

A Comparison of Alerting Strategies for Hemorrhage Identification during Prehospital Emergency Transport

Jianbo Liu-*IEEE Member*, Andrew T. Reisner, Shwetha Edla, and Jaques Reifman

Abstract—Early and accurate identification of physiological abnormalities is one feature of intelligent decision support. The ideal analytic strategy for identifying pathological states would be highly sensitive and highly specific, with minimal latency. In the field of manufacturing, there are well-established analytic strategies for statistical process control, whereby aberrancies in a manufacturing process are detected by monitoring and analyzing the process output. These include simple thresholding, the sequential probability ratio test (SPRT), risk-adjusted SPRT, and the cumulative sum method. In this report, we applied these strategies to continuously monitored prehospital vital-sign data from trauma patients during their helicopter transport to level I trauma centers, seeking to determine whether one strategy would be superior. We found that different configurations of each alerting strategy yielded widely different performances in terms of sensitivity, specificity, and average time to alert. Yet, comparing the different investigational analytic strategies, we observed substantial overlap among their different configurations, without any one analytic strategy yielding distinctly superior performance. In conclusion, performance did not depend as much on the specific analytic strategy as much as the configuration of each strategy. This implies that any analytic strategy must be carefully configured to yield the optimal performance (i.e., the optimal balance between sensitivity, specificity, and latency) for a specific use case. Conversely, this also implies that an alerting strategy optimized for one use case (e.g., long prehospital transport times) may not necessarily yield performance data that are optimized for another clinical application (e.g., short prehospital transport times, intensive care units, etc.).

I. INTRODUCTION

Real-time alerting of life-threatening conditions based on vital signs has the potential to help prehospital caregivers

This work was supported by the U.S. Department of Defense Medical Research and Development Program (Grant No. D10_I_AR_J6_773) and by the Combat Casualty Care Research Area Directorate of the U.S. Army Medical Research and Materiel Command (USAMRMC), Fort Detrick, MD.

J. Liu is with the Department of Defense Biotechnology High Performance Computing Software Applications Institute (BHS AI), Telemedicine and Advanced Technology Research Center (TATRC), U.S. Army Medical Research and Materiel Command (USAMRMC), Fort Detrick, MD 21702 USA (email: jliu2@bhsai.org).

A. T. Reisner is with the BHS AI, TATRC, USAMRMC, Fort Detrick, MD 21702 USA and with the Department of Emergency Medicine, Massachusetts General Hospital, Boston, MA 02114 USA (email: areisner@partners.org).

S. Edla is with the BHS AI, TATRC, USAMRMC, Fort Detrick, MD 21702 USA (corresponding author; phone: 301-619-1977; fax: 301-619-1983; email: sedla@bhsai.org).

J. Reifman is a senior research scientist and director of the BHS AI, TATRC, USAMRMC, Fort Detrick, MD 21702 USA (e-mail: jaques.reifman.civ@mail.mil).

better manage trauma patients and, via advance notification, to expedite time-sensitive interventions delivered at the receiving facilities. For instance, early transfusion of fresh frozen plasma (FFP) has been shown to be associated with improved outcomes for trauma patients with life-threatening hemorrhage [1]. In theory, prehospital alerting with advance radio notification could allow for the receiving trauma center to prepare FFP for immediate transfusion upon arrival.

Prehospital vital signs, however, can show considerable intra-individual fluctuations during the course of transport, due to transient stimuli, such as pain, fear, medications, movement, etc. [2]. These fluctuations can trigger false alarms when they (transiently) appear consistent with serious pathology. Moreover, they can obscure the evolution of the individual's true pathophysiology. When seeking to identify physiological abnormalities indicative of life-threatening pathology, an optimal alerting strategy would ignore transient, benign abnormalities, while remaining highly sensitive to the earliest physiological indicators of actual life-threatening pathology.

Classic test characteristics for diagnostic tests include sensitivity and specificity [3]. For alerts based on continuous monitoring over time, it is also important to consider the temporal behavior of the alert, because its accuracy may change as a function of time, and because some alerting algorithms may yield inconsistent output over time due to the aforementioned fluctuations in vital signs.

In prior work, we demonstrated that the sequential probability ratio test (SPRT) could be applied for post-processing of a multivariate classifier that identifies life-threatening hemorrhage in trauma patients based on patterns in heart rate (HR), systolic blood pressure (SBP), pulse pressure (PP), and respiratory rate (RR) [2]. The SPRT reduced the fraction of patients who triggered false alarms, but at the expense of some temporal latency for those who generated true alarms.

