorded nonzero SBP measurements and one set of HR measurements and excluded those who died before hospital admission (e.g., in the ED or en route) because resuscitation with PRBCs within 24 h was a criterion used to classify hemorrhage cases. All study protocols, including informed consent waivers, were reviewed and approved by the appropriate institutional review boards and by the Army Human Research Protections Office and were carried out in accordance with the ethical principles and guidelines applicable to human subjects protection regulations.

Definition of hemorrhage and control cases

Although there is no single gold-standard definition of hemorrhage, there are multiple clinical outcomes of importance related to patients with uncontrolled bleeding. As previously reported (24), here we defined a hemorrhage outcome as a trauma patient who received ≥1 unit of PRBCs within 24 h of hospital admission and had documented hemorrhagic injuries. Control cases included patients who did not receive blood or patients who received ≥1 unit of PRBCs within 24 h but had no documented hemorrhagic injuries. We also assumed that the status of each patient (hemorrhage or control) did not change during the data collection period. To automatically identify patients with documented hemorrhagic injuries, we defined terms and codes consistent with such injuries and searched the patients' records, after a research nurse had extracted patient data from the trauma registry. Specifically, we defined documented hemorrhagic injuries as recorded hemorrhage-control procedures (e.g., packing or suture of an artery) or injuries consistent with a hemorrhage outcome (e.g., major lacerations to internal organs or vessels or hemothorax) in the

TABLE 1. **Study population characteristics from three distinct studies, with data collected either during helicopter transport to a hospital (MHLF and BMF) or in a hospital (MGH)**

Demographic	MHLF	BMF	MGH	Total
Trauma patients, *n*	898	247	1,543	2,688
Patients with reliable vital signs, *n*	672	218	769	1,659
Hemorrhage patients, *n*	78	36	84	198
Control patients, *n*	594	182	685	1,461
Mean vital-sign length, min	23 (14)	19 (9)	21 (12)	21 (13)
Men, *n*	500	160	548	1,208
Women, *n*	172	58	221	451
Mean age, y	38 (15)	45 (20)	49 (21)	44 (20)
Method of injury				
Blunt, *n*	602	193	652	1,447
Penetrating, *n*	70	25	117	212

Values within parentheses denote one standard deviation.
BMF, Boston MedFlight; MGH, Massachusetts General Hospital; MHLF, Memorial Hermann Life Flight.

patient's case summary (182 positive terms [e.g., suture artery and amputation above knee] and 5 negative terms [e.g., not hemorrhagic]) or based on Abbreviated Injury Scale (AIS) codes (363 defined codes) or *International Classification of Disease, Ninth Revision* (ICD-9) codes (52 defined codes). Hence, our definitions of hemorrhage and control cases were systematically obtained and reproducible.

The APPRAISE-HRI algorithm

Figure 1 shows the three modules that together comprised the APPRAISE-HRI algorithm, which associated vital signs to the hemorrhage risk, the algorithm's end point. The inputs to the algorithm consisted of continuous vital-sign data, every 1 min, from each patient in the form of HR, SBP, and DBP from standard US Food and Drug Administration–cleared travel monitors, and the output was the corresponding risk level (i.e., HRI) for hemorrhage, which ranged from low (I) to average (II) to high (III). The first module preprocessed the three vital signs at every 1 min to ensure that (1) the vital signs were within the prescribed physiological range (i.e., spurious events were removed), (2) the data were timely and varying, (3) data variability was not excessive, (4) a minimum number of reliable measurements for each vital sign were available within a time window, and (5) physiological data variability was smoothed out (i.e., median HR and average BPs). We deemed the vital signs that met these criteria as "reliable" and continued to process them to the next module; otherwise, we waited and preprocessed vital signs for the next minute. The second module consisted of a multivariable linear regression model, which used HR, BP, and pulse pressure (PP; the difference between SBP and DBP) as inputs and generated an output, with large positive values indicating the highest likelihood of hemorrhage. The third module provided hemorrhage risk stratification, where we determined the thresholds between levels I to II and between levels II to III such that the fraction of patients in HRI-level II equaled 60% (arbitrarily selected) of our study population and the prevalence of hemorrhage cases equaled that of our study population.

