AWARD NUMBER: W81XWH-12-2-0019

TITLE: Assessment of Circulatory Dysfunction by Automated Processing of Vital Signs Data

PRINCIPAL INVESTIGATOR: Andrew Reisner MD

CONTRACTING ORGANIZATION: Massachusetts General Hospital
Boston , MA 02460

REPORT DATE: March 2019

TYPE OF REPORT: FINAL

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

# REPORT DOCUMENTATION PAGE

*Form Approved*
*OMB No. 0704-0188*



| 1. REPORT DATE | 2. REPORT TYPE | 3. DATES COVERED |
|---|---|---|
| MARCH 2019 | Final | 2 Mar 2016 - 1 Dec 2018 |

**4. TITLE AND SUBTITLE**

Assessment of Circulatory Dysfunction by Automated Processing of Vital Signs Data

**5a. CONTRACT NUMBER**
n/a

**5b. GRANT NUMBER**
W81XWH-12-2-0019

**5c. PROGRAM ELEMENT NUMBER**
n/a

**6. AUTHOR(S)**
Andrew Reisner MD

E-Mail: areisner@partners.org

**5d. PROJECT NUMBER**
n/a

**5e. TASK NUMBER**
n/a

**5f. WORK UNIT NUMBER**
n/a

**7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)**

Massachusetts General Hospital
Dept of Emergency Med
Zero Emerson Place Suite 3B
Boston, MA 0211

**8. PERFORMING ORGANIZATION REPORT NUMBER**

**9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)**
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

**10. SPONSOR/MONITOR'S ACRONYM(S)**
n/a

**11. SPONSOR/MONITOR'S REPORT NUMBER(S)**
n/a

**12. DISTRIBUTION / AVAILABILITY STATEMENT**
Approved for Public Release; Distribution Unlimited

**13. SUPPLEMENTARY NOTES**

**14. ABSTRACT** **Background**: Casualty care is challenging because caregivers may be inexperienced, or distracted by environmental dangers or multiple casualties. **Objective:** To provide clinical data for development and validation of a system that executes, in real time, automated decision-assist tools that accurately identify key trauma patient conditions and guide relevant life-saving interventions. This system will be comprised of novel "artificial intelligence" algorithms that only rely on data measured by standard patient transport monitors. **Specific Aims:** We will validate a, fully functional prototype of the decision-assist system, which can be provided to an industry partner for full productization. **Study Design**: We will prospectively trial these algorithms by making use of our operational, IRB-approved "plug-and-play" system for clinical field-testing of algorithms, presently in use on board Boston Medflight helicopters and the MGH Emergency Dept. **Relevance:** Because the necessary medical instrumentation, i.e., a standard travel monitor, is so very familiar to caregivers, these decision-assistance capabilities could be broadly deployed with a relative minimum of additional training, hardware acquisition, and up-front buy-in by clinicians.

**15. SUBJECT TERMS**

NONE LISTED

| 16. SECURITY CLASSIFICATION OF: | | | 17. LIMITATION OF ABSTRACT | 18. NUMBER OF PAGES | 19a. NAME OF RESPONSIBLE PERSON |
|---|---|---|---|---|---|
| n/a | | | | | USAMRMC |
| a. REPORT | b. ABSTRACT | c. THIS PAGE | | | 19b. TELEPHONE NUMBER *(include area code)* |
| U | U | U | UU | 146 | |

TABLE OF CONTENTS

## INTRODUCTION

### *1.1. State-of-the-art for patients with substantial bleeding*

Hemorrhage is recognized as the most important, treatable cause of death after injury [1, 2]. A recent series of reports [3, 4] has shown improved outcomes when trauma centers employ so-called "substantial bleeding protocols" [5].

Underlying these protocols are two basic strategies. First, damage-control resuscitation (DCR) includes aggressive measures to avoid coagulopathy (via permissive hypotension that slows blood loss, adequate restoration of coagulation factors via transfusion, and minimization of hypothermia). This is very important, because trauma-induced coagulopathy affects between 24 and 56% of critically injured patients [6, 7]. Second, DCR is paired with damage-control surgery, an operative strategy prioritizing early surgical control of bleeding, while sparing non-critical surgical repairs that are undertaken only after the patient has sufficiently recovered.

### *1.2. An unresolved question: when to initiate the "substantial bleeding protocol"*

The care protocols for substantially bleeding patients are resource intensive and time sensitive. It is therefore notable that there is no well-established, evidence-based method about when and how to activate the protocols. Consider two recent reports from centers that have demonstrated mortality benefit of these practices, where the protocols are initiated based on subjective assessments:

- Riskin et al. [3] reported that the Stanford Protocol is activated "*at the discretion of the attending physician*."

- Holcomb and Gumbert [5], who have previously shown mortality benefits of these protocols [4], remark that in their experience, activation is subjective: "The process of activating … varies with each institution. *Generally, on arrival to the emergency department, the attending trauma surgeon evaluates the patient's physiology and injury complex … Once a clinical diagnosis of substantial bleeding is made, the attending clinician activates* [the protocol]."

In theory, there are three reasons to seek an improved, objective method for the initiation of substantial bleeding protocols:

- Centers that are pioneering these protocols are staffed by thought-leaders in the trauma-care world, whose ability to make appropriate subjective decisions may be far superior to typical facilities that lack leading experts. Thus, the benefits of these protocols may not be accrued by centers or facilities that lack caregivers with the expertise necessary to apply them appropriately.

- Many of the measures are time sensitive. Yet, delaying initiation of the substantial bleeding protocol until after evaluation by a high-level specialist may result in preventable delays. For example, awaiting the evaluation of a trauma surgeon at the receiving hospital is at odds with the finding that the earlier plasma and platelets are received in the setting of substantial bleeding, the better the outcomes. Consider that the Stanford Medical Center [3] showed significant mortality benefit in a protocol that successfully reduced average time-to-first fresh frozen plasma (FFP) from 254 to 169 min. It seems entirely possible that FFP could be administered even earlier with potential to further benefit the patient.

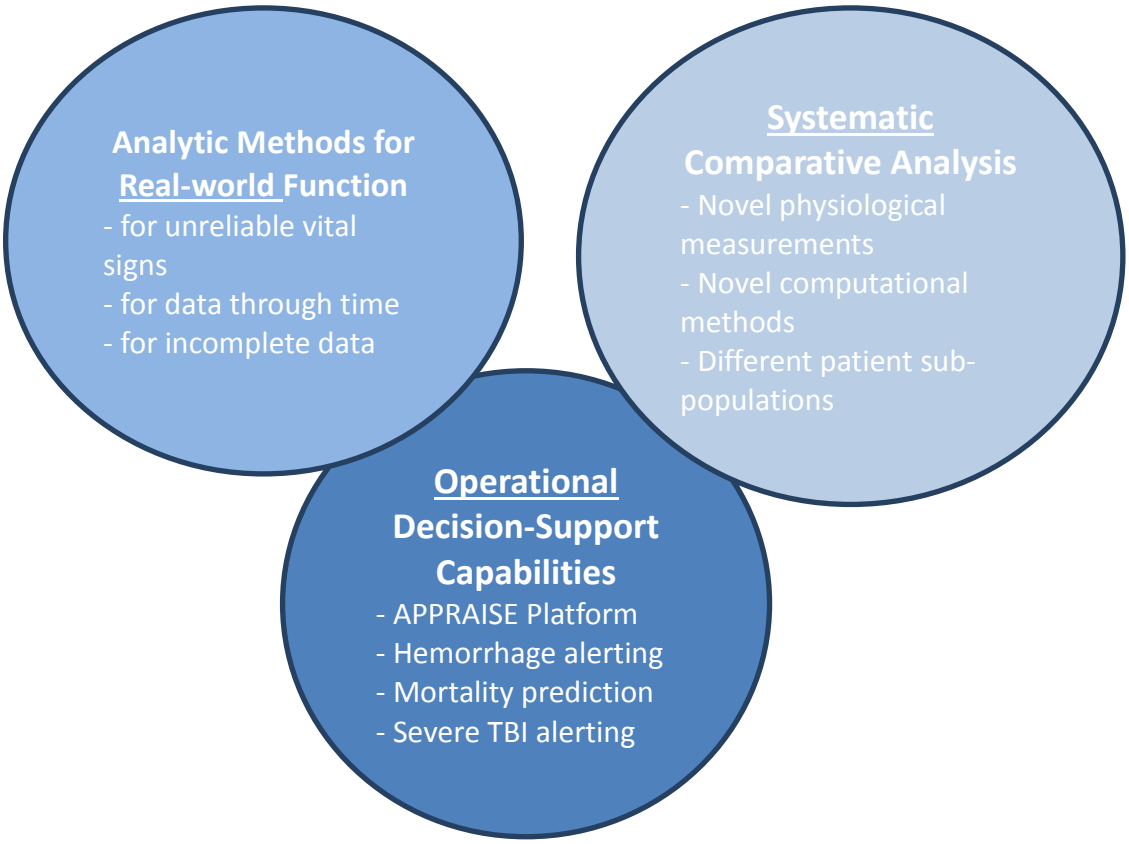
- Activations of the protocols after arrival at the receiving facility forestall pre-hospital interventions specific to the substantially bleeding patient (i.e., pre-hospital practices that should only be applied to

substantially bleeding patients but not to every trauma patient, either because the practices are too resource intensive or have an unfavorable risk-benefit profile in the non-bleeding patient).

*1.3. Is it possible to use an automated pre-hospital alerting system for detecting patients at high risk of substantial bleeding?*

Over the past decade, in close collaborator with the MRMC BHSAI, we have established that routine pre-hospital vital signs indicate trauma patients with substantial bleeding.

This proposal provided four categories of deliverables. First, we developed analytic computational algorithms to risk-stratify trauma patients so that patients with substantial bleeding would be identified early and treated optimally. Second, overlapping with the first category, we completed sets of comparative analyses to identify what physiological metrics (based on routine vital signs), and what ancillary metrics (based on novel sensors) could provide optimal operational value for the early identification of substantial bleeding in trauma patients. Third, we developed operational decision-support capabilities, culminating in a software system that provides real-time risk assessment of actual trauma patients and additional decision-support messaging. Fourth, we provided support to the BHSAI in preparing for FDA regulatory submissions related to our novel computational methodology.



BODY

In collaboration with the MRMC BHSAI, we produced four sets of deliverables.

First, we developed analytic methodologies that are effective during real-time performance. In the recent past, we have developed techniques for unreliable vital sign identification [8-10], decision-making of serial data through time [10,12-14], and incomplete data [11]. These capabilities were validated through two investigational foci: prehospital [13] and within the MGH ED. As well, we collaborated with the BHSAI to examine how these aforementioned techniques need to be customized to different environments (i.e., prehospital versus ED). We have completed our analysis of the prehospital dataset, and a report has been published [17]. In addition, we have collected a dataset with over 2000 trauma patients, and we have identified opportunities to make the analytic methodology even more clinically relevant. The first version algorithm yielded a binary output (indicating whether or not the patient is likely bleeding); this was the version tested prospectively on-board Boston Medflight. The second version algorithm yielded the probability that the patient is bleeding, and the probability that any given life-saving intervention will be necessary for that patient. The final version of the algorithm yielded an ordinal category of risk, where each higher risk level had a significantly higher association with a set of hemorrhage-related outcomes (see **Appendix I**). The final version was the version incorporated into our software with a functional user interface (see below).

Second, we completed a set of comparative analyses. Collaborating with the BHSAI, we investigated whether heart rate variability enhanced routine vital signs [18], and whether the CareGuide muscle oxygenation sensor enhances routine vital signs [26]. We also examined different analytic techniques for detecting abnormalities in time-series data (such as the cumulative-sum versus sequential probability ratio test [12,16]). Overall, here were the major findings of this scope of analysis:

- Almost all patients with major hemorrhage had patently abnormal vital signs *within 60 min* of monitoring (i.e., high sensitivity & high specificity using basic vital signs) [28].
- Within the first hour, differentiation of patients with and without hemorrhage on the basis of routine vital signs was improved with the application of statistical algorithms [17].
- Within the first hour, differentiation of patients with and without hemorrhage was also improved by tissue oximetery [27].
- There was no *additional* diagnostic value in heart rate variability [18], which supplements findings from preceding analyses by this collaborative team which found no *additional* value in SpO2 waveform analysis [29] nor in vital sign trend analysis [14].

Third, we developed operational decision-support software capabilities to make the computational techniques into operational tools suitable for real-time clinical use. As noted, our prehospital real-time use results have been described in a manuscript [17]. To this end, since the start of this project, we were awarded two US patents for this technology US 8,977,349 and 8,694,085. We have implemented a graphical user interface (GUI) for the system, which displays decision-support messages for supporting clinicians during the care of actual trauma patients (see **Appendix II**). We have developed a methodology for validating the system, which involves both objective performance metrics as well as subjective case reviews by clinicians to identify whether there are any episodes whereby clinicians judge the system to be confusing or misleading. To this end, we also developed a database that will support multiple key needs: offline simulation testing of the new APPRAISE operational system; clinical validation (for regulatory approval) of the APPRAISE system; and a clinical trial studying the benefit of the system on the performance of the trauma team.

Finally, we provided support for the FDA pre-submission application being prepared by the BHSAI. This support included providing documentation, technical details, and technical review of the algorithm methodology, the validation methodology, potential regulatory claims, and the risk assessment.

KEY RESEARCH ACCOMPLISHMENTS

1. We completed data analysis of our prehospital trial, and a manuscript that reports these findings was accepted for publication in the journal SHOCK [17]. Briefly, we found that the hemorrhage identification algorithm performed as well in prospective real-time use on-board prehospital helicopters as it did in its earlier development phase. This supports the value of the technology for Combat Casualty Care. For this work done in collaboration with MGH, the BHSAI team was awarded the prestigious Heyman Service to America medal.

2. In close collaboration with our TATRC/BHSAI colleagues, we published an analysis of the value of heart rate variability metrics for the identification of major hemorrhage in prehospital settings [18,19].

3. With our collaborators at TATRC/BHSAI, we reported in the Journal of Neurotrauma in the a new analytic technique that uses routine vital signs to identify prehospital trauma patients with the highest risk of fatal traumatic brain injury [20]. This could be used for adjusting prehospital care protocols based on TBI risk, and could be used for mobilizing neurosurgical resources at the receiving facilities.

4. With our collaborators at TATRC/BHSAI, we received two US patents for our computing platform, the APPRAISE system, "Collection and analysis of vital signs" (US Patent #8,694,085 and #8,977,349)

5. With our collaborators at TATRC/BHSAI, we made a series of conference presentations about technical aspects of our work [21-26].

6. We completed Emergency Dept data collection for testing an investigational configuration of the advanced sensor suite (specifically, standard vitals signs plus muscle O2 saturation). With our collaborators at TATRC/BHSAI, we reported the value of the CareGuide tissue sensor in trauma patients for early identification of life-threatening hemorrhage [27]. This manuscript was selected as key article for practicing clinicians in the prestigious New England Journal "Journal Watch" review.

7. With our collaborators at TATRC/BHSAI, we have completed our review of hemodymanic patterns in our database of > 2,000 trauma patient. We identified that there are no significant temporal trends in heart rate in trauma patients with life-threatening hemorrhage; that tachycardia is a weak but significant indicator of hemorrhagic injury; and that the vast majority of patients with life-threatening hemorrhage develop hypotension within 30 min from the onset of monitoring. This was published in the journal Injury [28].

8. With our collaborators at TATRC/BHSAI, we developed three iterations of our hemorrhage risk algorithm. The final version outputs an ordinal category, and each category carries a higher association with a set of different hemorrhage-related clinical outcomes. Evaluation of the final version is found in **Appendix I**.

9. We led the implementation of an enhanced database of > 2,000 trauma patient physiological data, and associated clinical details and outcomes data, for algorithm assessment, which can be used for offline assessment of any investigational algorithm, e.g., heart rate variability, CRI, etc. This was shared with our collaborators at TATRC/BHSAI. This database can support multiple key needs: offline simulation testing of the new APPRAISE operational system; clinical validation (for regulatory approval) of the APPRAISE system; and a clinical trial studying the benefit of the system on the performance of the trauma team.

10. We designed a graphical user interface for the APPRAISE system and implemented an operational software prototype. This software electronically interfaces with GE monitor; displays vital signs through time; displays hemorrhage risk through time; and displays conditional clinical advisory messages. Technical details of this software are available in **Appendix II**. We developed a clinical testing protocol that was approved by the local IRB, and passed the safety requirements of the hospital's biomedical engineering department.

11. We supported the preparation of an FDA pre-submission application being prepared by the BHSAI. This support includes providing documentation, technical details, and technical review of the algorithm methodology, the validation methodology, potential regulatory claims, and the risk assessment.

A compilation of the published manuscripts supported by this award are found in **Appendix III**.

REPORTABLE OUTCOMES

Below we report progress towards key outcomes.

| | |
|---|---|
| Quarter 1: | In this quarter, we added new pre-hospital patients transported to the BIDMC; initiated our ED clinical study; and made progress on algorithm development (specifically for TBI). |
| Quarter 2: | In this quarter, we nearly completing chart review for the new pre-hospital patients transported to the BIDMC; screened our 98$^{th}$ subject for the Emergency Dept study; successfully worked with the vendor to enhance the investigational sensor, and prepared a report about our algorithm for TBI diagnosis using standard vital signs. |
| Quarter 3: | In this quarter, we focused on retrospective data analysis (i.e., analysis of vital sign patterns for TBI patients); ongoing clinical studies (i.e., data collection for Boston Medflight patients and data collection of MGH ED patients using the investigational muscle O2 sensor), and development of new technology (i.e., real-time vital sign analysis in the MGH ED). |
| Quarter 4 | As of this quarter, prehospital data collection was completed and we undertook analysis (in collaboration with TATRC/BHSAI); analysis should be complete by next quarter and a manuscript submitted by the quarter thereafter; ED data collection continues successfully; our advancement plan suggests we will have a productive subsequent 12 months. |
| Quarter 5 (Year 2, Q1) | Prehospital data analysis is now largely complete and we are preparing reports and presentations of the findings which demonstrate that multivariate analysis of prehospital vital signs can identify patients with substantial bleeding long before arrival at the receiving facility; ED data collection continues to proceed as planned; in collaboration with BHSAI/TATRC, we have **successfully deployed** our real-time analysis system within the hospital Emergency Dept. |
| Quarter 6 (Year 2, Q2) | Prehospital data **analysis is complete** and a manuscript has been prepared for imminent submission. ED data collection proceeds and we have received approval for increasing the number of total subjects to ensure enough of the subjects have substantial bleeding. Preparation of reports has been a high-priority, spanning: *1*) diagnostic and prognostic value of the Glasgow Coma Scale for casualties with life-threatening traumatic brain injury (in submission);*2*) techniques for identifying abnormal patterns in time-series data (in preparation); *3*) comparative analysis of heart rate variability measures versus routine vital signs for the early identification of substantial bleeding (in preparation); and *4*) development and validation of the system for real-time analysis in the MGH ED (in preparation). |
| Quarter 7 (Year 2, Q3) | Prehospital data analysis is complete and a manuscript has been submitted. A second manuscript, about how a mathematical model can be used to assess for high-mortality traumatic brain injury on the basis of routine vital signs, **has been accepted to the Journal of Neurotrauma**. ED data collection proceeds and we expect to close enrollment in approximately 2 additional quarters. Preparation of reports has been a high-priority, spanning: *1*) techniques for identifying abnormal patterns in time-series data (in |

| | |
|---|---|
| | preparation); *2*) comparative analysis of heart rate variability measures versus routine vital signs for the early identification of substantial bleeding (in preparation); and *3*) development and validation of the system for real-time analysis in the MGH ED (in preparation); we anticipate these will be submitted in the upcoming quarter. |
| Quarter 8 (Year 2, Q4) | In terms of prehospital functionality, we have begun planning for a potential prospective, outcomes trial. For this, we have examined the data to better understand the APPRAISE system's functionality for patients who receive CPR (during a real-time trial, it will be important that the algorithm recognize when a patient is ineligible for automated hemorrhage identification, i.e., CPR) and we have held conversations with a potential industry partner, Zoll. **We received a US patent for our prehospital system US 8,694,085.** We have submitted three new papers (related to real-time analysis in the hospital; a comparison of different techniques for assessing continuous vital sign data to account for temporal variability that is unrelated to blood loss; and an examination of when, in trauma patients, the oscillation in sinus heart rate is coupled to respiration and when it is not. |
| Quarter 9 (Year 3, Q1) | This quarter, we have submitted a new paper (related to heart rate variability in identification of bleeding patients). Our prehospital report (about real-time vital signs automated analysis) is undergoing major revision at the behest of the journal editor. We have a total of 5 conference papers accepted this quarter: 3 papers accepted by the IEEE EMBC and 2 presentations accepted by MHSRS. We have collected CareGuide data in over 600 trauma patients and expect to complete subject enrollment in Sept 2014 (Quarter 10). We are engaged in discussions with MRMCs technology transfer office and Zoll medical about next steps for our hemorrhage detection algorithms. |
| Quarter 10 (Year 3, Q2) | This quarter, we have revised the two journal reports described in the prior quarter, and we await the decision of the journal reviewers and editors. We presented the 5 conference papers described in the prior quarter. We completed data enrollment for the CareGuide SmO2 sensor and we are undertaking analysis of the final results. We have commenced data collection for another 710 trauma subjects, archiving routine vital sign data for these patients, which will allow for testing of any algorithm that uses routine vital signs and/or pulse oximetry waveform analysis and/or heart rate variability analysis. We have proceeded to contract negotiations with Zoll medical for technology licensing and a cooperative research agreement. |
| Quarter 11 (Year 3, Q3) | This quarter, **the two journal reports described in the prior quarter were accepted for publication**. We have conducted data analysis on the CareGuide SmO2 sensor project described in the prior quarter, with plans to complete a manuscript in the next quarter. We have also worked with the vendor of CareGuide, RMI, to re-analyze our dataset using their reportedly improved SmO2 algorithm (for estimating SmO2 based on our pre-existing, archive of spectroscopy data from the CareGuide sensor in trauma patients). |

| | |
|---|---|
| Quarter 12 (Year 3, Q4) | This quarter, we prepared a manuscript about the CareGuide SmO2 sensor. As well, **we had two related conference presentations submitted and accepted**: to the annual Society of Academic Emergency Medicine (SAEM) meeting and the regional New England Research Directors' annual conference for SAEM. **We had a second patent awarded (Collection and analysis of vital signs US 8,977,349).** We have initiated data collection on physiological data for an additional 700 trauma patients within the Emergency Dept. |
| Quarter 13 (Year 4, Q1) | **The abstracts from Q12 were presented in this quarter**. The associated full journal manuscript was prepared and submitted, describing our experience evaluating the CareGuide tissue sensor in trauma patients. We initiated an enhanced, exhaustive investigation of automated statistical analysis for early detection and decision-support of trauma patients, seeking to validate the method throughout our complete archive of > 1,500 trauma patients. |
| Quarter 14 (Year 4, Q2) | **The manuscript (based on the abstracts presented in Q13) was accepted for publication in the peer-reviewed journal Academic Emergency Medicine.** We continue our enhanced investigation of automated statistical analysis of vital signs using our largest dataset; this will establish the version of the algorithm to be deployed in our outcomes trial. We have initiated development of the GUI, starting with defining the specifications, in close collaboration with BHSAI. We developed a prospective testing plan, in close collaboration with BHSAI. |
| Quarter 15 (Year 4, Q3) | Enhanced investigation of automated statistical analysis of vital signs has revealed two issues that need to be addressed prior to clinical deployment of the system. First, the SPRT results in a "persistence artifact" whereby low-level hemodynamic abnormalities will (given enough time) be determined to be hemorrhagic, even if the abnormality is mild and unchanging. Second, SPRT also causes latency that is an issue in the hospital, whereby hemorrhagic patients are frequently moved in-and-out of the resuscitation bay quickly, therefore its properties that were valuable during prehospital transport are potentially disadvantageous in the hospital. Based on the evaluation of this algorithm in > 1500 trauma patients, we have designed a modification to the algorithm which will give rise to the version 2.0. As well, we have recruited the software engineer and initiated design of the APPRAISE system for real-world clinical deployment. |
| Quarter 16 (Year 4, Q4) | We have developed a modified methodology for automated statistical analysis of vital signs. Preliminary testing (through simulation of real-time use, inputting archived trauma patient data from our curated dataset of > 1500 trauma patients) suggests that this methodology will enhance performance. We plan to finalize this analysis and document it in a peer-reviewed manuscript within the next quarter. In addition, we have made progress in the design and implementation of the APPRAISE system that is being built for true interaction with clinicians, i.e., clinicians will view the decision-support in real time and can use the information to potentially enhance patient care. Specifically, we have developed an |

| | |
|---|---|
| | implementation plan and initiated software coding, in consultation with our collaborators at the BHSAI. |
| Quarter 17 (Year 5, Q1) | This quarter has focused on preparations for deploying the real-time APPRAISE system with a goal for completing a functional system that is ready for deployment within Quarter 19 (noting that there are regulatory hurdles in terms of the hospital dept. of Biomed engineering and possibly IRB issues that may also affect the deployment date). The architecture of this system was finalized and implementation is underway, as is the v.0 version of the so-called message library of possible messages provided to caregivers. In parallel, we have examined the function of the new hemorrhage detection algorithm ("2.0") to examine, case-by-case in our dataset of >1800 trauma patients, what types of messages should be crafted. This will culminate in a manuscript ready for submission in Quarter 18. |
| Quarter 18 (Year 5, Q2) | Work continues as per Quarter 17. An initial version of the real-time system intended for testing has now been implemented and is undergoing reliability testing. The message library is being implemented into the real-time system. A clinical advisory group of physicians and nurses has been convened for ratifying the message library. A web-developer has been brought in to the program to improve the graphical user interface such that it is suitable for real-time clinical use. In parallel, the operation of the new hemorrhage detection algorithm ("2.0") has yielded a set of new results and authoring of a new report has been initiated. |
| Quarter 19 (Year 5, Q3) | The "alpha" version of the real-time APPRAISE system complete with GUI has been completed, including the graphical enhancements of the professional graphic designer. A protocol for testing this version (with clinicians blinded to the output) has been submitted to the IRB. The new hemorrhage detection algorithm has been further improved ("2.1") now including a new approach named the "time-adjusted binormal distribution" method and final validation is being completed of this algorithm, which is expected to lead to a journal report and possibly a patent application. **Another report, examining hemodynamic patterns during trauma patient deterioration, has been submitted for publication.** |
| Quarter 20 (Year 5, Q4) | The "alpha" version of the real-time APPRAISE system complete with GUI has been subjected to off-line non-clinical user acceptance reviews (a panel of clinicians including emergency medicine and trauma surgery, and including nurses and doctors-in-training), with important evolution of the system's specifications. Enhancements of the labeling and messages have been undertaken. As well, new enhancements, related to the necessity for clinicians to input information that is perceived as important by the panel, are being implemented. A revised algorithm optimized for real-time use is being implemented, in coordination with the MRMC BHSAI. The APPRAISE system was evaluated by the MGH Biomedical Engineering Dept. leading to safety requirements for both "blind" clinical testing (no display to clinicians) as well as future live clinical testing were developed. We are currently adapting the system to meet all safety requirements (related to issues such as network |

| | |
|---|---|
| | security, patient privacy, and electrical safety). We have also initiated planning for future necessary regulatory approvals, with a focus on the US FDA. The research report submitted last quarter, examining hemodynamic patterns during trauma patient deterioration, has been revised for publication. |
| Quarter 21 (Year 6, Q1) | The implementation of the revised algorithm (for detecting hemorrhage risk) has been completed and we are now initiating extensive validation of that new algorithm. As well, we have initiated redesign of the graphical user interface so that it is optimized for the characteristics of the new algorithm (which is based on risk strata, rather than a continuous output). We have initiated planning meetings to develop a regulatory clearance strategy, working closely with the USAMRMC BHSAI. We have initiated the implementation of a database that will support multiple key needs: offline simulation testing of the new APPRAISE operational system; clinical validation (for regulatory approval) of the APPRAISE system; and a clinical trial studying the benefit of the system on the performance of the trauma team. The research report being revised last quarter, examining hemodynamic patterns during trauma patient deterioration, has been completed and submitted for publication. |
| Quarter 22 (Year 6, Q2) | After revising the algorithm (now based on risk strata), we have started documenting its implementation and performance in a manner that is compliant with FDA regulations and software development standards. In parallel, we have invested significant effort to identify the optimal regulatory strategy. It should be understood that this level of specification and documentation of our system and preparation of the supporting clinical data is highly resource intensive. Specifically, we have begun exploring the potential claims for our software; how to validate those claims in an FDA-compliant manner (in terms of clinical investigation), how to scope the software (in terms of what functionality versus the trade-off of additional need for documentation/testing); and overall the documentation needed for our regulatory strategy. This has involved working with an FDA consultant. In parallel, we continue to work on the database started in the preceding quarter and the offline simulation started in the preceding quarter, and are on-track to complete these tasks in the first-half of the upcoming quarter. |
| Quarter 23 (Year 6, Q3) | A pilot system has been fully implemented including a revised graphical user display, a revised message engine/library of guidance messages, and a revised version of the hemorrhage-risk algorithm. This quarter was focused on revising the database of > 2000 trauma patients that is maintained in collaboration with the BHSAI, so that different parameters (from the individual datasets that make up the master database) are represented in consistent format, so that parameter values are verified, and so that parameter availability is documented and, if feasible, optimized by de novo chart review. The value of this new revised database is to enable a final phase of offline system software testing and validation, and user acceptance testing, which involves showing a panel of clinicians the various clinical situations |

| | |
|---|---|
| | and the output of the system through time, to ensure that the decision-support is always suitable for optimal patient care. We intend to conduct this final phase of testing (as well as associated software fixes and other optimizations) in the remainder of the funded time. |
| Quarter 24 (Year 6, Q4) | We completed reorganization and documentation of the previously collected datasets of vital signs and other clinical and outcome data from over 1500 trauma patients has been reorganized with improved annotation to allow improved simulation analysis of the APPRAISE system; specifically, studying minute-to-minute functionality and the clinical conditions and interventions associated with any unexpected system behavior. We started work to merge the software functionality that has been previously developed for bedside use and the hemorrhage risk assessment algorithm developed in the MATLAB environment. This has involved a new generation of "quality assurance" measures that are intended to suffice from a regulatory standpoint, for investigational human use and also obtaining 510(K) clearance for the software. This new functionality is generally more conservative, i.e., the algorithm is designed to become inoperable except when suitable data are available. |
| Quarter 25 (Year 7, Q1) | We worked to finalize the computational methodology for the "Vital Signs Risks Level" whereby trauma patients are risk-stratified into one of four ordinal categories; to finalize the validation methodology for the algorithm whereby each risk stratum has a demonstrably higher association with a set of hemorrhage-associated outcomes, and the risks for each stratum will not change through time; and to attempt to specify an exhaustive set of possible failure scenarios for the VRL; this set of possible failure scenarios will be used in final user acceptance testing. |
| Quarter 26 (Year 7, Q2) | We implemented the final computational methodology into the user interface; supported the BHSAI preparation of an FDA pre-submission filing for the APPRAISE computational decision-support system; and we reviewed patient files to quantify the frequency with which the aforementioned failure scenarios occurred, and the frequency with which the risk-assessment algorithm offered a demonstrable advantage over routine vital signs. |

## CONCLUSIONS

This technology development project is nearly complete, yielding a functional, well-validated system suitable for pilot testing during clinical care. Prehospital validation of the hemorrhage identification algorithm was a success, operating as specified. Two US patents have been awarded for the prehospital analysis system. A method for prehospital TBI assessment has been reported in the J Neurotrauma. A report about the validation of our prehospital system, APPRAISE, for detecting life-threatening hemorrhage was published. Our investigations have also shown that during the initial evaluation of actual trauma patients, a novel sensor, the CareGuide tissue oximeter, added significant diagnostic power beyond routine vital signs for identifying patients with life-threatening hemorrhage. We have now accumulated an archive of over 2000 trauma patients' electronic data, which is being used to enhance our decision-support algorithms. We studied patterns associated with hemorrhage, and have identified that almost all patients with life-threatening hemorrhage

develop hypotension within 30 minutes of monitoring, whereas our technology can identify high-risk patients before the onset of hypotension. We have implemented a pilot GUI for the aforementioned APPRAISE system which can be used in a new series of clinical investigations whereby investigational decision-support is made visible to clinicians, to study whether this new generation of technology improves clinical management of trauma patients. We are on track for the USAMRMC BSHAI to have all the scientific and technical support necessary for their preparation and submission of a FDA pre-submission application by the end of this funded project.

REFERENCES

(Research manuscripts in **BOLD** represent the scholarly work supported by grant W81XWH-12-2-0019; copies of reports are available in **Appendix II**)


[1] Cothren, C. C., E. E. Moore, H. B. Hedegaard, and K. Meng. Epidemiology of urban trauma deaths: a comprehensive reassessment 10 years later. World Journal of Surgery. 2007 Jul; 31(7):1507-1511.

[2] Sauaia, A., F. A. Moore, E. E. Moore, K. S. Moser, R. Brennan, R. A. Read, and P. T. Pons. Epidemiology of trauma deaths: a reassessment. Journal of Trauma. 1995 Feb; 38(2):185-193.

[3] Riskin, D. J., 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. Journal or the American College of Surgeons. 2009 Aug; 209(2):198-205.

[4] Cotton, B. A., 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. Annals of Surgery. 2011 Oct; 254(4):598-605.

[5] Holcomb, J. B., and S. Gumbert. Potential value of protocols in substantially bleeding trauma patients. Current Opinion in Anaesthesiology. 2013; 26(2):215-220.

[6] Curry, N., and P. W. Davis. What's new in resuscitation strategies for the patient with multiple trauma? Injury. 2012 Jul; 43(7):1021-1028.

[7] Floccard, B., L. Rugeri, A. Faure, M. Saint Denis, E. M. Boyle, O. Peguet, A. Levrat, C. Guillaume, G. Marcotte, A. Vulliez, E. Hautin, J. S. David, C. Negrier, and B. Allaouchiche. Early coagulopathy in trauma patients: an on-scene and hospital admission study. Injury. 2012 Jan; 43(1):26-32.

[8] Chen, L., T. McKenna, A. Reisner, and J. Reifman. Algorithms to qualify respiratory data collected during the transport of trauma patients. Physiological Measurement. 2006 Sep; 27(9):797-816.

[9] Yu, C., 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. Journal of the American Medical Informatics Association. 2006 May-Jun; 13(3):309-320.

**[10] Reisner, A. T., L. Chen, and J. Reifman. The association between vital signs and major hemorrhagic injury is significantly improved after controlling for sources of measurement variability. Journal of Critical Care. 2012 Oct; 27(5):533.e1-10.**

[11] Chen, L., T. M. McKenna, A. T. Reisner, A. Gribok, and J. Reifman. Decision tool for the early diagnosis of trauma patient hypovolemia. Journal of Biomedical Informatics. 2008 Jun; 41(3):469-478.

**[12] Chen, L., A. T. Reisner, X. Chen, A. Gribok, and J. Reifman. Are standard diagnostic test characteristics sufficient for the assessment of continual patient monitoring? Medical Decision Making. 2013 Feb; 33(2):225-234.**

[13] **Reisner, A. T., 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. Applied Clinical Informatics. 2013 August 21; 4:392-402.**

[14] Chen, L., A. T. Reisner, A. Gribok, and J. Reifman. Exploration of prehospital vital sign trends for the prediction of trauma outcomes. Prehospital Emergency Care. 2009 Jul-Sep; 13(3):286-294.

[15] Soller BR, Ryan KL, Rickards CA, Cooke WH, Yang Y, Soyemi OO, Crookes BA, Heard SO, Convertino VA. Oxygen saturation determined from deep muscle, not thenar tissue, is an early indicator of central hypovolemia in humans. Crit Care Med. 2008 Jan;36(1):176-82.

[16] Wald, A. Sequential tests of statistical hypotheses. The Annals of Mathematical Statistics. 1945; 16(2):117-186.