Yet if the goal of the alerting system is to provide the earliest possible identification of patients with life-threatening hemorrhage—to allow maximum time for preparation at the receiving hospital—this latency is sub-optimal. In the field of manufacturing, there are well-established analytic strategies for statistical process control, whereby aberrancies in a manufacturing process are detected by monitoring and analyzing the process output [4]. These include simple thresholding, the SPRT [5], the risk-adjusted SPRT (RASPT) [6], and the cumulative sum (CUSUM) method [4]. In this paper, we compared these alerting strategies for identifying hypovolemia based on prehospital vital signs during helicopter transport of trauma patients. Our

goal was to elucidate the achievable performance of the different investigational methods.

II. MATERIAL AND METHODS

A. Data Collection and Subject Selection

The study was based on physiological data collected with Institutional Review Board approval during helicopter transports of adult trauma patients (age ≥ 18 years) to several level I trauma centers via Memorial Hermann Life Flight (MHLF) between August 2001 and April 2004 [7], and Boston MedFlight (BMF) between February 2010 and December 2012. Propaq 206 patient monitors (Welch-Allyn, Beaverton, OR) recorded the data. The dataset consisted of physiological waveforms, such as electrocardiograms (ECGs), and vital signs, such as HR, RR, SBP, and diastolic blood pressure (DBP). We collected clinical outcome data, including demographics, prehospital interventions, in-hospital interventions, and injury descriptions, retrospectively via chart review at the receiving hospitals.

The study population consisted of patients with at least one blood pressure measurement. In the analysis, we excluded patients who died prior to hospital admission because resuscitation was often terminated before a large volume of packed red blood cells (PRBCs) could be administered. Our primary outcome was 24-hour PRBC transfusion volume in patients with explicitly documented hemorrhagic injury, such as laceration of solid organs, thoracic or intraperitoneal hematoma, vascular injury that required operative repair, or limb amputation. Patients who received blood transfusions without explicitly documented hemorrhagic injuries were excluded. Table 1 lists the characteristics of the study population.

B. Physiological Data Processing

Because of noise and artifacts that were commonly present in the physiological signals, we used automated quality assessment algorithms [8, 9] to identify clean and reliable measurements, which have been shown to offer superior diagnostic performance [10]. We used a previously developed ensemble classifier [11] to assess whether the patient had hypovolemia based on HR, RR, SBP and pulse

pressure (PP = SBP – DBP). The ensemble classifier is a set of linear regression models with one, two, or three input parameters which comprise all possible combinations of SBP, PP, HR, and RR. The ensemble classifier’s output is the average of the outputs of the set of regression models. The output generally ranged from 0 to 1, quantifying the similarity between the input vital-sign features and those of patients with hypovolemia. We re-applied the ensemble classifier every two minutes during the course of transport and used a moving window to smooth the vital-sign features before processing by the ensemble classifier.

C. Alerting Strategies

Statistical process control has been widely used in the industrial context, where quick detection of “out-of-control” process variation is essential for quality control [4]. We compared four commonly used alerting 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 , and an alert was raised when $y(t) < A$ for the first time, where $y(t)$ denotes the output of the ensemble classifier at time t . SPRT consisted of an upper limit A and a lower limit B , and 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. θ_0 and θ_1 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 , and 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 alert strategy by systematically varying the values of configurable parameters. Table 2 lists the configurable parameters for each alerting strategy and the range of values we explored for each parameter. We chose values to cover the full range of sensitivity and specificity from 0 to 100%. For each configuration, we applied the alerting 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 alert as detailed in Section II.D. We repeated the same analysis for different sizes of moving windows (2 minutes, 15 minutes, and 60 minutes).

D. Performance Measures

We defined massive transfusion as receipt of 9 or more units of PRBCs within the initial 24 hours. Routine test

TABLE 1. STUDY POPULATION CHARACTERISTICS

	Memorial Hermann Life Flight	Boston MedFlight
Population, n	646	209
Sex, male/female, n	479/167	155/54
Age (year), 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)
Prehospital airway intubation, n (%)	113 (17%)	80 (38%)
Prehospital GCS, median (IQR)	15 (13-15)	15 (8-15)
24-hour PRBC volume > 0 units, n (%)	75 (12%)	31 (15%)
24-hour PRBC volume ≥ 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. ALERTING STRATEGIES

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

characteristics [3] were computed for the prehospital diagnosis (alert) of subsequent massive transfusion. The mean and median times to alert were calculated for patients with massive transfusions. We also computed the specificity for patients who did not receive any PRBCs (i.e., < 1) within 24 hours.