APPRAISE-HRI training and testing

For the three data sets used to train and test the APPRAISE-HRI algorithm, we acquired HR every 1 to 2 s and BPs at multiminute intervals. Nevertheless, to train the algorithm, we only used the first reliable set of inputs (one mean value of SBP and PP, and one median value of HR) to make one decision (i.e., one output) per patient. Using the preprocessing module described previously, to obtain the first reliable set of SBP and PP inputs, we computed the mean value of the first two consecutive reliable BP measurements, with the requirement that the first SBP/DBP pair was collected up to 20 min before the current decision time and the second SBP/DBP pair up to 10 min before the current decision time. For the HR input, we computed the median of reliable values over the most recent minute from the current decision time, with the requirement that the most recent minute was within 4 min of the current time and contained at least 25% of reliable measurements. The algorithm repeated this procedure every minute and produced one output when the first reliable set of inputs met the requirements described previously. Accordingly, the algorithm makes a decision as fast as 2 min (if BP measurements were available every minute) up to N minutes from the first BP measurement (with $N < 20$).

When placed into operation, the APPRAISE-HRI algorithm will be used to continuously monitor a patient, providing a new HRI output every minute. Therefore, the algorithm was built to assess whether input data were available every minute and provide a new input to the algorithm whenever a new pair of reliable BP measurements and one new reliable HR measurement were available, following the same processing and requirements used in developing the algorithm described above. We tested the algorithm by analyzing (1) only the first output associated with the first reliable input and (2) a time series of outputs associated with the corresponding series of reliable inputs, to assess the stability of the results over time.

Statistical analysis

We summarized the data using mean and standard deviation (SD) or median and interquartile for continuous variables and count for categorical variables. To compare demographic and clinical data between the three studies, we used χ^2 tests for categorical variables and one-way ANOVA tests for continuous variables, which is recommended when there are three or more comparisons. Unless noted otherwise, we assessed the performance of the APPRAISE-HRI algorithm using one HRI output associated with the first reliable set of input vital-sign data. To assess the stability of the outputs over time, we compared the output corresponding to the first reliable input against the subsequent outputs, by computing the fraction of HRI outputs that changed by one risk level (e.g., from I to II) and by two risk levels (e.g., from III to I).

We assessed the algorithm by performing two separate analyses. First, we used the entire data set to train and test the algorithm, which allowed us to determine how well the model fit the data. Second, we performed a cross-validation using 100 realizations, where for each realization, we split the data into 60% for training and 40% for independently testing the algorithm, while fixing the incidence of hemorrhage in each partition of the data to preserve the incidence in the overall study population. For each of these two analyses, we quantified the ability of the APPRAISE-HRI algorithm to stratify hemorrhage by computing the hemorrhage risk, defined as the ratio of the number of hemorrhage cases to the number of total cases (hemorrhage and control) in a risk level, and its associated 95% confidence interval (CI), as well as the likelihood ratio (LR), defined as the ratio of the fraction of hemorrhage cases over the fraction of control cases in a risk level, and its associated 95% CI. For the cross-validation analysis, we reported the algorithm performance solely based on the testing data, which were not used to train the algorithm.