**[17] Liu J, Khitrov MY, Gates JD, Odom SR, Havens JM, de Moya MA, Wilkins K, Wedel SK, Kittell EO, Reifman J, Reisner AT. Automated analysis of vital signs to identify patients with substantial bleeding before hospital arrival: a feasibility study. Shock. 2015 May;43(5):429-36.**

**[18] Edla S, Reisner AT, Liu J, Convertino VA, Carter R, Reifman J. Is heart rate variability better than routine vital signs for prehospital identification of major hemorrhage? The American Journal of Emergency Medicine. 2015 February; 33(2):254-261.**

**[19] Edla S, Reisner AT, Liu J, Convertino VA, Carter R 3rd, Reifman J. In reply to "Utility of shock index calculation in hemorrhagic trauma". Am J Emerg Med. 2015;33(7):978-9.**
**[20] Reisner A, Chen X, Kumar K, Reifman J. 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 Mar.**
**[21] Military Health Systems Research Symposium 2014 Oral Presentation: "Can Prehospital Vital Signs Help Identify Traumatic Brain Injury?"**
**[22] Military Health Systems Research Symposium 2014 Oral Presentation: "Can an Intelligent Alerting System Identify Patients with Substantial Bleeding during Transport to a Trauma Center?"**
**[23] Chen, X., Reisner A.T., Chen L, Edla S, Reifman J. The matching of sinus arrhythmia to respiration: are trauma patients without serious injury comparable to healthy laboratory subjects? IEEE EMBC 2014 Aug.**
**[24] Smith, JB, Reisner AT, Edla S, Liu J, Liddle S, Reifman J. A platform for real-time acquisition and analysis of physiological data in hospital emergency departments. IEEE EMBC 2014 Aug.**
**[25] Liu, J., Reisner AT, Edla S, Reifman J. A comparison of alerting strategies for hemorrhage identification during prehospital emergency transport. IEEE EMBC 2014 Aug.**
**[26] Edla S, Reisner AT, Liu J, Rubin JT, Thorsen JE, Kittell EO, Yeh DD, Reifman J. Does muscle oxygen saturation (SmO2) improve the association of initial vital signs with major hemorrhage in trauma patients? Society of Academic Emergency Medicine. 2015; San Diego, USA.**
**[27] Reisner AT, Edla S, Liu J, Rubin JT, Thorsen JE, Kittell E, Smith JB, Yeh DD, Reifman J. Muscle oxygen saturation improves diagnostic association between initial vital signs and major hemorrhage: a prospective observational study. Academic Emergency Medicine. 2016 March.**
**[28] Reisner AT, Edla S, Liu J, Liu J, Khitrov MY, Reifman J. Tachycardic and non-tachycardic responses in trauma patients with haemorrhagic injuries. Injury. 2018 Apr.**
[29] Chen L, Reisner AT, Gribok A, Reifman J. Is respiration-induced variation in the photoplethysmogram associated with major hypovolemia in patients with acute traumatic injuries? Shock. 2010 Nov;34(5):455-60.

**Appendix I:**
**Key findings about the "VRL" (Vital-sign Risk Level), which was the hemorrhage risk category assigned to the subject by the final version of the automated algorithm, for each subject in our merged dataset of over 2,000 trauma patients**

Out of 66 with massive transfusion, 43 (65%) were identified as Level III, IV, or patent hypotension, in the initial 15 min:

- 21 massive transfusion cases detected on the basis of hypotension
- 11 massive transfusion cases detected (Level IV) prior to patent hypotension
- 12 massive transfusion cases detected (Level III) prior to patent hypotension
    - Most of these cases (8 of 12) were short durations of monitoring (<15 min)
    - Only a minority (4 of 12) detected (Level III) were longer duration with adequate data

Out of 66 with massive transfusion, 23 (35%) were ***not*** identified as Level III, IV, or patent hypotension, in the initial 15 min:

- 8 appeared to be truly "missed" based on reassuring vital signs
- 10 had some form of objectively incomplete vital signs (e.g.: absence of BP data; missing exogenous BP; or pulselessness)
- 4 massive transfusion cases were detected at a delayed time-point, after the initial 15 min
    - Most (3 of 4) had incomplete vital-sign data (e.g.: absence of any data; missing exogenous BP; or pulselessness) associated with the delay in detection
- We did not identify measurement artifact (e.g., spurious BP or HR that is deceptively normal) associated with any of the missed detections

Out of 10 patients with 24-hr RBC = 5 units, 4 (40%) were identified as Level III, IV, or patent hypotension, in the initial 15 min

- 1 of the hemorrhage cases detected on the basis of hypotension
- 3 of the hemorrhage cases had subtle "drift" patterns that were correctly detected by the VRL

Out of 10 patients with 24-hr RBC = 5 units, 6 (60%) were ***not*** identified as Level III, IV, or patent hypotension, in the initial 15 min

- 1 appeared to be truly "missed" based on reassuring vital signs
- 4 had some form of objectively incomplete vital signs (e.g.: absence of BP data; short records; missing exogenous BP; or pulselessness)
- 1 of the hemorrhage cases had subtle "drift" pattern correctly identified, but delayed after 15 min
- We did not identify measurement artifact (e.g., spurious BP or HR that is deceptively normal) associated with any of the missed detections

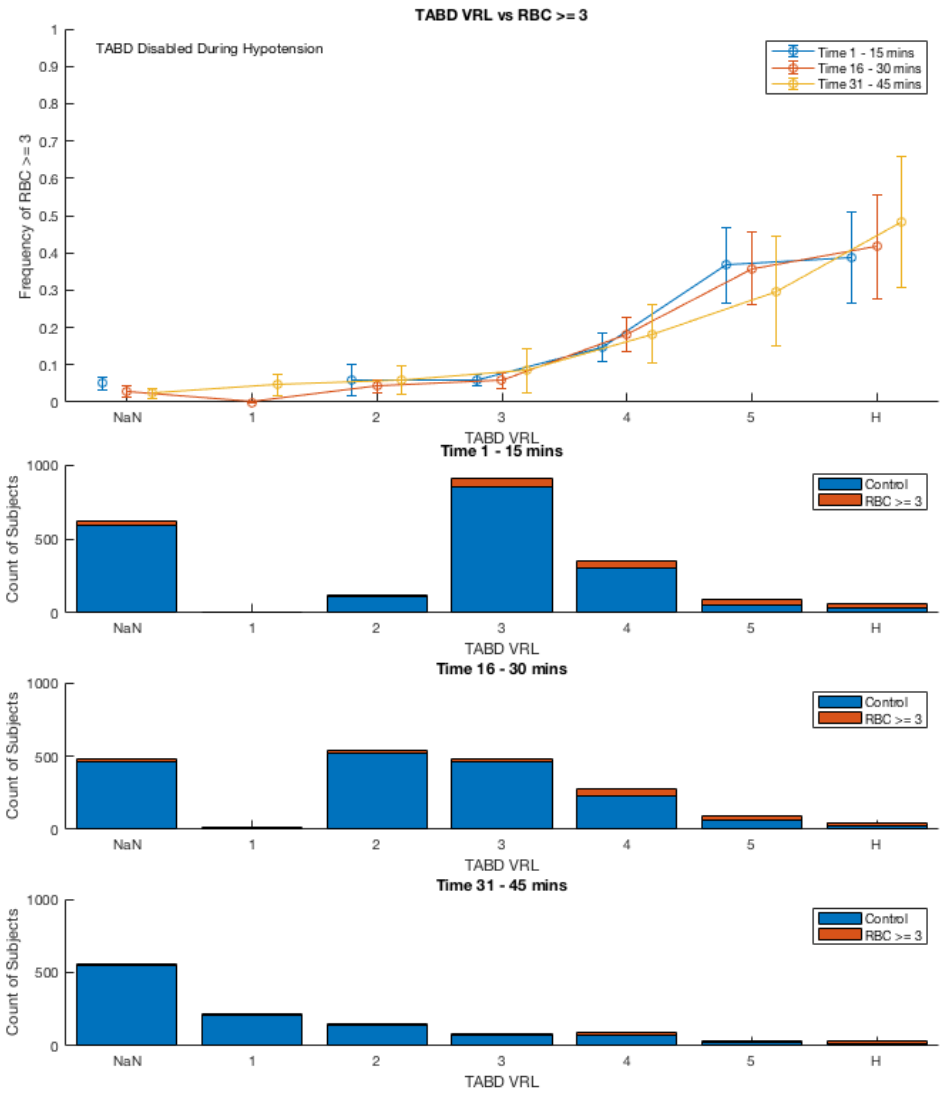
Observations

- The majority of hemorrhage cases were detected
    - Half of the detections were associated with patent hypotension
    - Half represent enhanced detection compared with basic hypotension threshold

- Objectively incomplete vital signs (e.g.: absence of BP data; missing exogenous BP; or pulselessness) is associated with the vast majority of missed or delayed detections
    - Incomplete data was identified in 14 of 66 massive transfusion patients with missed (n=10) or delayed (n=4) detection

    - Incomplete data was identified in 5 of 10 24-hr RBC=5 units patients with missed (n=4) or delayed (n=1) detection
    - *For many massive transfusion patients with correct detection < 15 min*, there was often a small delay between APPRAISE detection versus clinical recognition of significant hemorrhage, on the basis of additional information, e.g., visually apparent bleeding; pre-hospital hypotension or imaging
    - Incomplete data was identified in 8 of 66 massive transfusion cases with only Level III detection

- Missed detections in records with duration > 15 min and without missing information are relatively rare
    - 8 of 66 massive transfusion cases appeared to be truly "missed" based on reassuring vital signs
    - 1 of 10 24-hr RBC=5 units patients had reassuring vital signs
    - We did not identify measurement artifact (e.g., spurious BP or HR that is deceptively normal) associated with any of the missed detections



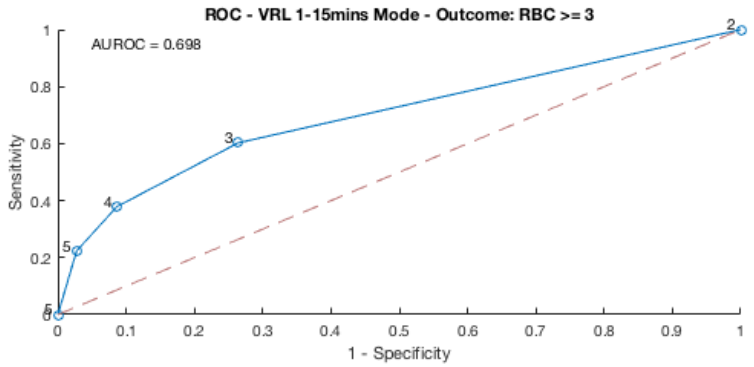

**VRL Positive Predictive Value (PPV) for *24-hr RBC >= 3 units***

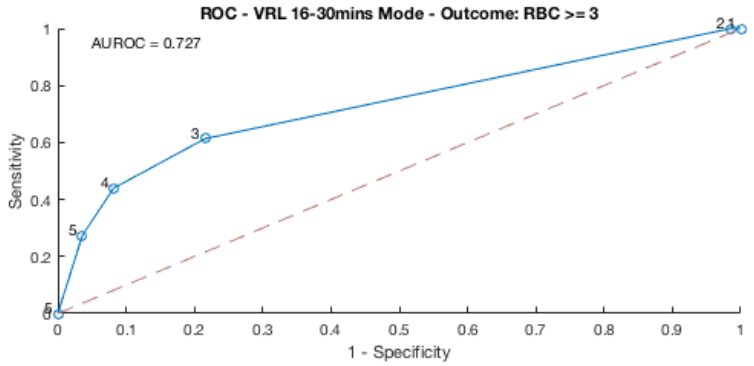
Top panel: PPV for each VRL at the end of 15 min (blue); 30 min (orange); and 45 min (yellow).

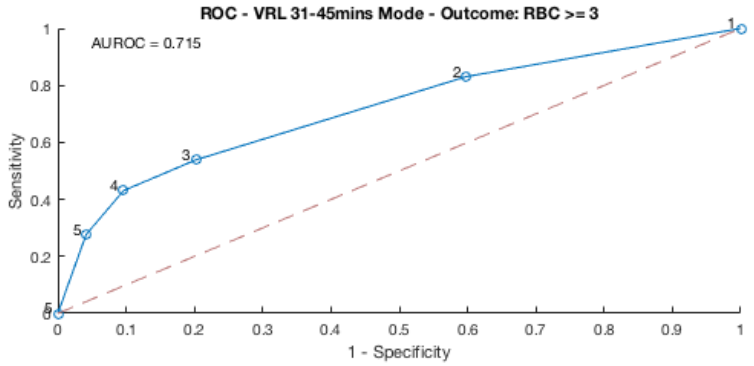
Bottom panels: Associated histograms for each VRL, showing total numbers of patients with RBC >= 3 units and RBC < 3 units at each VRL for the different time intervals of vital-sign data. Also shown are the total number of patients without VRL due to inadequate data ("NaN").

Key findings:

- Monotonically increasing PPV? Generally, yes [caveat: VRL 1 not distinct from VRL 2]
- Is PPV similar through time (i.e., similar PPV between each set of blue/orange/yellow)? Yes, after accounting for error bars



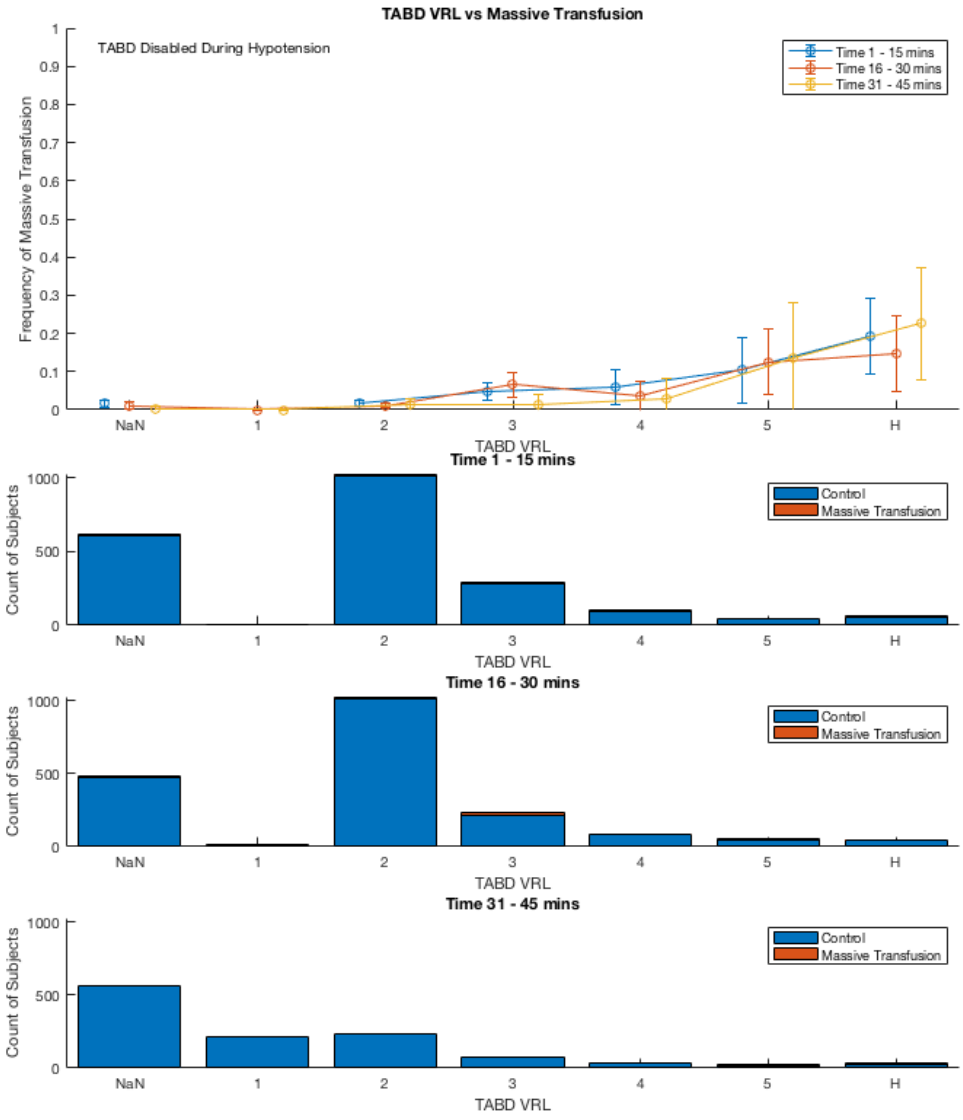




**<u>Receiver Operator Receiving Curves for VRLs and Outcome of *24-hr RBC >= 3 units*</u>**

Key finding:

- ROC AUC stable through time



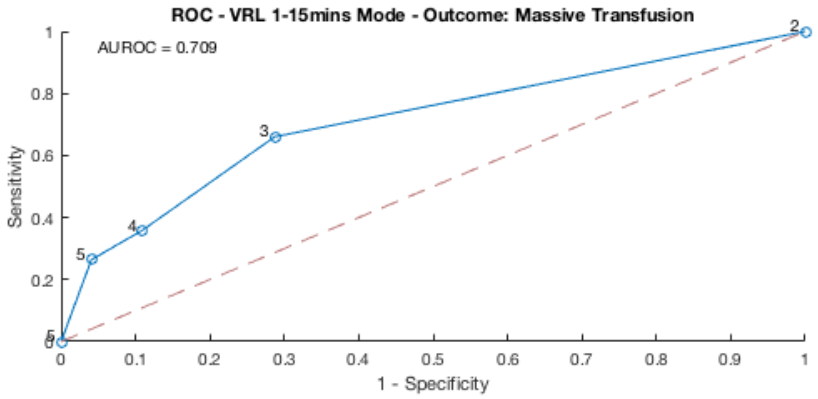

**VRL Positive Predictive Value (PPV) for *Massive Transfusion***

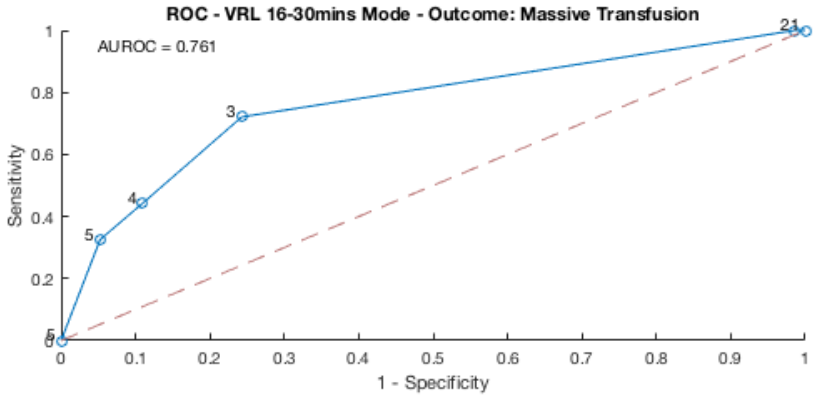
Top panel: PPV for each VRL at the end of 15 min (blue); 30 min (orange); and 45 min (yellow).

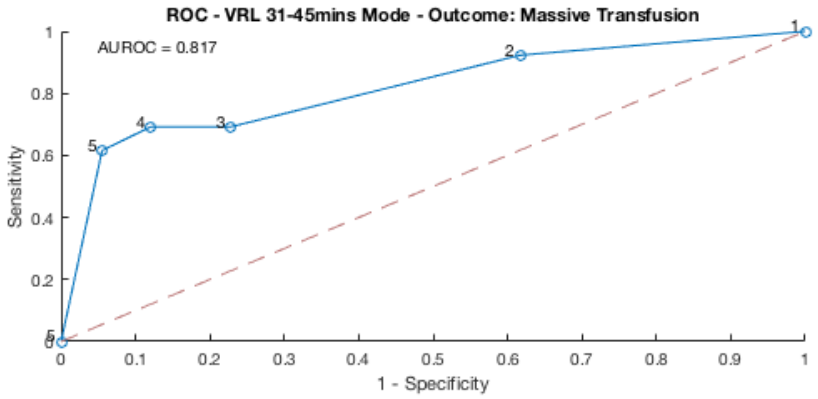
Bottom panels: Associated histograms for each VRL, showing total numbers of patients with massive transfusion and without massive transfusion at each VRL for the different time intervals of vital-sign data. Also shown are the total number of patients without VRL due to inadequate data ("NaN").

Key findings:

- Monotonically increasing PPV? Generally, yes [caveat: VRL 3 not distinct from VRL 4]
- Is PPV similar through time (i.e., similar PPV between each set of blue/orange/yellow)? Yes, after accounting for error bars



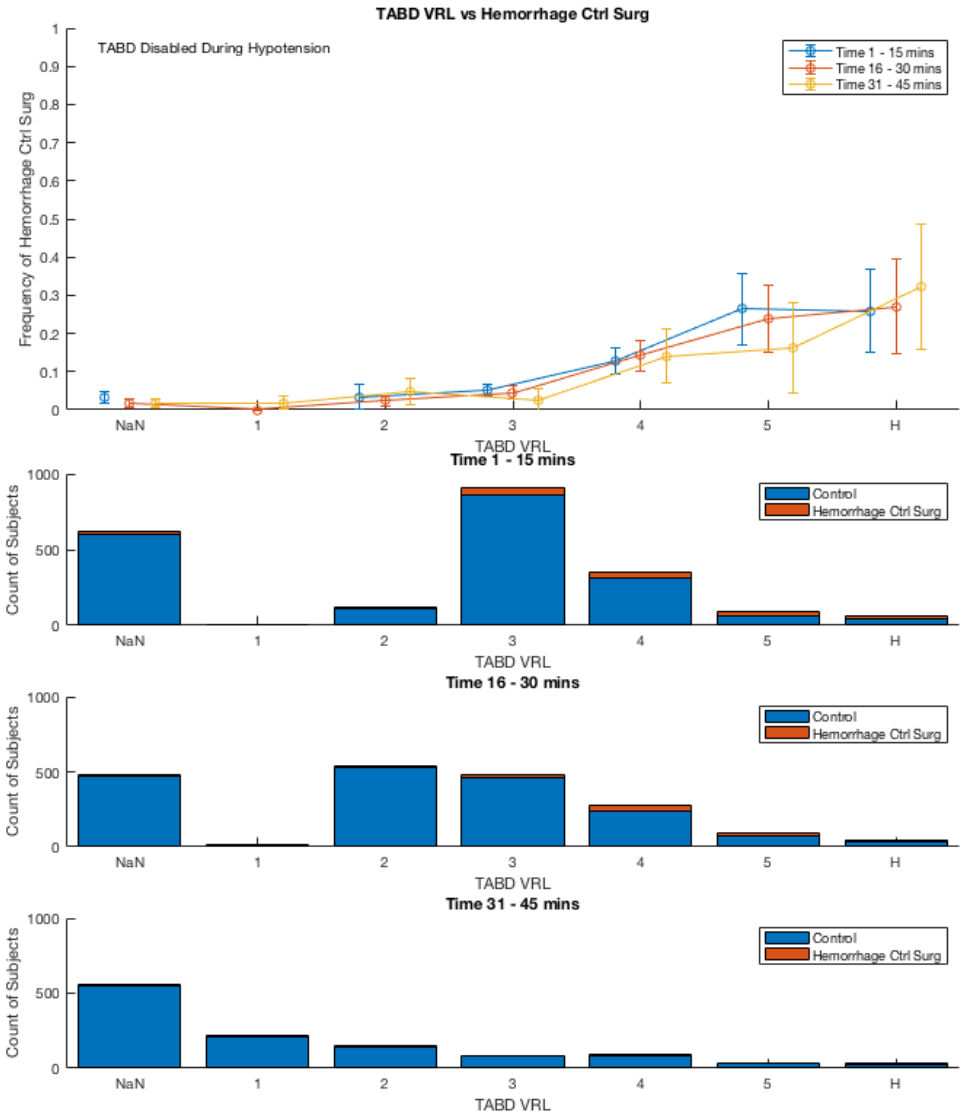




**<u>Receiver Operator Receiving Curves for VRLs and Outcome of *Massive Transfusion*</u>**

Key finding:

- ROC AUC appears to increase through time



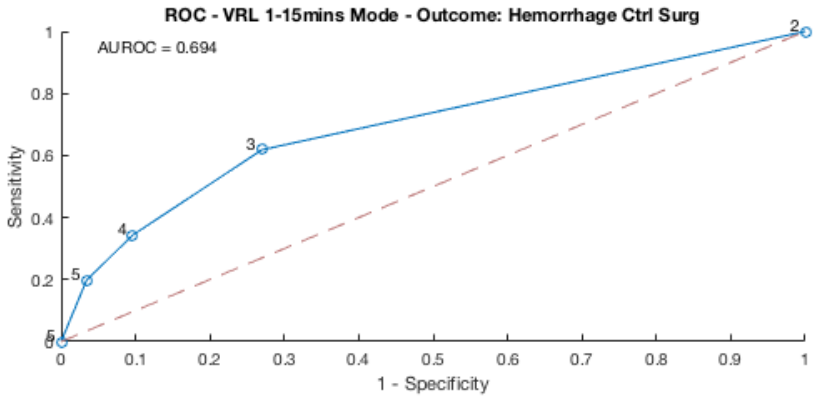

**VRL Positive Predictive Value (PPV) for *Hemorrhage-control Surgery (within 4 hrs of arrival)***

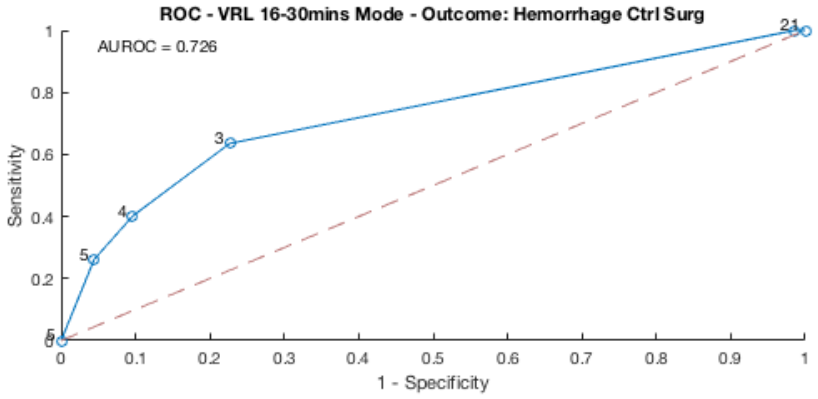
Top panel: PPV for each VRL at the end of 15 min (blue); 30 min (orange); and 45 min (yellow).

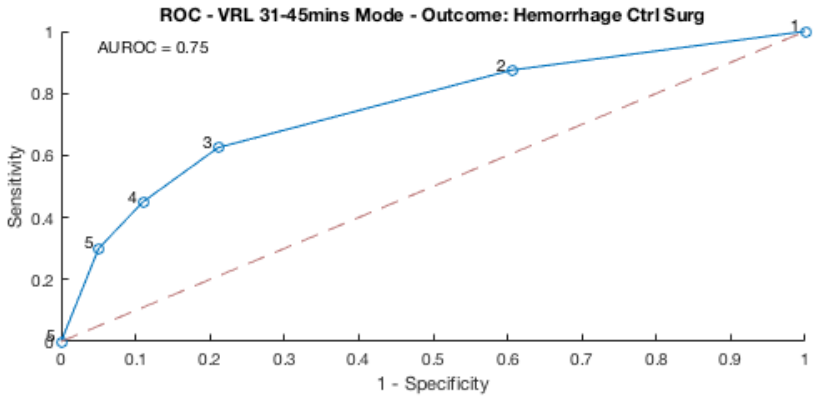
Bottom panels: Associated histograms for each VRL, showing total numbers of patients with hemorrhage-control surgery and without at each VRL for the different time intervals of vital-sign data. Also shown are the total number of patients without VRL due to inadequate data ("NaN").

Key findings:

- Monotonically increasing PPV? Generally, yes [caveat VRL 2 not distinct from VRL 3]
- Is PPV similar through time (i.e., similar PPV between each set of blue/orange/yellow)? Yes, after accounting for error bars




**<u>Receiver Operator Receiving Curves for VRLs and Outcome of *Hemorrhage-control Surgery (within 4 hrs of arrival)*</u>**

Key finding:

- ROC AUC appears to increase, slightly, through time

## Summary of Key Findings: VRL as ordinal rating of set of outcomes

- Each increased VRL *is* associated with increased risk for each of the trio of hemorrhage outcomes
- For each outcome, there exists a VRL pair with similar PPV (i.e, no ordinal increase) but *this is likely due to random chance*
    - 24-hr RBC >= 3 units: VRL 1 not dissimilar to VRL 2
    - Massive transfusion: VRL 3 not dissimilar to VRL 4
    - Hemorrhage-control surgery: VRL 2 not dissimilar to VRL 3
- Mortality was not predicted effectively by the VRL, likely due to head trauma (results not shown)
- Generally speaking, the results support that the VRL provides an ordinal risk level for set of hemorrhage outcomes
- Generally speaking, the results support that the risk associated with each VRL does *not* vary through time
- Increasing AUC through time indicates fewer patients with mid-tier VRL

# Appendix II
# APPRAISE GUI Software Prototype Documentation

Updated 3/27/17

Jeffrey Peterson



# INTRODUCTION

Automated clinical decision support systems may improve clinical protocol compliance in high acuity settings such as Trauma. To assess the effect of automated CDS in Trauma, the APPRAISE project aims to:

- Develop a protocol-guidance system and clinician-facing display
- Quantify errors of decision-making and protocol compliance
- Assess the impact of automated decision-support

The APPRAISE project is creating a prototype software application (titled APPRAISE) to display key information to clinical staff in the patient room as well as to persist data for retrospective analysis.

The APPRAISE software prototype application aims to evolve into a configurable protocol-guidance system for any diagnostic algorithm and any treatment protocol as the system matures.

# SYSTEM DESCRIPTION

The first implementation of the APPRAISE system will be in Acute Bay 1 (Trauma) of MGH's ED. The system will consist of hardware and software: a single computer physically located in Bay 1. The computer will be connected to the patient's physiological monitor, a GE Solar 8000i, via a serial connection. The system will feature a wall mounted, large format, commercial-grade monitor to display the user interface.

The software's user interface will display:

- Current **vital signs** sourced from the bedside monitor as well as vitals plotted over time to illustrate trends
- Current output of the APPRAISE **algorithm** computation (Hemorrhage Index) as well as a plot of the trend of the index
- Protocol guidance **messages** applicable to the current state of the patient
- **Timer** for the patient's length of stay


*Figure: A mockup of the APPRAISE user interface on a large format display*

In addition to the clinical display, the software will persist data required for retrospective analysis. The APPRAISE GUI system attempts to be highly configurable. The configuration options are specified in a file titled app_configuration.json. The file is included by default but can also be customized by the user.

The APPRAISE system will sit passively in the patient bay (i.e. no user inputs) collecting data and displaying the metrics, plots and guidance messages listed above. Interaction with the software and display will not be possible in the initial versions of the system. A future feature will be a discrete touchscreen monitor to accept input data.

# HARDWARE SETUP

The APPRAISE system consists of a computer, monitor, and serial connection infrastructure. Due to the Solar being mounted on a boom in the installation area (Acute Bay 1), the serial cable will be connected with “paired” serial-to-Ethernet adapters connected over a point-to-point Ethernet cable, shown in the figure below. Both the Solar and Computer will operate as if they are directly connected to one another with no knowledge of the serial-to-Ethernet adapters.

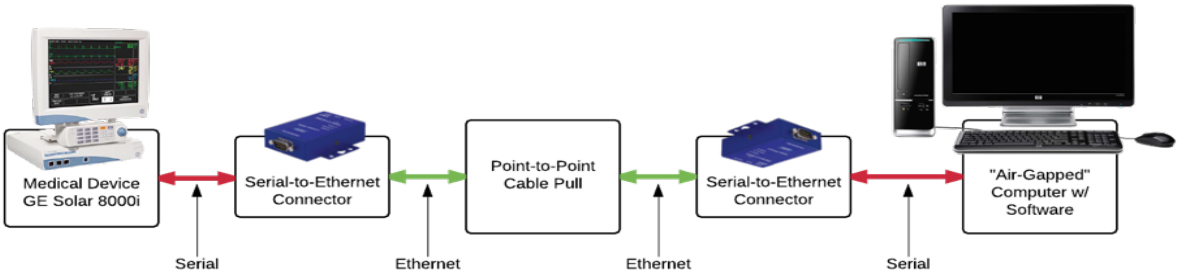



*Figure: The proposed installation for the serial connection to the GE Solar 8000i*

The display will be a large format flat screen - 48” Samsung DME Series - 24 Hour Rated Commercial LED Display. The screen will be mounted on the wall in proximity of the location typically used for charting.

*Figure: A rendering of the planned system installation in Acute Bay 1.*

Installation Inventory:

- GE Solar 8000i Patient Monitor (Existing)
- 2 x B&B Electronics VESP211 Serial-to-Ethernet Converters
- Samsung DM48E Commercial LED Display
- Dell Optiplex 7040 Micro Form Factor CTO serial port

# SOFTWARE SETUP

The APPRAISE application is a standalone, cross platform desktop application built using the Electron framework. The Electron framework provides a JavaScript interpreter (node.js) and DOM rendering engine (Chromium) to run a node.js-backed web view as a native application. This concept is similar to having a

dedicated, lightweight web browser that only has one website (a web browser without the web).

Electron uses a multi-process architecture with each view (or window) contained it's own process with a "main" controller process as described here. Electron's entry point is `main.js` which serves to start the APPRAISE GUI, the system tray menu, and configuration window. To open the APPRAISE window, we instruct the main process to open a new window from the resource location `index.html`. This HTML file has a script tag in the body with the source attribute pointing to `./index.bundle.js`, which is our GUI entry point. The file `./index.bundle.js` is a self contained bundle of all JS files for the GUI.

The APPRAISE GUI is a React-based webapp. APPRAISE is written with React's JSX and uses ES6+ language features such as the `import` keyword. For this code to run with Electron, it must first be "transpiled" to ES5 JS. To accomplish this, the babel.js library is invoked using a Webpack based build process. The configuration of this process is found in `webpack.config.js`, `webpack.config.main.js`, and `.babelrc`. It can be run using `` `./node_modules/.bin/webpack --config webpack.config.js` `` (or with the custom `package.json` script `` `npm run pack-app` ``). This process generates the `index.bundle.js` file.

The APPRAISE GUI utilizes the following libraries:

http://electron.atom.io/ - Electron runs the webapp as a native app
https://facebook.github.io/react/ - View library for writing UI components
http://redux.js.org/ - State container library that pairs well with React
https://github.com/EmergingTechnologyAdvisors/node-serialport - Serial port interaction from Node.js
https://d3js.org/ - Data visualization library for plots and animated graphics
https://babeljs.io/ - "Transpiling" JSX and ES6+ to ES5 JS
https://webpack.js.org/ - JS script bundling and task runner
https://github.com/electron/electron-rebuild - Compile Node modules for different architectures
https://github.com/electron-userland/electron-builder - Package and distribute Electron on different archtectures and operating systems
https://github.com/winstonjs/winston - Logging framework

See Appendix E for instructions installing and running APPRAISE.

# SOFTWARE COMPONENT FUNCTIONALITY

## Overview

The APPRAISE GUI system's core functionality is to record and calculate the necessary information to populate the user interface. The user interface consists of the following components: vital signs, hemorrhage algorithm plot, clinical protocol reminder messages, and length of stay timer. The raw vital signs data from the patient monitor is used to calculate Parameters (e.g. two minute median filtered heart rate) and Algorithms (e.g. probability of hemorrhagic injury). The Parameter and Algorithm data are in turn used to determine which clinical protocol Messages should be displayed, if any.

The data needed to plot the vitals signs, algorithm, and determine which messages to display are stored in a

tabular data structure titled the Flowsheet. The configuration information needed to populate the Flowsheet is found in a JSON file titled app_configuration.json.