III. RESULTS

We computed a total of 56,000 data points, where each data point consisted of the 1) sensitivity, 2) specificity, and 3) time to alert for each configuration of the four investigational strategies. These data points spanned the full range of sensitivities and specificities, from 0% to 100%. None of the four alerting strategies demonstrated any consistent, observable advantage. Alerting strategies that were more accurate overall tended to be less responsive and vice versa. Considering specific configurations of the four alerting strategies, besides the obvious trade-off between sensitivity and specificity, increased specificity generally was associated with increased mean time to alert. Because of space limitations, it is not possible to report all of these results, but it is possible to show representative subsets of the findings.

First, consider the trade-off between specificity and time to alert. Here, we examine one subset of results from one

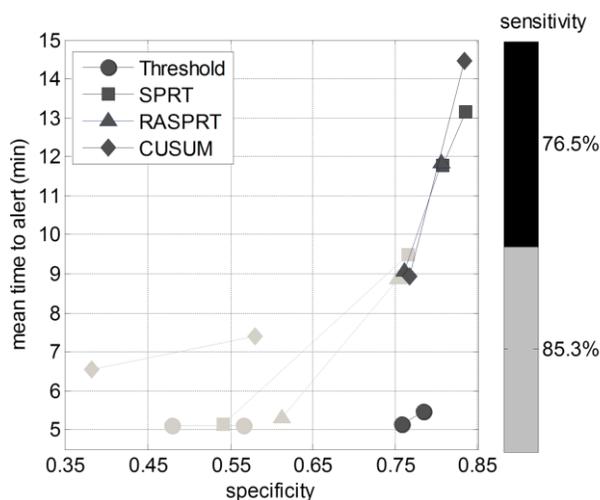


Figure 1. The trade-off between mean time to alert and specificity at fixed sensitivity levels of 76.5% and 85.3%. A 60-minute moving window was used to filter the vital-sign features. SPRT: sequential probability ratio test; RASPRT: risk-adjusted SPRT; CUSUM: cumulative sum.

fixed level of sensitivity (76.5%) with a moving window of 60 minutes. Among a set of 780 data points, we observed a wide spectrum of performance achieved by different configurations of each investigational alerting strategy, with substantial overlap between the four strategies, as illustrated in Fig. 1. There was no investigational strategy that offered distinctly superior performance.

Similarly, we may examine another subset of results from another fixed level of sensitivity (85.3%), again with a moving window of 60 minutes. In general, among a set of 280 data points, we observed lower specificity, and again, substantial overlap between the four investigational strategies (see Fig. 1).

Table 3 further shows the performance of various types of alerting strategies at a fixed sensitivity of 76.5% for various permutations of alerting strategies and window sizes. We chose 76.5% sensitivity because it represented an operating point of interest specific to our application. We chose the configuration of each permutation to maximize the specificity for patients who did not receive massive transfusions. The maximal specificity for SPRT, RASPRT, and CUSUM was higher than that of simple thresholding. This, however, came at a cost of increased time to alert. Among the three alerting strategies (SPRT, RASPRT, and CUSUM) that explicitly accumulate evidence before making a decision, RASPRT offered a shorter time to alert but had a slight decrease in maximal specificity. Overall, at the fixed sensitivity of 76.5%, higher maximal specificity tended to be associated with a longer time to alert.

The size of the moving window had a minimal impact on the diagnostic accuracy, and the specificity remained largely unchanged except in the case of simple thresholding. Further increasing the size of the moving window did not introduce sizable changes in the time to alert.

IV. DISCUSSION

In this report, we studied the performance of four different types of alerting strategies for diagnosing hypovolemia. None of the investigational strategies offered a distinct advantage in terms of accuracy versus responsiveness. Within each strategy, different configurations made it possible to trade-off between sensitivity, specificity, and time to alert. Configurations that were more accurate overall tended to be less responsive and vice versa.

Our results suggest that the nuanced differences among various alerting strategies were predominated by the fundamental trade-off between accuracy and responsiveness. Minor differences between these strategies, or whether a more elaborate alerting strategy (e.g., combination of two alerting strategies) could offer better performance, cannot be answered without a larger patient population.