RESULTS

Of the 2,688 patients, we excluded 986 who did not meet the inclusion criteria, lacked a nonzero BP measurement, or had missing vital-sign or clinical data, and 43 who did not survive to hospital admission (please see enrollment flowcharts in Figs. S1–S3 in Supplemental Digital Content, http://links.lww.com/SHK/B716), resulting in 1,659 evaluable patients with 540 h of continuous reliable vital signs, of which 1,461 were control cases (88% negative) and 198 were hemorrhage cases (12% positive; the prevalence of hemorrhage cases in our study population). Table 1 describes the evaluable trauma study population from the following three independent studies: MHLF, BMF, and MGH. The distributions of sex (men/women) and method of injury were similar across all three data sets. Trauma patients in the MHLF data set, however, were younger (38 [SD =15] years old, $P < 0.05$) compared with the BMF and MGH data sets. Although the BMF data set had the highest relative percent of hemorrhage cases (16.5%), it only accounted for 18% (n = 36) of the total hemorrhage cases, as it was the study with the fewest
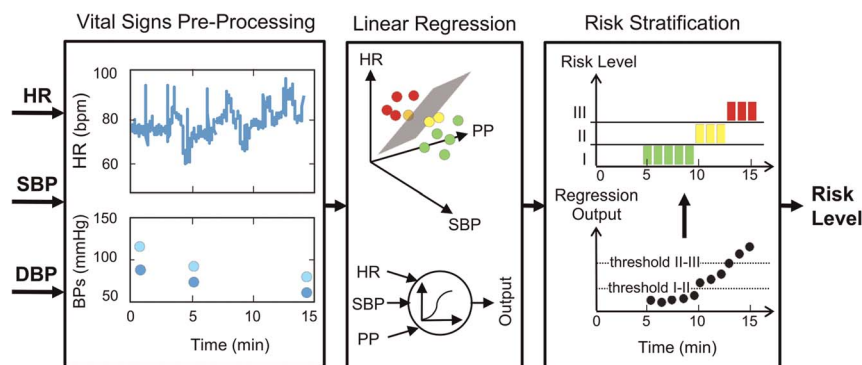

FIG. 1. The APPRAISE-HRI algorithm receives three vital signs as inputs (HR, SBP, and DBP) and provides a risk level as its output, through three software modules. The vital-signs preprocessing module discards unreliable physiological data and computes the PP (the difference between SBP and DBP). Reliable vital signs are provided to a multivariable linear regression model in the second module, which produces an output corresponding to the likelihood of hemorrhage. Finally, the last module maps the regression model output into three risk levels, stratifying hemorrhage risk into low (I), average (II), and high (III). DBP, diastolic blood pressure; HR, heart rate; PP, pulse pressure; SBP, systolic blood pressure.

TABLE 2. **Performance of the APPRAISE-HRI algorithm, using the first reliable set of inputs for each of the 1,659 trauma patients (198 hemorrhage cases and 1,461 control cases) to train the algorithm**

HRI level	Hemorrhage, n	Control, n	Total, n	Hemorrhage risk*	Likelihood ratio†
I	19	507	526	0.04 (0.02–0.06)	0.28 (0.13–0.43)
II	119	877	996	0.12 (0.10–0.14)	1.00 (0.85–1.15)
III	60	77	137	0.44 (0.35–0.53)	5.75 (3.57–7.93)
Total	198	1,461	1,659		

Values within parentheses denote 95% CIs.
*Number of hemorrhage cases / number of total cases, in a risk level.
†Fraction of hemorrhage cases (number of hemorrhage cases / number of total hemorrhage cases) over fraction of control cases (number of control cases / number of total control cases), in a risk level.
HRI, hemorrhage risk index.

patients with reliable vital signs (n = 218). Across all three studies, available Injury Severity Scores were higher in hemorrhage patients (22 [SD = 22], n = 197) compared with controls (13 [SD = 13], n = 1,096), suggesting that our definition of hemorrhage identified more severely injured patients.

In our first analysis, we used all 1,659 trauma patients (198 hemorrhage and 1,461 controls) to train (i.e., to fit) the APPRAISE-HRI algorithm, including the determination of the coefficients of the multivariable regression model and the thresholds for risk-level stratification between levels I to II and between levels II to III. Table 2 shows the number of hemorrhage and control cases categorized by the algorithm in each of the three risk levels, along with the computed hemorrhage risk and LR. Importantly, not only were the hemorrhage risks and LRs at least 3-fold different between risk levels, but the 95% CIs between levels did not overlap, indicating that the algorithm was capable of providing excellent risk stratification. We observed similar results for hemorrhage risk.