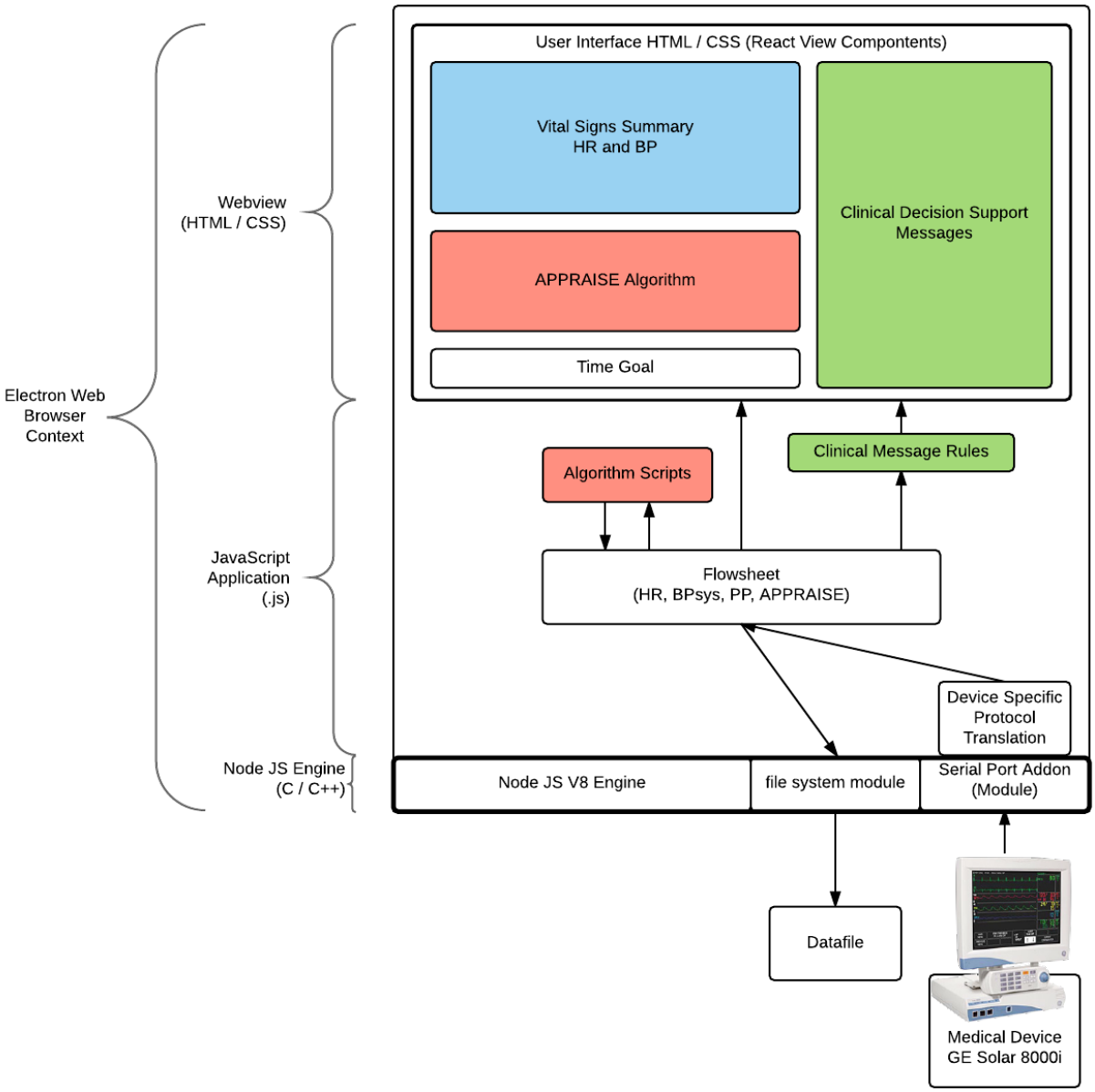



*Figure: A high-level block diagram of the data-flow in the APPRAISE GUI renderer process.*

The block diagram above illustrates the data flow for the APPRAISE system. Data enters the system at the Device Interface that connects to the GE Solar 8000i (or a playback interface). The raw data captured in the Flowsheet Buffer. The raw data is processed and stored in the Flowsheet. This is the central point of the application. The parameters and algorithms are calculated every time a new row is created. The Flowsheet is the source of data for the UI, including the plots and messages.

See Appendix D for an example of the application's state (a Redux data store).

## Flowsheet (Parameters and Algorithms)

One key piece of the APPRAISE GUI system's core functionality is to create a tabular record of data calculated from the input patient data. This tabular record is titled the Flowsheet. The columns of the Flowsheet correspond to Parameter and Algorithm values. Parameters represent data aggregated from the patient monitor such as `hrMedian2min` (the median value of the previous two minutes of heart rate data from the patient monitor) and `nibpSysCount` (the total count of the systolic noninvasive blood pressure measurements). Flowsheet columns are also used for Algorithms, such as `sysBp_diff` (the difference between the last two systolic blood pressure measurements.) Algorithms provide the user with a mechanism for creating columns that are calculations derived from one or more Parameters or even from other Algorithms.

Each patient encounter with the monitor creates a new Flowsheet. The individual patient encounters are titled Sessions. When the patient monitor detects a gap in data which can be reasonably associated with a new patient, the APPRAISE system will start a new Session with a new Flowsheet.

**Flowsheet for session 20170322_1539_playback=923**

| Time | hrMedian2min | hrMedian1min | hrCount | hrLastValue | hr2minStdDev |
|---|---|---|---|---|---|
| 3:39:26 PM | 129 | 129 | 8 | 132 | 2.345207879911715 |
| 3:39:36 PM | 132 | 132 | 18 | 133 | 2.5 |
| 3:39:46 PM | 133 | 133 | 28 | 135 | 3.027720567809173 |
| 3:39:56 PM | 135 | 135 | 39 | 135 | 2.8209789824564235 |
| 3:40:06 PM | 135 | 135 | 48 | 139 | 3.2605997019975734 |
| 3:40:16 PM | 136 | 136 | 58 | 142 | 3.6697821232795906 |
| 3:40:26 PM | 136 | 137 | 68 | 145 | 4.5803807203389315 |
| 3:40:37 PM | 137 | 140 | 79 | 144 | 4.960345879572592 |
| 3:40:46 PM | 139 | 142 | 88 | 145 | 5.058120363491784 |
| 3:40:56 PM | 140 | 144 | 98 | 145 | 5.172140875190822 |

*Figure: The first six columns of a Flowsheet for a data playback session.*

The rows of the Flowsheet are created once per interval of time. This time interval is configurable – the default value is 10 seconds. Since the Flowsheet is only interested in data from the patient monitor that is an aggregate of the raw data, the raw data needs to be held in between Flowsheet intervals. This holding area is called the FlowsheetBuffer. The FlowsheetBuffer has an array for each unique Metric as defined by

ISO 11073-10101 and maintained by NIST Harmonized Rosetta Terminology mapping. Metrics names not described by the standard are inferred from the MDC naming pattern. At the end of each row time interval (i.e. once every 10 seconds) the FlowsheetBuffer is read. The values in the buffer are reduced to a single value that is added to the appropriate table cell in the Flowsheet. Algorithms require the value of other table cells to calculate their values (e.g. hemorrhage index inputs: heart rate, systolic and diastolic NIBP) are evaluated after the Parameters and the resulting values added to the Flowsheet. Algorithms can use any column to its left (listed before it in the configuration) as input to its calculations.

## Messages

The resulting data in the Flowsheet is used to determine which Clinical Messages should be displayed. The clinical messages are short reminders, suggestions, or alerts that are intended to assist clinical staff in adhering to the Emergency Department's official protocols. Every 30 seconds, the Message Library is used to determine if a new message(s) should be added to the UI. Each message relies on three factors to determine if it should be added to the UI: if the message's logic is "active" (i.e. the boolean logic in the app_configuration.json file resolves to true), if the message has not been shown recently (specifically, has the message been displayed within the its refractory time), and if the message critical.

## Device Interface

The data input to the application is a serial port connection to the GE Solar 8000i. This is accomplished using the ge-serial-node repository. This JavaScript module will send the request packet to the patient monitor every two seconds (described in detail in Appendix A). When a full message is received, the module will translate the contents of the message as described in the GE serial specification manual. The translated responses from the monitor, now a JS object with human readable keys/values, are emitted from the module (ge-serial-node implements EventEmitter). The application state will listen for these events and dispatch a state update action when received (NEW_FLOWSHEET_BUFFER_VALUE). For testing and evaluation purposes, the serial interface can be replaced by a playback interface and a "fake" data interface. The playback interface will emit data from a patient data file. The fake data interface will emit heart rate and blood pressure data in the shape of an arbitrary sinusoid over time (HR and BP will simply oscillate up and down).

## app_configuration.json

This configuration file describes the parameters to be recorded (parameter_library), the algorithms to be calculated (algorithm_library), the messages to be displayed (message_library), and the UI plots (plots). The parameters and algorithms entered in the file directly map to the columns of the Flowsheet. Each entry in the message_library describes a clinical message to be displayed if the conditions are met. This is the central point of business logic in the system.

The following subsections will describe the configuration of each section of the app_configuration.json file.

*Message Library*

Messages help users adhere to clinical protocols and to operate the system, as described above. This array of objects contains the configuration of each message. A message is displayed if it is active and eligible. A message is active when its logical statement evaluates to true. A message is eligible when it has not been displayed within its refractory time (i.e. the minimum duration between a message being repeated). If one or more active and eligible messages are critical, all critical messages will be displayed. If multiple non-critical messages are active and eligible, the first message in order of the message library array will be used.

The message logic uses a simple, homegrown logical syntax to encode the clinical rulesets. The logic statements are recursively evaluated so multiple comparisons can be nested. The logic commonly starts with a tuple representing &&, ||, >, ≥, <, ≤, ==, or !==. These logic tuples must be an array of length three with first element being one of the following set: 'EQ', 'NEQ', 'GT', 'GTE', 'LT', 'LTE', 'AND', 'OR'. AND and OR length can be greater than three.

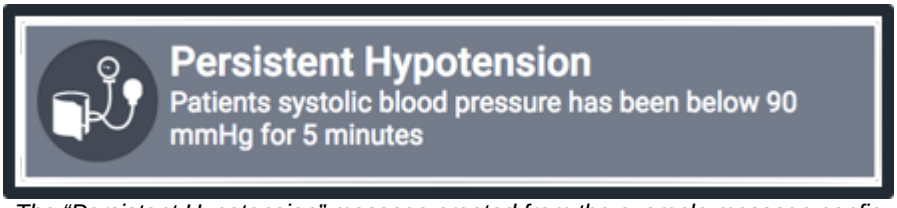



*Figure: The "Persistent Hypotension" message created from the example message configuration.*

Example:

```
{
     "name": "Persistent Hypotension",
"text": "Patients systolic blood pressure has been below 90 mmHg for 5 minutes",
     "critical": true,
     "refractory": 300000,
     "icon": "bp_cuff",
     "logic": ["LT", { "parameter": "nibpSys5minMax" }, 90]
}
```

*name* – The title of the message on the display.

*text* – The subtitle of the message on the display that provides more detail.

*critical* – Boolean - Is the message critical (displayed on the GUI without delay and highlighted with a brighter colored background)

*refractory* – The minimum time (in milliseconds) permitted between the same message being displayed again.

*icon* – The name of the icon to use. Corresponds to SVG files in app/resources/icons as well as FontAwesome icon names.

*logic* – The tuple representing the clinical ruleset that pertains to a protocol reminder message.

*logic[0]* – Must be [EQ, NEQ, GT, GTE, LT, LTE, AND, OR]

*logic[1..n]* – Must be a number, the string "elapsedTime", or an object with one key equal to "parameter", "algorithm", "variable", or "constant". The value of the object will be looked up the appropriate "library".

*Parameter Library*

Columns of the flowsheet – This array describes how the data from patient monitor should be collected, aggregated, and reduced to a cell on the Flowsheet. The parameters are referenced by a unique key. The message_library, algorithm_library, variable_library, and plots can reference the parameters by key.

Example:

```
{
    "key": "hrMedian2min",
    "metric": "MDC_ECG_HEART_RATE",
    "type": "parameter",
    "aggregation": "median",
    "time_filter": 120
}
```

*key* – The arbitrary titled assigned to parameter. Must be unique. This key is used by message_library, algorithm_library, variable_library, and plots to incorporate a parameter.

*metric* – The data type that this parameter corresponds to such as heart rate (`MDC_ECG_HEART_RATE`). The metric is a string that device interface will emit with each value. The device interface uses NIST hRTM (harmonized rosetta terminology mapping). If the data.metric from the serial interface packet matches this parameter name, the data is recorded in the flowsheet buffer.

*type* – [parameter, alarm, meta] Used to differentiate which area of the patient monitor data packet to read from. This is an artifact of the GE Serial interface and should be refactored out one day.

*subtype* – Only required when *type* is alarm or meta. Again used for reading from the GE serial data packet. Represents a tightly coupled feature that should be fixed.

*aggregation* – [average, median, stdDev, count, countAll, max, min] The array of data from the interface with the corresponding metric within the specified time window is aggregated with the method specified here. This method will determine how the parameter values in the flowsheet buffer should be reduced.

*time_filter* – [-1, 0, n, null] The time window the parameter would like to aggregate for a row in the flowsheet. The value of time_filter is similar to array.slice. To specify that the flowsheet row should be the result of a calculation using all data use *time_filter: 0*, the most recent measurement only use *time_filter: -1*, a duration of time such as prior 120 seconds use *time_filter: n*, or all data received since the last row was calculated, i.e. tPrev < window < t, use *time_filter: null*.

*Algorithm Library*

Columns of the flowsheet – The algorithm library lists the configurations of the columns that require more than what the Parameters are able to provide. Algorithms can use any column to its left (any Parameter or Algorithm listed before it in the configuration) as input. The algorithm scripts export a single default function that takes the parameters `config` and `flowsheet` (i.e. `function myAlgorithm(config, flowsheet)`). The `config` parameter is passed the `parameters` object as its argument and is used to provide mappings between names internal to the script and the Flowsheet as well as settings like thresholds.

Example:

```
{
    "key": "hypovolemia_risk",
    "parameters": {
        "hrKey": "hrMedian2min",
        "bpSysKey": "nibpSysRowAvg",
        "bpDiaKey": "nibpDiaRowAvg",
        "bpCountKey": "nibpSysCount"
    },
    "path": "appraise"
}
```

*key* – A unique key that is referenced by algorithms listed prior and message logic.

*parameters* – A configuration mapping that attempts to abstract the algorithm script from the parameter/algorithm Flowsheet column name. This object is passed into the js script as the algorithm config. Keys in the parameters object are unique to the algorithm script. It is used to match vital signs to flowsheet columns as well as to set thresholds. The algorithm script is responsible for enforcing proper parameter values.

*path* – A file path to the script containing the algorithm. The path omits the file extension ".js" and is relative to the `/local_modules/` directory.

*Variable Library*

Message logic, found in `message_library.logic`, can be very verbose and repetitive. There are often common logic statements that multiple messages leverage. The variable library provides the user with a method of abstracting common patterns that they would like to reuse. The variable library improves readability and composability or the message configuration. For example, the risk strata for the APPRAISE hemorrhage algorithm are verbose and used frequently. Below is the logic for `HEME_PROB_MEDIUM`, the 2nd highest of 4 levels of risk. This statement can be injected by referencing it with the "variable" keyword, i.e. `{ "variable": "HEME_PROB_MEDIUM" }` would substitute the "variable" property value with the corresponding entry in the variable library.

Example:

```
{
     "HEME_PROB_MEDIUM": ["AND",
          ["GTE",
               { "algorithm": "hypovolemia_risk" },
               { "constant": "HEME_PROB_MEDIUM_THRESHOLD" }
          ],
          ["LT",
               { "algorithm": "hypovolemia_risk" },
               { "constant": "HEME_PROB_HIGH_THRESHOLD" }
          ]
     ]
}
```

*Constant Library*

Similar to variables, when writing message logic it is extremely convenient to define your constants in one central location. This allows you to change thresholds or time goals that referenced throughout the app_configuration.json file from one location. The entries in the library are objects with a unique key that are referenced from message logic config. Message logic is found in `message_library.logic` and in `variable_library`. For example, `{ "constant": "HEME_PROB_MEDIUM_THRESHOLD" }` would inject the number 0.09. Constants are numbers but this is not enforced (yet).

Example:

```
{
"HEME_PROB_MEDIUM_THRESHOLD": 0.09
}
```

*Plots*

Configure the plots – Currently only “vitals_plot”, “algorithm_plot”, and “timer_plot” are supported. The APPRAISE system expects to find these three configurations only. For now, this configuration is tightly coupled to the functionality of the appraise-vitals-plot module.

Example:

```
"algorithm_plot": {
     "type": "algorithm_plot",
     "parameters": {
          "primary": "hypovolemia_risk"
     },
     "tickFrequency": 2000,
     "axis": {
          "x": {
               "mode": "30 Min"
          },
          "y": {
               "mode": "Fixed",
```

```
                "scale": "Log",
                "ticks": [0.1, 1, 10],
                "min": 0.07,
                "max": 14
            }
        },
        "series": [
            {
                "title": "appraisePath",
                "parameter": "primary",
                "legendLabel": "Hemorrhage Risk",
                "type": "path"
            },
            {
                "title": "appraisePathGlow",
                "parameter": "primary",
                "type": "path-glow"
            }
        ]
}
```

*parameters* – Object whose arbitrary keys correspond to the series.parameter value. The contents of each parameter entry must an object with key “key” and a value that matches one of the entries in the parameter_library.

*tickFrequency* – The time (in milliseconds) between plot updates. For example, 2000 would trigger the plots to update every two seconds.

*axis.X.mode* – [15 min, 30 min, 1 hr] Specify the value of the x time scale duration dropdown.

*axis.Y.mode* – [Fixed, Dynamic] Specify the value of the y scale dropdown to use either a fixed Y axis or a dynamic, auto-scaling Y axis. If Fixed is specified, include a min and max value.

*axis.Y.scale* – [Log, Linear] Choose desired Y axis scale. Remember zero is undefined on a Log scale when specifying tick marks and Y scale minimum.

*axis.Y.ticks* – Array of values to draw tick marks on the Y axis.

*series* – An array of objects that correspond to the series on the plot (i.e. how to represent the data - linear interpolation, points, symbols).

*series.title* – Unique label used internally to assign CSS classes and SVG groups.

*series.parameter* – Corresponds to one of the parameters values to determine which data the series will represent.

*series.legendLabel* – Legend label for the series (if falsey i.e. null, undefined, false the entry for the series will not be added to the legend)

*series.type* – [path, path-glow, circle, symbol, diff-bars, area] – Used to configure how the data should be represented. Linearly interpolated path, a circular point for each data point, a symbol for each data point. Diff-bars and area type require special data formats. Diff-bars require { t, value: [0, 1] } and area requires { t, sbp, dbp }.

*series.r* – If series.type = circle, describes the radius of the points.

*series.symbol* – [triangle-down, triangle-up] – If series.type = symbol, describes what type of symbol to use. Only two are implemented. Triangle-up (for diastolic data points) and triangle-down (for systolic data points).

## Application Configuration Validation

The application heavily relies on the validity and accuracy of the app_configuration.json file. The app_configuration.json file is examined in two ways: a schema check and a "business logic" check. This validation process is found in local_modules/app_config_validation.json.

This script first checks the given config against a JSON schema. The schema version "draft-04" described at http://json-schema.org. This schema is found in /app_configuration_schema.json.

The config validation script then checks the given config's "business logic" to enforce some assumptions in the software. For example, all parameter keys must be unique. The business logic validation includes:

- Message Logic Validation
    - EQ, NEQ, GT, GTE, LT and LTE tuple length === 3
    - parameter property values exist in parameter library
    - constant property values exist in constant library
    - algorithm property values exist in algorithm library
    - variable property values exist in variable library
- Parameter Library
    - parameter "key" values are unique
    - if type === alarm or meta, subtype is required
- Algorithm Library
    - algorithm "key" values are unique
- Constant Library
    - constant names are unique
- Variable Library
    - variable names are unique
    - logic validation
- Message Library
    - logic validation

## Application Configuration Window

In the host operating system's tray or dock, the APPRAISE GUI will place an application icon.

*Figure: APPRAISE GUI system tray icon*

Clicking on the icon will reveal a dropdown menu. Selecting the "Open Configuration Window" option will create a new application window with a configuration form. The form can be used to change the configuration of the application and trigger the application to restart with the Restart App button at the bottom.

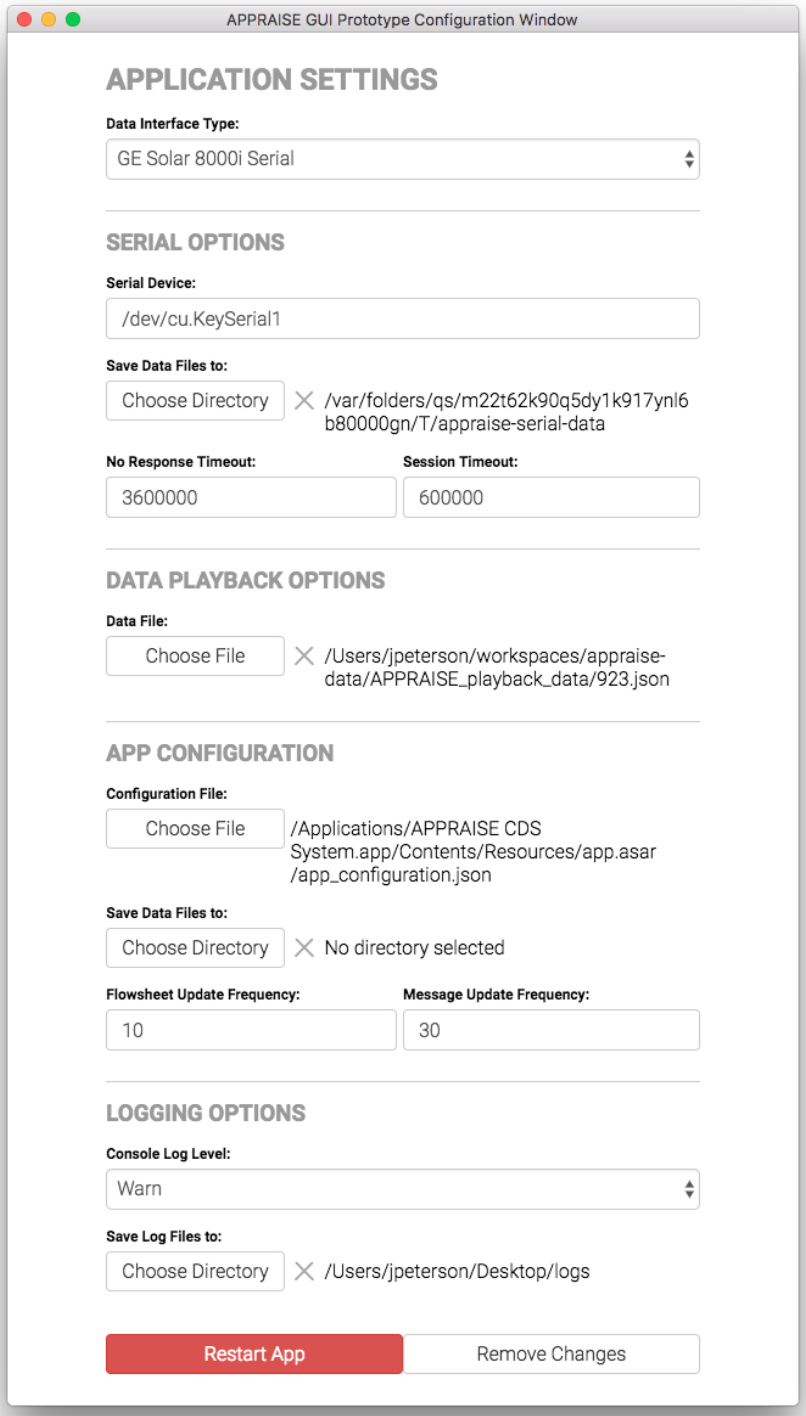



*Figure: APPRAISE GUI configuration window*

*Data Interface Type*: Choose where to retrieve incoming data. GE Solar 8000i Serial, Data File Playback, or Fake Data Generator.

### Serial Options

*Serial Device*: The physical port that the GE Solar is connected too. Typical /dev/tty.USB0 or similar.

*Save Data Files to*: All serial port traffic is saved to a file. Select which directory you'd like to save data too. Defaults to the OS specified temporary location.

*No Response Timeout*: Specify, in milliseconds, how long the software should send messages to the Solar without a response before giving up.

*Session Timeout*: Specify, in milliseconds, how long the software should wait, without receiving any vital signs data, before assuming the patient has left and starting a new session.

### Data Playback Options

*Data File*: Specify the data file to playback with the "Data Interface Type" of "Data File Playback".

### App Configuration

*Configuration File*: If you'd like to use a custom app_configuration.json file, specify the location of your custom file here.

*Save Data Files to*: Choose a directory to save application data. Every minute, the application saves the app state as described below in Data Persistence.

*Flowsheet Update Frequency*: Specify, in seconds, how frequently rows of the Flowsheet should be created.

*Message Update Frequency*: Specify, in seconds, how frequently messages should be calculated and potentially added to the display.

### Logging Options

*Console Log Level*: If a console is open, specify which level of log messages you would like printed. The options are: 'error', 'warn', 'info', 'verbose', 'debug', 'silly', and 'none'.

*Save Log Files to*: The software will persist a log file to the specified directory. In production mode, the file will contain level 'info', 'warn', and 'error' messages.

## Data Persistence

The APPRAISE GUI persists data required for retrospective analysis. The current data persistence strategy is simplistic and will ideally be replaced with a database layer when time allows. Current, the software saves a JSON encoded text file to the hard drive every 60 seconds. This file is overwritten with the updated data and state information. Specifically, the system records:

1. **application configuration** of the running system. This is typically read from a default file. This is the part of the system that contains the message library, parameter library, algorithm configurations, global variables, and plots configuration. The saved location is chosen by the user. If the configuration is changed, it should be noted. Config file is read only once at app startup.

2. **session start time**
3. **flowsheet** - tabular representation of calculated parameters and algorithms q10sec.
4. **flowsheet buffer** - arrays of "raw" data samples from the monitor for each connected parameter (i.e. ECG HR). A sample contains a timestamp, value, and sometimes metadata. NIBP alarms are included as a distinct parameter. The monitor is polled q2sec which would generate several data samples.
5. **message states** - an object that lists the times at which a message was "active" although not necessary displayed. This simply tracks when the message logic is true. Message display is calculated just prior to displaying a new message every 30 seconds.
6. **message display** - a list of the messages added to the UI with a timestamp (i.e. message ID 15 was displayed at 13:35). The six messages prior to any given time are the messages that were on the UI. This is updated q30sec after the message states.

In addition, the all data received from the GE Solar's serial port will be saved to a binary file. This can be re-parsed at a later time and represents a "comprehensive", encoded data record.

# APPENDIX II.A – SERIAL PORT INFORMATION

The serial interface to the GE Solar 8000i has only one published function: write all available data from the Solar to a connected computer. This sole write function is executed when the requesting computer queries the Solar with the appropriate instruction (a call and response paradigm). This is the only operation that the APPRAISE software uses and the only operation available in the GE serial interface specification. The query sent to the Solar by APPRAISE, titled `requestPacket`, is described in exhaustive detail below.

The data fields the Solar will transmit are not configurable. The Solar will transmit the data for all the connected parameters, if queried. It will not, however, transmit the admitted patient's name through the serial interface.

The Solar's serial interface does not have the capability to affect the operation of the device. Common actions, like discharging a patient or zeroing a transducer, are not possible via the serial interface. (It should be noted that the vendor may provide unpublished mechanisms for accomplishing these actions, but the 'function' and 'subfunction' codes are totally unknown. There is no way to know how to do this and the software does not allow the user to affect this in any way).

## GE's Requirements for Serial Connection to Solar 8000i

The GE Solar 8000i Serial Specification lists several requirements for connecting software to the monitor.

Requirement: *"Packet structure: Data bits 8; Parity None; Stopbits 1; Speed 9600 bps"*
The serial software used in APPRAISE is configured to communicate via standard "9600 8N1" format. This is shown with an excerpt from APPRAISE software code:

```
new SerialPort(this.serialportDevice, {
     baudrate: 9600,
     databits: 8,
     stopbits: 1,
     parser: serialport.parsers.raw,
   }, false);
```

Requirement: *"Polling frequency: Not more than once every 2 seconds"*
The minimum duration between requests from the software to the Solar is never less than 2 seconds. The only instances that the request packet is sent to the Solar monitor are when:

- The app initially is opened
- Two seconds after the receipt of a valid response from the Solar
- A configurable, recurring time (default 5 seconds) after an unanswered request (this polling is meant to detect when a disconnected serial cable is reconnected). This polling continues every 5 seconds until a configurable timeout (configurable; default 10 minutes).

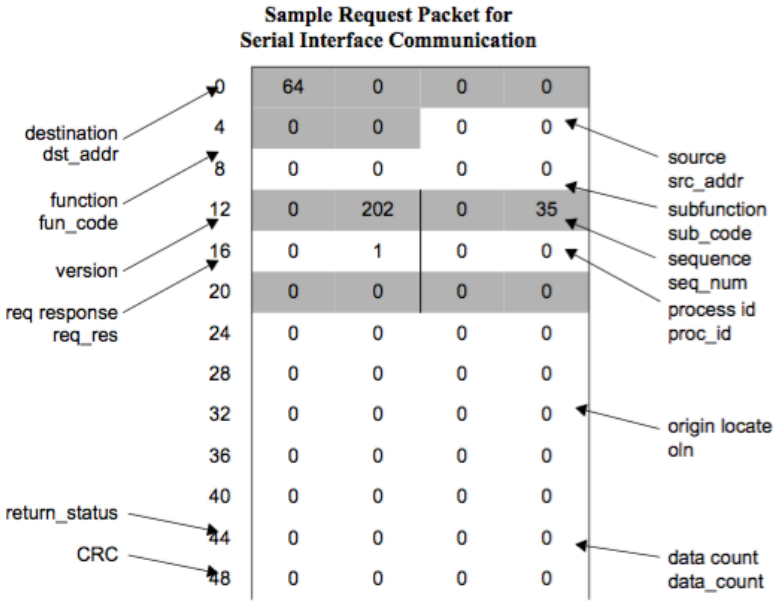
Laboratory testing has been done and found that the Solar will function soundly at polling frequencies as high as 250ms. The 2 second frequency mandated by the vendor, and adhered to by the APPRAISE software, is a generous duration that should not threaten the performance of the monitor.

Additionally, the GE Solar 8000i appends its serial response messages with a cyclic redundancy check (specifically a two byte CRC-16). This given CRC is compared to a calculated CRC value. The message is only accepted if the CRCs match.

## Request Packet Content

Guaranteeing the software only sends the proper command to the Solar is of the highest importance. The following section aims to 1) describe, in exhaustive detail, the exact request packet format used for communication with the GE Solar monitor and 2) show that the APPRAISE software is precisely adhering to this format.

GE Solar 8000i Serial Specification Request Packet Definition
Copied from *Serial Interface Data Services - Service Manual 2001005-128A*




*(PLEASE NOTE: The sloppy arrow placements are present in the official manual.)*

GE Solar 8000i Serial Specification Request Packet Content Description
Adapted from *Serial Interface Data Services - Service Manual 2001005-128A*

`dst_addr` - The first byte of the destination address is always set to 64 when using the serial port connection.
`src_addr` - The source address is not necessary. Set it to 0.
`fun_code` - The function code specifies what action the server is to perform. To simply read data from the server a function code of 202 would be used.
`sub_code` - The subfunction code further defines the request being sent to the server (Solar). The subfunction code 35 is sent to request polled parameters.
`version` - Determines the message structure to be used. For serial port communications this value is always set to 1.

`seq_num` - Not used, set to 0.
`req_res` - Not used, set to 0.
`proc_id` - Not used, set to 0.
`oln` - Not used, set to 0.
`return_status` - Not used, set to 0.
`data_count` - Not used, set to 0.
`CRC` - Used to ensure the monitor's response message is not corrupted during transmission (bit flips, missing bytes, faulty cable, etc)

GE Solar 8000i Serial Specification Request Packet struct
Copied from *Serial Interface Data Services - Service Manual 2001005-128A*

```
typedef struct sbedside_msg_def
{
 UTINY dst_addr[6];           /* destination address */
 UTINY src_addr[6];           /* source address */
 COUNT fun_code;              /* function code */
 COUNT sub_code;              /* subfunction code */
 COUNT version;               /* version of bed_msg */
 COUNT seq_num;               /* response sequence number */
 COUNT req_res;               /* request response flag */
 COUNT proc_id;               /* requestors process id */
 UTINY oln[32];               /* origin location name */
 COUNT return_status;         /* return status */
 COUNT data_count;            /* following message data count */
 COUNT CRC;
}
```

APPRAISE Application Serial Request Packet Generation Code:

```
const requestPacket = Buffer.alloc(104);
requestPacket[0] = 64;
requestPacket[13] = 202;
requestPacket[15] = 35;
requestPacket[17] = 1;
```

The `requestPacket` is "hard coded". There is no way for it to be altered before or during operation of the APPRAISE application. This is the only `Buffer` sent to the serial port at any time. There is no logic or code path that can alter what is sent to the serial port or the content of the `Buffer` `requestPacket`.

### Testing the `requestPacket`:

The accuracy of this operation was verified by connecting the serial cable to a terminal emulation program "screen" on a second computer. This emulation program observed `requestPacket` being properly emitted from the serial port during operation of the APPRAISE software.

# APPENDIX II.B – COMPUTER SECURITY AND INFORMATION SAFETY

Due to guidance issued by the MGH IRB and MGH Biomed department, the following requirements apply to the APPRAISE GUI pilot installation at MGH.

The “client” computer running the APPRAISE software will be password protected. There is no user input into the initial revision of the software so the keyboard and mouse will be removed. The software will run in kiosk mode, meaning the user desktop will not be accessible to users.

The computer will be collecting anonymous (de-identified) patient vital signs. There will be absolutely no patient identifying information on the computer. Additionally, the hard drive will be encrypted.

The APPRAISE computer will not be connected to a network. The only external connection will be to the serial interface of the Solar. This configuration will prevent the computer from being compromised by a network-based attack vector, which is the most common method a computer receiving malware or a virus. To be safe, the computer will have a firewall configured and up-to-date anti-virus running.

# APPENDIX II.C – SOFTWARE GLOSSARY

*Algorithm* - The APPRAISE GUI makes several calculations based on vital signs parameter data which are collectively titled Algorithms. These Algorithms, along with Parameters, are recorded as columns in the Flowsheet. Any Flowsheet data that requires calculation or other parameter/algorithm as input are considered Algorithms. This includes, most notably, the hemorrhage risk index. For example, addition Algorithms include hemProbTransitionMediumToHigh which takes the hemorrhage index as input and returns true if it has risen over the High threshold.

*algorithm_library* - Configured in app_configuration.json and entered into the application state at startup. The order matters - each Algorithm can take any Algorithm or Parameter listed before it as input. Algorithm key must be unique, as enforced by app_config_validation.js. Keys listed on the "parameters" property are unique to the script found at the "path" property value.

*Constant* - When writing "logic" for the Messages in the app_configuration.json, it is very convenient to have an abstraction for reusing common thresholds or times. This enables the system thresholds to be adjusted in one location. Similar to Variables, the Constants allow numbers to be inserted into message.logic or other variables. For example, "constant: INITIAL_BP_GOAL_TIME" will insert the time allotted for taking the first cuff based blood pressure before the BP is considered late.

*constant_library* - Configured in app_configuration.json and entered into the application state at startup. Contains the definitions of the configuration Constants. It is accessed whenever a key in a logic statement is "constant".

*Device Interface* - The interface between the APPRAISE application and a medical device. The Device Interface is responsible for IO with the device, parsing the bytes, translation of the data to a common format, and determining when a patient is connected to delineate Sessions. The APPRAISE system currently interfaces with the GE Solar 8000i via a serial port connection only. A data playback interface and a fake data generator are also available.

*Flowsheet* - A tabular record of clinical data for each patient Session. The columns correspond to the parameters described in the parameter_library and algorithms described in the algorithm_library. The rows are calculated every n seconds (default: 10 seconds).

