Automated Analysis of Vital Signs Identified Patients with Substantial Bleeding Prior to Hospital Arrival

Jaques Reifman,[1] Jianbo Liu,[1] Maxim Y. Khitrov,[1] Shwetha Edla,[1] Andrew T. Reisner[1,2]
[1]Department of Defense Biotechnology High Performance Computing Software Applications Institute, Telemedicine and Advanced Technology Research Center, U.S. Army Medical Research and Materiel Command, Fort Detrick, MD 21702
UNITED STATES

[2]Department of Emergency Medicine, Massachusetts General Hospital, Boston, MA 02114
UNITED STATES

jaques.reifman.civ@mail.mil; jliu2@bhsai.org; mkhitrov@bhsai.org; sedla@bhsai.org; areisner@partners.org

ABSTRACT

Uncontrolled bleeding is the leading cause of preventable death on the battlefield. For the recent conflicts in Iraq and Afghanistan, it has been reported that as many as 22% of such casualties could potentially survive. Protocols for substantial bleeding, typically activated after the patient's arrival in a hospital, are known to improve trauma outcomes. Early identification of patients with substantial bleeding could facilitate faster implementation of these protocols, thereby improving patient outcomes. Over the last decade, our interdisciplinary research team has been developing technologies to automatically diagnose hemorrhage in trauma casualties, culminating with the first and only deployment of an automated emergency care decision system on board active air ambulances: the APPRAISE system, a hardware/software platform for automated, real-time analysis of vital-sign data. After developing the APPRAISE system using data from trauma patients transported by Memorial Hermann Life Flight (MHLF), we field-tested it on two active Boston MedFlight (BMF) helicopters during emergency transport of adult trauma patients to three Level 1 trauma centers between February 2010 and December 2012. Between the MHLF and BMF populations, we observed that there were significant differences in terms of vital signs as a function of 24-hr blood transfusion requirements. Despite these differences, the APPRAISE system provided consistent determination of whether or not patients were bleeding. We found that the automated APPRAISE system using a multivariate classifier could automatically diagnose casualties in need of massive blood transfusion with 78% sensitivity and 90% specificity within 6-10 min (median time) after the start of transport to a trauma center. In addition to casualty triage and evacuation decision-making, this capability could be useful to expedite preparedness at medical treatment facilities for receiving patients with substantial blood loss.

1.0 INTRODUCTION

In military casualties, early identification of life-threatening bleeding is of singular importance because it is a primary cause of fatality, and yet life-threatening bleeding may be effectively treated when surgery and blood resuscitation are provided sufficiently quickly after injury [1, 2]. Standard field assessment of casualties includes measuring vital signs, which has been criticized as being inadequately sensitive to life-threatening hemorrhage.

Over the past decade, our group has investigated methods for improving the usefulness of routine vital signs using novel pattern-recognition algorithms that could be deployed in field settings with relative minimum expense and new training. In a prior NATO report [3], we summarized our work involving the development of algorithms that automatically identify unreliable vital-sign measurements and perform multivariate pattern-

recognition, while tolerating missing data and data variability through time. In addition, we described the development of a specialized platform for field-testing the algorithms during prehospital operations and performed initial prospective evaluation.

Here, we report our subsequent progress. We compare the performance of the algorithms in a new dataset versus the original dataset used to develop the algorithms (both datasets collected during air transport of civilian trauma casualties) and examine three key investigational questions: *1*) To what degree were there consistent vital-sign patterns associated with life-threatening hemorrhage? *2*) Could an automated algorithm consistently identify life-threatening hemorrhage using only vital-sign data? and *3*) How sensitive would the algorithm's performance be to different methods of temporal analysis?

2.0 VITAL-SIGN PATTERNS ASSOCIATED WITH LIFE-THREATENING HEMORRHAGE

Here, we compare two datasets of vital signs collected during air transport of civilian trauma casualties. The goal is to understand whether there are consistent prehospital patterns that can provide indication of life-threatening hemorrhage.

2.1 Methods: Vital-sign Patterns and Life-threatening Hemorrhage

2.1.1 Setting and Study Population

We examined a convenience sample of adult (≥ 18 years) trauma patients transported by air emergency medical service to several participating Level 1 trauma centers. With Institutional Review Board approval, we collected a prospective dataset from Boston MedFlight (BMF; Bedford, MA) and compared the findings with an archival dataset originally collected from Memorial Hermann Life Flight (MHLF; Houston, TX) by Cooke et al. [4] and Holcomb et al. [5]. In both datasets, we analyzed all subjects with at least one recorded non-zero systolic blood pressure (SBP). Patients who died prior to hospital admission (e.g., in the emergency department) were excluded from analysis, because resuscitation was often terminated before large-volume packed red blood cell (PRBC) transfusion could be completed, regardless of whether or not the patient had significant hypovolemia.

Our primary study outcome was 24-hr PRBC transfusion volume in patients with hemorrhagic injury, defined as a documented hemorrhagic injury that unequivocally caused some loss of blood volume (i.e., laceration or fracture of a solid organ, thoracic or intraperitoneal hematoma, vascular injury that required operative repair, or limb amputation). We excluded patients who received PRBCs, but lacked a documented hemorrhagic injury from the primary analysis. In a secondary analysis, we studied all patients who received PRBC transfusion regardless of injury.