Based on the algorithm's categorization of hemorrhage and control cases in each of the three risk levels, we computed the median value of the three vital signs (HR, SBP, and PP) as well as the units of transfused PRBCs in 24 h for each HRI (Table 3, Fig. 2). Consistent with the trends described in the American College of Surgeons Bulletin (30), HR and PRBC values increased with increasing HRI levels while SBP and PP decreased. When compared to HRI-level I, HRI-level III subjects demonstrated elevated HR (+50 beats per minute), decreased SBP (−56 mm Hg), and decreased PP (−37 mm Hg). Reassuringly, on average, the required resuscitation (units of PRBCs in 24 h) also increased with HRI level.

Importantly, to assess the performance of the APPRAISE-HRI algorithm over time, for each patient, we compared the first HRI decision against subsequent decisions made every time a new reliable set of vital signs was available. Across the 1,659 patients, the decision changed by one HRI level for 431 patients (69 hemorrhage) and by two levels for 6 patients (2 hemorrhage). In total, we performed 5,127 such comparisons, where in nearly 80% of the comparisons (4,012) the HRI level never changed relative to the first decision. Reassuringly, in only 0.2% of the comparisons (10), the decisions changed from HRI I to III or from HRI III to I.

In our second analysis, to assess whether the algorithm overfitted the training data, we performed a cross-validation analysis based on 100 realizations, where in each realization we first randomly partitioned the data into 60% of the patients for training the algorithm and the remaining 40% for independent testing, while preserving the prevalence of hemorrhage (12%) in the underlying study population (Table 4). We then trained the multivariable regression algorithm and computed the risk-level thresholds of the stratification module (Fig. 1). Lastly, we assessed the algorithm's performance by processing the testing patients through the trained algorithm, counting the number of patients categorized into each of the three risk levels, and computing the hemorrhage risk and the LR. We repeated this procedure 100 times and reported the mean and SD over the realizations. Table 4 shows that the computed statistics are very similar to those obtained when we used all the data to fit the algorithm (Table 2).

DISCUSSION

Here, we demonstrated that the APPRAISE-HRI algorithm preprocesses vital-sign data and fuses the input into an HRI output that, consistent with our hypothesis, stratifies the risk for

TABLE 3. **Vital signs and blood transfusion median (interquartile) values for each of the three risk levels based on the APPRAISE-HRI algorithm's categorization of hemorrhage and control cases**

Assessment	HRI level I	HRI level II	HRI level III
Total patients/hemorrhage, n	526/19	996/119	137/60
Hemorrhage risk*	0.04 (0.03–0.06)	0.12 (0.10–0.14)	0.43 (0.35–0.50)
Likelihood ratio†	0.28 (0.16–0.45)	1.00 (0.78–1.28)	5.74 (3.90–8.44)
HR,‡ bpm	76 (66–87)	94 (82–109)	126 (110–140)
SBP,‡ mm Hg	158 (146–174)	129 (118–142)	102 (85–115)
PP,‡ mm Hg	75 (67–86)	54 (47–62)	38 (32–44)
24-h PRBCs,§ U	2 (1–4)	3 (2–6)	4 (3–9)