*FlowsheetBuffer* - The Flowsheet rows are calculated every 10 seconds. During the 10 seconds between Flowsheet row calculations, data received from the Device Interface that match a metric needed for a Parameter are added to the FlowsheetBuffer. When a new Flowsheet row is calculated, the FlowsheetBuffer is read and all appropriate values are reduced to value for a cell in the Flowsheet table. The keys of the FlowsheetBuffer correspond to the intersection of Metric names that the APPRAISE GUI is interested in receiving and Metric names that have been provided by the Device Interface.

*Hemorrhage Index* - A novel algorithm for assessing the likelihood of a patient requiring blood products (RBCs ≥ 1) based on their current vital signs (HR and BP). This algorithm is displayed on the UI and used to determine which reminder messages are applicable.

*local_modules* - A directory in the code repository where a variety of scripts are located, including the algorithms, the playback and fake data interfaces, and the app_configuration validation.

*Message* - A card displayed on the user interface that contains a picture/icon, title and short message. The message content relates to the current state of the patient (Flowsheet). The messages relate to clinical protocol adherence (e.g. “No crystalloid resuscitation”) and technical functionality (e.g. “Unable to determine hemodynamic state: no blood pressure available”).

*message_library* - The messages that the system will display when the message criteria are met. The library is read from the file system from the app_configuration.json file. The message_library is stored in the app state. The message_library entries contain the title, content, and a logical statement that describes when to display the message. The message_library is described in detail above.

*Metric* - Metrics are the unique data types provided by the Device Interface. Metric names are defined by ISO 11073-10101 and maintained by NIST Harmonized Rosetta Terminology Mapping (hRTM). These names provide a standardized approach to encoding physiological concepts. For example, the patient’s heart rate, as measured by an electrocardiogram in beats per minute, is encoded as “MDC_ECG_HEART_RATE” with units “MDC_DIM_BEAT_PER_MIN”.

*metric_list* - An array of Metric names created when the parameter_library is imported that represents all Metrics the APPRAISE GUI is interested in receiving. This list is used to determine whether an incoming message from the Device Interface should be added to the Flowsheet Buffer.

*Parameter* - A Parameter is a summarization of a measurement made by the device connected to the Device Interface (e.g. `hrMedian2min` which would correspond to the median value of the heart rate measurements from the previous two minutes).

*parameter_library* - Configured in app_configuration.json and entered into the application state at startup. The parameter_library describes the Parameters to record in the Flowsheet (the columns). The parameters in the parameter_library contain a parameter name that corresponds to the metric title in the data from the Device Interface as well as the method to reduce the parameter data (average, median, count, etc).

*Session* - A unique patient encounter. The Device Interface will determine if there has been a gap in the data stream from the patient. A gap in the data is defined by either the patient discard button being pressed on the monitor or a 10 minute time continuous interval where no new ECG HR, SpO2, or NIBP data are received. When a gap is detected, the current session is set to no_patient. When a new valid measurement is received, the Session name is set to the date and time.

*Variable* - When writing “logic” for the Messages in the app_configuration.json, it is very convenient to have an abstraction for reusing branches of your logic statement. Similar to Constants, the Variables allow logic statements to be inserted into message.logic or other variables. The variables follow the same syntax as message.logic. For example, “algorithm: FAILED BP” is used to encapsulate the logic for when blood pressure measurements are absent or have failed.

*variable_library* - Configured in app_configuration.json and entered into the application state at startup. Contains the definitions of the configuration Variables. It is accessed whenever a key in a logic statement is algorithm.

# APPENDIX II.D – APPLICATION STATE EXAMPLE

The top level keys are: currentSession, sessions, parameterLibrary, variableLibrary, algorithmLibrary, messageLibrary, plots, metricList, and constantLibrary. The libraries and plots come directly from app_configuration.json (with) and are described


```
{
  "currentSession": "20161206_1226_playback=23",
  "sessions": {
    "20161206_1226_playback=23": {
      "startTime": 1481045162422,
      "flowsheet": [
        {
          "t": 1481045190003,
          "hrMedian2min": 114,
          "hrMedian1min": 114,
          "hrCount": 24,
          "hrLastValue": 126,
          "hr2minStdDev": 6.619494567815079,
          "hypovolemia_risk": null,
          "sysBp_diff": null,
          "diaBp_diff": null,
          "high_heme_prob_time": null,
          "dbpTransition120": false,
          "hemProbTransitionMedium": false,
          "hemProbTransitionHigh": false
        },
        { ... },
        { ... },
        { ... },
        // ...truncated for clarity
      ],
      "flowsheetBuffer": {
        "MDC_ECG_HEART_RATE": [
          {
            "value": 103,
            "t": 1481045166445,
            "meta": {
              "highLimitViolation": false,
              "leadFail": false,
              "lowLimitViolation": false
            }
          },
          {
            "value": 109,
            "t": 1481045167445,
            "meta": {
              "highLimitViolation": false,
              "leadFail": false,
              "lowLimitViolation": false
            }
          }
          { ... },
          { ... },
          { ... },
          // ...truncated for clarity
        ]
      },
      "messageStates": [
        {
```

```
        "id": 0,
        "active": true,
        "startTime": [
          1481045190003
        ]
      }
      // ...truncated for clarity
    ],
    "messageDisplay": [
      {
        "id": 0,
        "displayTime": 1481045190003
      }
      // ...truncated for clarity
    ]
  }
},
"parameterLibrary": {
  "hrMedian2min": {
    "metric": "MDC_ECG_HEART_RATE",
    "type": "parameter",
    "subtype": "value",
    "aggregation": "median",
    "time_filter": 120000
  },
  "hrMedian1min": { ... },
  "fakeParameter": { ... },
  "hrCount": { ... },
  "hrLastValue": { ... },
  "hr2minStdDev": { ... },
  "hrLeadsFail2minCount": { ... },
  "nibpSys1minAvg": { ... },
  "nibpDia1minAvg": { ... },
  "nibpMean1minAvg": { ... },
  "nibpSysRowAvg": { ... },
  "nibpDiaRowAvg": { ... },
  "nibpSysLastValue": { ... },
  "nibpDiaLastValue": { ... },
  "nibpSysCount": { ... },
  "nibpDiaCount": { ... },
  "nibpSysCount10min": { ... },
  "nibpFailedAttempts1minCount": { ... },
  "nibpFailedAttemptsCount": { ... },
  "spo22minMedian": { ... },
  "spo22minStdDev": { ... },
  "spo2SigQual2minAvg": { ... }
},
"variableLibrary": {
  "HEME_PROB_LOW": ["LT",
    { "algorithm": "hypovolemia_risk" },
    { "constant": "HEME_PROB_MEDIUM_THRESHOLD" }
  ],
  "HEME_PROB_MEDIUM": ["AND",
    ["GTE",
      { "algorithm": "hypovolemia_risk" },
      { "constant": "HEME_PROB_MEDIUM_THRESHOLD" }
    ],
    ["LT",
      { "algorithm": "hypovolemia_risk" },
      { "constant": "HEME_PROB_HIGH_THRESHOLD" }
    ]
  ],
  // ...truncated for clarity
```

```
  },
  "algorithmLibrary": [
    {
      "key": "hypovolemia_risk",
      "parameters": {
        "hrKey": "hrMedian2min",
        "bpSysKey": "nibpSysRowAvg",
        "bpDiaKey": "nibpDiaRowAvg",
        "bpCountKey": "nibpSysCount"
      },
      "path": "appraise"
    },
    {
      "key": "sysBp_diff",
      "parameters": {
        "key": "nibpSysRowAvg"
      },
      "path": "diff"
    }
    // ...truncated for clarity
  ],
  "messageLibrary": {
    "0": {
      "name": "For Trauma Patients Only",
      "text": "New session: APPRAISE protocol reminder system",
      "category": "general",
      "critical": true,
      "icon": "caution_symbol",
      "logic": ["LT", "elapsedTime", 10800000],
      "id": 0
    },
    "1": {
      "name": "Prepare for Transport",
      "text": "If high-risk patient: fluid bags, travel monitor, etc. should be readied
for travel",
      "category": "general",
      "critical": false,
      "icon": "patient_transport",
      "logic": ["AND",
        ["GT", "elapsedTime", 960000],
        ["LT", "elapsedTime", 1200000],
        { "variable": "HEME_PROB_MEDIUM" }
      ],
      "id": 1
    },
    "2": {
      "name": "Repeat Cuff BP",
      "text": "Most recent DBP may not be accurate (too high); repeat cuff BP to
confirm",
      "category": "monitoring",
      "critical": false,
      "refractory": 600000,
      "icon": "bp_cuff",
      "logic": ["OR",
        ["AND",
          ["GTE", { "parameter": "nibpDiaLastValue" }, 120],
          ["EQ", { "parameter": "nibpDiaCount" }, 1]
        ],
        { "algorithm": "dbpTransition120" }
      ],
      "id": 2
    },
    // ...truncated for clarity
```

```
  },
  "plots": {
    "vitals_plot": { ... },
    "algorithm_plot": {
      "type": "algorithm_plot",
      "parameters": {
        "primary": "hypovolemia_risk"
      },
      "tickFrequency": 2000,
      "axis": {
        "x": {
          "mode": "30 Min"
        },
        "y": {
          "mode": "Fixed",
          "scale": "Linear",
          "ticks": [
            1,
            50,
            100
          ],
          "min": 0,
          "max": 100
        }
      },
      "series": [
        {
          "title": "appraisePath",
          "parameter": "primary",
          "legendLabel": "Probability (%)",
          "type": "path",
          "data": []
        },
        {
          "title": "appraisePoints",
          "parameter": "primary",
          "type": "circle",
          "r": 2,
          "data": []
        }
      ],
      "containerId": "algorithmPlot"
    }
  },
  "metricList": [
    "MDC_ECG_HEART_RATE",
    "THIS_DOESNT_EXIST",
    "MDC_ALARM_ECG_HEART_RATE",
    "MDC_PRESS_CUFF_SYSTOLIC",
    "MDC_PRESS_CUFF_DIASTOLIC",
    "MDC_PRESS_CUFF_MEAN",
    "MDC_ALARM_PRESS_CUFF",
    "MDC_PULS_OXIM_SAT_O2"
  ],
  "constant_library": {
    "HEME_PROB_MEDIUM_THRESHOLD": 0.09,
    "HEME_PROB_HIGH_THRESHOLD": 1,
    "HEME_PROB_VERYHIGH_THRESHOLD": 7,
    "CT_OR_GOAL_TIME": 1200000,
    "INITIAL_BP_GOAL_TIME": 600000
  }
}
```

# APPENDIX II.E – INSTALLATION INSTRUCTIONS

## Prerequisites:

1. Install Node.js from https://nodejs.org/en/ or a package manager of your choosing
2. We need to be able to compile binaries on our system. It is required by the node-gyp tool to enable compiling node.js "native addon" code (e.g. C/C++ code to communicate with a serial-port). Installation instructions can be found here: https://github.com/nodejs/node-gyp#installation
    a. Regardless of platform, Python 2.7 should be installed.
    b. Unix: you will need `make` and a compiler like `gcc`
    c. MacOS: install `make` and `gcc` via Command Line Tools (Xcode -> Preferences -> Downloads)
    d. Windows: Additional instructions for installation in Windows can be found here: https://github.com/nodejs/node-gyp/wiki/Visual-Studio-2010-Setup
    e. You are not required to install `node-gyp` globally (`-g`) but `node-gyp` recommends it

## Install and Run APPRAISE (development):

1. Copy and extract the source code zip file to a location of your choice (we'll call it the `project` directory from here on)
2. Open a terminal or command prompt and navigate to the `project/appraise` directory. This is the main directory of our application.
3. Install required `/node_modules/` via `` `npm install` ``. This will install all dependencies and devDependencies listed in the `package.json` file.
    a. Additionally, this step will compile node addon binaries for the current platform architecture automatically using the `postinstall` script found in `package.json`. (This is where node-gyp is invoked under the hood of electron-rebuild)
4. Run the software with `` `npm run go` ``. This will invoke the `go` script in `package.json`.
    a. `npm run go` will invoke a code packaging routine via webpack then start the software
        i. Specifically, `webpack.config.js` will use `app/index.js` to create `app/index.bundle.js` and `app/index.bundle.js.map` as defined in the webpack config file
    b. Why is `cross-env NODE_ENV=` needed?

i. `NODE_ENV` affects the webpack builds as well as file paths in main.js

ii. If running from a local development environment (as described above), the code is packaged and run with the NODE_ENV environment variable set to 'development'. This points file paths to the local directory and does not "minify" the code.

iii. When the code is packaged and installed as an application (as opposed to running from terminal) the NODE_ENV is set to 'production'. This minifies the code and points files to an `.asar` archive in application.

iv. Trying to run the software with the wrong NODE_ENV (i.e. locally in production or application in development) will potentially produce errors.

5. Click on the  icon in system tray or menu bar and select "Open Configuration Window". The under "Data Interface Type:" select "Fake Data Generator" then press "Restart App" at the bottom. This will replace the serial port interface with arbitrary vital signs data to allow you use the system with the bedside monitor.

## Creating an Installer (production):

1. In a terminal or command prompt, navigate to the appraise repository and enter `` `npm run package` ``

a. This command uses electron-builder. It will compile code with your current architecture as the target.

Journal of Critical Care (2012) **27**, 533.e1–533.e10

ELSEVIER

Journal of Critical Care

# The association between vital signs and major hemorrhagic injury is significantly improved after controlling for sources of measurement variability☆

**Andrew T. Reisner MD[a,b,*], Liangyou Chen PhD[a], Jaques Reifman PhD[a,*]**

[a]*Bioinformatics Cell, Telemedicine and Advanced Technology Research Center (TATRC), US Army Medical Research and Materiel Command (USAMRMC), ATTN: MCMR-TT, 504 Scott Street, Fort Detrick, MD 21702, USA*
[b]*Massachusetts General Hospital Department of Emergency Medicine, Boston, MA 02114, USA*

**Keywords:**
Vital signs;
Decision support systems;
Clinical;
Hemorrhage;
Data interpretation;
Statistical;
Emergency care;
Prehospital;
Trauma

**Abstract**

**Purpose:** Measurement error and transient variability affect vital signs. These issues are inconsistently considered in published reports and clinical practice. We investigated the association between major hemorrhagic injury and vital signs, successively applying analytic techniques that excluded unreliable measurements, reduced transient variation, and then controlled for ambiguity in individual vital signs through multivariate analysis.

**Methods:** Vital sign data from 671 adult prehospital trauma patients were analyzed retrospectively. Computer algorithms were used to identify and exclude unreliable data and to apply time averaging. An ensemble classifier was developed and tested by cross-validation. Primary outcome was hemorrhagic injury plus red cell transfusion. Areas under receiver operating characteristic curves (ROC AUCs) were compared by the test of DeLong et al.

**Results:** Of initial vital signs, systolic blood pressure (BP) had the highest ROC AUC of 0.71 (95% confidence interval, 0.64-0.78). The ROC AUCs improved after excluding unreliable data, significantly for heart rate and respiratory rate but not significantly for BP. Time averaging to reduce temporal variability further increased AUCs, significantly for BP and not significantly for heart rate and respiratory rate. The ensemble classifier yielded a final ROC AUC of 0.84 (95% confidence interval, 0.80-0.89) in cross-validation.

**Conclusions:** Techniques to reduce variability in vital sign data can lead to significantly improved diagnostic performance. Failure to consider such variability could significantly reduce clinical effectiveness or confound research investigations.



☆ Competing interests: Dr Andrew Reisner participated in a customer advisory meeting for General Electric Healthcare in 2008 and has received speaking fees from Masimo Corp. Dr Reisner is a coinvestigator on a US National Institutes of Health Bioengineering Research Partnership that includes Philips Healthcare.

* Corresponding author. Tel.: +1 301 619 7915; fax: +1 301 619 1983.
*E-mail addresses:* areisner@partners.org (A.T. Reisner), lchen@bioanalysis.org (L. Chen), jaques.reifman@us.army.mil (J. Reifman).

0883-9441/$ – see front matter © 2012 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.jcrc.2012.01.006

## 1. Introduction

Vital signs measurement is a routine aspect of clinical practice and research protocols. Although it is known that transient variability and measurement error can, in principle, affect the accuracy of vital signs, what is unknown is the extent to which these factors affect diagnostic capabilities in actual clinical practice. Vital signs fluctuate through time because of transient perturbations (eg, medication boluses, bouts of pain, anxiety, coughing) as well as natural steady-state variability. In addition, the accuracy of vital sign data is affected by clinicians' technique [1]. For example, accurate blood pressure (BP) measurement using a cuff requires proper fit and positioning of the cuff, a relaxed and properly positioned extremity, and the absence of patient motion [2]. Significant discrepancies have been reported between different methods of measuring noninvasive BP [3]. Similarly, respiratory rate (RR) measurement is prone to technical error, whether measured by a clinician [4] or by a bedside monitor via impedance pneumography (IP) [5]. In one report, both triage nurses' measurements of RR and electronic measurement of RR revealed poor sensitivity for bradypnea and tachypnea, and the authors referred to RR as "the vexatious vital"[4]. Heart rate (HR) monitored by electrocardiography (ECG) can be unreliable, that is, if electrodes are improperly affixed, and false arrhythmia alarms are commonplace [6]. Multiple authors have called into question the value of HR in assessing the hemodynamic state of a patient because of its variable relationship with hypovolemia [7,8]. Finally, it is worth noting that the accuracy of vital sign data may vary considerably for different makes and models of measurement devices [9-11].

The extent to which these factors affect diagnostic capabilities in actual practice is relevant to the design and interpretation of clinical investigations. If vital sign data were often polluted by inaccuracies, then there would be a bias toward the null hypothesis, where positive study effects might be masked (ie, type II study errors). Alternatively, failure to describe key methodology that improved vital sign accuracy (eg, superior equipment, training, or study protocols) would make it harder for others to replicate a successful study. Consider that some reports support the usefulness of prehospital severity scores for trauma patients [12-14], whereas other studies found those scores ineffective [15,16]. In these examples, the reports lacked any explicit consideration of the measurement apparatus, clinical protocols, and quality assurance processes related to vital sign measurements; and inconsistency in how vital signs were measured could have contributed to the heterogeneous findings. More broadly, there are diverse sets of conflicting reports with a shared failure to detail vital sign measurement methodology, for example, the risk of volume resuscitation of trauma patients with uncontrolled hemorrhage [17,18], the benefit of rapid response teams for inpatients with physiologic deterioration [19-22], and the benefit of early goal-directed resuscitation for septic shock [23]. It is possible that different approaches to vital sign measurements contributed to the inconsistencies of the reports' findings.

We investigated the association between standard vital signs and major hemorrhagic injury in a population of prehospital trauma patients using computational techniques that excluded unreliable measurements, reduced transient perturbations, and reduced ambiguity of individual vital signs. We compared these results with conventional analyses. The findings are applicable to the clinical evaluation of hemorrhage, which is the single most treatable cause of mortality in trauma patients [24,25]. Moreover, the findings may relate to a range of applications because the extent to which different analytic methods yield significantly different results indicates the importance of considering these factors in clinical practice and research studies.

## 2. Materials and methods

### 2.1. Clinical data collection

This was a retrospective analysis of a database, originally collected and analyzed by Cooke et al [26] with institutional review board approval, of trauma patients during transport by air ambulance from the scene of injury to a level I trauma center [26]. Between August 2001 and April 2004, the following physiologic data were measured in a convenience sample by Propaq 206EL monitors (Protocol Systems, Beaverton, Ore) and archived using a networked personal digital assistant: ECG and IP recorded at 182 and 23 Hz, respectively; the corresponding HR and RR output at 1-second intervals; and systolic BP (SBP) and diastolic BP (DBP) measured intermittently at multiminute intervals. Clinical data were collected during retrospective chart review, including demographics, prehospital interventions, hospital treatments, and injury descriptions. Subsequently, vital sign data from 788 patients were uploaded to our data warehousing system [27]. Protected health information was not included.

All data analysis was performed using MATLAB v7 (MathWorks, Natick, Mass).

### 2.2. Vital sign reliability

For each vital sign value, reliable data were identified by automated algorithms that rated each datum on an integer scale of 0 to 3 from least reliable to most reliable. Vital sign data rated 2 or 3 were considered reliable; otherwise, they were unreliable. Detailed descriptions of these algorithms have been previously reported [28-30]. Here, we provide an overview of the methodology. The algorithms analyze moving windows of physiologic data. The algorithms rate the reliability of vital signs computed

from the data windows based on (1) a computerized assessment of the ECG or IP waveforms' reliability and (2) a comparison between the rates output by the Propaq 206EL vs an independent calculation of the HR or RR performed by the algorithm. In practice, when waveforms demonstrate clear, rhythmic beats or breaths and the rates output by the Propaq 206EL match the algorithms' own calculations, then the corresponding HR or RR is rated as reliable. Conversely, when the waveforms are noisy with irregular, heterogeneous beats or breaths and/or there were major discrepancies between the rates output by the Propaq 206EL vs the algorithms' own calculations, then the HR or RR is rated as unreliable. The underlying rationale is the assumption that clean ECG or IP waveforms lead to reliable HR or RR measurements and that HR or RR tends to be reliable when 2 independent calculation methods yield similar results.

In prior validation, the reliability rating of RR using the automated algorithms typically concurred with clinicians who independently applied the reliability criteria to a set of test cases [28,30]. In 99% of the test cases, the automated algorithm agreed with the clinician RR rating (±1 level), where high RR reliability ratings were found to be associated with smaller differences between computer-calculated and human-calculated RR (average differences of 1.7 and 8.1 breaths per minute for the best and worst RR reliability ratings, respectively). Likewise, there was close agreement (within ±5 beats per minute) between computer-calculated and human-calculated HR in 97% of the test cases rated 2 or 3 by the automated HR reliability algorithm [30].

The BP reliability algorithm determined if the ratio between SBP, DBP, and mean pressure is physiologic and if the HR measured by the inflatable oscillometric cuff matches the ECG HR [29]. The algorithm does not attempt to distinguish between unequal HRs because of motion artifact vs unequal HRs because of nonperfusing electrical beats, for example, premature contractions; in the latter case, it would be possible for reliable BP data to be misclassified as unreliable.

## 2.3. Subject selection

The primary study population consisted of patients with any reliable vital sign datum within the initial 15 minutes of prehospital monitoring. We also studied 3 subgroups: patients with pairs of at least 1 reliable and 1 unreliable (*a*) HR, (*b*) RR, and (*c*) BP. In the primary analysis, we excluded the "ambiguous outcome" patients who received red blood cell (RBC) transfusions but lacked documented injuries that were indisputably hemorrhagic (see below). These cases were reincluded in sensitivity analyses (see below). Also excluded were the few patients who died before any diagnostic imaging or surgical exploration, when it could not be determined whether the patient died to major hemorrhage vs other critical pathology.

## 2.4. Primary outcome

*Major hemorrhagic injury* was defined as a documented injury that unequivocally causes some loss of blood volume (laceration or fracture of a solid organ, thoracic or abdominal hematoma, vascular injury that required operative repair, or limb amputation) and RBC transfusion within 24 hours.

## 2.5. Comparison of reliable vs unreliable vital signs

We computed the patients' proportions of reliable vital signs (median and interquartile range). For the 3 subgroups with at least 1 reliable and 1 unreliable vital sign—HR, RR, and BP—we computed each patient's mean of the reliable and of the unreliable data and compared the population mean of the subjects' means with Student *t* test for paired data (note that the *t* test is valid for normal and nonnormal distributions as long as there are enough subjects per distribution, eg, 30 or more [31]).

To compare diagnostic performance, we repeated the following statistical computation 100 times for each vital sign: from each patient, we randomly selected 1 reliable and 1 unreliable measurement, then computed receiver operating characteristic (ROC) curves for the selected reliable and the unreliable data using the method of DeLong et al [32]. We computed the difference between the areas under those curves (ROC AUCs) and averaged the results from the 100 cycles. This methodology avoided biases due to those patients with a surplus of measurements and unequal ratios of reliable vs unreliable measurements between patients.

## 2.6. Association between vital signs and major hemorrhagic injury within the initial 15 minutes

For each vital sign, we computed the univariate ROC AUC for (*a*) the first nonzero value, (*b*) the first reliable value, (*c*) the last reliable value, and (*d*) the average of all reliable values within 15 minutes.

We performed multivariate analysis using ensemble classification, a collection of multivariate regression models. Each of the models within the ensemble is a standard linear regression model, and their outputs are simply averaged to yield the ensemble classifier output [33]. Ensemble classification is able to classify subjects with incomplete data, as is explained below. This property was important because many patients lacked reliable data for every vital sign.

Each regression model within the ensemble used 1, 2, or 3 of the following parameters: HR, RR, SBP, and SBP − DBP. The final ensemble was composed of all possible combinations (14 total regression models). We applied cross-validation, randomly partitioning 50% of the study population for classifier training. Each model was trained using the subset of patients who possessed at least 1 reliable measurement of each model parameter within the initial 15 minutes, using the average of all reliable values from the initial 15 minutes. Next,

**Table 1** Population description

| Characteristic | Study population |
|---|---|
| Population size, n | 671 |
| Male/female, n [a] | 498/172 |
| Age (y), mean (SD) | 38 (15) |
| Blunt injury, n (%) | 596 (89) |
| Mortality, n (%) | 41 (6) |
| Prehospital intubation, n (%) | 115 (17) |
| Major hemorrhagic injury, n (%) | 78 (12) |
| % Reliable HR for patient, median (IQR) | 62 (4-84) |
| % Reliable RR for patient, median (IQR) | 16 (0-45) |
| % Reliable SBP for patient, median (IQR) | 100 (83-100) |

Patients with at least 1 reliable vital sign datum within 15 minutes after exclusion of cases who received RBC transfusions but lacked documented injuries that were indisputably hemorrhagic (see text for details). IQR indicates interquartile range.

[a] Sex unknown for 1 patient in the database.

we tested the ensemble classifier in the remaining 50% of the patients. For each patient, we only used those regression models for which the patient had the necessary reliable data during the initial 15 minutes and used the models' average output as the final output. This process was repeated for 100 cycles, each time randomly repartitioning the patients into training/testing sets. We computed the mean ROC AUC of those 100 testing cycles.

### 2.7. Sensitivity analyses

We repeated the ensemble classification using 4 alternative methodologies: (*a*) reinclusion of the "ambiguous outcome" patients, treating them as nonhemorrhage control cases; (*b*) redefining "major hemorrhagic injury" as a documented hemorrhagic injury, as above, plus RBC transfusion *or* at least 3 L of crystalloid infusion; (*c*) redefining "major hemorrhagic injury" as the receipt of at least 5 U of RBC regardless of the documented injuries; and (*d*) using reliable vital sign data from only the initial 10 minutes.

## 3. Results

The database had 788 records with at least 1 nonzero vital sign datum. One hundred seventeen cases were excluded (105 were "ambiguous outcome" cases subsequently reintroduced in the sensitivity analysis described below, whereas 12 lacked any reliable vital sign data). Table 1 shows the population characteristics, with 12% having major hemorrhagic injury, 17% with prehospital intubation, and 6% overall mortality. Respiratory rate data had the lowest rate of reliability, whereas BP data had the highest.

Table 2 shows reliable data compared with unreliable data. Unreliable measurements of HR, RR, and SBP all had significantly elevated values vs their reliable counterparts and tended to have reduced ROC AUCs.

Table 3 reports the cumulative diagnostic yields of the investigative techniques. The ROC AUCs were significantly improved for initial HR and initial RR when reliability was considered. The ROC AUCs were significantly improved for SBP when the average of all its reliable values was used, whereas these were nonsignificantly increased for the average of reliable HR or RR. (In regard to the effects of mechanical ventilation on RR, the average of all reliable RR yielded an ROC AUC of 0.72 [95% confidence interval {CI}, 0.62-0.80] for spontaneously breathing patients and 0.64 [95% CI, 0.46-0.78] for mechanically ventilated patients.)

Applied to all 671 patients in the study population, the ensemble classifier yielded an ROC AUC of 0.84 (95% CI, 0.80-0.89) in cross-validation. This AUC was significantly greater than any univariate vital sign. The classifier could identify 36% of major hemorrhagic injury cases with greater

**Table 2** Reliable compared with unreliable vital signs

| Vital sign | Population with at least 1 reliable and 1 unreliable vital sign, n | Patients with major hemorrhagic injury, n (%) | Patients' proportion of reliable data (%), median (IQR) | Reliable data, mean (SD) | Unreliable data, mean (SD) | Reliable vs unreliable data, *P* value (Student *t* test) | Unreliable vital signs: ΔROC AUC for Dx of major hemorrhagic injury, mean (upper/lower range) |
|---|---|---|---|---|---|---|---|
| HR | 632 | 72 (11) | 65 (7-85) | 95 (20) | 99 (20) | <.001 [a] | −0.02 (−0.05/+ 0.01) |
| RR | 388 | 52 (13) | 39 (20-61) | 27 (7) | 37 (17) | <.001 [a] | −0.11 (−0.18/−0.03) |
| SBP | 217 | 34 (16) | 75 (67-86) | 127 (22) | 138 (37) | <.001 [b] | −0.12 (−0.21/−0.03) |
| DBP | 221 | 34 (15) | 75 (67-86) | 72 (15) | 76 (75) | NS | −0.02 (−0.09/+ 0.04) |

Populations included only those patients determined to have at least 1 reliable and 1 unreliable vital sign measurement, according to the reliability algorithms, at any time during their transport. Shown are the patients' means of reliable vs unreliable data for all patients (computing first the mean of each patient and then computing the mean of the patients' means). Student *t* test for paired data was used to test for significant differences between patients' means. Finally, the change in ROC AUC in the diagnosis of major hemorrhagic injury is shown, when one random unreliable measurement was used in place of a random reliable measurement (see text for details of this calculation). NS indicates not significant. Dx indicates diagnosis.

[a] Reliable vs unreliable data are also significant ($P < .001$) in hemorrhage cases alone and in control cases alone.

[b] Reliable vs unreliable data are also significant in hemorrhage cases alone ($P = .01$) and in control cases alone ($P < .001$).

**Table 3** Areas under receiver operating characteristic curves for the diagnosis of major hemorrhagic injury with application of vital sign reliability criteria, time averaging, and multivariate (ensemble) classification

| Vital sign | Population | ROC AUC (95% CI) | | | |
|---|---|---|---|---|---|
| | | First nonzero | First reliable | Last reliable | All reliable |
| HR | At least 1 reliable HR (n = 625) | 0.60 (0.53-0.68) | 0.71 (0.63-0.77)[a] | 0.72 (0.65-0.78)[a] | 0.73 (0.65-0.79)[a] |
| RR | At least 1 reliable RR and intubated (n = 85) | 0.52 (0.46-0.58) | 0.64 (0.55-0.72)[a] | 0.63 (0.53-0.71) | 0.67 (0.58-0.75)[a] |
| RR | At least 1 reliable RR and spontaneous breathing (n = 313) | 0.55 (0.48-0.61) | 0.64 (0.53-0.74) | 0.68 (0.56-0.77)[a] | 0.72 (0.62-0.80)[a] |
| SBP | At least 1 reliable SBP (n = 648) | 0.71 (0.64-0.78) | 0.74 (0.67-0.80) | 0.77 (0.70-0.83) | 0.79 (0.73-0.85)[a,b] |
| DBP | At least 1 reliable DBP (n = 648) | 0.62 (0.54-0.69) | 0.64 (0.56-0.71) | 0.64 (0.56-0.71) | 0.63 (0.55-0.71) |
| Ensemble classifier | At least 1 reliable HR or reliable RR or reliable SBP (n = 671) | NA | NA | NA | 0.84 (0.80-0.89)[c] |

Ensemble classification was applied to the overall study population. For RR, results are also provided separately for intubated patients and for spontaneously breathing patients. The method of DeLong [32] for paired data was used to test for statistically significant differences of ROC AUCs. NA indicates not applicable.

[a] ROC AUC significantly ($P < .05$) increased vs ROC AUC for "first nonzero" value.

[b] ROC AUC significantly ($P < .05$) increased vs ROC AUC for "first reliable" data.

[c] Ensemble ROC AUC significantly increased vs ROC AUC for "all reliable" HR data ($P < .001$), "all reliable" RR data ($P < .001$), "all reliable" SBP data ($P < .05$), and "all reliable" DBP data ($P < .001$).

than 60% positive predictive value (PPV) and greater than 85% of hemorrhage cases with 24% PPV (Fig. 1).

The sensitivity analyses yielded the following ROC AUCs for major hemorrhagic injury, which were similar to the primary analysis: (*a*) inclusion of the ambiguous outcome patients, 0.82 (95% CI, 0.77-0.87); (*b*) use of RBC transfusion or at least 3 L of crystalloid infusion as the outcome, 0.83 (95% CI, 0.79-0.87); (*c*) inclusion of

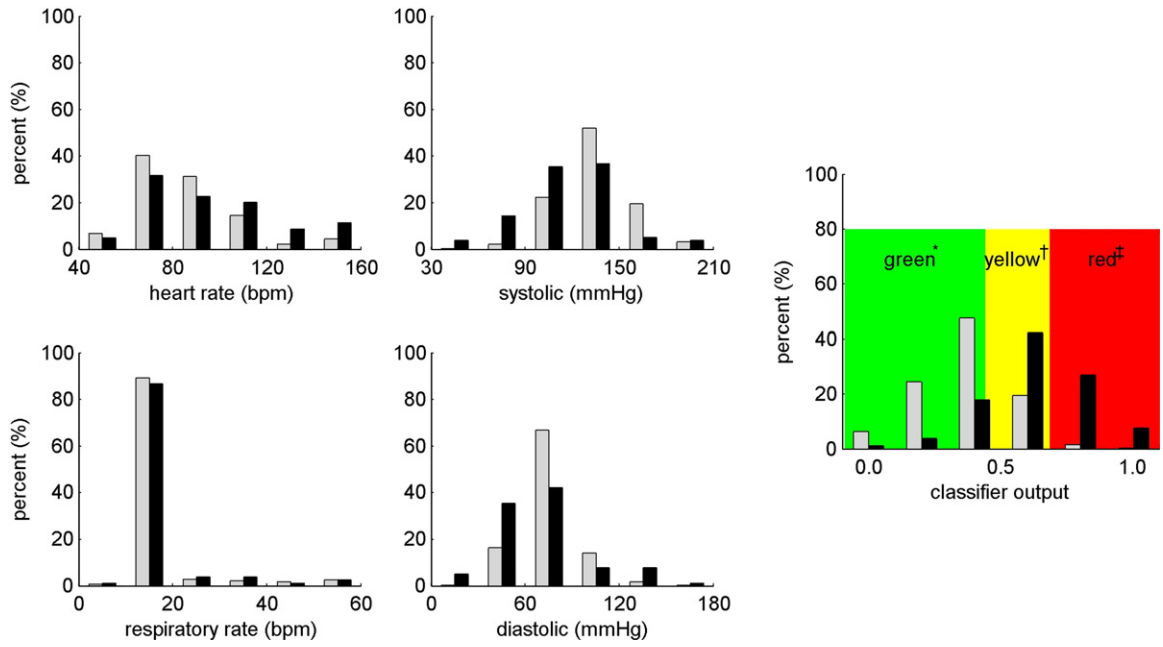



**Fig. 1** Histograms of basic vital signs and of the multivariate ensemble classifier for major hemorrhagic injury cases vs control cases. Histograms for each basic vital sign (HR, RR, SBP, and DBP) using the first nonzero value and the output of the multivariate ensemble classifier (using cross-validation with distinct training/testing data; see text for details). Patient populations for each histogram correspond to the populations in Table 3, whereas multivariate ensemble classification was applied to the entire study population. Right: Ensemble output (testing data) averaged from 100 iterations of cross-validation. Using one cutoff, ensemble classification yielded a sensitivity of greater than 85% and a PPV of 24%; patients below this threshold lie in the green background field. Using an alternative cutoff, ensemble classification offered a sensitivity of 36% and a PPV of greater than 60%; patients above this threshold lie in the red background field. *Green zone: 383 control cases and 11 major hemorrhagic injury cases; †yellow zone: 192 control cases and 39 major hemorrhagic injury cases; ‡red zone: 18 control cases and 28 major hemorrhagic injury cases.

ambiguous outcome patients and changing the outcome to the receipt of at least 5 U of RBC, 0.81 (95% CI, 0.76-0.86); and (*d*) use of only the initial 10 minutes, 0.81 (95% CI, 0.76-0.86).