2.1.2 Vital-sign Data Processing

For the prospective cohort, we deployed the APPRAISE system (Automated Processing of the Physiological Registry for Assessment of Injury Severity [6]; see Figure 1) onto two active BMF helicopters between February 5, 2010, and December 31, 2012. The APPRAISE system consists of a Propaq 206 patient monitor (Welch-Allyn, Beaverton, OR) networked to a GoBook ultra-compact ruggedized personal computer (General Dynamics Itronix, Sunrise, FL) running analytic algorithms developed for this research project [6]. The APPRAISE software *1*) created an electronic record of the Propaq data, *2*) analyzed the vital-sign data in real time using algorithms described below, and *3*) archived the results. The results of the automated analysis were not visible to

the flight crew so that the investigational system would not affect clinical decision-making (this was a matter of human subject protection for a diagnostic system that had not yet been validated during clinical operation).

Figure 1: The hardware components of the APPRAISE system in a disassembled state. The GoBook personal computer (General Dynamics Itronix, Sunrise, FL) on the right is connected to the Propaq 206 patient monitor (Welch-Allyn, Beaverton, OR) on the left through an RS-232 serial cable. During field operations, the personal computer was affixed to the top surface of the Propaq monitor using nylon strapping and velcro (not pictured).

The retrospective data originally had been collected on board MHLF helicopters between August 2001 and April 2004 using a personal digital assistant networked to a Propaq 206 patient monitor to archive the vital-sign data [4, 5]. Subsequently, those data were uploaded to our data warehousing system [7] and analyzed offline.

We analyzed the prospective and the retrospective Propaq 206 data using the exact same computational methodology, applied to the following independent vital-sign variables: heart rate (HR), respiratory rate (RR), SBP, and pulse pressure (PP; the difference between SBP and diastolic BP). HR and RR were measured continuously by the Propaq 206 monitor via electrocardiography (ECG) and impedance pneumography (IP), respectively. SBP and PP were measured by oscillometry at multi-minute intervals. We used automated algorithms to identify and exclude unreliable vital-sign measurements. The HR and RR reliability algorithms involved analysis of ECG and IP waveforms; this allowed us to discriminate between a clean source signal versus an unreliable segment due to signal artifacts [8, 9]. The SBP and PP reliability algorithms assessed signal quality by *1*) analyzing the relationship between systolic, diastolic, and mean arterial pressures, and *2*) comparing HR as measured by ECG versus HR as measured by oscillometry [10]. These automated algorithms, which have been shown to agree with human experts' opinions [8, 9], can significantly increase the diagnostic value of vital signs by removing spurious measurements [10, 11].

2.1.3 Clinical Outcomes

For the BMF dataset, a research nurse collected patient attributes and outcome data via retrospective chart review of the receiving hospitals' medical records (i.e., Beth Israel Deaconess Medical Center, the Brigham and Women's Hospital, and the Massachusetts General Hospital). We obtained injury severity scores from each hospital's trauma registry. For the MHLF dataset, a chart review was conducted by the original study authors [4, 5].

2.1.4 Statistical Analysis

We computed the median and interquartile ranges of HR, RR, SBP, and PP as a function of 24-hr PRBC volume and, using the Wilcoxon rank-sum test, we tested for differences between BMF and MHLF, and between those with different PRBC transfusion volumes.

2.2 Results: Vital-sign Patterns and Life-threatening Hemorrhage

Of the 999 patients with electronic data available (MHLF: 757, BMF: 242) we excluded 22 who lacked a non-zero blood pressure measurement (MHLF: 20, BMF: 2) and 33 who did not survive to admission (MHLF: 27, BMF 6). Also, there were 89 patients who received 24-hr PRBC transfusion without documented hemorrhagic injuries (MHLF: 64, BMF 25). Table 1 describes the primary study population (MHLF: 646, BMF 209).

Table 1: Study population characteristics.

	Memorial Hermann Life Flight	Boston MedFlight
Population, n	646	209
Sex, male, n (%)	479 (74)	155 (74)
Age, yr, mean (SD)	38 (15)	45 (20)
Blunt, n (%)	577 (89)	188 (90)
Penetrating, n (%)	61 (9)	21 (10)
ISS, median (IQR)	16 (9-34)	16 (9-26)
Interhospital transfer, n (%)	0 (0)	103 (49)
Prehospital airway intubation, n (%)	111 (17)	80 (38)
Prehospital GCS, median (IQR)	15 (13-15)	15 (8-15)
24-hr PRBC vol ≥ 1 unit, n (%)	75 (12)	31 (15)
24-hr PRBC vol ≥ 3 units, n (%)	57 (9)	18 (9)
24-hr PRBC vol ≥ 9 units, n (%)	25 (4)	9 (4)
Survival to discharge, n (%)	608 (94)	191 (91)

GCS: Glasgow coma scale; IQR: interquartile range; ISS: injury severity score; PRBC: packed red blood cell; SD: standard deviation.

Table 2 reports time-averaged prehospital vital signs as a function of 24-hr PRBC transfusion volume. For pooled patients in the two studies with large 24-hr PRBC volumes (≥ 3 units), each of the time-averaged vital

signs—HR, RR, SBP, and PP—were significantly different than for patients with zero 24-hr PRBC volume. Between the two study populations, there were subtle differences in vital signs. In patients with hemorrhage, MHLF patients had higher HR and RR, and also had a trend towards higher SBP, as compared with BMF.

Table 2: Time-averaged prehospital vital signs as a function of subsequent 24-hr PRBC transfusion volume.