*Number of hemorrhage cases / number of total cases, in a risk level. Values within parentheses denote 95% CIs.
†Fraction of hemorrhage cases (number of hemorrhage cases / number of total hemorrhage cases) over fraction of control cases (number of control cases / number of total control cases), in a risk level. Values within parentheses denote 95% CIs.
‡Based on the entire record of each patient.
§For patients who received ≥1 unit of PRBCs.
HR, heart rate; HRI, hemorrhage risk index; PP, pulse pressure; PRBC, packed red blood cell; SBP, systolic blood pressure.
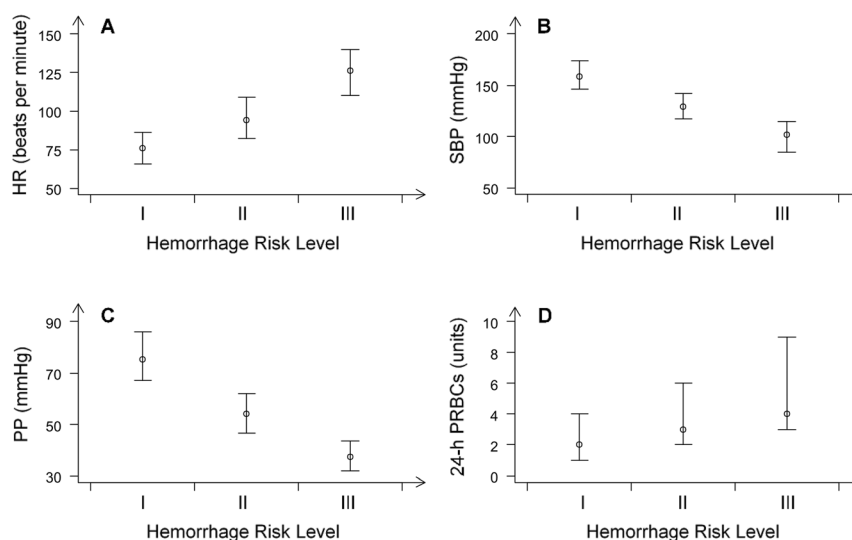
FIG. 2. **Vital signs and blood transfusion (median and interquartile values) associated with each of the three risk levels of the APPRAISE-HRI algorithm's categorization of hemorrhage and control cases.** A, Heart rate. B, Systolic blood pressure. C, Pulse pressure. D, Units of transfused PRBCs for patients who received ≥1 unit of PRBCs. Plots for HR, SBP, and PP are based on the entire record of each patient. HR, heart rate; PP, pulse pressure; PRBC, packed red blood cell; SBP, systolic blood pressure.

hemorrhage outcomes into the following three distinct risk levels: HRI-level I indicates low risk of hemorrhage (at least 2-fold lower than the prevalence in our trauma population), HRI-level II indicates average risk of hemorrhage (akin to the prevalence in our trauma population), and HRI-level III indicates high risk of hemorrhage (at least 2-fold higher). Accordingly, these APPRAISE alarms can be used as additional information for medics who could be inexperienced, tired, distracted, or simply overwhelmed by a large number of casualties. The APPRAISE-HRI algorithm is intended as an adjunct to other physical vital-sign parameters and clinical judgment and is not intended to direct decisions independently. Because the three trauma patient data sets used to train and test the algorithm do not have the granularity to unambiguously identify the patients who had subtle signs of distress from those who did not, or to clearly identify patients who had visible bleeding from those who did not, the end point of the algorithm is to detect hemorrhage risk, regardless of whether bleeding is visible. Our results suggest that the algorithm will have utility in the military operational environment.

Similar to military trauma casualties, the population in this study primarily included males (73% *vs.* approximately 85% in the military), with an average age (<50 years old) within the range of US active-duty service members (Table 1). In addition, we collected 54% of our trauma-casualty data (MHLF and BMF) in the prehospital setting during helicopter transport to a hospital, which is the environment closest to battlefield conditions, allowing us to test the ruggedness of our algorithm and its ability to stratify

trauma patients with vital signs collected in a challenging, unstructured environment.

Our study population consisted of patients with both blunt and penetrating injury. While the distribution of cases with blunt and penetrating trauma in civilian and military settings is not expected to be the same (there is a larger fraction of blunt trauma in the civilian setting as compared to the battlefield), we expect that our algorithm will be able to stratify the risk of hemorrhage for both types of trauma in either setting. Based on the National Trauma Data Bank of the American College of Surgeons (31) and the US Joint Theater Trauma Registry (32), the rate of penetrating injury in the civilian population is approximately 11%, whereas approximately 70% of battlefield wounds are by penetrating mechanism. Approximately 13% of the trauma cases in our study population were penetrating, matching the National Trauma Data Bank. To demonstrate that the algorithm can stratify hemorrhage for both types of trauma, we separated the blunt and penetrating cases in our study population into two data sets and separately retrained the algorithm using each data set, to obtain hemorrhage risk and LR results (Tables S1 and S2 in Supplemental Digital Content, http://links.lww.com/SHK/B716) very similar to those in Table 2, where we used the entire data set. We also obtained similar results when we stratified the patients in our study population by gender (female and male; Tables S3 and S4 in Supplemental Digital Content, http://links.lww.com/SHK/B716) and by age (<35 and ≥35 years old; Tables S5 and S6 in Supplemental Digital Content, http://links.lww.com/SHK/B716).