## 4. Discussion

We found that accounting for measurement error and physiologic variability can significantly improve the association between vital signs and major hemorrhagic injury. Vital signs may be more informative about a trauma patient's circulatory state than previously appreciated in reports that did not explicitly consider these factors [26,34-36]. Moreover, these findings may inform the design and interpretation of a range of clinical trials that involve vital signs and how vital signs are used in clinical practice. The implications are cautionary, suggesting that such factors are important to consider. At the same time, these findings also suggest potential solutions.

The computational techniques used in this analysis have been previously described [28-30,33,37,38]. Here, the techniques were integrated to determine their cumulative effects in a population of trauma patients. These techniques significantly improved the association of vital signs and major hemorrhagic injury without the need for consideration of the patients' baseline vital signs, administration of medications, anatomical location of the injury, age, or mechanism of injury. Applied cumulatively, diagnostic performance exceeded prior reports on the individual techniques [33]. The vital sign patterns correctly classified by these techniques were not always self-evident by eye (eg, Fig. 2).

### 4.1. Clinical implications

We have shown that reliable vital sign data have a significantly higher association with a life-threatening pathophysiology, even as unreliable measurements were commonplace (Table 1). These findings support the adherence to proper vital sign measurement techniques; even better than excluding unreliable data, as was done in this retrospective study, would be reducing unreliable measurements in the first place. When procuring monitoring apparatus, it would be desirable to prioritize makes and models that possess maximum accuracy [11-13]. In addition, the study implies a potential benefit to continuing training of clinical staff to enhance the diagnostic value of vital sign

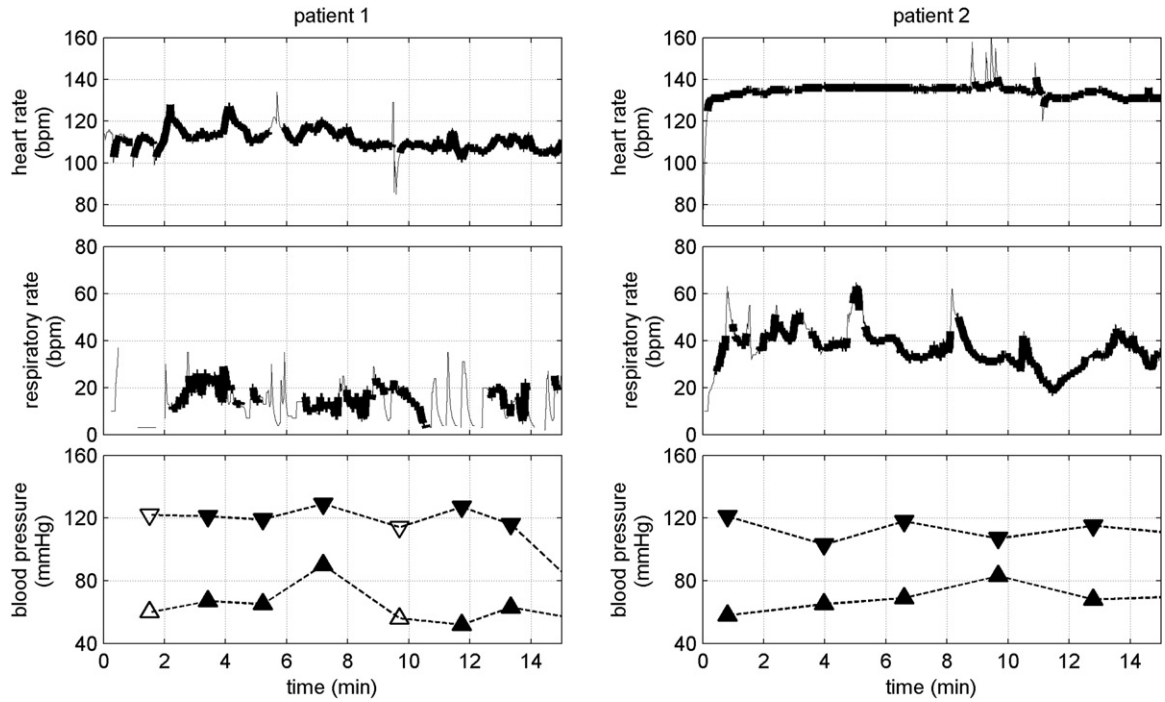



**Fig. 2** Case examples for which the presence or absence of major hemorrhagic injuries can be identified by patterns in the vital signs. Both cases had HRs of more than 120 beats per minute and normotension. In patient 1, the ensemble multivariate classifier—which weighs the HR, RR, and systolic and pulse pressures—indicated that major hemorrhagic injury was unlikely (ie, the classifier output lay in the low-risk green zone, with a 97% negative predictive value; see Fig. 1). Patient 1 did not require RBC transfusion and was diagnosed with a cerebral contusion and a neck injury without major vascular involvement. In patient 2, the ensemble multivariate classifier indicated that major hemorrhagic injury was probable (ie, the classifier output lay in the high-risk red zone, with a 60% PPV; see Fig. 1). Patient 2 had a grade III liver laceration; a fractured, disrupted pelvis; and a femur fracture and received 3 U of RBC. Thin lines and open triangles indicate unreliable data according to the automated algorithms; thick lines and solid triangles indicate reliable data.

measurement. Sources of unreliable vital sign data include poor electrode placement (eg, chest hair causing poor skin adhesion), excessive patient movement, and poor placement of BP cuffs. It would be truly revealing to study, prospectively, which sources of error are the most problematic and whether the association between vital signs and pathology can be enhanced through focused training.

Certain techniques suggested by this report might be applied at the bedside to assess the state of the casualty. For example, when patients arrive at the hospital, clinicians expecting obvious vital sign trends might be misled because we have found that transient perturbations may mask the underlying trends and that measurements made at the end of transport are not necessarily more useful than the preceding prehospital measurements. Shapiro et al [39] and Lipsky et al [40] reported that, among patients who arrived normotensive in the emergency department, one or more episodes of preceding hypotension were associated with higher acuity. Our findings suggest that, in addition to the most recent measurements, clinicians should consider the time average of recent data, which we have shown can be significantly more diagnostic.

This raises the question of what duration of time window is optimal for computation of average values of recent vital sign data, for example, 5, 15, or 60 minutes. The goal of the time averaging is to filter out transient perturbations; but if the time window gets too large, then time averaging can actually obscure trends developing in later data. Therefore, it is important that the time window should not be too large. We speculate that averaging over more than 15 minutes may not be diagnostically optimal, but this is difficult to answer definitively with the current data set because the records are of such heterogeneous duration.

Simultaneous consideration of multiple vital signs can also improve the value of the data. For instance, low BP could represent significant blood loss, the patient's normal baseline, or reduced adrenergic tone. Tachycardia and tachypnea suggest the former, normal rate and respiration suggest baseline physiology, and bradycardia and bradypnea suggest sympatholysis. Clinicians may be unable to mentally compute a multivariate statistical model; but a simple multivariate metric, such as the shock index (the ratio of HR and SBP [41,42]), can be applied at the bedside.

## 4.2. Research implications

We demonstrated that accounting for sources of measurement variability can yield significantly different results when analyzing vital sign data. Accordingly, we recommend the following steps for clinical research involving vital sign data: (*a*) report the make and model of any monitoring equipment used and, when available, provide accuracy citations [12,13,43]; (*b*) report relevant in-service training, or its absence, of the clinical staff; (*c*) keep the measurement environment as consistent as possible to reduce transient variability, or else use the average of several measurements; and (*d*) consider the use of validated clinical scores or propensity scores to supplement or replace individual vital signs.

In addition, we note that there has been academic interest in novel types of physiologic sensors intended to improve patient monitoring. The cost and effort necessary to adopt new sensor modalities might be weighed against the findings in this report, which are that standard vital signs can be significantly improved through application of some simple techniques. Academically, we suggest that new monitoring modalities should be directly compared against conventional monitoring, with consideration given to the sources of variability highlighted here.

## 4.3. Specific findings

Systolic BP was the best univariate predictor. We [37] and others [44,45] have previously found that prehospital trauma patients demonstrate substantial temporal variability. We reduced the effects of transient perturbations by using the time average of serial vital sign measurements, which yielded significantly higher ROC AUCs for SBP, higher than either the initial or final prehospital SBP. Diastolic BP alone was a weak predictor; but we found that it provides additional information independent of SBP because it is useful to compute pulse pressure, the difference between SBP and DBP [33]. In spontaneously breathing patients, reliable RR was a useful predictor of hemorrhage. This finding was anticipated by classic physiologic reports that demonstrated that blood flow to the carotid body chemoreceptors is reduced in early hemorrhage because of compensatory vasoconstriction. "Stagnant" hypoxia then develops in the chemoreceptors, triggering an increased respiratory drive and tachypnea [46-49]. Interestingly, this RR reliability algorithm was not originally developed to diagnose major hemorrhagic injury per se, but to identify intervals in the IP that matched clinicians' opinions that the respiratory waveform was rhythmic and consistent [28]. Used as a diagnostic tool, we found that reliable RR data were significantly more diagnostic than unreliable RR. We observed that unreliable RR was often falsely elevated (ie, biased) because of motion artifacts in the pneumogram that were incorrectly counted as additional breaths.

Only a subset of patients (59%) had a complete set of reliable vital signs within 15 minutes. This was consistent with prior reports that unreliable vital sign data are all too typical in clinical practice [1,2,4-6]. To deal with missing data, we used an ensemble classifier for multivariate classification, which was significantly better than univariate classification. In a prior report, the ensemble classifier was applied to a moving 2-minute window of vital sign data [33]. That approach was not as successful because, in any given 2-minute window, there was an exaggerated proportion of missing data and there was major minute-to-minute variability that, here, we successfully filtered out by time averaging over 15 minutes (see above). In addition, the

current ensemble uses pulse pressure instead of DBP and does not incorporate oxygen saturation, thus excluding weak univariate predictors.

### 4.4. Automated diagnostic algorithms

It is technically feasible to run this investigation's analysis algorithms in real time, automatically distinguishing between normovolemic vs hemorrhagic vital sign patterns. We speculate that such automated, continuous analysis could improve the quality and safety of any monitored patient, especially when the clinical staff is distracted or inexperienced. In addition, protocols for triage or resuscitation could be considered using the algorithm's output as a starting point that may be more clinically valid than any sole vital sign. Lastly, in some cases, the algorithm could enhance the judgment of the clinician (eg, cases such as in Fig. 2). Similar types of automated analysis of vital sign data may likewise prove useful for other clinical applications, such as early detection of acute deterioration of hospital ward patients [50].

### 4.5. Limitations

There are several factors to consider in terms of the internal validity of this study. First, there is no gold standard definition to retrospectively distinguish true hemorrhagic injury vs minor (or non) hemorrhagic injures. We therefore analyzed several alternative outcome definitions. The similar results, regardless of the specific definition, suggest that the findings were not an artifact of the outcome definition but will be similar given any reasonable definition of hemorrhagic injury (note that our database did not contain parameters such as base deficit and pH). Our findings would be further strengthened if future investigations demonstrate comparable findings given additional end points and pathologic processes.

As a second limitation, the present findings depended on our algorithms to identify reliable vital signs; and the results might be different with different algorithms. However, in developing these algorithms, we found that most analytic methodologies that we explored yielded similar results because, in practice, the different algorithms only differed about borderline cases, a minority of the data set [51]. In most of the cases, which were clearly reliable (eg, HR based on very clean ECG) or clearly unreliable (eg, HR based on very noisy ECG), different versions of the algorithms that we explored yielded consistent ratings of vital sign reliability. (Note that these reliability algorithms were not a priori developed to diagnose major hemorrhage but to match clinicians' opinions regarding whether waveform segments were clean with well-defined heartbeats [27] or breaths [28].)

Third, the data set was notable in that many patients were missing a full set of reliable data. However, we contend that this is a salient finding of the study, rather than a limitation, because it emphasizes the prevalence of unreliable vital sign data. At the same time, it did not hamper the univariate analyses because there were suitably large populations for each analysis. Finally, for the multivariate analysis, we were able to report a valid ROC AUC for the broadest study population (any patient with at least 1 reliable vital sign within the first 15 minutes) by using an ensemble classifier, which can tolerate missing data. The performance of the ensemble classifier was assessed through cross-validation, that is, with distinct training and testing patient populations.

In terms of the external validity of the study, the issues that we studied have been previously recognized [1,2,4-6]. This report offers a novel, quantitative analysis of their magnitude of effect in actual prehospital practice. It is not certain to what extent the quantitative results of this analysis will apply to different clinical settings, for example, emergency department vs hospital ward vs ground EMS, and different make and model of patient monitors. Likewise, there may be salient differences given alternative populations, for example, patients older in age with a higher rate of $\beta$-blocker medication. However, the study population of this report was reasonably large (>600 subjects); and such considerations were outside its scope. This analysis provides a prima facie demonstration that each of the factors is important and that specific strategies can significantly alter diagnostic test characteristics of routine clinical data. Further work is warranted to explore these factors in a diversity of clinical arenas and populations.

## 5. Conclusion

The study is notable for quantifying the magnitude of the effect of physiologic variability and measurement error on a diagnostic application of vital signs. These sources of variability were commonplace in this clinical data analysis. Techniques that accounted for the variability yielded significantly improved diagnostic test characteristics. Vital sign data are often treated uncritically in published reports. The findings here suggest that these factors should be carefully considered when using vital signs in clinical practice or research protocols.

## Acknowledgments

Funding/support:

This work was funded by the Combat Casualty Care Research Area Directorate of the US Army Medical Research and Materiel Command, Fort Detrick, MD.

Role of sponsor:

The sponsor did not participate in the design and conduct of the study; collection, management, analysis, and interpretation of the data; or preparation, review, or approval of the manuscript.

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 US Army or of the US Department of Defense. This article has been approved for public release with unlimited distribution.

## References


[1] Edmonds ZV, Mower WR, Lovato LM, et al. The reliability of vital sign measurements. Ann Emerg Med 2002;39:233-7.

[2] Jones DW, Appel LJ, Sheps SG, et al. Measuring blood pressure accurately: new and persistent challenges. JAMA 2003;289:1027-30.

[3] Davis JW, Davis IC, Bennink LD, et al. Are automated blood pressure measurements accurate in trauma patients? J Trauma 2003; 55:860-3.

[4] Lovett PB, Buchwald JM, Sturmann K, et al. The vexatious vital: neither clinical measurements by nurses nor an electronic monitor provides accurate measurements of respiratory rate in triage. Ann Emerg Med 2005;45:68-76.

[5] Friesdorf W, Konichezky S, Gross-Alltag F, et al. Data quality of bedside monitoring in an intensive care unit. Int J Clin Monit Comput 1994;11:123-8.

[6] Chambrin MC. Alarms in the intensive care unit: how can the number of false alarms be reduced? Crit Care 2001;5:184-8.

[7] Brasel KJ, Guse C, Gentilello LM, et al. Heart rate: is it truly a vital sign? J Trauma 2007;62:812-7.

[8] Victorino GP, Battistella FD, Wisner DH. Does tachycardia correlate with hypotension after trauma? J Am Coll Surg 2003;196:679-84.

[9] Staessen JA, Fagard R, Thijs L, et al. A consensus view on the technique of ambulatory blood pressure monitoring. The Fourth International Consensus Conference on 24-Hour Ambulatory Blood Pressure Monitoring. Hypertension 1995;26:912-8.

[10] O'Brien E, Waeber B, Parati G, et al. Blood pressure measuring devices: recommendations of the European Society of Hypertension. BMJ 2001;322:531-6.

[11] Clayton DG, Webb RK, Ralston AC, et al. A comparison of the performance of 20 pulse oximeters under conditions of poor perfusion. Anaesthesia 1991;46:3-10.

[12] Clemmer TP, Orme Jr JF, Thomas FO, et al. Outcome of critically injured patients treated at level I trauma centers versus full-service community hospitals. Crit Care Med 1985;13:861-3.

[13] Hedges JR, Feero S, Moore B, et al. Comparison of prehospital trauma triage instruments in a semirural population. J Emerg Med 1987;5: 197-208.

[14] Koehler JJ, Malafa SA, Hillesland J, et al. A multicenter validation of the prehospital index. Ann Emerg Med 1987;16:380-5.

[15] Baxt WG, Berry CC, Epperson MD, et al. The failure of prehospital trauma prediction rules to classify trauma patients accurately. Ann Emerg Med 1989;18:1-8.

[16] Emerman CL, Shade B, Kubincanek J. A comparison of EMT judgment and prehospital trauma triage instruments. J Trauma 1991;31:1369-75.

[17] Bickell WH, Wall Jr MJ, Pepe PE, et al. Immediate versus delayed fluid resuscitation for hypotensive patients with penetrating torso injuries. N Engl J Med 1994;331:1105-9.

[18] Dutton RP, Mackenzie CF, Scalea TM. Hypotensive resuscitation during active hemorrhage: impact on in-hospital mortality. J Trauma 2002;52:1141-6.

[19] Buist MD, Moore GE, Bernard SA, et al. Effects of a medical emergency team on reduction of incidence of and mortality from unexpected cardiac arrests in hospital: preliminary study. BMJ 2002;324:387-90.

[20] Dacey MJ, Mirza ER, Wilcox V, et al. The effect of a rapid response team on major clinical outcome measures in a community hospital. Crit Care Med 2007;35:2076-82.

[21] Hillman K, Chen J, Cretikos M, et al. Introduction of the medical emergency team (MET) system: a cluster-randomised controlled trial. Lancet 2005;365:2091-7.

[22] Kenward G, Castle N, Hodgetts T, et al. Evaluation of a medical emergency team one year after implementation. Resuscitation 2004;61:257-63.

[23] Jones AE, Brown MD, Trzeciak S, et al. The effect of a quantitative resuscitation strategy on mortality in patients with sepsis: a meta-analysis. Crit Care Med 2008;36:2734-9.

[24] Peng R, Chang C, Gilmore D, et al. Epidemiology of immediate and early trauma deaths at an urban level I trauma center. Am Surg 1998;64:950-4.

[25] Sauaia A, Moore FA, Moore EE, et al. Epidemiology of trauma deaths: a reassessment. J Trauma 1995;38:185-93.

[26] Cooke WH, Salinas J, Convertino VA, et al. Heart rate variability and its association with mortality in prehospital trauma patients. J Trauma 2006;60:363-70.

[27] McKenna TM, Bawa G, Kumar K, et al. The physiology analysis system: an integrated approach for warehousing, management and analysis of time-series physiology data. Comput Methods Programs Biomed 2007;86:62-72.

[28] Chen L, McKenna TM, Reisner AT, et al. Algorithms to qualify respiratory data collected during the transport of trauma patients. Physiol Meas 2006;27:797-816.

[29] Reisner AT, Chen L, McKenna TM, et al. Automatically-computed prehospital severity scores are equivalent to scores based on medic documentation. J Trauma 2008;65:915-23.

[30] 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;13:309-20.

[31] Lumley T, Diehr P, Emerson S, et al. The importance of the normality assumption in large public health data sets. Annu Rev Public Health 2002;23:151-69.

[32] DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics 1988;44:837-45.

[33] Chen L, McKenna TM, Reisner AT, et al. Decision tool for the early diagnosis of trauma patient hypovolemia. J Biomed Inform 2008;41: 469-78.

[34] Luna GK, Eddy AC, Copass M. The sensitivity of vital signs in identifying major thoracoabdominal hemorrhage. Am J Surg 1989;157:512-5.

[35] McGee S, Abernethy III WB, Simel DL. The rational clinical examination. Is this patient hypovolemic? JAMA 1999;281:1022-9.

[36] West JG, Murdock MA, Baldwin LC, et al. A method for evaluating field triage criteria. J Trauma 1986;26:655-9.

[37] Chen L, Reisner AT, Gribok A, et al. Exploration of prehospital vital sign trends for the prediction of trauma outcomes. Prehosp Emerg Care 2009;13:286-94.

[38] Chen L, Reisner AT, Gribok A, et al. Can we improve the clinical utility of respiratory rate as a monitored vital sign? Shock 2009;31:574-80.

[39] Shapiro NI, Kociszewski C, Harrison T, Chang Y, Wedel SK, Thomas SH. Isolated prehospital hypotension after traumatic injuries: a predictor of mortality? J Emerg Med 2003;25:175-9.

[40] Lipsky AM, Gausche-Hill M, Henneman PL, et al. Prehospital hypotension is a predictor of the need for an emergent, therapeutic operation in trauma patients with normal systolic blood pressure in the emergency department. J Trauma 2006;61:1228-33.

[41] King RW, Plewa MC, Buderer NM, et al. Shock index as a marker for significant injury in trauma patients. Acad Emerg Med 1996; 3:1041-5.

[42] Zarzaur BL, Croce MA, Fischer PE, et al. New vitals after injury: shock index for the young and age × shock index for the old. J Surg Res 2008;147:229-36.

[43] Runcie CJ, Reeve WG, Reidy J, et al. Blood pressure measurement during transport. A comparison of direct and oscillotonometric readings in critically ill patients. Anaesthesia 1990;45:659-65.
[44] Clemmer TP, Orme Jr JF, Thomas F, et al. Prospective evaluation of the CRAMS scale for triaging major trauma. J Trauma 1985;25:188-91.
[45] Morris Jr JA, Auerbach PS, Marshall GA, et al. The Trauma Score as a triage tool in the prehospital setting. JAMA 1986;256:1319-25.
[46] D'Silva JL, Gill D, Mendel D. The effects of acute haemorrhage on respiration in the cat. J Physiol 1966;187:369-77.
[47] Kenney RA, Neil E. The contribution of aortic chemoceptor mechanisms to the maintenance of arterial blood pressure of cats and dogs after haemorrhage. J Physiol 1951;112:223-8.
[48] Landgren S, Neil E. Chemoreceptor impulse activity following haemorrhage. Acta Physiol Scand 1951;23:158-67.
[49] Winder CV. Combination of hypoxic and hypercapnic stimulation at the carotid body. Am J Physiol 1942;136:200-6.
[50] Hravnak M, Edwards L, Clontz A, et al. Defining the incidence of cardiorespiratory instability in patients in step-down units using an electronic integrated monitoring system. Arch Intern Med 2008;168: 1300-8.
[51] Chen X, Chen L, Reisner AT, et al. Using confidence intervals to assess the reliability of instantaneous heart rate and respiratory rate: The 33rd Annual EMBS Conference, IEEE Engineering in Medicine and Biology Society and Biomedical Society. Boston, MA; 2011.

# Are Standard Diagnostic Test Characteristics Sufficient for the Assessment of Continual Patient Monitoring?

*Liangyou Chen, PhD, Andrew T. Reisner, MD, Xiaoxiao Chen, PhD, Andrei Gribok, PhD, Jaques Reifman, PhD*

***Background.*** *For diagnostic processes involving continual measurements from a single patient, conventional test characteristics, such as sensitivity and specificity, do not consider decision consistency, which might be a distinct, clinically relevant test characteristic.* ***Objective****. The authors investigated the performance of a decision-support classifier for the diagnosis of traumatic injury with blood loss, implemented with three different data-processing methods. For each method, they computed standard diagnostic test characteristics and novel metrics related to decision consistency and latency.* ***Setting.*** *Prehospital air ambulance transport.* ***Patients.*** *A total of 557 trauma patients.* ***Design****. Continually monitored vital-sign data from 279 patients (50%) were randomly selected for classifier development, and the remaining were used for testing. Three data-processing methods were evaluated over 16 min of patient monitoring: a 2-min moving window, time averaging, and postprocessing with the sequential probability ratio test (SPRT).* ***Measurements****. Sensitivity and specificity were computed. Consistency was quantified through cumulative counts of decision changes over time and the fraction of patients affected by false alarms. Latency was evaluated by the fraction of patients without a decision.* ***Results****. All 3 methods showed very similar final sensitivities and specificities. Yet, there were significant differences in terms of the fraction of patients affected by false alarms, decision changes through time, and latency. For instance, use of the SPRT led to a 75% reduction in the number of decision changes and a 36% reduction in the number of patients affected by false alarms, at the expense of 3% unresolved final decisions.* ***Conclusion****. The proposed metrics of decision consistency and decision latency provided additional information beyond what could be obtained from test sensitivity and specificity and are likely to be clinically relevant in some applications involving temporal decision making.* ***Key words:*** *continual patient monitoring; decision-support algorithm; sequential probability ratio test; physiological data.* ***(Med Decis Making 2013;33:225–234)***

Continual physiological monitoring is standard practice in many health care arenas, e.g., hospital wards and operating rooms, where vital-sign data are measured repeatedly so that if instability occurs it can be detected and treated promptly. However, false alarms are a major problem because standard alarms are triggered when certain parameter thresholds are reached.[1–3] All too often, the abnormality that triggers an alarm is either a measurement artifact or a benign transient event. Yet, when false alarms occur frequently, there is a deleterious effect on patients in that caregivers may be slow to respond to alarms with low positive predictive value.[4]

Received 10 August 2011 from DoD Biotechnology HPC Software Applications Institute (BHSAI), Telemedicine and Advanced Technology Research Center (TATRC), Fort Detrick, MD (LC, ATR, XC, AG, JR), and Massachusetts General Hospital Department of Emergency Medicine, Boston, MA (ATR). This work was performed at BHSAI, TATRC, US Army Medical Research and Materiel Command (USAMRMC), Fort Detrick, MD 21702. This work was partially supported by the Combat Casualty Care Research Area Directorate of the USAMRMC, Fort Detrick, MD. The funding agreement ensured the authors' independence in designing the study, interpreting the data, writing, and publishing the report. The following author is employed by the sponsor: Jaques Reifman is a U.S. Department of the Army Senior Research Scientist. 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 article has been approved for public release with unlimited distribution. Revision accepted for publication 24 March 2012.

DOI: 10.1177/0272989X12451059

Address correspondence to Jaques Reifman, BHSAI, TATRC, ATTN: MCMR-TT, 504 Scott Street, Fort Detrick, MD 21702-5012; e-mail: jaques.reifman@us.army.mil.

In this report, we considered a set of analytic methods for detecting abnormalities from continual physiological data and examined how the techniques compared through time. We examined whether standard test characteristics (sensitivity and specificity) were adequate for describing the resultant alarm behaviors from one time interval to the next. Specifically, we developed metrics to measure the temporal stability of test decisions, which we termed *consistency,* and examined the extent to which patient alarms were consistent through time. Our intent was to describe whether alarms tended to reoccur in the same patients from one time period to the next (on whom the clinical staff would be able to focus attention) or if (false) alarms were distributed throughout the entire monitored population (so that many disparate patients would—unnecessarily—require attention as the alarms were triggered).

We focused on several basic methods for pre- and postprocessing of continual vital-sign data into and out of a core alarm algorithm. Analytic methods for identifying irregularities from a set of time-series data have been well established in the manufacturing quality control literature. Methods dealing with this problem include the sequential probability ratio test (SPRT),[5,6] which evaluates the likelihood ratio of 2 hypotheses based on sequentially available evidences. Alternatives include the control chart method,[7,8] the belief-modeling method,[9] and other Bayesian-based methods.[10,11] Among these methods, the SPRT is simple to calculate and, for given false-positive and false-negative probabilities, requires the smallest number of samples to achieve a decision (unless the statistical model is grossly incorrect).[5]

Our goal was to investigate if conventional test characteristics were adequate for assessing the basic performance of an alarm or if it was also necessary to consider its temporal consistency. In a comparative analysis, we employed 3 methods for pre- and postprocessing of continual data into and out of our core alarm algorithm. Compared with a 2-min moving window, we examined if additional time averaging and the SPRT could alter the overall accuracy, the temporal consistency, and the latency of the algorithm output. The core alarm algorithm was a multivariate classifier for the diagnosis of traumatic injury with blood loss, given data from a standard prehospital patient monitor.[12] This analysis has implications for any diagnostic process involving continual vital-sign measurements from a single patient.

## METHODS

This is a retrospective, comparative analysis, based on a previously reported ensemble classifier,[12] which provides automated detection of traumatic injury with blood loss in prehospital patients based on basic vital signs. We simulated 3 methods to process real-time data during the initial 16 min of prehospital patient transportation. The moving window method involved a moving 2-min analysis window; at every moment in time, the classifier was applied to the most recent 2 min of physiological data. The time-averaging method analyzed all available data from a given patient, from the onset of the data record to the current time (up to a maximum of 16 min). In the SPRT method, we applied the SPRT to the output of the classifier.

### Trauma Patient Data

The physiological time-series data were collected from 643 trauma-injured patients during their first 16 min of helicopter transport to a trauma center.[13] The time-series variables were measured by ProPaq 206EL vital-sign monitors (Protocol Systems, Beaverton, OR) and consisted of electrocardiogram, photoplethysmogram, and respiratory waveform signals recorded at various frequencies and their corresponding monitor-calculated vital signs, including heart rate (HR), respiratory rate (RR), and arterial oxygen saturation ($SaO_2$), recorded at 1-s intervals, and systolic (SBP), mean, and diastolic (DBP) blood pressures collected intermittently at multiminute intervals.

We performed chart reviews to determine whether the transported trauma patients had traumatic injury with blood loss. Traumatic injury with blood loss was defined as requirement of a blood transfusion within 24 h upon arrival at the trauma center and also documentation of an explicitly hemorrhagic injury, either a) laceration of solid organs, b) thoracic or abdominal hematomas, c) explicit vascular injury and operative repair, or d) limb amputation. Patients who received blood but did not meet the documented injury criteria (60 cases), and patients who died before arrival at the hospital (26 cases) were excluded from the analyses because of uncertainty about whether they truly suffered traumatic injury with blood loss. Thus, we used a total of 557 patients, of which 61 were categorized as patients with traumatic injury and blood loss and the remaining 496 as controls.

### Decision-Support Classifier: Training

The ensemble classifier aggregates 25 least-squares linear classifiers, each trained with a different subset of 5 input variables (HR, RR, $SaO_2$, SBP, and DBP) and with target values of 0.0 and 1.0, standing for control and traumatic injury with blood loss outcomes, respectively, to generate an (arithmetic) average output that can be used to discriminate the 2 outcomes.[12] We assigned ensemble-averaged outputs of $\leq 0.5$ as control outcomes and outputs of $> 0.5$ as traumatic injury with blood loss. The ensemble classifier has been shown to provide more consistent performance than a single linear classifier, and importantly, it accommodates missing data, providing an output as long as any 1 of the 5 inputs is available.[12]

We randomly selected 50% of the study population (279 patients; 248 controls and 31 patients with traumatic injury and blood loss) to train (i.e., develop) the classifier. In prior studies,[14] we found that prehospital vital-sign data are very noisy, and hence, we developed algorithms that automatically assess the reliability of each vital sign used as input to the classifier.[15–17] We also reported that reliable data are diagnostically superior to unreliable data.[15,18] In another study,[14] we found that there are no major time-series trends in these vital-sign data, and averaging the most reliable data during transport yielded the best discriminatory performance. Consequently, we used the average value of the most reliable training data points from the first 16 min of transport time as input to train the ensemble classifier.

### Evaluation of the Moving Window, Time-Averaging, and SPRT Methods

We investigated 3 methods to pre- and postprocess the ensemble classifier data. In each method, 1) the first 2 min of transport vital-sign data were used as a buffer where no classifications were made; 2) every 1 s we averaged the most reliable available vital-sign data (HR, RR, etc.) over a specified time window, input the averaged values to the classifier, and obtained an output; and 3) every 15 s, we averaged the previous 15 classifier outputs to generate a decision. The 3 methods differed on the length of the preprocessing time window of the classifier input data in item 2 (above) and on any additional postprocessing in the classifier outputs in item 3.

For the moving window, we averaged the classifier inputs over a 2-min time window and compared the averaged decision every 15 s with a 0.5 threshold. The time-averaging method differed from the first method in that the length of the time window for averaging the vital-sign input data grew continually up to the current decision time so that all available data were considered for each decision. In the SPRT method, the classifier outputs were further processed as inputs to the SPRT to generate a SPRT decision (or no decision), as described below.

### The Sequential Probability Ratio Test

We investigated the ability of Wald's SPRT[5,6] to consider the sequential nature and postprocess the outputs of the ensemble classifier while balancing decision accuracy, consistency, and latency. Given a sequence of outputs $Y_1, Y_2, \ldots$ not necessarily independent from the ensemble classifier, so that $Y = N(\mu_Y, \sigma_Y^2)$ is a normal Gaussian process with an unknown mean $\mu_Y$ and a known constant variance $\sigma_Y^2$, the SPRT classifies a patient as control or traumatic injury with blood loss, or makes no decision, based on hypothesis testing. Note that $\sigma_Y^2$ was estimated as the variance of the ensemble classifier outputs at the end of the transport, i.e., at 16 min, and was kept fixed throughout the analysis. The SPRT tests a null hypothesis ($H_0$) that $\mu_Y = \mu_0$ against an alternative hypothesis ($H_1$) that $\mu_Y = \mu_1$, where $\mu_0$ and $\mu_1$ denote the arithmetic mean values of the classifier outputs for the control and traumatic injury with blood loss cases, respectively, with $\mu_0 < \mu_1$. If we let $p_0$ and $p_1$ be the probability density functions governing the two hypotheses, $H_0$ and $H_1$, respectively, then the observed likelihood ratio at decision time $J$ can be represented as $l_J = \prod_{j=1}^{J} \frac{p_1(Y_j)}{p_0(Y_j)}$, with $J = 1, 2, \ldots$.

According to Wald's SPRT methodology,[5] we

$$\begin{gathered}\text{accept } H_0 \text{ (control)}, \text{if } \log(l_J) < B; \text{or} \\ \text{accept } H_1 \text{ (traumatic injury with blood loss)}, \\ \text{if } \log(l_J) > A; \text{or} \\ \text{continue to decision time } J+1, \text{if } B \leq \log(l_J) \leq A,\end{gathered} \quad (1)$$

where $A$ and $B$ are constants, with $0 < B < A < \infty$, chosen using Wald's criteria,[5] as to yield nominal false-positive probability ($\alpha$; $0.0 < \alpha < 0.5$) and nominal false-negative probability ($\beta$; $0.0 < \beta < 0.5$) as follows:

$$\begin{aligned} A &= \log\frac{1-\beta}{\alpha}, \text{ and} \\ B &= \log\frac{\beta}{1-\alpha}. \end{aligned} \quad (2)$$

When $\alpha$ and $\beta$ are relatively small (e.g., $< 0.05$), the SPRT tends to delay making a decision until

additional corroborating classifier outputs become available. Conversely, when α and β are large (e.g., ≈ 0.5), the SPRT makes quicker, albeit less accurate, decisions. Thus, by appropriately selecting these two parameters, we can balance decision accuracy, consistency, and delay. To this end, we determined the nominal probabilities α and β by minimizing a cost function ¢, which linearly combined the accuracy of the decisions, defined by its sensitivity ($S_e$) and specificity ($S_p$), at the end of the transport (i.e., at 16 min); the cumulative incidences of decision changes ($D_c$; from control to traumatic injury with blood loss and vice versa) over the 16 min of transport time; and the fraction of patients with no decision ($N_d$) at the end of the transport. Accordingly, we defined ¢ as follows:

$$¢ = \frac{1-S_e}{0.05} + \frac{1-S_p}{0.05} + \frac{D_c}{10} + \frac{N_d}{0.01}, \quad (3)$$

where the rescaling factors of the summands were empirically obtained through SPRT trial simulations on the training data so to normalize the effect of each of the four summands on ¢.

Under the Gaussian model, the log-likelihood ratio $\log(l_J)$ in equation 1 can be recursively calculated as follows:

$$\log(l_{J+1}) = \log(l_J) + \frac{\mu_1 - \mu_0}{{\sigma_Y}^2}\left(Y_{J+1} - \frac{\mu_1 + \mu_0}{2}\right), \; J = 0, 1, 2, \ldots, \quad (4)$$

where the initial log-likelihood $\log(l_0)$ can be selected arbitrarily and was set to 0.0 in this study. While it has been shown that the SPRT achieves a selected confidence in the shortest decision time,[5] it may not always arrive at a decision. However, when a decision was made, we noted the decision, stuck to it, and restarted the SPRT process from that time point until a new decision emerged.