		24-hr PRBC volume, units			
		0	1 – 2	3 – 8	≥ 9
Total patients, n	**All**	**749**	**31**	**41**	**34**
	MHLF	571	18	32	25
	BMF	178	13	9	9
HR,bpm	**All**	**90 (78–104)**	**105 (85–116)†**	**97 (87–128)††**	**120 (92–136)†††**
	MHLF	92 (80–105)***	113 (103–117)*	101 (89–133)	122 (94–138)
	BMF	84 (73–99)***	89 (75–105)*	92 (82–101)	93 (89–120)
RR, bpm	**All**	**25 (22–28)**	**27 (23–31)**	**28 (24–35)††**	**28 (24–35)††**
	MHLF	25 (22–29)	29 (25–33)	29 (24–36)	33 (26–38)*
	BMF	24 (22–28)	24 (21–27)	27 (22–29)	26 (24–27)*
SBP, mmHg	**All**	**134 (122–149)**	**118 (112–134)††**	**106 (94–117)†††**	**112 (87–125)†††**
	MHLF	134 (122–148)	117 (104–131)	107 (93–118)	118 (91–125)
	BMF	132 (119–152)	122 (115–141)	102 (97–115)	93 (79–115)
PP, mmHg	**All**	**57 (49–66)**	**51 (42–57)††**	**44 (34–48)†††**	**34 (28–49)†††**
	MHLF	57 (50–66)	46 (41–53)*	42 (35–47)	35 (28–50)
	BMF	58 (48–70)	57 (50–68)*	44 (33–62)	31 (28–41)

Each entry represents median (interquartile range).

Significantly different versus 24-hr PRBC volume = 0: $^{\dagger}p < 0.05$, $^{\dagger\dagger}p < 0.01$, $^{\dagger\dagger\dagger}p < 0.001$ by Wilcoxon rank-sum test.

Significantly different MHLF versus BMF: $^{*}p < 0.05$, $^{***}p < 0.001$ by Wilcoxon rank-sum test.

BMF: Boston MedFlight; HR: heart rate; MHLF: Memorial Hermann Life Flight; PP: pulse pressure (SBP-diastolic blood pressure); PRBC: packed red blood cell; RR: respiratory rate; SBP: systolic blood pressure.

2.3 Discussion: Vital-sign Patterns and Life-threatening Hemorrhage

In both datasets of prehospital trauma casualties, MHLF and BMF, there were significant differences associated with blood transfusion requirement, for every one of the routine vital signs. However, there were also significant differences between the two datasets, which represent different physiological responses to blood loss. Specifically, the patients in the BMF dataset appeared to exhibit less sympathetic compensation: less tachycardia, less tachypnea, and increased pulse pressure, but overall, a trend toward more hypotension. By contrast, the patients in the MHLF dataset appeared to exhibit greater sympathetic compensation: more tachycardia, more tachypnea, and a trend toward less overall hypotension.

The major implication of these findings is that *individual* vital signs have an *inconsistent relationship* with transfusion requirement, which supports the conventional wisdom that individual vital signs may not be reliable indicators of which trauma patients are at high-risk for bleeding to death. However, in principle, a multivariate classifier could provide a more consistent classification of vital signs for purposes of identifying patients with major hemorrhage.

3.0 CAN AN AUTOMATED ALGORITHM CONSISTENTLY IDENTIFY VITAL-SIGN PATTERNS ASSOCIATED WITH LIFE-THREATENING HEMORRHAGE?

In principle, if there are different types of compensation to blood loss (e.g., more sympathetic compensation with tachycardia versus less sympathetic compensation with greater hypotension), then a multivariate classifier could provide a more consistent classification of vital signs.

3.1 Methods: Automated Algorithms and Life-threatening Hemorrhage

3.1.1 Multivariate Classification

Figure 2 describes the methodology for automated identification of life-threatening hemorrhage using multivariate classification. First, we processed the vital signs to exclude unreliable measurements using automated algorithms as described in Section 2.1.2.