TABLE 4. **Cross-validation performance of the APPRAISE-HRI algorithm, using the first reliable set of inputs for each of the 663 patients in the testing set (79 with hemorrhage; for a 12% prevalence rate)**

HRI level	Hemorrhage, n	Control, n	Total, n	Hemorrhage risk	Likelihood ratio
I	8 (2)	197 (17)	205 (17)	0.04 (0.01)	0.29 (0.08)
II	49 (5)	355 (14)	404 (14)	0.12 (0.01)	1.03 (0.12)
III	22 (5)	32 (9)	54 (14)	0.41 (0.05)	5.26 (1.04)
Total	79	584	663		

Entries indicate mean value (one SD) over 100 random realizations using 60% of the study population to train and 40% to independently test the algorithm. HRI, hemorrhage risk index.

In the prehospital setting, vital-sign data vary over time due to changes in the physiological status of the patient as they stabilize, deteriorate, or encounter monitoring equipment failures. To assess the consistency of the APPRAISE-HRI algorithm over time, for each patient, we compared the time series of HRI outputs with the first output and found that, over the 1,659 patients and 540 h of time records, on average, only 20% of the outputs changed to the immediately adjacent risk level, with a negligible fraction of cases (0.2%) changing by two risk levels, from low risk to high risk, or *vice versa*. These results provide sufficient evidence that the algorithm is stable over normal fluctuations of vital signs over time, even inside a moving helicopter.

During mass casualties and triage management, the deployment of mobile vital-sign monitors with embedded APPRAISE software would allow the fusion of real-time physiological inputs and provide overwhelmed caregivers with important notifications about a patient's status and risk of hemorrhage. It is not the intent for the APPRAISE alarm (i.e., output) to decrease the amount of time and resources devoted to casualties who are at a lower risk of hemorrhage (i.e., HRI-level I or II), if these assets are plentiful, but rather to alert the caregiver to immediately focus on patients who are at the greatest risk and have the greatest need of treatment or evacuation (i.e., HRI-level III). Because the alarm is consistent over time, with only a few readings the caregiver could make definitive decisions about evacuation priority for a large number of subjects at a mass casualty checkpoint. Furthermore, because the APPRAISE stratification of trauma patients also correlates with increased need for resuscitation (Table 3, Fig. 2), real-time evacuee status throughout the echelons of care could provide notifications to activate trauma teams, prepare blood products, and ready capabilities related to damage-control surgery. Thus, the APPRAISE-HRI algorithm provides a simple way to maximize the value of continuous vital-sign monitoring by identifying those patients at greatest risk for hemorrhage and in need of resuscitation.

To map vital signs to hemorrhage risk, we used a supervised machine-learning algorithm in the form of a multivariable linear regression model. In previous efforts, we investigated more "sophisticated" AI algorithms (e.g., artificial neural networks and support vector machines (20)) and found that the performance of these algorithms was similar to that of the simpler regression model, hence our decision to use a multivariable linear regression model in our work.

Our study has limitations. There is no single gold- standard definition of hemorrhage; however, we used 363 AIS codes, 52 ICD-9 codes, as well as 182 positive and 5 negative case-summary terms in combination with receipt of ≥1 unit of PRBCs within 24 h of hospital admission to define positive hemorrhage cases. While these codes and terms represent a surrogate indication of hemorrhage (e.g., hemorrhage control procedures consistent with hemorrhagic injury), many of these patients also demonstrated textual indications of estimated blood loss in the case summary. Accordingly, we removed subjects who died before 24 h to ensure the clinical outcomes/documentation were available to classify subjects, but this undoubtedly removed a few subjects who likely died from massive hemorrhage, dismemberment, or traumatic brain injury. However, with 60% of cases falling into the average risk category (the same as the overall trauma population), the model could be viewed as under calling bleeding 10% of the time (the fraction of hemorrhage cases categorized in HRI-level I) and