### Investigational Metrics

We compared the performance of the 3 data-processing methods using testing data from 278 patients where we evaluated the accuracy, latency, and consistency (in a sense to be defined) of the methods in aggregate using the following 5 performance metrics:

1. Sensitivity: at any given time *t*, the fraction of patients with traumatic injury and blood loss who were correctly identified by the algorithm at time *t*;
2. Specificity: at any given time *t*, the fraction of control patients who were correctly identified by the algorithm at time *t*;
3. No decisions: at any given time *t*, the fraction of patients without a decision out of the total number of patients;
4. Cumulative decision changes: the cumulative count up through time *t* of decision changes $D_c$; and
5. False-alarm-affected patients: the fraction of control patients incorrectly identified as having traumatic injury with blood loss, at or before time *t*, out of the total number of patients.

Every 2 min, from 2 to 16 min of transport time, we performed statistical tests of significance with pairwise comparisons between the investigational methods (i.e., moving window, time averaging, and SPRT). For proportions (sensitivity, specificity, no decisions, and false-alarm-affected patients), we employed Liddell's exact test.[19] The counts of total decision changes throughout the population cannot be statistically evaluated, so we also computed the total decision changes *per subject* and applied the Wilcoxon signed–rank test to the distributions. For all statistical tests, we considered a $P$ value of $< 0.05$ to be statistically significant.

## RESULTS

Figure 1 illustrates the continual output of the 3 data-processing methods, the moving window, time-averaging, and SPRT methods, in monitoring 4 control subjects (panel A) and 3 subjects with traumatic injury and blood loss (panel B). Each tile in the figure represents a 15-s outcome decision, with red (or dark) representing traumatic injury with blood loss decisions, green (or medium gray) control, and yellow (or light gray) no decisions. The selected control subjects illustrate different patterns in outcome decisions that we observed in the 248 control subjects in the testing data. For example, for subject 364, all 3 methods made correct and consistent control decisions over the 16-min transport time. For subject 607, each method generated some false-positive (i.e., false traumatic injury with blood loss) decisions. However, the moving window method generated the most frequent number of decision changes (3 changes from control to traumatic injury with blood loss and 3 from traumatic injury with blood loss to control, for a total of 6 decision changes), while the other 2 methods generated 2 decision changes each. For the third subject (640), unlike the other 2 methods, the SPRT method avoided making incorrect decisions (and decision changes), but the decision was delayed by more than 4 min. Finally, for subject 749, the SPRT was not able to make a definite decision during the

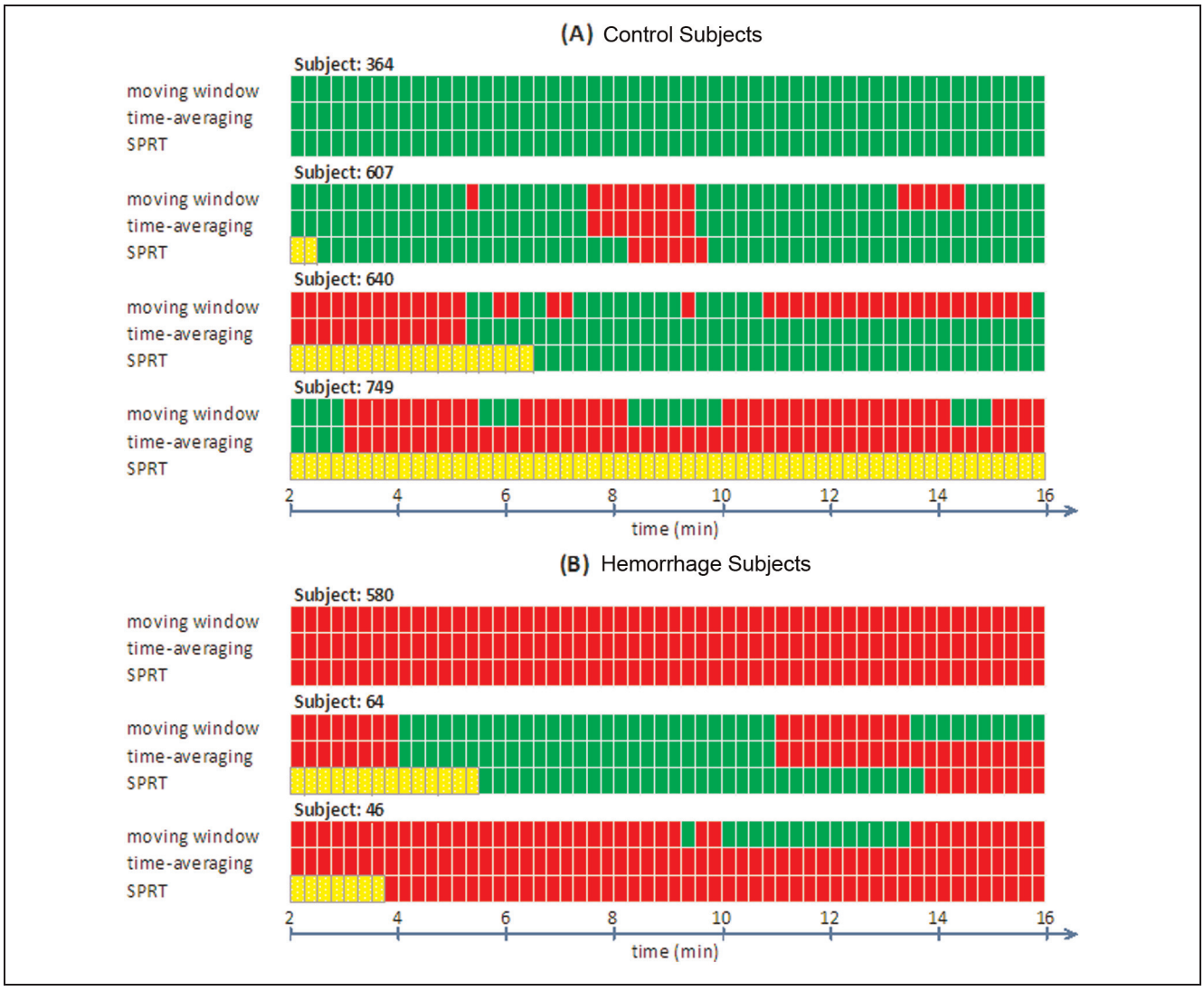



*Figure 1 Continual outcome decisions over the 16 min of transport time for each of the 3 data-processing methods. (A) Selected pattern for 4 control subjects, and (B) 3 subjects with traumatic injury and blood loss. Each tile represents a 15-s outcome decision, with red (or dark) representing traumatic injury with blood loss decisions, green (or medium gray) control, and yellow (or light gray) no decisions. SPRT, sequential probability ratio test.*

16-min transport time, while the other 2 methods generated decision changes and mostly incorrect decisions.

Panel B illustrates 3 patterns of decisions observed within the 31 patients in the testing set with traumatic injury and blood loss: for subject 580, all methods generated a consistent decision; for subject 64, the methods generated intermittent false-negative (i.e., false control) decisions, with the moving window method yielding an incorrect decision at 16 min; and for subject 46, all methods generated the correct final decision—however, the moving window produced decision changes and some incorrect decisions, while the SPRT did not produce a decision until almost 4 min.

Figure 2 illustrates the performance of the methods based on the 5 metrics (sensitivity, specificity, no decisions, cumulative decision changes, and false-alarm-affected patients) used to evaluate the accuracies, latencies, and consistencies of the methods for the 278 testing subjects over the 16-min transport time. Each of the 3 methods—moving window, time averaging, and SPRT—yielded comparable performance in terms of sensitivity and specificity at the end of the

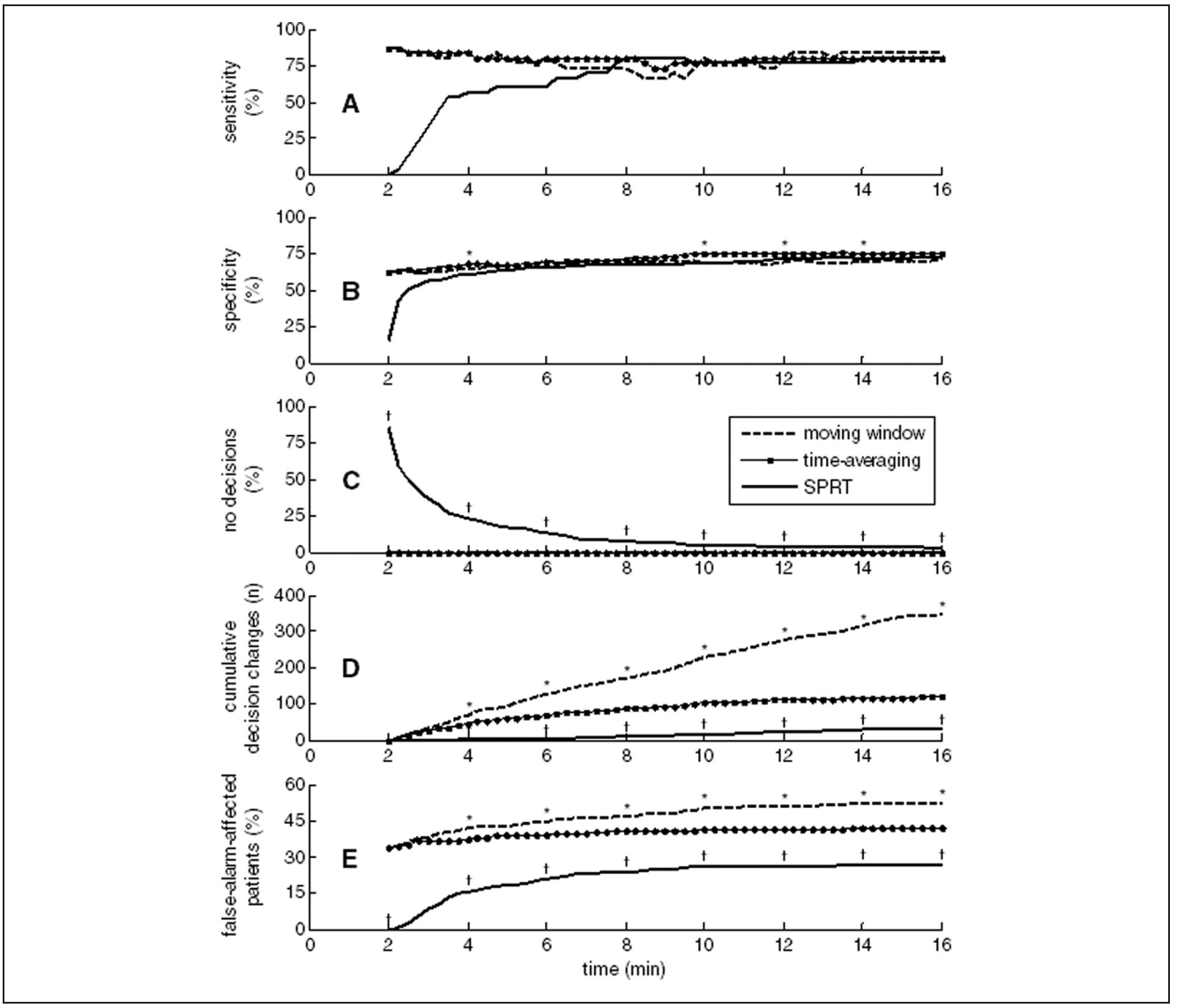



*Figure 2 Comparison of 3 data-processing decision methods for the 278 testing subjects analyzed over the 16-min transport time based on 5 performance metrics: (A) sensitivity, (B) specificity, (C) fraction of patients with no decisions, (D) cumulative number of decision changes, and (E) false-alarm-affected patients. Pairwise tests of significance were performed every 2 min. Proportions were compared by Liddell's exact test (panels A–C, E). Panel D illustrates cumulative count of total-population decision changes, and the Wilcoxon signed–rank test was applied to the per patient counts of decision changes.*$^{*}P < 0.05$, *time averaging v. moving window.* $^{\dagger}P < 0.05$, *SPRT v. both moving window and time averaging.*

transport time (sensitivity: 83%, 80%, and 80%, respectively; specificity: 71%, 75%, and 73%, respectively). Note that the SPRT method provided relatively low sensitivity and specificity ($\leq 60\%$) during the first 6 min of transport because of a large fraction of patients without SPRT decisions (see panel C). For instance, at 2 min, fewer than 25% of the patients had a decision rendered by the SPRT, and consequently, the corresponding sensitivity was also less than 25%. The SPRT method failed to make a decision at 16 min for 8 subjects (or 3% of the subjects), while the other 2 methods showed no decision latency (panel C).

In terms of consistency of decisions, the SPRT demonstrated a significantly reduced fraction of

false-alarm-affected patients throughout and at the end of the 16-min transport, compared with both other methods—27% of the subjects, which was 36% lower than the time-averaging method (42% of the subjects) and 48% lower than the moving window method (52% of the subjects; panel E). The SPRT also consistently generated fewer numbers of decision changes over time (29 total decision changes v. 118 for the time-averaging method and 348 for the moving window method; panel D).

The time-averaging method was more consistent than the moving window method, with significantly fewer false-alarm-affected patients and average decision changes *per patient*. The time-averaging method did not demonstrate the latency of the SPRT method.

## DISCUSSION

In this article, we studied the accuracy, consistency, and latency of a decision-support classifier employing three different data-processing methods for the continual prehospital diagnosis of traumatic injury with blood loss in 557 trauma patients. It is striking that all methods showed very similar sensitivities and specificities yet very different temporal behaviors. For instance, Wald's SPRT was much more consistent, generating false alarms in significantly fewer patients, with significantly fewer decision changes.

There are 2 major implications. First, for some continual monitoring applications, standard test characteristics, e.g., sensitivity and specificity, are insufficient for describing the performance of a classifier because they do not describe if false alarms occur repeatedly in a limited subpopulation or if false alarms are evenly distributed throughout a population. Second, as a corollary, it is apparent that pre- and postprocessing of time-series data can significantly alter temporal consistency, as was seen in the application of time averaging and of the SPRT, a classic technique intended for precisely this type of application.

*Insufficiency of standard test characteristics for describing the continual performance of a classifier.* For the continual monitoring of patients, standard test characteristics do not consider the sequential nature of the algorithm's decisions when there are repeated decisions being made on each subject. For example, while 2 binary decision classifiers may have similar overall sensitivity and specificity, 1 may be less stable than the other, continually ''flipping'' its decisions through time (which is naturally exacerbated the more that a classifier is sensitive to transient noise in the signal). We found this exact phenomenon in our data set: After 5 to 10 min, the 3 investigational methods had similar sensitivities and specificities, but there were significant differences in the total number of patients affected by a false alarm. Using the SPRT significantly reduced the fraction of false-alarm-affected patients by approximately half, compared to the moving window method.

We speculate that this effect was notable in this analysis because the prehospital vital signs showed considerable intrasubject variability through time, with sizable fluctuations in HR, blood pressure, etc., during the course of prehospital transport.[14] Comparable fluctuations in the prehospital vital signs of trauma patients have been observed in other prehospital studies as well,[20–22] which may be physiological responses to episodic stimuli (e.g., pain and fear), to episodic therapies (e.g., fluids), or to underlying pathology, as well as some degree of routine biological variability and measurement error.

In general, are standard diagnostic test characteristics sufficient for the assessment of continual patient monitoring, or is it appropriate to quantify classifier consistency? It is likely that the frequency of decision changes in diagnostic classification is dependent on the classifier evaluation frequency, the temporal fluctuations in the diagnostic data, and the proximity of the classifier output to the decision boundary. Presumably, there is a continuum of diagnostic applications in terms of the classifier consistency through time. If the diagnostic data are temporally stable during intervals of disease and health, then standard test characteristics are likely sufficient. At the other extreme, if the diagnostic data fluctuate through time, then the diagnostic classification will also fluctuate through time, and it may be illuminating to consider metrics of consistency (as we have done in this report) in addition to standard test characteristics. In many reports, continual classifiers are evaluated without explicit consideration of their performance and consistency through time, such as reports by our group[12] and by others.[23–25] It is likely that, at least for a subset of continual monitoring applications, standard diagnostic test characteristics are insufficient and it would be valuable to consider consistency to quantify clinically relevant properties of the diagnostic test.

In addition, evaluating a temporal classifier through time can reveal if performance changes because of temporal disease progression. Presumably, it is easier to diagnose blood loss or septic shock as the pathology progresses, due to the spectrum

effect (e.g., when a diagnostic test performs better in a study population with more severe disease. Consider that the sensitivity of a hypothetical dip-test for leukocyte esterase in the diagnosis of urinary tract infection may be higher in patients of an underserved population, who tend to receive evaluation later in the course of disease, rather than in patients of an affluent population, who are promptly evaluated after the earliest symptoms). Spectrum effects also affect the temporal consistency of a diagnostic classifier, because small fluctuations in diagnostic data for a borderline case would be more likely to affect diagnostic classification (e.g., during early stages of blood loss). By contrast, cases with more advanced pathology will often have more frankly abnormal diagnostic data, and so temporal fluctuations are unlikely to alter diagnostic classification. That diagnostic classification may become easier as the disease process progresses is often well recognized. For instance, Cuthbertson[26] reported test characteristics for an investigative early warning score over hourly intervals, e.g., 1 h prior to patient acute deterioration, 2 h prior, etc. However, it was not reported to what extent the true and false alarms occurred in the same patients hour by hour, i.e., consistency. In this report, we describe the minute-by-minute performance of an investigational algorithm during the initial 16 min of prehospital transportation, including the temporal variation of decision changes in the same patients and the fraction of total patients affected by some of these changes. At least in our application, the additional statistics provide information beyond standard test characteristics, perhaps in part because we examined data measured soon after traumatic injury.

*Pre- and postprocessing of time series alters performance of an automated continual classifier.* Pre- and postprocessing of time-series data is appropriate for removing noise that occurs over faster time scales than the process of interest, thus enhancing the underlying signal. In this study, the narrow 2-min moving window caused a large number of patients to trigger false alarms (24% more than the time-averaging approach and 93% more than the SPRT approach). Failure of developers of monitoring algorithms to explicitly consider classifier output stability, or consistency, through time will presumably exacerbate the well-described problem of false alarms in medical monitoring systems[1–4] and will likely decrease the incentive for caregivers to adopt novel decision-support technologies. Conversely, excessively stable classifiers are also problematic, causing unacceptable latency when a patient's state does change. The challenge is to optimize the tradeoffs between classifier accuracy, consistency, and latency.

Consider time averaging. As long as the noise in the time series has no major bias, this is a practical technique for filtering out measurement error and transient physiological events. For a monitoring algorithm, the time-averaging window should be shorter than the onset time of the disease of interest. In other words, time averaging over 15 min may be useful when seeking hemorrhage physiology, although time averaging over 60 min might be too large a window, causing unacceptable latency to the detection of hemorrhage physiology that can progress in less than an hour. In this report, the time-averaging method was able to improve decision consistency (with 66% fewer decision changes) and reduce false-alarm-affected patients (with 20% fewer false-alarm-affected patients) compared with the simple 2-min moving window method.

A prior report corroborates this principle: that it is often possible to reduce false alarms at the expense of clinically acceptable latency. In monitoring children at home by pulse oximetry, Gelinas and others[27] suggested that the rate of hypoxia alarms ($SpO2 < 85\%$) could be reduced from 3.6 to 0.2 alarms per night without missing any clinically significant events, simply by requiring a 10-s duration of hypoxia (rather than alarming the instant that the hypoxia threshold was met).

*The SPRT: a classic technique that can improve temporal consistency during continual monitoring.* One classic application of the SPRT is for the evaluation of a shipment of manufactured components. Components are measured 1 by 1 until a SPRT decision is rendered that the set of components is within (or outside of) the acceptable tolerances. Our investigational algorithm is analogous in that measurements were taken repeatedly from 1 trauma patient, and the SPRT was used to decide whether the patient was within (or outside of) the range of vital signs typical of patients with traumatic injury and blood loss. Of course, given a shipment of components, individual measurements are statistically independent, while there is temporal correlation when measurements are repeated in the same patient. Regardless, our findings suggest that the SPRT is suitable for improving the consistency of the investigational classifier based on continual physiological data.

In the medical area, the SPRT has been previously applied to the performance monitoring of clinical teams[26–30] (to continually monitor the surgical outcome rate and ensure it does not deviate from the expected success rate), routine surveillance of drug

safety[31] (to continually monitor whether a new vaccine is safe over a period of time), and determination of early stopping criteria of clinical trials[32,33] (to allow the trial to be stopped as soon as the information accumulated is considered sufficient to reach a conclusion). Our results demonstrated that the SPRT may be effective for continual physiological monitoring, in the reduction of false-alarm-affected patients (36% fewer patients than the time-averaging method) and overall decision changes (75% fewer decision changes). The tradeoff was the occurrence of some decision latency because, unlike the other investigative methods, the SPRT can yield an ''undecided'' output (see Figure 2). Indeed, for several cases (3% of the total), there was never a diagnostic decision generated when applying the SPRT. For applications in which such a tradeoff is acceptable, the SPRT is optimal in the sense that, mathematically, it guarantees the smallest number of samples to achieve a decision for given false-positive and false-negative probabilities.[5] The performance of the SPRT depends on the selected nominal probabilities $\alpha$ and $\beta$, which can be set either arbitrarily or by optimizing certain cost function during classifier training. Properly chosen $\alpha$ and $\beta$ may improve the sensitivity and specificity, and decrease the cumulative incidences of decision changes, with acceptable final unresolved decisions. However, improperly chosen $\alpha$ and $\beta$ may significantly downgrade the sensitivity or the specificity. As well, when we first attempted to optimize the SPRT with a cost function customized wholly to yield small false-positive $\alpha$ and false-negative $\beta$ probabilities, we improved the final accuracy but simultaneously increased the unresolved decisions to 40% on the testing data. In the end, the cost function defined in equation 3 provided a simple yet effective tool to balance accuracy, consistency, and latency.

This tradeoff between latency and consistency may limit the application of the SPRT in the detection of conditions that involve an imminent threat to life, e.g., cardiac tachyarrhythmia. However, in the monitoring of early disease states, when some latency is acceptable, e.g., early hemorrhage detection,[12] sepsis detection,[25,34] or other early warning functionality,[23,24] we suggest that the SPRT may provide a means to improve classifier stability and to reduce false alarms, without any necessary loss in decision accuracy.

*Identification of traumatic injury with blood loss via continual physiological monitoring.* The potential usefulness of the diagnostic classifier described in this report is not the focus of this study, and an assessment of potential clinical value must be tempered by the fact that the analysis is retrospective, based on post hoc classification as to whether each subject had traumatic injury with blood loss. Having said that, we believe that there is potential clinical value to the methodological application of conventional and commonsense analysis techniques to standard vital-sign data, e.g., noise rejection, time averaging, and multivariate classification. We previously found that automated techniques are diagnostically equivalent to prehospital severity scores based on medics' documentation.[15] In this case, we focused on the identification of hemorrhage because blood loss is 1 of the 2 primary reasons why trauma patients die,[35,36] but in many cases it can be treated effectively with blood transfusion and surgical hemorrhage control. We speculate that formal quantitative analysis of continual vital signs may be able to supplement today's convention, which relies on informal clinician judgments to integrate vital-sign data with other important clinical data. For instance, automated algorithms during prehospital care could be useful for triage and to aid the receiving hospital to efficiently mobilize proper resources, such as surgical teams and units of blood. Similar techniques could identify hospitalized patients who suffer unexpected episodes of blood loss during convalescence, e.g., early warning systems. However, actual performance and clinical usefulness must be prospectively assessed, and the optimal approach to decision support for trauma patients (e.g., attempt to identify any patients with traumatic injury and blood loss v. attempt to identify patients with uncontrolled, ongoing blood loss) involves open questions that are not addressed in this analysis.

## CONCLUSION

Over time, all 3 methods converged to demonstrate very similar diagnostic accuracy (i.e., sensitivity and specificity). However, their consistency was significantly different. The SPRT significantly reduced the total number of patients affected by false alarms, but with significantly greater latency, compared with the moving window method and the time-averaging method. Time averaging showed significantly fewer patients affected by false alarms compared with moving window, and without latency. These findings highlight how there are continual monitoring applications for which the proposed test characteristics provide additional, useful information. Metrics of consistency and latency can demonstrate additional properties that are likely relevant to clinical practice.

## REFERENCES


1. Korniewicz DM, Clark T, David Y. A national online survey on the effectiveness of clinical alarms. Am J Crit Care. 2008;17(1): 36–41.

2. Imhoff M, Kuhls S. Alarm algorithms in critical care monitoring. Anesth Analg. 2006;102(5):1525–37.

3. Chambrin MC. Alarms in the intensive care unit: how can the number of false alarms be reduced? Crit Care. 2001;5(4):184–8.

4. Lawless ST. Crying wolf: false alarms in a pediatric intensive care unit. Crit Care Med. 1994;22:981–5.

5. Wald A. Sequential tests of statistical hypotheses. Ann Math Statist. 1945;16(2):117–86.

6. Wald A. Sequential Analysis. New York: John Wiley and Sons; 1947.

7. Reynolds MR, Kim K. Multivariate control charts for monitoring the process mean and variability using sequential sampling. Seq Anal. 2007;26(3):283–315.

8. Bogowicz P, Flores-Mir C, Major PW, Heo G. Sequential analysis applied to clinical trials in dentistry: a systematic review. Evid Based Dent. 2008;9(2):55–60.

9. Rácz A. Comments on the sequential probability ratio testing methods. Ann Nucl Energy. 1996;23(11):919–34.

10. Carlin BP, Chaloner K, Church T, Louis TA, Matts JP. Bayesian approaches for monitoring clinical trials with an application to toxoplasmic encephalitis prophylaxis. Statistician. 1993;42: 355–67.

11. Brown M. Bayesian detection of changes of a Poisson process monitored at discrete time points where the arrival rates are unknown. Seq Anal. 2008;27(1):68–77.

12. Chen L, McKenna TM, Reisner AT, Gribok A, Reifman J. Decision tool for the early diagnosis of trauma-patient hypovolemia. J Biomed Inform. 2008;41:469–78.

13. 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):363–70.

14. Chen L, Reisner AT, Gribok A, Reifman J. Exploration of prehospital vital-sign trends for the prediction of trauma outcomes. Prehosp Emerg Care. 2009;13(3):286–94.

15. Reisner AT, Chen L, McKenna TM, Reifman J. Automatically-computed prehospital severity scores are equivalent to scores based on medic documentation. J Trauma. 2008;65(4):915–23.

16. Yu C, Liu Z, McKenna T, Reisner AT, Reifman J. A method for automatic identification of reliable heart rates calculated from ECG and PPG waveforms. J Am Med Inform Assoc. 2006;13(3):309–20.

17. Chen L, McKenna TM, Reisner AT, Reifman J. Algorithms to qualify respiratory data collected during the transport of trauma patients. Physiol Meas. 2006;27(9):797–816.

18. Chen L, Reisner AT, Gribok A, McKenna TM, Reifman J. Can we improve the clinical utility of respiratory rate as a monitored vital sign? Shock. 2009;31(6):574–80.

19. Liddell FDK. Simplified exact analysis of case-referent studies: matched pairs; dichotomous exposure. J Epidemiol Community Health. 1983;37:82–4.

20. Rhee KJ, Willits NH, Turner JE, Ward RE. Trauma Score change during transport: is it predictive of mortality? Am J Emerg Med. 1987;5:353–6.

21. Shapiro NI, Kociszewski C, Harrison T, Chang Y, Wedel SK, Thomas SH. Isolated prehospital hypotension after traumatic injuries: a predictor of mortality? J Emerg Med. 2003;25:175–9.

22. Lipsky AM, Gausche-Hill M, Henneman PL, et al. Prehospital hypotension is a predictor of the need for an emergent, therapeutic operation in trauma patients with normal systolic blood pressure in the emergency department. J Trauma. 2006;61:1228–33.

23. Duckitt RW, Buxton-Thomas R, Walker J, et al. Worthing physiological scoring system: derivation and validation of a physiological early-warning system for medical admissions. An observational, population-based single-centre study. Br J Anaesth. 2007;98(6): 769–74.

24. Subbe CP, Slater A, Menon D, Gemmell L. Validation of physiological scoring systems in the accident and emergency department. Emerg Med J. 2006;23(11):841–5.

25. Peres Bota D, Melot C, Lopes Ferreira F, Vincent JL. Infection Probability Score (IPS): A method to help assess the probability of infection in critically ill patients. Crit Care Med. 2003;31(11): 2579–84.

26. Cuthbertson BH. Optimising early warning scoring systems. Resuscitation. 2008;77(2):153–4.

27. Gelinas JF, Davis GM, Arlegui C, Cote A. Prolonged, documented home-monitoring of oxygenation in infants and children. Pediatr Pulmonol. 2008;43(3):288–96.

28. Sibanda N, Lewsey JD, van der Meulen JH, Stringer MD. Continuous monitoring tools for pediatric surgical outcomes: an example using biliary atresia. J Pediatr Surg. 2007;42(11):1919–25.

29. Poloniecki J, Sismanidis C, Bland M, Jones P. Retrospective cohort study of false alarm rates associated with a series of heart operations: the case for hospital mortality monitoring groups. BMJ. 2004;2004:328–75.

30. Reynolds MR, Stoumbos ZG. The SPRT chart for monitoring a proportion. IIE Trans. 1998;30(6):545–61.

31. Musonda P, Hocine MN, Andrews NJ, Tubert-Bitter P, Farrington CP. Monitoring vaccine safety using case series cumulative sum charts. Vaccine. 2008;26(42):5358–67.

32. Sebille V, Bellissant E. Comparison of four sequential methods allowing for early stopping of comparative clinical trials. Clin Sci (Lond). 2000;98(5):569–78.

33. Golhar DY, Pollock SM. Sequential analysis for diagnosing diabetes. Med Decis Making. 1987;7(1):47–51.

34. Griffin MP, O'Shea TM, Bissonette EA, Harrell FE, Jr., Lake DE, Moorman JR. Abnormal heart rate characteristics preceding neonatal sepsis and sepsis-like illness. Pediatr Res. 2003;53(6):920–6.

35. Sauaia A, Moore FA, Moore EE, Moser KS, Brennan R, Read RA, Pons PT. Epidemiology of trauma deaths: a reassessment. J Trauma. 1995;38:185–93.

36. Peng R, Chang C, Gilmore D, Bongard F. Epidemiology of immediate and early trauma deaths at an urban level I trauma center. Am Surg, 1998;64:950–4.

# The Matching of Sinus Arrhythmia to Respiration: Are Trauma Patients without Serious Injury Comparable to Healthy Laboratory Subjects?

Xiaoxiao Chen, Andrew T. Reisner, Liangyou Chen, Shwetha Edla, and Jaques Reifman

*Abstract*— **We sought to better understand the physiology underlying the metrics of heart rate variability (HRV) in trauma patients without serious injury, compared to healthy laboratory controls. In trauma patients without serious injury (110 subjects, 470 2-min data segments), we studied the correlation between sinus arrhythmia (SA) rate, heart rate (HR), and respiratory rate (RR). Most segments with $2.4 \leq HR/RR < 4.8$ exhibited SA-RR matching, whereas rate matching was absent in 81% of the segments with $HR/RR < 2.4$ and in 86% of the segments with $HR/RR \geq 4.8$. The findings were comparable, in some cases remarkably so, to previous reports from healthy laboratory subjects. The presence (or absence) of SA-RR matching, when SA is largely controlled by respiration, can be anticipated in this trauma population. This work provides a valuable step towards the definition of patterns of HRV found in trauma patients with and without life-threatening injury.**

## I. Introduction

We sought to better understand the physiology that underlies metrics of heart rate variability (HRV). Respiration is a predominant determinant of HRV. Hence it has been argued that it is crucial to consider the relationship between HRV and respiration when interpreting HRV data [1]. In some circumstances, the frequency of sinus rhythm variation (sinus arrhythmia [SA]; rhythmic fluctuations in heart rate [HR]) is *wholly* driven by the respiratory rate (RR), such that their rates will be identical [2-5]. The amplitude of SA is also correlated, inversely, to RR [3]. Rate matching between the SA oscillation rate and RR can be so tight that some research protocols accept the SA rate as a proxy measurement for RR [2, 4]. Yet the SA rate and RR are not equal under all physiological conditions. For instance, during exertion or bradycardia, the SA rate is driven by non-respiratory factors, even in subjects with otherwise normal autonomic control systems [6, 7].

This physiology has major implications for comprehending why a particular HRV pattern would be diagnostically associated with a particular disease state or a healthy state. Consider the case of a patient exhibiting an atypical HRV pattern. This atypical HRV could be caused by any of the following: *1)* an atypical respiratory pattern driving an atypical SA pattern via a typical autonomic control system; *2)* an atypical driver (i.e., non-respiratory) causing an atypical SA pattern; or *3)* an atypical control system directly producing an atypical SA pattern. Interestingly, most prior reports in trauma patients ascribed distinctive HRV patterns to differences in autonomic tone, without detailed consideration of the underlying causes, such as respiratory or non-respiratory drivers [8-11].

The present study is intended to better understand the causes of HRV patterns in trauma patients, here focusing on neurologically intact, hemodynamically stable patients. We seek to answer the following questions: *First, under what conditions are the SA rate and the RR tightly matched? Second, are the findings consistent with reports of healthy laboratory subjects?* To address this, we explored a population of patients monitored during transport to the hospital after an episode of physical trauma. We examined how the relationship between SA rate and RR changed as a function of HR, RR, and their ratio (*HR/RR*).

## II. Methods

### A. Clinical Data Collection

Physiological data for this study was collected from 898 trauma patients during medical helicopter transport between August 2001 and April 2004 from the scene of injury to the level I unit at the Memorial Hermann Hospital in Houston, TX [8]. Additional attribute data were collected retrospectively via chart review. The time-series variables were measured by Propaq 206EL vital-sign monitors (Welch Allyn, Skaneateles Falls, NY), downloaded to an attached personal digital assistant, and ultimately stored in our database. Physiological data included the electrocardiogram (ECG; sampled at 182 Hz), a respiratory waveform (an impedance pneumograph, IP, measured through the ECG leads and sampled at 23 Hz), their corresponding monitor-computed HR and RR (recorded at 1-s intervals), and other

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.

X. Chen was with the Department of Defense Biotechnology High Performance Computing Software Applications Institute (BHSAI), Telemedicine and Advanced Technology Research Center (TATRC), USAMRMC, Fort Detrick, MD 21702 USA.

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

L. Chen was with the BHSAI, TATRC, USAMRMC, Fort Detrick, MD 21702 USA.

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

J. Reifman is a Senior Research Scientist and Director of the BHSAI, TATRC, USAMRMC, Fort Detrick, MD 21702 USA (e-mail: jaques.reifman.civ@mail.mil).

standard vital-sign data. Patient attribute data included demographics, injury descriptions, pre-hospital interventions, and hospital treatments. Data collection and analysis was performed with the approval of both the local and the United States Army's human subjects Institutional Review boards (the latter at Fort Detrick, MD).

### B. Study Population

We selected relatively healthy subjects for analysis according to the following attributes: no major hemorrhage that required the transfusion of red blood cells, no prehospital or hospital intubation, head abbreviated injury scale equal to 0, and Glasgow coma scale of 13 or higher.

In these subjects, we split time-synchronized ECG and IP waveforms into successive 2-min data segments and only analyzed those with reliable waveforms based on our previously developed quality index, which rated the waveforms as reliable if they were clean with rhythmic and consistent beats or breaths [12, 13]. Visual inspection to ensure that the ECG contained no ectopic beats resulted in the exclusion of a total of five 2-min data segments, all from the same subject. In total, 470 2-min recordings from 110 subjects (age, mean ± standard deviation): 39 ± 12 yr, age range: 18-76 yr, 86 men and 24 women, median of three data segments per subject) formed the study dataset.

### C. Estimation of SA Rate, HR, and RR

For each 2-min ECG and IP waveform, we computed second-by-second HR and RR values using automated computer algorithms that have been previously reported and demonstrated to match human experts' estimation [12, 13].