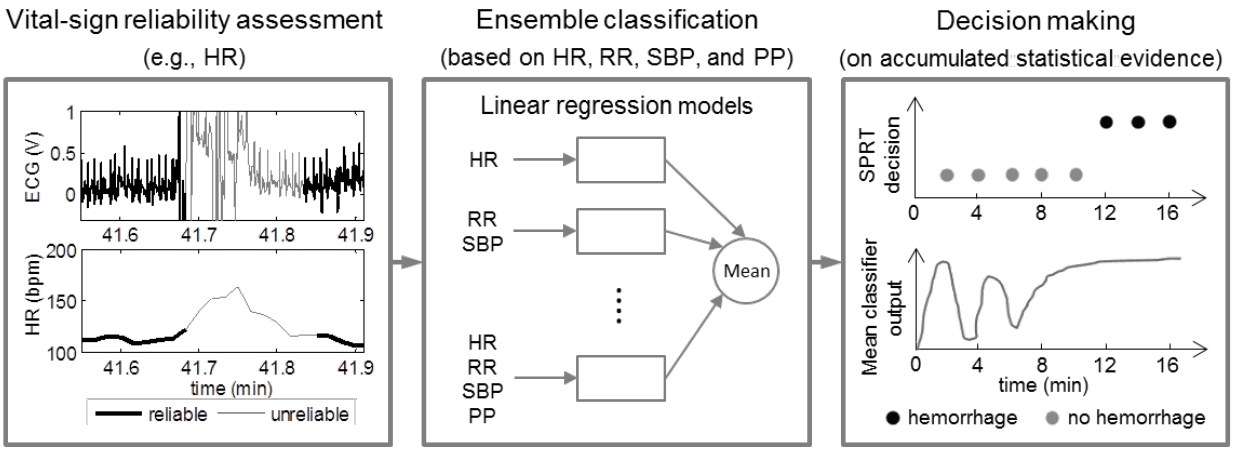

Figure 2: Analytic methodology for hemorrhage identification. In the first step (left panel), algorithms were used to identify, and exclude, unreliable vital signs. In the second step (middle panel), ensemble classification was applied, which consisted of a set of different linear regression models, that were subsequently averaged together. Ensemble classification is useful when missing data are commonplace: different regression models contain different combinations of the vital signs and it is possible to omit any of those models that contain a missing input parameter. In the third step (right panel), the mean ensemble classifier output was evaluated by the SPRT, a statistical test of whether or not measurements repeated over time are consistent with a control distribution (e.g., non-hemorrhagic patient) or with a different (e.g., hemorrhagic patient) distribution. bpm: beats per minute; ECG: electrocardiography; HR: heart rate; PP: pulse pressure; RR: respiratory rate; SBP: systolic blood pressure; SPRT: sequential probability ratio test; V: volt.

Second, we applied an ensemble classifier, which is a set of multivariate regression models whose numerical outputs were averaged to yield the final output. Compared with routine multivariate regression, an ensemble classifier can provide two advantages. First, the ensemble can still classify patients even if there are missing vital signs. Second, it can offer more consistent performance from one dataset to the next [12, 13].

Originally, we trained the ensemble's multivariate regression models (i.e., set the weights for the input variables) for a binary outcome as per Chen et al. [12], using the initial 15 min of vital-sign data from each MHLF subject. The binary outcome was whether patients received ≥ 1 PRBCs for an unambiguous hemorrhagic injury, or not. This model training yielded a classifier that, on the basis of the input vital signs, quantified whether the pattern was similar to the population with hemorrhage (output closer to 1) or to the non-hemorrhagic control population (output closer to 0).

This ensemble classifier was re-applied to each patient's data every 2 minutes.

- For the BMF dataset, this was done in real time during actual patient transport onboard medical helicopters, using a specialized computing platform [6].
- For the MHLF dataset, we performed the analysis retrospectively, applying the algorithms at every 2-min mark of the patient's electronic record, simulating real-time application.

In both studies, every time the ensemble classifier was applied (i.e., every 2 min), we analyzed the time-averaged value of all reliable HR, RR, SBP, and PP measured since the beginning of the record, and up to the time of analysis[1]. The rationale for analyzing data reaching back to the start of the mission arose from prior analysis suggesting that prehospital vital signs contained enormous variability—likely due to pain, medications, or other transient stimuli—and that time-averaging was an effective method to remove some of the confounding data perturbations [14].

Finally, we used the Wald's Sequential Probability Ratio Test (SPRT) [15] to determine whether to issue an automated "hemorrhage high-risk" notification on the basis of the accumulated evidence from the ensemble classifier outputs. The SPRT classifies data through time and determines whether repeated measurement samples are consistent with one statistical distribution (e.g., a normal population) versus a second statistical distribution (e.g., an abnormal population) [15]. Thresholds for the SPRT were set as per [16], where the SPRT was shown to reduce false alarms at the expense of some alarm latency.

3.1.2 Statistical Analysis

We computed the proportion of patients who received a hemorrhage notification as a function of 24-hr PRBC volume. For comparison, we also computed the proportion of patients with other hemodynamic abnormalities: initial SBP < 110 mmHg, any prehospital SBP < 90 mmHg, or any prehospital Shock Index (SI = HR/SBP) ≥ 1.4. We tested for significant differences between those proportions using McNemar's test.

3.2 Results: Automated Algorithms and Life-threatening Hemorrhage

Table 3 shows the relationship between incidence of APPRAISE hemorrhage notification and 24-hr PRBC transfusion volume. With increasing 24-hr PRBC transfusion volume, the proportion of APPRAISE notification of positive subjects exhibited an increasing trend in both the MHLF and BMF studies. In the pooled dataset (MHLF and BMF), we found that the sensitivity of APPRAISE notification for 24-hr PRBC transfusion volume

[1] For example, at t = 6 min, all vital-sign data from t = 0 to t = 6 min were analyzed. At t = 8 min, all vital sign data from t = 0 to t = 8 min were analyzed.

≥ 9 units was significantly higher than SI ≥ 1.4 (p = 0.014; 76% vs. 59%), initial SBP < 110 mmHg (p = 0.007; 76% vs. 50%), and any hypotension, i.e., SBP < 90 mmHg (p = 0.007; 76% vs. 50%). Also, the sensitivities of APPRAISE notification for 24-hr PRBC transfusion volume ≥ 9 units was similar for the MHLF versus BMF datasets.

In the pooled dataset (MHLF and BMF), we found that the specificity of the APPRAISE system for 24-hr PRBC transfusion volume = 0 units (i.e., no blood transfusion at all) was not significantly different from initial SBP < 110 mmHg (87% vs. 88%) or any prehospital SI ≥ 1.4 (87% vs. 88%). Compared to any prehospital SBP < 90 mmHg, APPRAISE notification showed a significantly lower specificity (p < 0.05; 87% vs. 90%), though the absolute magnitude of the difference was 3%.

Table 3: Prehospital APPRAISE hemorrhage notification incidence as a function of 24-hr PRBC transfusion volume.