uninformative 60% of the time. Thus, the model really identifies the risk of bleeding in approximately one third of patients with the most life-threatening vital signs. Our study population included blunt trauma patients with both hemorrhage and traumatic brain injury. In such cases, it is possible that traumatic brain injury masked the value of certain vital signs, such as a decrease in SBP (33), which would otherwise be observed in cases of hemorrhage alone, resulting in a potential misclassification by the APPRAISE-HRI algorithm. As another limitation, before the APPRAISE-HRI algorithm can be used to help triage combat casualties, it must first be cleared by the Food and Drug Administration. We recently started the required regulatory process, where among many other steps, we will validate the algorithm using an independent data set.

In conclusion, the APPRAISE-HRI algorithm provides a simple way to leverage physiological data collected with standard vital-sign monitors to readily assess which trauma patients are at the greatest risk of hemorrhage. Alerts provided en route or at casualty collection points would notify healthcare providers to pay more attention to specific casualties that are at a high risk of hemorrhage, thereby optimizing decision making related to triage, treatment, and evacuation.

ACKNOWLEDGMENTS

REFERENCES

1. Gurney JM, Spinella PC. Blood transfusion management in the severely bleeding military patient. *Curr Opin Anaesthesiol.* 2018;31(2):207–214.
2. Davis JS, Satahoo SS, Butler FK, et al. An analysis of prehospital deaths: who can we save? *J Trauma Acute Care Surg.* 2014;77(2):213–218.
3. Eastridge BJ, Hardin M, Cantrell J, et al. Died of wounds on the battlefield: causation and implications for improving combat casualty care. *J Trauma.* 2011; 71(1 suppl):S4–S8.
4. Eastridge BJ, Mabry RL, Seguin P, et al. Death on the battlefield (2001–2011): implications for the future of combat casualty care. *J Trauma Acute Care Surg.* 2012; 73(6 suppl 5):S431–S437.
5. Sauaia A, Moore FA, Moore EE, et al. Epidemiology of trauma deaths: a reassessment. *J Trauma.* 1995;38(2):185–193.
6. Chang R, Holcomb JB. Optimal fluid therapy for traumatic hemorrhagic shock. *Crit Care Clin.* 2017;33(1):15–36.
7. Holcomb JB, Fox EE, Wade CE, PROMMTT Study Group. The PRospective Observational Multicenter Major Trauma Transfusion (PROMMTT) study. *J Trauma Acute Care Surg.* 2013;75(1 suppl 1):S1–S2.
8. Holcomb JB, del Junco DJ, Fox EE, et al. The PRospective, Observational, Multicenter, Major Trauma Transfusion (PROMMTT) study: comparative effectiveness of a time-varying treatment with competing risks. *JAMA Surg.* 2013;148(2):127–136.
9. Holcomb JB, Donathan DP, Cotton BA, et al. Prehospital transfusion of plasma and red blood cells in trauma patients. *Prehosp Emerg Care.* 2015;19(1):1–9.
10. Holcomb JB, Tilley BC, Baraniuk S, et al. Transfusion of plasma, platelets, and red blood cells in a 1:1:1 vs a 1:1:2 ratio and mortality in patients with severe trauma: the PROPPR randomized clinical trial. *JAMA.* 2015;313(5):471–482.
11. Curry N, Davis PW. What's new in resuscitation strategies for the patient with multiple trauma? *Injury.* 2012;43(7):1021–1028.
12. Mazuchowski EL, Kotwal RS, Janak JC, et al. Mortality review of US Special Operations Command battle-injured fatalities. *J Trauma Acute Care Surg.* 2020;88(5):686–695.
13. Holcomb JB, Salinas J, McManus JM, et al. Manual vital signs reliably predict need for life-saving interventions in trauma patients. *J Trauma.* 2005;59(4):821–829.
14. Convertino VA, Ryan KL, Rickards CA, et al. Physiological and medical monitoring for en route care of combat casualties. *J Trauma.* 2008;64(4 Suppl):S342–S353.
15. Yu C, Liu Z, McKenna T, et al. A method for automatic identification of reliable heart rates calculated from ECG and PPG waveforms. *J Am Med Inform Assoc.* 2006;13(3):309–320.