We used the following method to construct the R-R interval (RRI) time series used to estimate the SA rate. First, we upsampled each ECG segment to 2000 Hz by cubic spline interpolation and detected R-wave time locations in the upsampled ECG using the method described in [13]. Second, we calculated RRIs as the difference between the time locations of successive R-waves, i.e., $RRI_i = R_{i+1} - R_i$ (i = 1, …, N-1; where N is the total number of R-waves), and located them at time location $R_{i+1}$. Third, we transformed the unevenly spaced RRI time series into an evenly spaced one with a sampling frequency of 23 Hz (the same as that of the IP waveforms) using cubic spline interpolation. Next, to count the SA cycles within the RRI waveform, we treated the RRI time series as a form of respiratory waveform and applied our previously developed RR estimation and reliability algorithms to compute the second-by-second SA rate and determine whether the waveform was of adequate reliability [12]. Finally, we averaged the reliable SA rate and corresponding HR and RR within the same time period for each 2-min recording and performed an analysis based on the mean SA rate, HR, and RR. Because the HR and RR were estimated from reliable ECG and IP waveforms, no further reliability filtering was implemented.

### D. Determination of SA-RR Matching

Although the overall relationship between SA and respiration can be mathematically quantified by coherence and cross-approximate entropy [14, 15], the results may be difficult to interpret. For a simple and practical approach to determine whether or not each 2-min data segment showed SA-RR matching, we visually inspected the 2-min normalized RRI time series and IP waveform pairs by examining every non-overlapping 15-s data segment, and identified whether there was a consistent pattern of alteration between each SA oscillation and each respiratory oscillation. If at least 75% of consecutive 15-s RRI and IP waveform pairs exhibited alternating SA and RR oscillations, we considered the whole 2-min data segment to represent an SA-RR matching case. Otherwise, it was considered as not rate matched.

Visual determination of SA-RR matching was based on the judgment of a single investigator and objectively corroborated using automated algorithms to calculate the difference between the SA rate and RR (confirming that for matched segments, the difference between SA rate and RR was within ±5 cycles per minute [cpm]).

### E. Data Analysis

To quantify the agreement between SA rate and RR, we calculated the Pearson's correlation coefficient ($r_p$) and the concordance correlation coefficient ($r_c$) between SA rate and RR. While the well-known $r_p$ quantifies the linear relationship between two variables regardless of the slope and $x$-intercept of the regression line, $r_c$ quantifies the linear relationship with respect to the identity line [16] and is thus a better metric to measure the degree to which two variables are equal to each other. Next, we computed the percentage of data segments that lack SA-RR matching within each HR, RR, and *HR/RR* range. The 95% confidence intervals (CIs) of the percentages were also calculated [17].

## III. Results

In this dataset, $r_p$ between the SA rate and RR was 0.43, and $r_c$ was 0.39, reflecting a significant but moderate overall correlation. Of the 110 subjects under study, 43% of the subjects exhibited SA-RR matching for each of their 2-min data segments, 27% of the subjects lacked SA-RR matching for each of their 2-min data segments, and the remaining 30% of the subjects exhibited a mix of present and absent SA-RR matching in different 2-min data segments. For the data segments that exhibited matching via visual inspection, we found a high agreement between the automatically computed SA rate and RR (the difference between the SA rate and RR was within ±5 cpm for 93% of those segments).

Fig. 1 illustrates the SA-RR relationship using three selected pairs of sample ECG, RRI, and IP waveforms. A tight SA-RR matching with $2.4 \leq HR/RR < 4.8$ (*left*), lack of SA-RR matching with a higher (than RR) SA rate and $HR/RR \geq 4.8$ (*middle*), and lack of SA-RR matching with a lower (than RR) SA rate and $HR/RR < 2.4$ (*right)* were some of the typical patterns observed.

Fig. 2 shows the percentage (along with 95% CIs) of data segments that lacked SA-RR matching in different *HR/RR*, RR, and HR ranges. Fig. 2A shows that both low and high *HR/RR* values were associated with a high fraction of data segments that lacked SA-RR matching. When $HR/RR < 2.4$, 81% of the 2-min data segments lacked SA-RR matching;

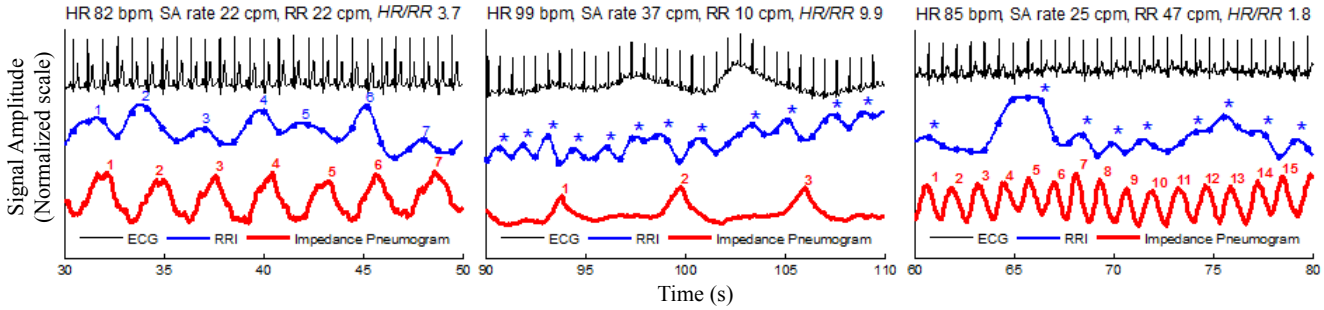



Figure 1. Examples of ECG, RRI, and impedance pneumogram waveforms. *Left*: Rate matching between SA and respiration. *Middle*: Absence of SA-RR matching (with tachycardia and bradypnea). *Right*: Absence of SA-RR matching (with HR almost double RR). Symbols above the respiratory and RRI waveforms (* and numerals, respectively) denote distinct oscillations that were identified by automated computer algorithms. ECG: electrocardiogram, HR: heart rate, RR: respiratory rate, RRI: R-R interval time series, SA: sinus arrhythmia.

when *HR/RR* ≥ 4.8, 86% of the 2-min data segments lacked SA-RR matching. We also found independent associations between rate matching and RR, as well as HR. Figs. 2B and 2C show that ≥ 57% of the 2-min data segments with RR ≥ 30 cpm, and 55% of the 2-min data segments with HR < 60 beats per minute (bpm), respectively, lacked SA-RR matching.

## IV. Discussion

In this study, in a population of relatively healthy patients (i.e., no hemorrhage nor serious neurological injury) early after major trauma, we investigated when SA oscillation was predominantly driven by respiration and we proposed a simple metric that can determine when SA-RR matching is likely.

We found a lack of SA-RR matching when RR was elevated ≥ 28 cpm (Fig. 2), where the SA rate tended to be lower than the RR. This likely reflected the inability of the sinoatrial node to oscillate fast enough to keep up with rapid respiration, as reported in previous studies [18, 19] wherein the transfer function between vagal nerve impulses and the sinoatrial node rate exhibited the characteristics of a low pass filter with a cutoff frequency of ~0.5 Hz, or 30 cpm. Above this cut-off, the SA rate cannot keep up with RR. This was very close to our cut-off of 28 cpm suggesting comparable SA rate cut-offs in both uninjured trauma patients and healthy laboratory subjects.

In this dataset, there was an absence of SA-RR matching when HR was low, e.g., < 60 bpm. This is related to cardiac aliasing [7]. Cardiac aliasing is mathematically inevitable unless HR is equal to or greater than twice RR (i.e., *HR/RR* ≥ 2), because it requires at least two heart periods for each respiratory cycle to establish an oscillation (an oscillation requires, at minimum, one shorter interval that alternates with a second, longer interval).

When HR was elevated, e.g., HR > 100 bpm, the SA rate exceeded RR in approximately 50% the cases, whereas in the other 50% there was SA-RR matching. The association between tachycardia and reduced SA-RR matching was previously observed in athletes during exercise, who exhibited rapid SA rates (> RR) [6]. This phenomenon was attributed to the fact that the cardiovascular system that coupled respiration to HR had nonlinear components and that harmonics of RR could appear in the output HR time series [20]. Furthermore, in normal subjects, the cardiac vagal system served as strong, fast negative feedback, attenuating the harmonics in the HR time series. However, in young athletes during exercise, as well as in heart transplant patients, the vagal control was either minimal or absent, and a higher (than RR) SA rate was observed. It was concluded that elevated SA rate was thus an indicator of reduced vagal control of the heart.

In terms of anticipating whether or not rate matching would occur in our dataset, it was more effective to consider the ratio *HR/RR* than to look for the presence of tachycardia alone (Fig. 2). What we found in terms of *HR/RR* versus rate matching was wholly consistent with a prior report by Cysarz et al. [2], wherein an $r_c$ of 0.64 was observed between SA rate and RR within a laboratory population (for a population with 3.0 < *HR/RR* < 8.7, approximately). For the comparable subset of our study population who had 3.0 < *HR/RR* < 8.7, we found a rather similar result with $r_c$ = 0.60.

In contrast, the correlation between SA rate and RR was reported to be $r_c$ = 0.95 (when 6 cpm < RR < 30 cpm) in [4]. For a comparable subset of our study population who had 6 cpm < RR < 30 cpm, we obtained an $r_c$ = 0.27, which is far lower than the value reported in [4]. Does this mean our findings are inconsistent? Not necessarily. The study in [4] did not report the HR (unlike [2]); It is entirely possible, if not likely, that our population had a relative elevation in HR. Also, in [4] subjects were studied during supine rest in a laboratory, whereas we studied acute trauma patients during prehospital care. This underlies one of our major findings, that it is necessary to consider HR and RR simultaneously when trying to determine whether SA-RR matching is likely to occur in a typical population.

In general, the *HR/RR* metric provided a compact summary of all of our aforementioned findings: (a) when 2.4 ≤ *HR/RR* < 4.8, SA and respiration were typically rate matched; (b) when *HR/RR* was high (i.e., ≥ 4.8), there might be high HR indicating vagal withdrawal and the resultant elevated SA rate [20]; and (c) when *HR/RR* was low (i.e., <

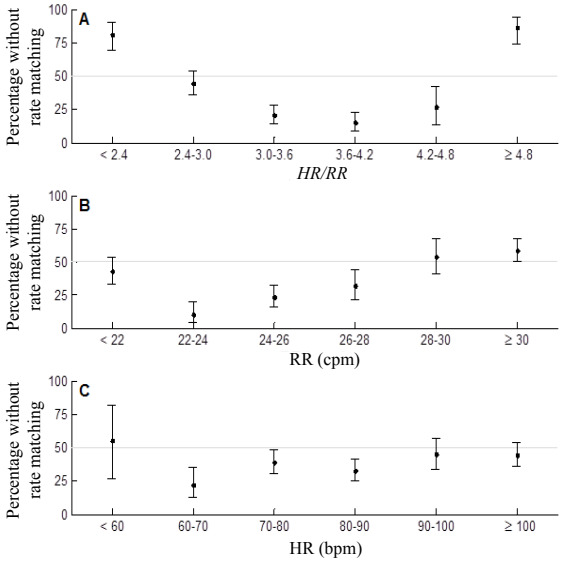



Figure 2. The percentage of data segments that lacked SA-RR matching for different *(A) HR/RR*, (*B*) RR, and (*C*) HR ranges. The vertical bars represent the 95% confidence intervals. The grey gridline indicates 50% of data segments. The largest percentage of data segments that lacked SA-RR matching was observed when *HR/RR* < 2.4 or *HR/RR* ≥ 4.8. bpm: beats per min, cpm: cycles per min, HR: heart rate, RR: respiratory rate, SA: sinus arrhythmia.

2.4), there were two phenomena that caused the absence of SA-RR matching. First, the RR might be so elevated that the sinoatrial node could not keep up with the rapid respiratory oscillations [18]. Second, cardiac aliasing would likely have occurred.

Our findings support the validity of laboratory-based investigation as a model for actual trauma patients, and confirm that respiration is frequently the predominant driver of SA in trauma patients without major injuries. A second implication relates to studies that investigate whether or not SA rate monitoring can serve as a suitable RR proxy [2, 4]. Our findings suggest that, in a population similar to these trauma patients, this methodology will work provided that HR is neither too fast nor too slow and there is no tachypnea. The *HR/RR* metric might have anticipated the findings in [4], wherein a very high correlation was reported between SA rate and RR at rest, as well as the findings in [2], wherein reduced correlation was seen in subjects during low levels of exercise. Finally, we expect that our findings may be informative to future studies into the determinants of HRV in trauma patients, by providing a better understanding of those trauma patients without serious injury.

## 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] T. Buchner, "HRV strongly depends on breathing. Are we questioning the right suspect?," in *Conf. Proc. IEEE Engineering in Medicine and Biology Society*, Boston, MA, 2011, pp. 7739-7742.

[2] D. Cysarz, R. Zerm, H. Bettermann, M. Fruhwirth, M. Moser, and M. Kroz, "Comparison of respiratory rates derived from heart rate variability, ECG amplitude, and nasal/oral airflow," *Ann. Biomed. Eng.,* vol. 36, pp. 2085-2094, 2008.

[3] J. A. Hirsch and B. Bishop, "Respiratory sinus arrhythmia in humans: how breathing pattern modulates heart rate," *Am. J. Physiol. Heart Circ. Physiol.,* vol. 241, pp. H620-H629, 1981.

[4] A. Schafer and K. W. Kratky, "Estimation of breathing rate from respiratory sinus arrhythmia: comparison of various methods," *Ann. Biomed. Eng.,* vol. 36, pp. 476-485, 2008.

[5] J. F. Thayer, J. J. R. Sollers, E. Ruiz-Padial, and J. Vila, "Estimating respiratory frequency from autoregressive spectral analysis of heart period," *IEEE Eng. Med. Biol. Mag.,* vol. 21, pp. 41-45, 2002.

[6] L. Bernardi, F. Salvucci, R. Suardi, P. L. Solda, A. Calciati, S. Perlini, C. Falcone, and L. Ricciardi, "Evidence for an intrinsic mechanism regulating heart rate variability in the transplanted and the intact heart during submaximal dynamic exercise?," *Cardiovasc. Res.,* vol. 24, pp. 969-981, 1990.

[7] H. Witte, U. Zwiener, M. Rother, and S. Glaser, "Evidence of a previously undescribed form of respiratory sinus arrhythmia (RSA)--the physiological manifestation of "cardiac aliasing"," *Pflugers Arch.,* vol. 412, pp. 442-444, 1988.

[8] 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 preshospital trauma patients," *J. Trauma,* vol. 60, pp. 363-370, 2006.

[9] E. L. Grogan, J. A. Morris, P. R. Norris, D. J. France, A. Ozdas, R. A. Stiles, P. A. Harris, B. M. Dawant, and T. Speroff, "Reduced heart rate volatility: an early predictor of death in trauma patients," *Ann. Surg.,* vol. 240, pp. 547-554, 2004.

[10] K. G. Proctor, S. A. Atapattu, and R. C. Duncan, "Heart rate variability index in trauma patients," *J. Trauma,* vol. 63, pp. 33-43, 2007.

[11] R. J. Winchell and D. B. Hoyt, "Analysis of heart-rate variability: a noninvasive predictor of death and poor outcome in patients with severe head injury," *J. Trauma,* vol. 43, pp. 927-933, Dec 1997.

[12] L. Chen, T. McKenna, A. T. Reisner, and J. Reifman, "Algorithms to qualify respiratory data collected during the transport of trauma patients," *Physiol. Meas.,* vol. 27, pp. 797-816, 2006.

[13] 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, 2006.

[14] K. J. Bar, A. Schuhmacher, S. Hofels, S. Schulz, A. Voss, V. K. Yeragani, W. Maier, and A. Zobel, "Reduced cardio-respiratory coupling after treatment with nortriptyline in contrast to S-citalopram," *J. Affect. Disord.,* vol. 127, pp. 266-273, Dec 2010.

[15] S. Berger, M. K. Boettger, M. Tancer, S. M. Guinjoan, V. K. Yeragani, and K. J. Bar, "Reduced cardio-respiratory coupling indicates suppression of vagal activity in healthy relatives of patients with schizophrenia," *Prog. Neuropsychopharmacol. Biol. Psychiatry,* vol. 34, pp. 406-411, Mar 17 2010.

[16] L. Lin, "A concordance correlation coefficient to evaluate reproducibility," *Biometrics,* vol. 45, pp. 255-268, 1989.

[17] T. D. Ross, "Accurate confidence intervals for binomial proportion and Poisson rate estimation," *Comput. Biol. Med.,* vol. 33, pp. 509-531, Nov 2003.

[18] R. D. Berger, J. P. Saul, and R. J. Cohen, "Transfer function analysis of autonomic regulation. I. Canine atrial rate response," *Am. J. Physiol. Heart Circ. Physiol.,* vol. 256, pp. H142-H152, 1989.

[19] J. P. Saul, R. D. Berger, M. H. Chen, and R. J. Cohen, "Transfer function analysis of autonomic regulation. II. Respiratory sinus arrhythmia," *Am. J. Physiol. Heart Circ. Physiol.,* vol. 256, pp. H153-H161, 1989.

[20] E. Toledo, I. Pinhas, D. Aravot, and S. Akselrod, "Very high frequency oscillations in the heart rate and blood pressure of heart transplant patients," *Med. Biol. Eng. Comput.,* vol. 41, pp. 432-438, 2003.

# 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 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 (RASPRT) [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

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 (BHSAI), 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 BHSAI, 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 BHSAI, 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 BHSAI, TATRC, USAMRMC, Fort Detrick, MD 21702 USA (e-mail: jaques.reifman.civ@mail.mil).

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.

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.

### 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. $\theta_0$ and $\theta_1$ were estimated from the MHLF dataset. RASPRT was exactly the same as SPRT, except that the probability density functions $f\big(y(t);\theta_0(t)\big)$ and $f\big(y(t);\theta_1(t)\big)$ were time varying depending on the availability of the vital signs at each time instant t (15 pairs of $\theta_0$ and $\theta_1$ 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 2. ALERTING STRATEGIES

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

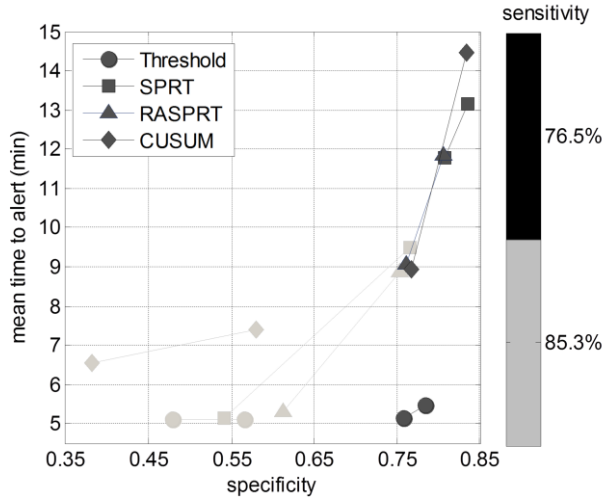



Figure 1. The trade-off between mean time to alert and specificity at fixed sensitivity levels of 76.5% and 85.3%. A 60-minute moving window was used to filter the vital-sign features. SPRT: sequential probability ratio test; RASPRT: risk-adjusted SPRT; CUSUM: cumulative sum.

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.

JOURNAL OF NEUROTRAUMA 31:906–913 (May 15, 2014)

DOI: 10.1089/neu.2013.3128

# Prehospital Heart Rate and Blood Pressure Increase the Positive Predictive Value of the Glasgow Coma Scale for High-Mortality Traumatic Brain Injury

Andrew Reisner,[1,2] Xiaoxiao Chen,[1] Kamal Kumar,[1] and Jaques Reifman[1]




## Abstract

We hypothesized that vital signs could be used to improve the association between a trauma patient's prehospital Glasgow Coma Scale (GCS) score and his or her clinical condition. Previously, abnormally low and high blood pressures have both been associated with higher mortality for patients with traumatic brain injury (TBI). We undertook a retrospective analysis of 1384 adult prehospital trauma patients. Vital-sign data were electronically archived and analyzed. We examined the relative risk of severe head Abbreviated Injury Scale (AIS) 5–6 as a function of the GCS, systolic blood pressure (SBP), heart rate (HR), and respiratory rate (RR). We created multi-variate logistic regression models and, using DeLong's test, compared their area under receiver operating characteristic curves (ROC AUCs) for three outcomes: head AIS 5–6, all-cause mortality, and either head AIS 5–6 or neurosurgical procedure. We found significant bimodal relationships between head AIS 5–6 versus SBP and HR, but not RR. When the GCS was $<15$, ROC AUCs were significantly higher for a multi-variate regression model (GCS, SBP, and HR) versus GCS alone. In particular, patients with abnormalities in all parameters (GCS, SBP, and HR) were significantly more likely to have high-mortality TBI versus those with abnormalities in GCS alone. This could be useful for mobilizing resources (e.g., neurosurgeons and operating rooms at the receiving hospital) and might enable new prehospital management protocols where therapies are selected based on TBI mortality risk.


**Key words:** blood pressure; Glasgow Coma Scale; heart rate; prehospital; traumatic brain injury

## Introduction

The Glasgow Coma Scale (GCS) was developed to standardize the assessment of coma and impaired consciousness after traumatic brain injury (TBI).[1] As originally intended, the GCS is to be assessed only after hemodynamic resuscitation and in the absence of pharmacologic sedation, paralysis, or other forms of chemical intoxication.[2] The GCS was an innovation, providing an objective method for measuring patients' global brain function. In the absence of established alternatives, use of the GCS spread to a multitude of applications outside the researchers' original intent. For example, in national guidelines for trauma patient management, a below-normal prehospital GCS is one criterion for emergency medical service (EMS) transport from the field directly to a level 1 trauma center[3] and for emergency tracheal intubation after traumatic injury.[4] In addition, the GCS is often relied upon in prehospital research to help control for degree of TBI (for example, see Davis and colleagues[5]), even though it was not originally intended, nor validated, for this clinical context. Indeed, the GCS is currently being used for non-TBI patients, for instance, to measure brain function in meningitis[6] and hypothyroidism[7] cases. Overall, the GCS has evolved to become a near-universal measure for global mental function, despite its original intent and validation in TBI patients subsequent to stabilization.

Unsurprisingly, because the GCS is applied in different ways distinct from its original intent, there is growing recognition that it may not be optimal for all these applications[8] and that such widespread, inconsistent application of the GCS can cloud its interpretation.[9,10] Fundamentally, the provisos for the classic GCS—measurement after hemodynamic resuscitation and in the absence of intoxication—are incompatible with clinical decision making or research into early trauma care. Further, EMS caregivers may have less capacity for careful clinical evaluation, which may be one factor why significant discrepancies between prehospital GCS versus emergency department (ED) GCS have been reported.[11] For the early stages of trauma care, it has been suggested that a simplified score, for example, either the motor-only score, or the ''alert, voice, pain, unresponsive'' rating, would offer reliability and convenience without much loss of clinical accuracy, because the three GCS subscales are largely redundant

[1]Biotechnology High Performance Computing Software Applications Institute, Telemedicine and Advanced Technology Research Center, U.S. Army Medical Research and Materiel Command, Ft. Detrick, Maryland.
[2]Department of Emergency Medicine, Massachusetts General Hospital, Boston, Massachusetts.

(because of the high correlation between the three).[8] However, a simplified coma score is decidedly low resolution (i.e., patients are stratified into just a small number of severity levels) and does not overcome another deficiency: mediocre outcome prediction.[8]

Overall, there is an unanswered need for an accurate, practical method of assessing the severity of TBI during early trauma patient care, which could be employed in clinical decision algorithms or research investigations. A series of reports has suggested that blood pressure (BP) offers prognostic information relevant to TBI. It is intuitive that, for the initial evaluation of the trauma patient, a low GCS and low BP are correlates of mortality;[12,13] the relationship between a low GCS, hypotension, and higher mortality has been quantified in classic prehospital severity scores, such as the trauma score,[14] the prehospital index,[15] and the ''circulation, respiration, abdomen, motor, speech'' (CRAMS) score.[16]

At the same time, hypertension is recognized as another correlate of mortality in TBI patients. In a population of TBI patients, Zafar and colleagues[17] found that high and low BP were both associated with increased short-term mortality. These findings were consistent with an earlier analysis of TBI patients by Butcher and colleagues[18] and another report of an association between elevated BP and reduced survival in trauma patients.[19] Our group has previously examined how real-time computerized vital-sign analysis can enhance prehospital recognition of hemorrhagic hypovolemia,[20,21] and we decided to investigate whether the findings of Zafar and colleagues[17] and Butcher and colleagues[18] could be diagnostically applicable to a general prehospital trauma population (i.e., patients without and with TBI). In addition to examining BP, we also sought to explore the diagnostic significance of other vital signs, such as heart rate (HR) and respiratory rate (RR), in the early evaluation of TBI, because abnormalities such as bradycardia and bradypnea are hallmarks of severe TBI. We hypothesized that it would be possible to use routine vital signs to improve the correlation between the prehospital GCS and high-mortality TBI. Accordingly, we undertook a retrospective analysis of a prehospital trauma patient database to test the hypothesis.

## Methods

### *Clinical data collection*

This was a retrospective analysis of clinical data originally collected and analyzed by Cooke and colleagues,[22] with institutional review board approval, of trauma patients during transport by air ambulance from the scene of injury to a level 1 trauma center. In a convenience sample of prehospital trauma patients, vital-sign data were obtained using a Propaq 206EL monitor (Welch Allyn, Beaverton, OR) between August 2001 and April 2004 and using a PIC 50 monitor (Welch Allyn) between March 2005 and May 2007. The following data were archived using a networked personal digital assistant: electrocardiogram (ECG) and associated continuous HR; impedance pneumogram (IP) and associated continuous RR; and systolic BP (SBP) and diastolic BP (DBP) measured intermittently at multi-minute intervals. Prospectively, prehospital GCS was assessed and documented by EMS caregivers (paramedics and critical care flight nurses) as per routine clinical operations; no focused training related to GCS assessment was provided to these EMS providers as part of this investigation. Retrospectively, clinical data for analysis were collated by chart review, including demographics, prehospital GCS, prehospital interventions, hospital treatments, coded injury descriptions (Abbreviated Injury Scale; AIS), and overall outcomes (mortality). The complete investigational data set was subsequently uploaded to our data warehousing system.[23] Protected health information was not included. All subsequent data analyses were performed using MATLAB (version 7; MathWorks, Natick, MA).

### *Vital-sign data processing*

Vital signs and other physiological data measured during prehospital clinical operations are often corrupted by artifacts. Previous research has demonstrated that automated computer algorithms can identify and remove unreliable data, leading to significant improvements in the association between vital signs and traumatic hemorrhage.[20,24,25]

Here, we used the same validated methodology, summarized as follows: For each vital-sign value, reliable data were identified by automated algorithms that rated each datum on an integer scale from least to most reliable.[24,26,27] HR and RR reliability algorithms involved analysis of ECG and IP waveforms, respectively.[26,27] Briefly, when the waveforms were clean with rhythmic, consistent beats or breaths, the corresponding rates tended to be rated as reliable. Conversely, when the waveforms were noisy with irregular, heterogeneous beats or breaths, the rates were rated as unreliable. In previous validation, these algorithms' ratings of ECG and IP waveforms and reliability of the corresponding HR and RR typically concurred with the opinion of clinicians. The BP reliability algorithm determined whether the ratios between SBP, DBP, and mean pressure were physiological and whether the HR measured by the inflatable oscillometric cuff matched the ECG HR.[24]

For the TBI analysis, we studied the mean of the reliable HR, RR, and SBP in the initial 15-min transportation, so each patient had no more than one HR, RR, and SBP datum. In previous research investigating the relationship between prehospital vital signs and clinically significant blood loss, we found that taking the average over 15 min was an effective measure to reduce transient variability and unreliable measurements, leading to an improved association with clinical outcomes.[20,28]

### *Subject selection*

For analyses involving HR as an independent variable, we studied patients with an available GCS score and at least one reliable HR value in the initial 15 min of transportation. For analyses involving RR as an independent variable, we studied patients with an available GCS score and at least one reliable RR value in the initial 15 min of transportation (and further examined a subset of patients who were spontaneously breathing, i.e., nonintubated). For analyses involving SBP as an independent variable, we studied patients with an available GCS score and at least one reliable SBP value in the initial 15-min transportation. For the multi-variate analyses, we studied patients with available GCS scores and at least one reliable HR value and one reliable SBP value in the initial 15 min of transportation. We excluded the 1 patient who left against medical advice, because his injuries and outcome were unknown.

### *Definition of high-mortality traumatic brain injury for investigation of diagnostic test characteristics*

It can be difficult to differentiate between patients who died as a result of TBI versus coexistent injuries and other clinical factors. Therefore, we investigated three parallel definitions of high-mortality TBI. Our assumption was that any valid study finding should be consistent for any reasonable definition of high-mortality TBI (i.e., consistent across all three outcome definitions).

Our primary outcome definition was head AIS score of 5 or 6. The AIS is a well-validated, widely used scoring system that assigns a score from 0 to 6 based on the anatomic injury pattern, rating how likely the patient is to die from the injury.[29] The specific AIS cutoff (i.e., 5–6) for high-mortality TBI was selected post hoc after a preliminary analysis to identify the discriminatory capability of SBP, HR, and RR as a function of specific AIS scores. For that

preliminary analysis, we calculated the relative risk of head AIS 3–6, 4–6, and 5–6 as a function of different ranges of SBP ($x \geq SBP > x + 25$ mm Hg), HR ($x \geq HR > x + 20$ beats/min), and RR ($x \geq RR > x + 5$ breaths/min). Relative risk was defined as the risk for patients within the range (i.e., ratio of positive cases to total cases within the range) divided by the risk for patients outside of the range (i.e., ratio of positive cases to total cases outside of the range). Confidence intervals (CIs) were computed as per Daly.[30] For testing the significance of relative risks for different ranges, we compared each against the relative risks of specific reference ranges (reference ranges for SBP, 100–125 mm Hg; HR, 80–100 beats/min; RR, 30–35 breaths/min) using the method of Altman and Bland.[31] This analysis led us to define the primary outcome as head AIS 5–6.

We explored two secondary definitions of high-mortality TBI. Specifically, we examined all-cause mortality. We also examined head AIS 5–6 *or* documented neurosurgical procedure (''head AIS 5–6/procedure'') as a secondary outcome for those cases in which a neurosurgical intervention was performed that may have prevented an otherwise fatal TBI.

### *Diagnostic test characteristics of Glasgow Coma Scale, systolic blood pressure, and heart rate for high-mortality traumatic brain injury*

After the preliminary analysis suggested that there was no significant association between RR and TBI, RR was excluded from further analysis.

We investigated the diagnostic performance of the following:

- GCS alone as the independent variable.
- SBP and HR; to accommodate their bimodal relationship with TBI (i.e., both high and low values of SBP and HR are associated with an increased risk of TBI), we used relative risks as follows. The preliminary analysis (detailed above) yielded relative risk as a function of each SBP and HR. We fitted a cubic spline to these curves, obtaining a mathematical expression for TBI relative risk as a function of SBP or HR values. We then used the computed relative risks ($SBP_{Risk}$ and $HR_{Risk}$) as inputs to multi-variate logistical regression models trained to predict each of the investigational outcomes.
- A multivariate logistic regression model using all three investigational predictors (GCS, $SBP_{Risk}$, and $HR_{Risk}$).

We computed receiver-operating characteristic (ROC) curves for the investigational outcomes to evaluate their diagnostic performance. We compared the areas under each ROC curve (ROC AUC) using DeLong and colleagues' method[32] (significance level of $p < 0.05$).

## Results

Our data set contained 1384 subjects with at least one nonzero vital-sign datum. We identified 1289 subjects with at least one reliable HR value in the first 15 min, 649 with at least one reliable RR value (of these, 499 were spontaneously breathing), 1247 with at least one reliable SBP value, and 1158 with at least one reliable SBP and one reliable HR value. Table 1 shows the characteristics of these 1247 subjects, as well as the two subpopulations ($GCS < 15$ and $GCS \leq 8$) analyzed in the multi-variate analysis. GCS from the ED was available for 88% of the study population. Compared with prehospital GCS, average ED GCS was 0.03 points lower and the standard deviation of their differences was 2.1.

Unless otherwise specified, we used $p < 0.05$ for significant results reported below.

### *Relative risk of traumatic brain injury as a function of prehospital vital signs*

When we computed the relative risk of high-mortality TBI (i.e., head AIS 5–6, 4–6, and 3–6) as a function of prehospital SBP, we found the following:

- The relative risks given *low* SBP (< 100 mm Hg) were significantly different than the relative risks given SBP within the reference range for all three head AIS cutoffs (i.e., AIS 5–6, 4–6, and 3–6). Relative risks given SBP < 100 mm Hg were 2.6 (95% CI, 1.3–5.2), 2.0 (1.3–3.0), and 1.5 (1.1–2.0) for head AIS 5–6, 4–6, and 3–6, respectively. Given SBP within the reference range (125 mm Hg ≥ SBP > 100 mm Hg), relative risks were 0.5 (0.2–1.1), 0.7 (0.5–1.0), and 0.9 (0.7–1.2), respectively.

Table 1. Population Description for the Overall Study Population and Key Subpopulations

| *Population* / *Characteristics* | *Any GCS* | *GCS < 15* | *GCS ≤ 8* |
|---|---|---|---|
| Population size (*n*) | 1,158 | 530 | 225 |
| Men (%) | 836 (72) | 374 (71) | 162 (72) |
| Women (%) | 319 (28) | 154 (29) | 62 (28) |
| Mean age, years | 38 (15) | 36 (15) | 36 (15) |
| Blunt injury (%) | 1,012 (87) | 478 (90) | 191 (85) |
| Penetrating injury (%) | 125 (11) | 44 (8) | 31 (14) |
| Mortality (%) | 82 (7) | 73 (14)[a] | 65 (29)[a] |
| Tracheal intubation (%) | 253 (22) | 236 (45)[a] | 203 (90)[a] |
| 24-h PRBC vol ≥ 1 (%) | 220 (19) | 122 (23)[a] | 75 (33)[a] |
| 24-h PRBC vol ≥ 1 and hemorrhagic injury (%) | 106 (9) | 50 (9) | 31 (14)[a] |
| 24-h PRBC vol ≥ 4 (%) | 109 (9) | 64 (12)[a] | 42 (19)[a] |
| 24-h PRBC vol ≥ 4 and hemorrhagic injury (%) | 66 (6) | 34 (6) | 23 (10)[a] |
| Head AIS 3 (%) | 116 (10) | 87 (16)[a] | 49 (22)[a] |
| Head AIS 4 (%) | 76 (7) | 64 (12)[a] | 41 (18)[a] |
| Head AIS 5–6 (%) | 41 (4) | 40 (8)[a] | 35 (16)[a] |
| Intracranial pressure monitoring or craniotomy (%) | 57 (5) | 56 (11)[a] | 45 (20)[a] |

''Hemorrhagic injury'' was a documented laceration or fracture of a solid organ, a thoracic or abdominal hematoma, a vascular injury that required operative repair, or a limb amputation.

[a]Subpopulation significantly different from the ''Any GCS'' study population (using chi-squared test for proportion data; Student's *t*-test for mean age).

AIS, abbreviated injury scale; GCS, Glasgow Coma Scale; PRBC, packed red blood cells.

- The relative risks given *high* SBP ($\geq 175$ mm Hg) were significantly different than the relative risks given HR within the reference range for head AIS 5–6 and for AIS 4–6 (but not for AIS 3–6). Relative risks given SBP $\geq 175$ mm Hg were 3.6 (95% CI, 1.5–8.7), 1.6 (0.8–3.3), and 1.3 (0.8–2.2), for head AIS 5–6, 4–6, and 3–6, respectively.

When we computed the relative risk of high-mortality TBI as a function of prehospital HR, we found the following:

- The relative risks given *low* HR ($<60$ beats/min) were significantly different than the relative risks given HR within the reference range for all three head AIS cutoffs (i.e., AIS 5–6, 4–6, and 3–6). Relative risks given HR $<60$ beats/min were 5.9 (95% CI, 2.9–12.3), 3.0 (1.8–5.1), and 2.0 (1.3–3.0) for head AIS 5–6, 4–6, and 3–6, respectively. Given HR within the reference range (100 beats/min $\geq$ HR $>80$ beats/min), relative risks were 0.7 (0.4–1.3), 0.7 (0.5–1.0), and 0.8 (0.6–1.0), respectively.
- The relative risks given *high* HR ($\geq 120$ beats/min) were not significantly different than the relative risks given HR within the reference range for head AIS 5–6, but they were significantly different for head AIS 4–6 and for AIS 3–6. Relative risks given HR $\geq 120$ beats/min were 1.0 (95% CI, 0.5–2.2), 1.2 (0.8–1.9), and 1.2 (0.9–1.6), for head AIS 5–6, 4–6, and 3–6, respectively.