	24-hr PRBC volume, units				Total
	0	1 to 2	3 to 8	≥ 9	
Total patients, n	**749**	**31**	**41**	**34**	**855**
MHLF patients, n	571	18	32	25	646
BMF patients, n	178	13	9	9	209
Hemorrhage notification, n (%)	**96 (13)**	**12 (39)**	**26 (63)**	**26 (76)**	
MHLF, n (%)	79 (14)	9 (50)	22 (69)	19 (76)	
BMF, n (%)	17 (10)	3 (23)	4 (44)	7 (78)	
Initial SBP < 110 mmHg, n (%)	**87 (12)**	**9 (29)**	**22 (54)**	**17 (50)**	
MHLF, n (%)	67 (12)	5 (28)	18 (56)	11 (44)	
BMF, n (%)	20 (11)	4 (31)	4 (44)	6 (67)	
Any SBP < 90 mmHg, n (%)	**73 (10)**	**9 (29)**	**24 (59)**	**17 (50)**	
MHLF, n (%)	51 (9)	6 (33)	18 (56)	11 (44)	
BMF, n (%)	22 (12)	3 (23)	6 (67)	6 (67)	
Any SI ≥ 1.4, n (%)	**92 (12)**	**8 (26)**	**21 (51)**	**20 (59)**	
MHLF, n (%)	70 (12)	6 (33)	18 (56)	14 (56)	
BMF, n (%)	22 (12)	2 (15)	3 (33)	6 (67)	

BMF: Boston MedFlight; HR: heart rate; MHLF: Memorial Hermann Life Flight; PRBC: packed red blood cell; SBP: systolic blood pressure; SI: shock index.

3.3 Discussion: Automated Algorithms and Life-threatening Hemorrhage

At a rudimentary level, this study suggests that patients with massive 24-hr blood transfusion requirements demonstrated identifiable hypovolemic physiology before hospital arrival.

In Section 2, it was shown that patient populations may have varied responses to hemorrhage, with some patients demonstrating greater sympathetic compensation (i.e., greater tachycardia and less hypotension) and others with

less compensation. Despite the differences between the vital signs in the BMF versus MHLF datasets, the multivariate classifier provided very consistent performance across both.

The finding that, during the preliminary evaluation of a trauma patient, their vital signs are useful for predicting life-threatening hemorrhage is consistent with other prediction rules for massive transfusion where hypotension and tachycardia are recognized as predictive factors for massive transfusion (i.e., Refs. 17-19). Unlike the other prediction rules, the APPRAISE system only involves vital-sign data analyzed during prehospital transport. Essential to its performance is a focus on analyzing multiple vital-sign measurements, rather than a single set.

The median notification time after the start time of transport was 6 min for MHLF and 10 min for BMF. The median notification time before arrival at the hospital was 17 min for MHLF and 52 min for BMF, and the difference was largely due to shorter transport times for MHLF (the median transport time for subjects with 24-hr PRBC volume $\geq$ 9 units was 25 min for MHLF and 66 min for BMF). Combining the two populations, APPRAISE notification occurred in the first half of the transportation in 73% of the cases.

Overall, here are the key implications:

- The automated analysis of vital signs allowed for significantly improved sensitivity for life-threatening hemorrhage without any clinically significant increase in false alarms. This supports the conclusion that any trauma management protocol that uses vital signs for decision-making (e.g., for activating the trauma team or activating an operating room or initiating resuscitation) could be enhanced by using automated analysis, rather than a single vital-sign criterion (e.g., SBP < 90 mmHg).
- A second potential advantage of the automated system is that it requires less cognitive effort by the clinicians. We speculate that use of an automated system could allow caregivers to focus on other aspects of bedside care and situational awareness, rather than focus on the vital-sign monitor patterns.
- A third potential advantage of the automated system is that it could be valuable, providing consistency and vigilance, even when caregivers are inexperienced, tired or distracted.

An expanded treatment of these findings was reported in Ref. 20.

4.0 HOW SENSITIVE IS THE ALGORITHM'S PERFORMANCE TO DIFFERENT METHODS OF ANALYZING VITAL-SIGN DATA THROUGH TIME?

In the aforementioned analysis, we used SPRT as a statistical test to determine whether the vital-sign patterns through time were abnormal or not. As noted above, this method successfully identified casualties with hemorrhage after a median of 6 – 10 min. Yet, this also meant that there was a substantial subset who required greater than 10 min of vital-sign monitoring for hemorrhage identification.

When decision-making must be done in less than 10 min, then this latency is sub-optimal. In the field of manufacturing, the SPRT [15] is one of several well-established analytic strategies for statistical process control, whereby aberrancies in a manufacturing process are detected by monitoring and analyzing the process output [21]. These include simple thresholding, the risk-adjusted SPRT (RASPRT) [22], and the cumulative sum (CUSUM) method [21].

In this section, we compare these classification strategies, to elucidate the achievable performance of the different methods.

4.1 Methods: Analyzing Vital-sign Data through Time

Statistical process control has been widely used in manufacturing processes where quick detection of "out-of-control" process variation is essential for quality control [21]. We compared four commonly used notification strategies based on the output of the ensemble classifier over time.

The simple thresholding used in our analysis consisted of a single upper limit A, where an alert was raised when $y(t) > A$ for the first time, with $y(t)$ denoting the output of the ensemble classifier at time t.