16. Chen L, McKenna T, Reisner A, et al. Algorithms to qualify respiratory data collected during the transport of trauma patients. *Physiol Meas*. 2006;27(9): 797–816.

17. Reisner AT, Chen L, McKenna TM, et al. Automatically-computed prehospital severity scores are equivalent to scores based on medic documentation. *J Trauma*. 2008;65(4):915–923.

18. Chen L, Reisner AT, Gribok A, et al. Can we improve the clinical utility of respiratory rate as a monitored vital sign? *Shock*. 2009;31(6):574–580.

19. Reisner AT, Chen L, Reifman J. The association between vital signs and major hemorrhagic injury is significantly improved after controlling for sources of measurement variability. *J Crit Care*. 2012;27(5):533.e1–533.e10.

20. Chen L, McKenna TM, Reisner AT, et al. Decision tool for the early diagnosis of trauma patient hypovolemia. *J Biomed Inform*. 2008;41(3):469–478.

21. Reisner AT, Edla S, Liu J, et al. Tachycardic and non-tachycardic responses in trauma patients with haemorrhagic injuries. *Injury*. 2018;49(9):1654–1660.

22. Chen L, Reisner AT, Gribok A, et al. Exploration of prehospital vital sign trends for the prediction of trauma outcomes. *Prehosp Emerg Care*. 2009;13(3):286–294.

23. Chen L, Reisner AT, Chen X, et al. Are standard diagnostic test characteristics sufficient for the assessment of continual patient monitoring? *Med Decis Making*. 2013;33(2):225–234.

24. Reisner AT, Edla S, Liu J, et al. Muscle oxygen saturation improves diagnostic association between initial vital signs and major hemorrhage: a prospective observational study. *Acad Emerg Med*. 2016;23(3):353–357.

25. Edla S, Reisner AT, Liu J, et al. Is heart rate variability better than routine vital signs for prehospital identification of major hemorrhage? *Am J Emerg Med*. 2015;33(2): 254–261.

26. Liu J, Khitrov MY, Gates JD, et al. Automated analysis of vital signs to identify patients with substantial bleeding before hospital arrival: a feasibility study. *Shock*. 2015;43(5):429–436.

27. Reisner AT, Khitrov MY, Chen L, et al. Development and validation of a portable platform for deploying decision-support algorithms in prehospital settings. *Appl Clin Inform*. 2013;4(3):392–402.

28. Joint Publication 4-02, Joint Health Services (11 December 2017). Available at: https://www.jcs.mil/Portals/36/Documents/Doctrine/pubs/jp4_02ch1.pdf. Accessed May 29, 2023.

29. Cooke WH, Salinas J, Convertino VA, et al. Heart rate variability and its association with mortality in prehospital trauma patients. *J Trauma*. 2006;60(2):363–370.

30. Henry S. ATLS 10th edition offers new insights into managing trauma patients. *Bull Am Coll Surg*. 2018;103(6):15–22.

31. National Trauma Data Bank Annual Report 2005. Available at: https://www.facs.org/media/3mjnbmxr/ntdbannualreport2005.pdf. Accessed January 11, 2023.

32. Eastridge BJ, Costanzo G, Jenkins D, et al. Impact of joint theater trauma system initiatives on battlefield injury outcomes. *Am J Surg*. 2009;198(6):852–857.

33. Reisner A, Chen X, Kumar K, et al. Prehospital heart rate and blood pressure increase the positive predictive value of the Glasgow Coma Scale for high-mortality traumatic brain injury. *J Neurotrauma*. 2014;31(10):906–913.