It seems likely that the fundamental trade-off between accuracy and responsiveness was imposed by the innate characteristics of the vital-sign time series, with substantial fluctuations not directly related to hypovolemia (e.g., due to pain or medication therapy [2]) that could trigger a false alert. Techniques that tolerate transient fluctuations without alerting reduced the incidence of false alarms but were slower to react to early changes indicative of true

TABLE 3. PERFORMANCE OF CONTROL CHARTS AT A FIXED SENSITIVITY OF 76.5%

Alerting strategies	Size of moving window, minutes	Specificity for 24-hour PRBC < 9 (95% CI), %	Specificity for 24-hour PRBC < 1 (95% CI), %	Median time to alert, minutes	Mean time to alert, minutes
Simple thresholding	2	73 (70, 76)	77 (74, 80)	4	7
	15	79 (76, 82)	83 (80, 85)	2	8
	60	78 (75, 81)	82 (79, 84)	2	5
Sequential probability ratio test (SPRT)	2	84 (81, 86)	88 (85, 90)	12	14
	15	84 (81, 86)	88 (85, 90)	10	13
	60	84 (81, 86)	87 (85, 90)	9	13
Risk-adjusted SPRT (RASPRT)	2	83 (81, 86)	87 (84, 89)	11	14
	15	81 (78, 84)	85 (82, 87)	6	11
	60	81 (78, 83)	84 (81, 87)	5	11
Cumulative sum (CUSUM)	2	82 (79, 85)	86 (83, 89)	14	15
	15	84 (81, 86)	87 (85, 90)	10	13
	60	83 (81, 86)	87 (85, 90)	11	14

CI: confidence interval; PRBC: packed red blood cell

hypovolemia. Our findings suggest that none of the investigative methods were able to overcome this fundamental trade-off, and that a reasonably designed alerting strategy must simply balance accuracy versus responsiveness; it may not be possible to simultaneously excel at both by any large margin.

The optimal balance between accuracy and responsiveness may need to be customized to a clinical use case. Consider a prehospital alerting system intended to trigger labor-intensive preparations at the receiving trauma center (e.g., clearing operating rooms, mobilizing surgeons and blood products, etc.). At least 15 minutes of advance warning would be desirable, while false alarms would be costly, squandering the time of busy staff. If the typical (hypothetical) flight was 45 minutes, then an alerting strategy that afforded high specificity despite 13-14 minutes of latency would be appropriate (e.g., the SPRT; see Table 3). But if the typical (hypothetical) flight was 20 minutes, then it would be more appropriate to apply simple thresholding, with its median alert time < 5 minutes.

These findings have implications beyond prehospital decision support. Generally, medical alerts may be beneficial if they are configured for specific clinical uses. For an operating room or intensive care unit, when there is already a clinician at the bedside (and therefore an alert carries a low operational cost) it may be appropriate to employ very early alerts. By contrast, for ward patients, if an alert mobilizes a full rapid response team (at a high operational cost), it may be worth a degree of latency to reduce false alarms. For each application, the cost of latency should be weighed against the cost of false alerts.

In conclusion, we found that the investigational strategies offered a wide spectrum of performance levels, and the performance spectra from different strategies often overlapped substantially. Our findings suggest that the optimization of an alerting strategy requires careful examination of both clinical requirements and patient data characteristics, and caution needs to be exercised when applying the same configuration to a different clinical setting.

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.

REFERENCES

- [1] 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, pp. 198-205, Aug 2009.
- [2] 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, pp. 225-234, Feb 2013.
- [3] M. S. Pepe, *The Statistical Evaluation of Medical Tests for Classification and Prediction*. New York: Oxford University Press, 2004.
- [4] D. C. Montgomery, *Introduction to statistical quality control*, Sixth ed. New York: Wiley, 2009.
- [5] A. Wald, "Sequential tests of statistical hypotheses," *Ann Math Stat*, vol. 16, pp. 117-186, 1945.
- [6] 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, pp. 147-170, Mar 2003.
- [7] 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, pp. 821-828, Oct 2005.
- [8] 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, pp. 309-320, May-Jun 2006.
- [9] 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, pp. 797-816, Sep 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, pp. 915-923, Oct 2008.
- [11] 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, pp. 469-478, Jun 2008.