SPRT consisted of an upper limit A and a lower limit B, where the system issued an alert when the accumulated log likelihood ratio $LLR(t)$ exceeded the upper limit A. We calculated $LLR(t)$ as follows:

$$LLR(t) = LLR(t-1) + log\frac{f(y(t);\theta_1)}{f(y(t);\theta_0)},$$

but if $LLR(t) < B$, then $LLR(t)$ was reset to zero, where $f(y(t);\theta_0)$ and $f(y(t);\theta_1)$ denoted the probability density functions governing the null hypothesis (e.g., control) and alternative hypothesis (e.g., hypovolemia), respectively. $\theta_0 = (\mu_0, \sigma_0^2)$ and $\theta_1 = (\mu_1, \sigma_1^2)$ represent the mean and variance of the probability density functions governing the null and alternative hypotheses, respectively, which were estimated from the MHLF dataset.

RASPRT was exactly the same as SPRT, except that the probability density functions $f(y(t);\theta_0(t))$ and $f(y(t);\theta_1(t))$ were time varying depending on the availability of the vital signs at each time instant t (15 pairs of θ_0 and θ_1 were estimated from the MHLF dataset for 15 possible scenarios of vital-sign availability).

CUSUM consisted of an upper limit A and an offset w, where the system issued an alert when the accumulated $CUSUM(t)$ exceeded A. $CUSUM(t)$ was computed as follows:

$$CUSUM(t) = \max(CUSUM(t-1) + y(t) - w, 0).$$

We investigated the performance of each notification strategy by systematically varying the values of configurable parameters. Table 4 lists the configurable parameters for each notification strategy and the range of values we explored for each parameter. We chose the range of values to cover the full range of sensitivity and specificity from 0 to 100%. For each configuration, we applied the notification strategy to each patient using the ensemble classifier output over the course of the entire transport. We recorded the decision and then computed the sensitivity, specificity, and mean/median time to notification. We repeated the same analysis for different sizes of moving windows (2 min, 15 min, and 60 min).

Table 4: Notification strategies.

	Parameters	Range explored
Simple thresholding	1. Upper limit A 2. Window size L	$0 < A < 1$ L = 2, 15, 60 min
Sequential probability ratio testing (SPRT)	1. Upper limit A 2. Lower limit B 3. Window size L	$-2.2 < A < 6.9$ $-6.9 < B < 2.2$ L = 2, 15, 60 min
Risk-adjusted SPRT (RASPRT)	1. Upper limit A 2. Lower limit B 3. Window size L	$-2.2 < A < 6.9$ $-6.9 < B < 2.2$ L = 2, 15, 60 min
Cumulative sum (CUSUM)	1. Upper limit A 2. Offset w 3. Window size L	$0 < A < 1$ $0 < w < 1$ L = 2, 15, 60 min

We explored four investigational strategies to account for the substantial minute-to-minute fluctuations in the likelihood that a patient is bleeding. Each statistical strategy had several parameters to set, which determined their performance and resultant diagnostic test characteristics, in terms of sensitivity, specificity, and time to alert. Those parameters, and the range of values explored, are listed in the table.

4.2 Results: Analyzing Vital-sign Data through Time

We computed a total of 56,000 data points, where for each data point we calculated the *1*) sensitivity, *2*) specificity, and *3*) time to notification for one configuration of each of the four investigational strategies. These data points spanned the full range of sensitivities and specificities, from 0% to 100%. Because of space limitations, it is not possible to report all of these results, but it is possible to show representative findings. Figure 3 illustrates some of the trade-offs that we observed, exploring the four investigational methods for two levels of sensitivity (~75% and ~85%).

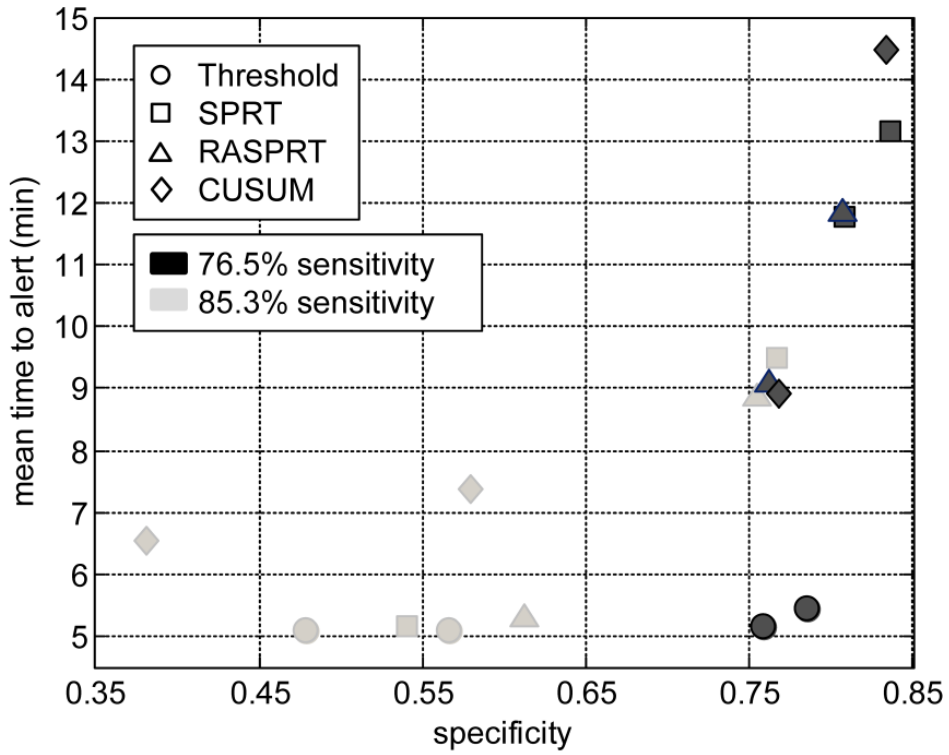

Figure 3: The trade-off between mean time to alert and specificity at fixed sensitivity levels of 76.5% and 85.3%. The four investigational strategies yielded a spectrum of results varying in sensitivity, specificity, and time to alert (depending on the setting of parameter values; see Table 4). Above, we illustrate results for two arbitrary levels of sensitivity (sensitivity of 76.5% and 85.3%). For each level of sensitivity and investigational strategy, we plot two results representing the minimum and maximum specificity (and corresponding times to alert) that were observed as we methodically explored the constellation of different parameter values for each investigational strategy. This figure illustrates the inevitable trade-offs between sensitivity, specificity, and time to alert, and that no one strategy was consistently superior to the others. CUSUM: cumulative sum; RASPRT: risk-adjusted SPRT; SPRT: sequential probability ratio test.

The key findings are as follows:

- None of the four classification strategies demonstrated any consistent, observable advantage. Classification strategies that were more accurate overall tended to be not as responsive (i.e., had a greater time to alert) and vice versa. We observed well-known trade-offs between sensitivity and specificity. In addition, we observed that increasing specificity was associated with increasing mean time to notification.
- At the ~75% sensitivity, the optimal classifier was arguably the simple threshold: it offered a similar specificity as the other methods, but with minimal time latency (see Figure 3).
- For higher sensitivity, ~85%, the simple threshold required a reduced value of upper limit A, which meant more false alarms (i.e., a reduced specificity). At this higher level of sensitivity, it was possible to reduce false alarms by relying on SPRT or RASPRT, but these methods came at the cost of ~5 min in additional notification latency.

4.3 Discussion: Analyzing Vital-sign Data through Time

Different methods of classification through time yielded different diagnostic test characteristics. No method was clearly superior. Instead, the methods offered different trade-offs.

Our initial algorithm was intended to analyze patients during prehospital transport. In the majority of the cases, the algorithms were able to identify hemorrhage long before hospital arrival. The use of SPRT was therefore appropriate for this application: it greatly reduced "false alarms," and the latency of ~5 min was acceptable considering that the transport times were significantly longer.

Conversely, for some other applications (e.g., assessment of casualties immediately upon arrival) this latency might be suboptimal. Our findings suggest that it would be possible to detect hemorrhage patients earlier, but the trade-off would either be reduced sensitivity and/or specificity.

These findings were presented at the 2014 IEEE Engineering in Medicine and Biology Society annual meeting [23].

5.0 CONCLUSION

Our work to date has demonstrated that, using well-known statistical techniques, it is possible to automate the analysis of vital signs in trauma patients and significantly improve the identification of life-threatening hemorrhage, compared to the use of simple thresholds for individual vital signs, e.g., SBP < 90 mmHg. Moreover, this approach does not lead to clinically significant increases in false alarms, it is fully automatable, and it would require a minimum of new sensors and training. The method is based on linear classification, and so its performance is "transparent" (i.e., the basis for its classification is readily apparent by examining the underlying vital signs, unlike a neural network black box).

Perhaps most significantly, the method has now been successfully validated prospectively during actual trauma patient care, which suggests that the technology is indeed viable for clinical operations. Future investigation will be focused on evaluating where this new capability provides clinical or operational benefit.

6.0 ACKNOWLEDGEMENTS

This work was sponsored by the U.S. Department of Defense Medical Research and Development Program and by the Combat Casualty Care Research Area Directorate of the U.S. Army Medical Research and Materiel Command, Fort Detrick, MD.

7.0 DISCLAIMER

The opinions and assertions contained herein are the private views of the authors and are not to be construed as official or as reflecting the views of the U.S. Army or of the U.S. Department of Defense. This paper has been approved for public release with unlimited distribution.

8.0 REFERENCES

[1] B. A. Cotton, N. Reddy, Q. M. Hatch, E. LeFebvre, C. E. Wade, R. A. Kozar, B. S. Gill, R. Albarado, M. K. McNutt, and J. B. Holcomb, "Damage control resuscitation is associated with a reduction in

resuscitation volumes and improvement in survival in 390 damage control laparotomy patients," *Ann Surg*, vol. 254, no. 4, pp. 598-605, Oct. 2011.

[2] D. J. Riskin, T. C. Tsai, L. Riskin, T. Hernandez-Boussard, M. Purtill, P. M. Maggio, D. A. Spain, and S. I. Brundage, "Massive transfusion protocols: the role of aggressive resuscitation versus product ratio in mortality reduction," *J Am Coll Surg*, vol. 209, no. 2, pp. 198-205, Aug. 2009.

[3] J. Reifman, L. Chen, M. Y. Khitrov, and A. T. Reisner, "Automated decision-support technologies for prehospital care of trauma casualties," in *Proc RTO HFM Symposium on NATO Human Use of Advanced Technologies and New Procedures in Medical Field Operations*, Essen, Germany, pp. 1-14, 2010.

[4] W. H. Cooke, J. Salinas, V. A. Convertino, D. A. Ludwig, D. Hinds, J. H. Duke, F. A. Moore, and J. B. Holcomb, "Heart rate variability and its association with mortality in prehospital trauma patients," *J Trauma*, vol. 60, no. 2, pp. 363-70, Feb. 2006.

[5] J. B. Holcomb, J. Salinas, J. M. McManus, C. C. Miller, W. H. Cooke, and V. A. Convertino, "Manual vital signs reliably predict need for life-saving interventions in trauma patients," *J Trauma*, vol. 59, no. 4, pp. 821-8, Oct. 2005.

[6] A. T. Reisner, M. Y. Khitrov, L. Chen, A. Blood, K. Wilkins, W. Doyle, S. Wilcox, T. Denison, and J. Reifman, "Development and validation of a portable platform for deploying decision-support algorithms in prehospital settings," *Appl Clin Inform*, vol. 4, no. 3, pp. 392-402, Aug. 2013.

[7] T. M. McKenna, G. Bawa, K. Kumar, and J. Reifman, "The physiology analysis system: an integrated approach for warehousing, management and analysis of time-series physiology data," *Comput Meth Prog Bio*, vol. 86, no. 1, pp. 62-72, Apr. 2007.

[8] L. Chen, T. McKenna, A. Reisner, and J. Reifman, "Algorithms to qualify respiratory data collected during the transport of trauma patients," *Physiol Meas*, vol. 27, no. 9, pp. 797-816, Sep. 2006.

[9] C. Yu, Z. Liu, T. McKenna, A. T. Reisner, and J. Reifman, "A method for automatic identification of reliable heart rates calculated from ECG and PPG waveforms," *J Am Med Inform Assoc*, vol. 13, no. 3, pp. 309-20, May-Jun. 2006.

[10] A. T. Reisner, L. Chen, T. M. McKenna, and J. Reifman, "Automatically-computed prehospital severity scores are equivalent to scores based on medic documentation," *J Trauma*, vol. 65, no. 4, pp. 915-23, Oct. 2008.

[11] L. Chen, A. T. Reisner, A. Gribok, T. M. McKenna, and J. Reifman, "Can we improve the clinical utility of respiratory rate as a monitored vital sign?," *Shock*, vol. 31, no. 6, pp. 574-80, Jun. 2009.

[12] L. Chen, T. M. McKenna, A. T. Reisner, A. Gribok, and J. Reifman, "Decision tool for the early diagnosis of trauma patient hypovolemia," *J Biomed Inform*, vol. 41, no. 3, pp. 469-478, Jun. 2008.

[13] T. G. Dietterich, "Ensemble Methods in Machine Learning," in *Proc 1st Int Workshop on Multiple Classifier Systems*, Cagliari, Italy, pp. 1-15, 2000.

[14] L. Chen, A. T. Reisner, A. Gribok, and J. Reifman, "Exploration of prehospital vital sign trends for the prediction of trauma outcomes," *Prehosp Emerg Care*, vol. 13, no. 3, pp. 286-294, Jul-Sep. 2009.

[15] A. Wald, "Sequential tests of statistical hypotheses," *Ann Math Stat*, vol. 16, pp. 117-186, 1945.

[16] L. Chen, A. T. Reisner, X. Chen, A. Gribok, and J. Reifman, "Are standard diagnostic test characteristics sufficient for the assessment of continual patient monitoring?," *Med Decis Making*, vol. 33, no. 2, pp. 225-234, Feb. 2013.

[17] D. F. McLaughlin, S. E. Niles, J. Salinas, J. G. Perkins, E. D. Cox, C. E. Wade, and J. B. Holcomb, "A predictive model for massive transfusion in combat casualty patients," *J Trauma*, vol. 64, no. 2 Suppl, pp. S57-63, Feb. 2008.

[18] T. C. Nunez, I. V. Voskresensky, L. A. Dossett, R. Shinall, W. D. Dutton, and B. A. Cotton, "Early prediction of massive transfusion in trauma: simple as ABC (assessment of blood consumption)?," *J Trauma*, vol. 66, no. 2, pp. 346-52, Feb. 2009.

[19] N. Yucel, R. Lefering, M. Maegele, M. Vorweg, T. Tjardes, S. Ruchholtz, E. A. Neugebauer, F. Wappler, B. Bouillon, and D. Rixen, "Trauma Associated Severe Hemorrhage (TASH)-Score: probability of mass transfusion as surrogate for life threatening hemorrhage after multiple trauma," *J Trauma*, vol. 60, no. 6, pp. 1228-36, Jun. 2006.

[20] J. Liu, M. Y. Khitrov, J. D. Gates, S. R. Odom, J. M. Havens, M. A. de Moya, K. Wilkins, S. K. Wedel, E. O. Kittell, J. Reifman, and A. T. Reisner, "Automated analysis of vital signs to identify patients with substantial bleeding prior to hospital arrival: a feasibility study," *Shock*, vol. 43, no. 5, pp. 429-36, Feb. 2015.

[21] D. C. Montgomery, *Introduction to statistical quality control*, Sixth ed. New York: Wiley, 2009.

[22] O. A. Grigg, V. T. Farewell, and D. J. Spiegelhalter, "Use of risk-adjusted CUSUM and RSPRT charts for monitoring in medical contexts," *Stat Methods Med Res*, vol. 12, no. 2, pp. 147-70, Mar. 2003.

[23] J. Liu, A. T. Reisner, S. Edla, and J. Reifman, "A comparison of alerting strategies for hemorrhage identification during prehospital emergency transport," in *Proc Conf IEEE Eng Med Bio Soc*, Chicago, IL, USA, pp. 2670-3, 2014.