When we computed the relative risk of high-mortality TBI as a function of prehospital RR, we found no statistically significant risks. This absence of significant findings persisted through all definitions of high-mortality TBI (head AIS 5–6, 4–6, and 3–6). Moreover, there were no significant findings related to RR for either the entire study population or the subset of patients who were spontaneously breathing (no airway intubation).

Figure 1 shows the relative risk of head AIS 5–6 as a function of different ranges of SBP, HR, and RR.

### *Multi-variate regression models*

- *Multi-variate models that included $SBP_{Risk}$ and $HR_{Risk}$*: For the models that did not include the GCS, we found that $HR_{Risk}$ was a significant term in all of the nine investigated multi-variate regression models, whereas $SBP_{Risk}$ was a significant term in eight of nine of the models (see Table 2).
- *Multi-variate models that included GCS, $SBP_{Risk}$, and $HR_{Risk}$*: For the models that did include GCS, we found that $HR_{Risk}$ was a significant term in eight of the nine investigated multi-variate regression models, with elevated statistical significance ($p<0.01$) in the three models for head AIS 5–6/procedure. $SBP_{Risk}$ was a significant term in six of the nine multivariate models (head AIS 5–6 and all-cause mortality), but not in the three models for head AIS 5–6/procedure (see Table 2).

### *Diagnostic test characteristics of the Glasgow Coma Scale alone versus multi-variate regression models*

- For the overall study population ($n=1158$): GCS provided rather good ROC AUCs, and the multi-variate models, including GCS, $SBP_{Risk}$, and $HR_{Risk}$, did not offer a significant increase in ROC AUCs versus the GCS used alone (see Table 2).
- For patients with GCS $<15$: The GCS alone was less discriminatory, that is, it yielded lower ROC AUCs. The models that included the GCS, $SBP_{Risk}$, and $HR_{Risk}$ offered significant improvements over the GCS for all three definitions of high-mortality TBI: head AIS 5–6 (ROC AUC $+0.04$); all-cause mortality (ROC AUC $+0.03$); and head AIS 5–6/procedure (ROC AUC $+0.03$; see Table 2).
- For patients with GCS $\leq 8$: In this subpopulation, the GCS alone was less discriminatory for high-mortality TBI. The models that included the GCS, $SBP_{Risk}$, and $HR_{Risk}$ offered significant improvements over the GCS for all three outcomes, head AIS 5–6 (ROC AUC $+0.12$), mortality (ROC AUC $+0.09$), and AIS 5–6/procedure (ROC AUC $+0.07$; see Table 2).

### *Example of improved risk stratification using the Glasgow Coma Scale, heart rate, and systolic blood pressure versus the Glasgow Coma Scale alone*

Figure 2 shows another distinction between the prehospital GCS alone and the multi-variate model.

- For GCS, its positive predictive value (PPV) for high-mortality TBI gradually increased as the GCS score grew more abnormal.
- The multi-variate regression model was different from the GCS (see Fig. 2). Like the GCS, the PPV for the multivariate model rose gradually as the model output was more abnormal. Unlike the GCS, there was an apparent threshold above which the PPV rose quite steeply and above which it demonstrated significantly higher PPV for high-mortality TBI, as compared with the GCS alone. This implies that patients with a combination of an abnormal GCS, abnormal BP (too high or too low), and abnormal HR (too high or too low) had a $>50\%$ probability of high-mortality TBI.

## Discussion

In this report, we investigated whether vital signs could be used to improve prehospital GCS as a diagnostic indicator of high-mortality TBI. Improved prehospital identification of high-mortality TBI could be valuable, guiding prehospital protocols and mobilizing resources at the receiving facility in an efficient manner. It could also offer an improved tool for research.

This investigation built on previous reports observing a distinct bimodal relationship between BP and clinical outcomes in TBI patient populations.[17,18] Consistent with those earlier reports, we identified a bimodal relationship between prehospital BP and the study outcomes as well as several novel findings:

- We identified a bimodal relationship between high-mortality TBI and prehospital HR.
- We found that the GCS, as documented by the EMS caregivers, did not provide additional risk stratification once the GCS was $\leq 8$ (in other words, all patients with GCS $\leq 8$ had similar rates of high-mortality TBI; see Table 2).
- By combining the GCS and prehospital vital signs in a multivariate model, after accounting for the bimodal relationships, it was possible to improve the identification of the highest-mortality TBI. For example, it was possible to identify an extremely high-risk subgroup that evidenced $>50\%$ probability of all-cause mortality. In contrast, the lowest GCS did not offer such positive predictive value (see Fig. 2). Of course, a clinical score involving relative risk calculation plus multi-variate regression is not feasible for a bedside clinician, but it is well within the capabilities of emerging information technologies (for examples, see previous reports[20,21]).

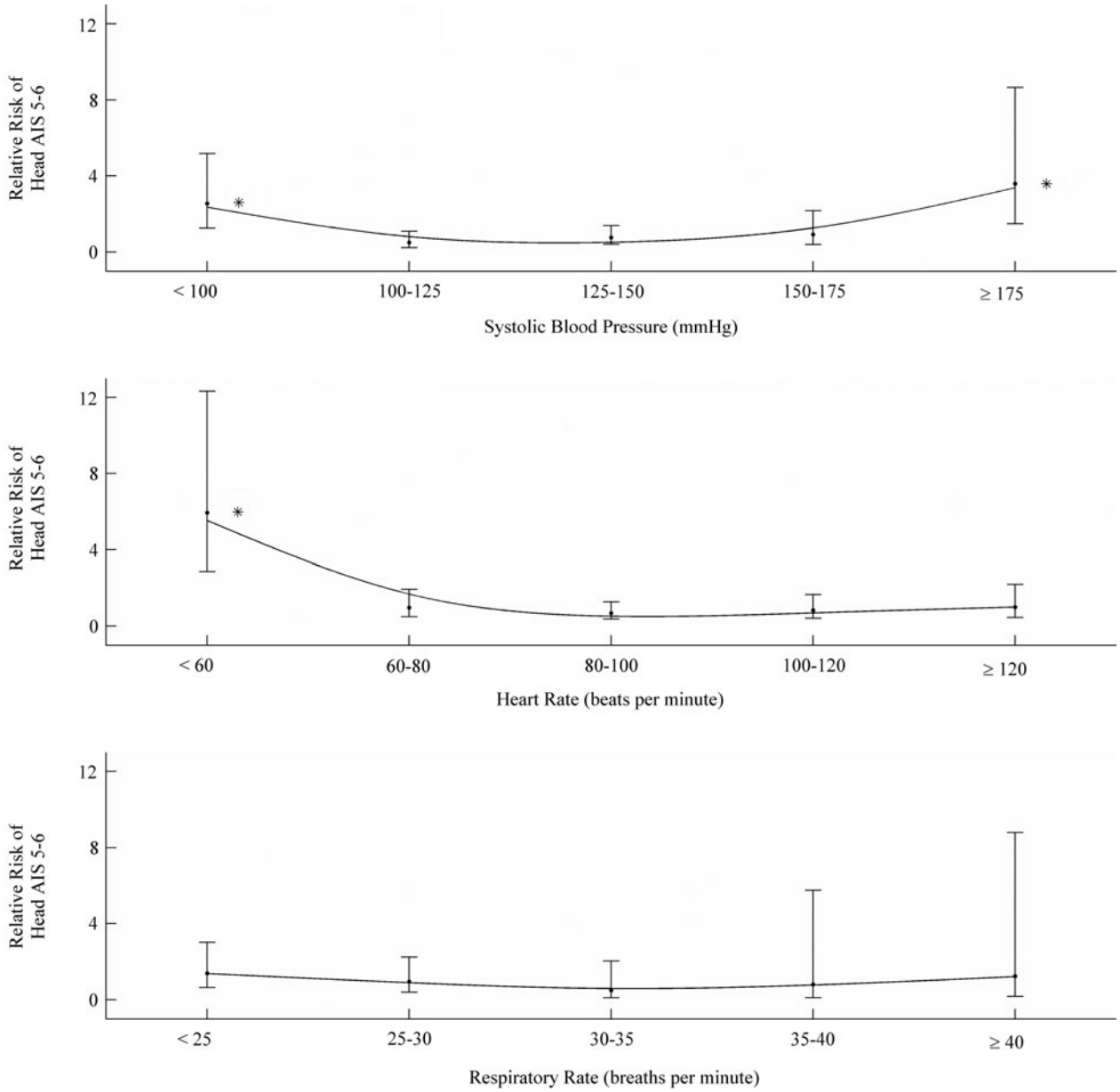



**FIG. 1.** Prehospital vital signs versus relative risk of head abbreviated injury scale (AIS) 5–6. Error bars signify 95% confidence interval. Solid line indicates the cubic spline fit to data. *Statistically significant difference ($p < 0.05$) from the baseline reference range (reference ranges: systolic blood pressure, 100–125 mm Hg; heart rate, 80–100 beats/min; respiratory rate, 30–35 breaths/min).



### *Hemodynamics of high-mortality traumatic brain injury: examining the bimodal relationships*

What was the basis of the bimodal relationship between hemodynamics and high-mortality TBI? High BP and low HR are hallmarks of the Cushing reflex, a well-known hemodynamic response to elevated intracranial pressure. The association between low BP and high HR in high-mortality TBI is not as clear. Major mechanism polytrauma is likely a root cause, which can result in hemorrhagic hypovolemia (and hypotension and tachycardia) as well as high-mortality TBI. This is suggested by Table 1, where it is apparent that the subpopulation with lower GCS had higher rates of blood transfusion coincident with explicitly hemorrhagic injuries. In addition, in a few cases, the low BP could be a correlate of spinal shock. Finally, the association between mortality and hypotension may be causal, in that low BP causes secondary harm to the injured brain. Overall, it seems likely that the basis of the hypotension and/or tachycardia relationship with TBI is multi-factorial.

We hypothesized that there would be an association between RR and TBI. However, we did not identify any significant relationships involving RR. Possibly, the patients in this data set who had respiratory depression tended to receive early tracheal intubation, and so their RR was under control of the caregivers, not the patient's own depressed respiratory drive.

We found it necessary to consider the bimodal relationships between TBI risk versus SBP and HR. When we first performed a routine regression analysis on TBI versus BP and HR, without accounting for the bimodal relationship, we did not find linear correlations because the TBI cases with abnormally high values

Table 2. Comparison of Areas Under Receiver-Operating Characteristic Curves of the Investigative Parameters

| Investigated variables \ Population | ROC AUC (95% CI) | | |
|---|---|---|---|
| | *Any GCS*<br>*Total = 1,158 subjects* | *GCS < 15*<br>*Total = 530 subjects* | *GCS ≤ 8*<br>*Total = 225 subjects* |
| Head AIS 5–6 | Cases = 41, Controls = 1,117 | Cases = 40, Controls = 490 | Cases = 35, Controls = 190 |
| GCS | 0.90 (0.86–0.93) | 0.80 (0.76–0.85) | 0.59 (0.52–0.66) |
| $SBP_{Risk}$, $HR_{Risk}$ | 0.64 (0.54–0.73)[a,b,c] | 0.68 (0.58–0.76)[a,b,c] | 0.65 (0.54–0.74)[a,b] |
| GCS, $SBP_{Risk}$, $HR_{Risk}$ | 0.91 (0.85–0.94)[a,b] | 0.84 (0.78–0.89)[a,b,c] | 0.71 (0.61–0.79)[a,b,c] |
| All-cause mortality | Cases = 82, Controls = 1,076 | Cases = 73, Controls = 457 | Cases = 65, Controls = 160 |
| GCS | 0.85 (0.80–0.90) | 0.82 (0.77–0.86) | 0.65 (0.59–0.70) |
| $SBP_{Risk}$, $HR_{Risk}$ | 0.68 (0.61–0.74)[a,b,c] | 0.66 (0.59–0.73)[a,b,c] | 0.66 (0.58–0.74)[a,b] |
| GCS, $SBP_{Risk}$, $HR_{Risk}$ | 0.88 (0.83–0.91)[a,c] | 0.85 (0.80–0.89)[a,b,c] | 0.74 (0.66–0.81)[a,b,c] |
| Head AIS 5–6/procedure | Cases = 83, Controls = 1,075 | Cases = 81, Controls = 449 | Cases = 68, Controls = 157 |
| GCS | 0.89 (0.86–0.92) | 0.78 (0.73–0.82) | 0.55 (0.49–0.62) |
| $SBP_{Risk}$, $HR_{Risk}$ | 0.62 (0.55–0.68)[a,b,c] | 0.63 (0.56–0.69)[a,b,c] | 0.59 (0.50–0.67)[b] |
| GCS, $SBP_{Risk}$, $HR_{Risk}$ | 0.90 (0.86–0.93)[b] | 0.81 (0.77–0.86)[b,c] | 0.62 (0.54–0.70)[b] |

Shown are comparisons of areas under receiver operating characteristic curves (ROC AUCs) of the following investigative variables: 1) GCS; 2) the multi-variate regression model using the relative risk of traumatic brain injury computed from SBP ($SBP_{Risk}$) and from HR ($HR_{Risk}$); and 3) the multi-variate regression model using GCS, $SBP_{Risk}$, and $HR_{Risk}$.

[a]$SBP_{Risk}$ term was statistically significant in the multi-variate regression model.

[b]$HR_{Risk}$ term was statistically significant in the multi-variate regression model.

[c]ROC AUC was statistically significantly different from that of GCS.

AIS, abbreviated injury scale; CI, confidence interval; GCS, Glasgow Coma Scale; ROC, receiver operating characteristic; AUC, area under the curve; SBP, systolic blood pressure; HR, heart rate.



cancelled out TBI cases with abnormally low values. Interestingly, most of the classic prehospital severity scores (e.g., the trauma score,[14] the prehospital index,[15] the CRAMS score,[16] and the newer Glasgow Coma Scale, Age, and Systolic Blood Pressure[13]/Mechanism, Glasgow Coma Scale, Age, and Arterial Pressure scores[12]) do not account for the prognostic value of hypertension and bradycardia in patients with TBI. We speculate that superior overall severity scores could be developed by accounting for the bimodal relationship between hemodynamics and high-mortality TBI.

### *Clinical implications*

Although originally intended for use after initial resuscitation, the GCS has been adopted for the earliest stages of trauma care, although it is not optimal for that context (see Introduction). Indeed, in this study, we found that below a cutoff of 8, the prehospital GCS offered no additional discriminatory value (see ROC AUCs in Table 2). Given the lowest level of prehospital GCS (= 3), the likelihood of mortality was 37% (see Fig. 2). In contrast, using prehospital HR and SBP with the GCS, it was possible to further risk stratify patients with a prehospital GCS ≤ 8. For example, given those patients with a low GCS and abnormal HR and SBP, we found a *probability* (> 50%) of mortality risk (see Fig. 2).

How could a superior tool for early estimation of TBI mortality risk be useful? First, this measure could be applied in research to control for TBI severity in biostatistical analyses of prehospital care (for instance, studies such as Davis and colleagues[5]). Second, such a tool may be useful for triage applications, such as determining which patients should receive priority care or involvement of a neurosurgeon at the earliest juncture. Third, improved risk stratification could be used to develop more rational clinical decision-making algorithms for prehospital management. For example, transport speed might be the highest priority (e.g., no delay for tracheal intubation) for patients with the highest TBI risk, in case the patient requires immediate decompressive neurosurgery, and osmotherapy may be useful, whereas permissive hypotension[33] would be contraindicated. In contrast, for patients with depressed consciousness who are unlikely to have high-mortality TBI based on the patterns of GCS, HR, and SBP, it may prove judicious to delay transportation long enough to secure the airway and protect against aspiration, and permissive hypotension could offer more benefit than risk. In summary, based on a patient's quantitative risk of high-mortality TBI, different prehospital interventions may offer different risk-benefit profiles. Further research is warranted into novel prehospital protocols in which decision making is dependent on the quantitative risk of life-threatening TBI.

As a practical matter, the multi-variate analysis used in this investigation requires computer analysis; such a tool is well within today's in- and prehospital capabilities. Indeed, our research team has currently deployed such automated computational devices, networked to a Propaq 206 patient monitor (Welch Allyn), on board Boston Medflight helicopters for prospective trials of advanced decision-support algorithms.[21] In practice, after the GCS score was electronically documented by a caregiver, the informatics system could process the patient's recent BP and HR measurements, automatically identify and exclude any unreliable vital-sign values, compute $HR_{Risk}$ and $SBP_{Risk}$, and output the multi-variate regression model result.

### *Limitations*

In a few of the multi-variate models, either the $SBP_{Risk}$ term or the $HR_{Risk}$ term did not reach statistical significance. However, we do not consider each and every outcome as a distinct hypothesis. Rather, we are testing the overall hypothesis that information from $SBP_{Risk}$ and $HR_{Risk}$ can significantly improve on the ability of the GCS to identify patients with high-mortality TBI. The results shown in Table 2 demonstrate a consistent pattern supporting this hypothesis in patients with a GCS < 15. In contrast, GCS = 15 was

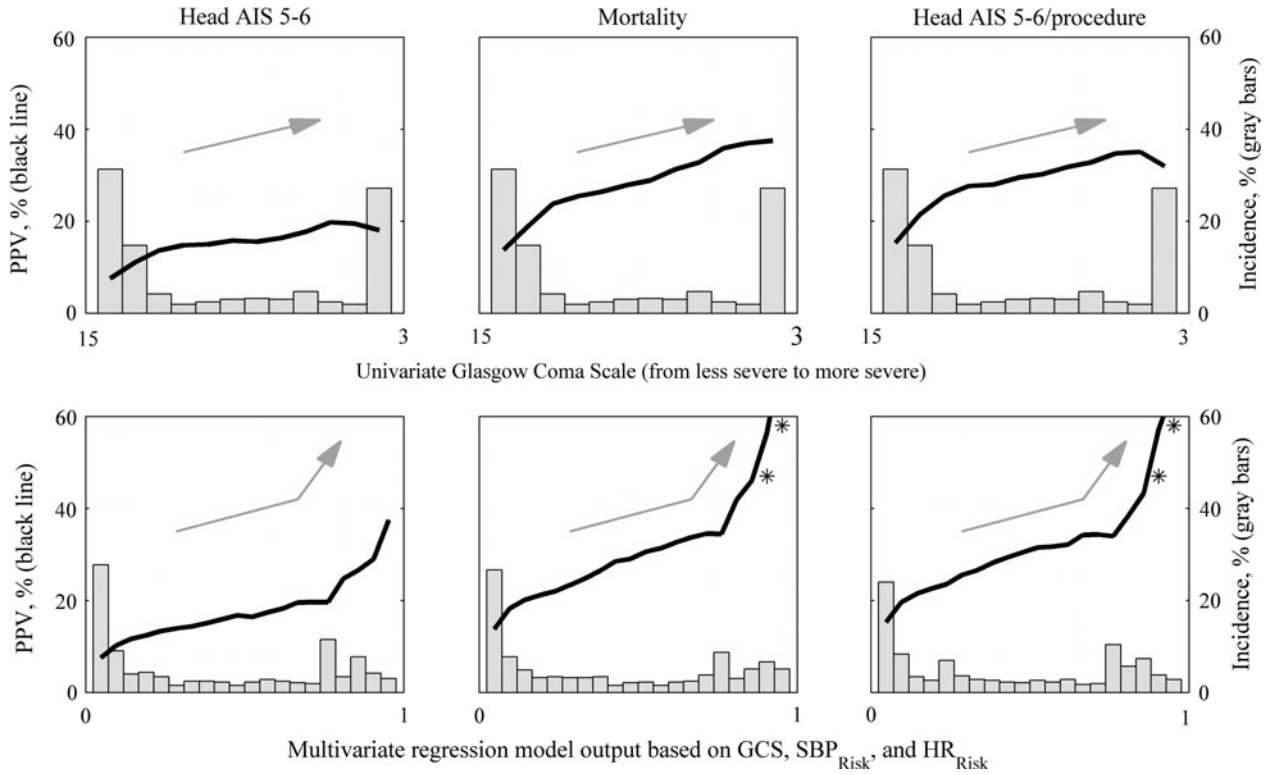



**FIG. 2.** Positive predictive value (PPV) of Glasgow Coma Scale (GCS) versus PPV of multi-variate regression model. Compared with GCS alone (top row), the multi-variate regression model (bottom row) better identified a subset of patients at elevated risk of high-mortality traumatic brain injury (TBI), either defined by head AIS 5–6 (left column), all-cause mortality (middle column), or head AIS 5–6/procedure (right column). Solid black lines show the PPV of high-mortality TBI for patients ≤ GCS score (top row) or ≥ multivariate model output (bottom row). Distribution of patients with each GCS score is also shown (shaded bars, top row) as is the distribution of patients with different ranges of regression model outputs (shaded bars, bottom row). *Multi-variate classifier's PPV was significantly greater than PPV of GCS = 3 ($p < 0.05$, chi-square test). SBP$_{Risk}$, systolic blood pressure relative risk; HR$_{Risk}$, heart rate relative risk.

very effective at identifying the majority of patients who had a very low risk of high-mortality TBI; the associated ROC AUC was quite high, without much room for improvement.

Second, this analysis focused on early identification of high-mortality TBI, because optimizing survival is a primary goal of prehospital and early hospital care. However, preventing disability is as important to consider as survival. Whether BP and HR offer prognostic information about functional neurological outcome was not addressed in this analysis, and future investigation into this important question is warranted.

Third, this analysis focused on the GCS measured by EMS caregivers. Our findings may not extend to the classic GCS carefully measured in-hospital after patient stabilization and the elimination of intoxicants, which are often different from prehospital measurements.[9,10] Prehospital conditions are more demanding, whereas staffing is often limited to a couple of caregivers, so nuanced GCS scoring is unlikely to be a high priority: The field medic will probably not heed the difference between flexion versus extension motor responses when it is obvious that the patient has time-sensitive injuries.

The final limitation relates to the precise numerical results of our analysis. For some HR and SBP ranges, there were not enough cases for a tight estimation of the associated relative risk. A larger data set would presumably yield a more accurate quantification of the HR and SBP bimodal relationship and the optimal coefficients for the multi-variate model. All the same, our findings were qualitatively consistent with findings from other reports.[17,18] We suggest that the overall findings of this report are likely valid.

## Conclusion

We found that the prehospital GCS alone was unable to effectively distinguish between trauma patients with moderate risk versus the highest risk of high-mortality TBI. A multi-variate regression model with three terms—GCS, SBP, and HR—offered significantly improved test performance after accounting for the bimodal relationships between TBI versus SBP and HR. This score could be useful for guiding operations at the receiving hospital (e.g., early consultation by a neurosurgeon and readying an operating room). Further, we speculate that improved methods for the prehospital assessment of TBI risk could facilitate new prehospital management practices.

## Acknowledgments

This work was supported by the Defense Health Program and the Combat Casualty Care Research Area Directorate of the U.S. Army Medical Research and Materiel Command (Fort Detrick, MD). The authors thank Christopher Jamieson for his support in formatting the manuscript and Kathleen McGuire Gilbert for editing the manuscript.

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 article has been approved for public release with unlimited distribution.

### Author Disclosure Statement

No competing financial interests exist.

### References


1. Teasdale, G., and Jennett, B. (1974). Assessment of coma and impaired consciousness. A practical scale. Lancet 2, 81–84.
2. The Brain Trauma Foundation. The American Association of Neurological Surgeons. The Joint Section on Neurotrauma and Critical Care. (2000). Glasgow Coma Scale score. J. Neurotrauma 17, 563–571.
3. Sasser, S.M., Hunt, R.C., Faul, M., Sugerman, D., Pearson, W.S., Dulski, T., Wald, M.M., Jurkovich, G.J., Newgard, C.D., and Lerner, E.B.; Centers for Disease Control and Prevention. (2012). Guidelines for field triage of injured patients: recommendations of the National Expert Panel on Field Triage, 2011. MMWR Recomm. Rep. 61, 1–20.
4. Dunham, C.M., Barraco, R.D., Clark, D.E., Daley, B.J., Davis, F.E., 3rd, Gibbs, M.A., Knuth, T., Letarte, P.B., Luchette, F.A., Omert, L., Weireter, L.J., and Wiles, C.E., 3rd; EAST Practice Management Guidelines Work Group. (2003). Guidelines for emergency tracheal intubation immediately after traumatic injury. J. Trauma 55, 162–179.
5. Davis, D.P., Peay, J., Sise, M.J., Vilke, G.M., Kennedy, F., Eastman, A.B., Velky, T., and Hoyt, D.B. (2005). The impact of prehospital endotracheal intubation on outcome in moderate to severe traumatic brain injury. J. Trauma 58, 933–939.
6. Magazzini, S., Nazerian, P., Vanni, S., Paladini, B., Pepe, G., Casanova, B., Crugnola, C., and Grifoni, S. (2012). Clinical picture of meningitis in the adult patient and its relationship with age. Intern. Emerg. Med. 7, 359–364.
7. Mathew, V., Misgar, R.A., Ghosh, S., Mukhopadhyay, P., Roychowdhury, P., Pandit, K., Mukhopadhyay, S., and Chowdhury, S. (2011). Myxedema coma: a new look into an old crisis. J. Thyroid Res. 2011, 493462.
8. Green, S.M. (2011). Cheerio, laddie! Bidding farewell to the Glasgow Coma Scale. Ann. Emerg. Med. 58, 427–430.
9. Zuercher, M., Ummenhofer, W., Baltussen, A., and Walder, B. (2009). The use of Glasgow Coma Scale in injury assessment: a critical review. Brain Inj. 23, 371–384.
10. Marion, D.W., and Carlier, P.M. (1994). Problems with initial Glasgow Coma Scale assessment caused by prehospital treatment of patients with head injuries: results of a national survey. J. Trauma 36, 89–95.
11. Kerby, J.D., MacLennan, P.A., Burton, J.N., McGwin, G., Jr., and Rue, L.W., 3rd. (2007). Agreement between prehospital and emergency department Glasgow Coma scores. J. Trauma 63, 1026–1031.
12. Sartorius, D., Le Manach, Y., David, J.S., Rancurel, E., Smail, N., Thicoipe, M., Wiel, E., Ricard-Hibon, A., Berthier, F., Gueugniaud, P.Y., and Riou, B. (2010). Mechanism, Glasgow Coma Scale, age, and arterial pressure (MGAP): a new simple prehospital triage score to predict mortality in trauma patients. Crit. Care Med. 38, 831–837.
13. Kondo, Y., Abe, T., Kohshi, K., Tokuda, Y., Cook, E.F., and Kukita, I. (2011). Revised trauma scoring system to predict in-hospital mortality in the emergency department: Glasgow Coma Scale, age, and systolic blood pressure score. Crit. Care 15, R191.
14. Champion, H.R., Sacco, W.J., Copes, W.S., Gann, D.S., Gennarelli, T.A., and Flanagan, M.E. (1989). A revision of the trauma score. J. Trauma 29, 623–629.
15. Koehler, J.J., Baer, L.J., Malafa, S.A., Meindertsma, M.S., Navitskas, N.R., and Huizenga, J.E. (1986). Prehospital Index: a scoring system for field triage of trauma victims. Ann. Emerg. Med. 15, 178–182.
16. Gormican, S.P. (1982). CRAMS scale: field triage of trauma victims. Ann. Emerg. Med. 11, 132–135.
17. Zafar, S.N., Millham, F.H., Chang, Y., Fikry, K., Alam, H.B., King, D.R., Velmahos, G.C., and de Moya, M.A. (2011). Presenting blood pressure in traumatic brain injury: a bimodal distribution of death. J. Trauma 71, 1179–1184.
18. Butcher, I., Maas, A.I., Lu, J., Marmarou, A., Murray, G.D., Mushkudiani, N.A., McHugh, G.S., and Steyerberg, E.W. (2007). Prognostic value of admission blood pressure in traumatic brain injury: results from the IMPACT study. J. Neurotrauma 24, 294–302.
19. Ley, E.J., Singer, M.B., Clond, M.A., Gangi, A., Mirocha, J., Bukur, M., Brown, C.V., and Salim, A. (2011). Elevated admission systolic blood pressure after blunt trauma predicts delayed pneumonia and mortality. J. Trauma 71, 1689–1693.
20. Reisner, A.T., Chen, L., and Reifman, J. (2012). The association between vital signs and major hemorrhagic injury is significantly improved after controlling for sources of measurement variability. J. Crit. Care 27, 533.e1–533.e10.
21. Reisner, A.T., Khitrov, M.Y., Chen, L., Blood, A., Wilkins, K., Doyle, W., Wilcox, S., Denison, T., and Reifman, J. (2013). Development and validation of a portable platform for deploying decision-support algorithms in prehospital settings. Appl. Clin. Inform. 4, 392–402.
22. Cooke, W.H., Salinas, J., Convertino, V.A., Ludwig, D.A., Hinds, D., Duke, J.H., Moore, F.A., and Holcomb, J.B. (2006). Heart rate variability and its association with mortality in prehospital trauma patients. J. Trauma 60, 363–370.
23. McKenna, T.M., Bawa, G., Kumar, K., and Reifman, J. (2007). The physiology analysis system: an integrated approach for warehousing, management and analysis of time-series physiology data. Comput. Methods Programs Biomed. 86, 62–72.
24. Reisner, A.T., Chen, L., McKenna, T.M., and Reifman, J. (2008). Automatically-computed prehospital severity scores are equivalent to scores based on medic documentation. J. Trauma 65, 915–923.
25. Hug, C.W., Clifford, G.D., and Reisner, A.T. (2011). Clinician blood pressure documentation of stable intensive care patients: an intelligent archiving agent has a higher association with future hypotension. Crit. Care Med. 39, 1006–1014.
26. Chen, L., McKenna, T., Reisner, A., and Reifman, J. (2006). Algorithms to qualify respiratory data collected during the transport of trauma patients. Physiol. Meas. 27, 797–816.
27. Yu, C., Liu, Z., McKenna, T., Reisner, A.T., and Reifman, J. (2006). A method for automatic identification of reliable heart rates calculated from ECG and PPG waveforms. J. Am. Med. Inform. Assoc. 13, 309–320.
28. Chen, L., Reisner, A.T., Gribok, A., and Reifman, J. (2009). Exploration of prehospital vital sign trends for the prediction of trauma outcomes. Prehosp. Emerg. Care 13, 286–294.
29. Garthe, E., States, J.D., and Mango, N.K. (1999). Abbreviated injury scale unification: the case for a unified injury system for global use. J. Trauma 47, 309–323.
30. Daly, L.E. (1998). Confidence limits made easy: interval estimation using a substitution method. Am. J. Epidemiol. 147, 783–790.
31. Altman, D.G., and Bland, J.M. (2003). Interaction revisited: the difference between two estimates. BMJ 326, 219.
32. DeLong, E.R., DeLong, D.M., and Clarke-Pearson, D.L. (1988). Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics 44, 837–845.
33. Kobayashi, L., Costantini, T.W., and Coimbra, R. (2012). Hypovolemic shock resuscitation. Surg. Clin. North. Am. 92, 1403–1423.



Address correspondence to:
*Jaques Reifman, PhD*
*Biotechnology HPC Institute*
*U.S. Army Medical Research and Materiel Command*
*MCMR-TT*
*504 Scott Street*
*Fort Detrick, MD 21702*

*E-mail:* jaques.reifman.civ@mail.mil

# A Platform for Real-time Acquisition and Analysis of Physiological Data in Hospital Emergency Departments

Jason B. Smith-*IEEE Member*, Andrew T. Reisner, Shwetha Edla, Jianbo Liu, Stephanie Liddle, and Jaques Reifman

***Abstract*— An opportunity exists for automated clinical decision support, in which raw source data from a conventional physiological monitoring system are continuously streamed to an independent analysis platform. Such a system would enable a wider range of functionality than offered by the source monitoring system. Although vendor solutions for this purpose are emerging, we developed our own system in order to control the expense and to permit forensic analysis of the internal core functionality of the system. In this report, we describe a platform that can provide decision support for trauma patients in an Emergency Department (ED). System evaluation spanned 39 days, and included a total of 2200 patient session hrs of real-time monitoring. We highlight the technical issues that we confronted, including protection of the core monitoring network, the real-time communication of electronic medical data, and the reliability of the real-time analysis. Detailing these nuanced technical issues may be valuable to other software developers or for those interested in investing in a vendor solution for similar functionality.**

## I. Introduction

Automated alerting and decision support are possible when hospitalized patients receive continuous monitoring, such as physiological alarms embedded within core monitoring systems. A newer form of decision support also exists in which vital-sign data are fed (e.g., using Health Level Seven [HL7] standards) to electronic medical records (EMRs). However, EMR decision-support algorithms usually operate on clinical data that have been reviewed and verified by clinicians (as opposed to raw source data from the monitoring network).

A third opportunity for implementing decision support exists, in which the raw source data from the monitoring network are continuously streamed to an independent analysis platform, enabling a wider range of functionality than offered by the source monitoring system [1]. A number of emerging vendor solutions are available to support such functionality, including the BedMasterEx (BM) Data Acquisition Software (Excel Medical Electronics Inc., Jupiter, FL) with its StreamingAnalytics platform powered by IBM InfoSphere Streams (IBM, Yorktown Heights, NY); Bernoulli Enterprise Software (Cardiopulmonary Corp., Milford, CT); and DocBox (DocBox Inc., Newton, MA).

Our research team was interested in implementing a set of investigational decision-support algorithms that analyze streaming physiological data in realtime. Our focus is the development of decision support for trauma patients, and this project is named for its intended purpose: Automated Processing of the Physiological Registry for Assessment of Injury Severity in the Emergency Department (APPRAISE-ED). APPRAISE-ED is a follow-up to a similar system previously developed for prehospital air ambulances [2].

Also in prior work, we studied whether the BM system for data acquisition would have any identifiable harmful effects on the hospital's core monitoring network [3]. In that preliminary work, our testing did not reveal any deleterious impact on the core monitoring network, although a telephone poll of customers using the vendor's product revealed that a majority experienced at least one episode of unanticipated failure to archive data due to the difficulties in managing a distributed, network-based data acquisition system. This suggested that the system was safe and effective, but that the support and oversight necessary for the product were often underestimated by novice users.

The next step for our research team was in-hospital real-time data analysis, requiring a platform for acquisition and analysis of physiological signals for automated decision support. We considered using one of the aforementioned vendor solutions, but decided to develop our own solution because these were (in our subjective assessment) relatively expensive, without an established track record of good performance, and also lacked sufficient documentation for us to decisively evaluate their core functionality.

In this report, we describe our system, its evaluation, and the lessons we learned. For those who are considering the development or the purchase of such a system, this report may help them to better understand several key underlying

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. B. Smith is with the Department of Defense Biotechnology High Performance Computing Software Applications Institute (BHSAI), Telemedicine and Advanced Technology Research Center (TATRC), USAMRMC, Fort Detrick, MD 21702 USA (e-mail: jsmith@bhsai.org).

A. T. Reisner is with the BHSAI, TATRC, USAMRMC, Fort Detrick, MD 21702 USA and with the Department of Emergency Medicine, Massachusetts General Hospital, Boston, MA 02114 USA (corresponding author; phone: 617-726-2241; e-mail: areisner@partners.org).

S. Edla is with the BHSAI, TATRC, USAMRMC, Fort Detrick, MD 21702 USA (e-mail: sedla@bhsai.org).

J. Liu is with the BHSAI, TATRC, USAMRMC, Fort Detrick, MD 21702 USA (e-mail: jliu2@bhsai.org).

S. Liddle was with the Department of Biomedical Engineering, Massachusetts General Hospital, Boston, MA 02114 USA.

J. Reifman is a Senior Research Scientist and Director of the BHSAI, TATRC, USAMRMC, Fort Detrick, MD 21702 USA (e-mail: jaques.reifman.civ@mail.mil).

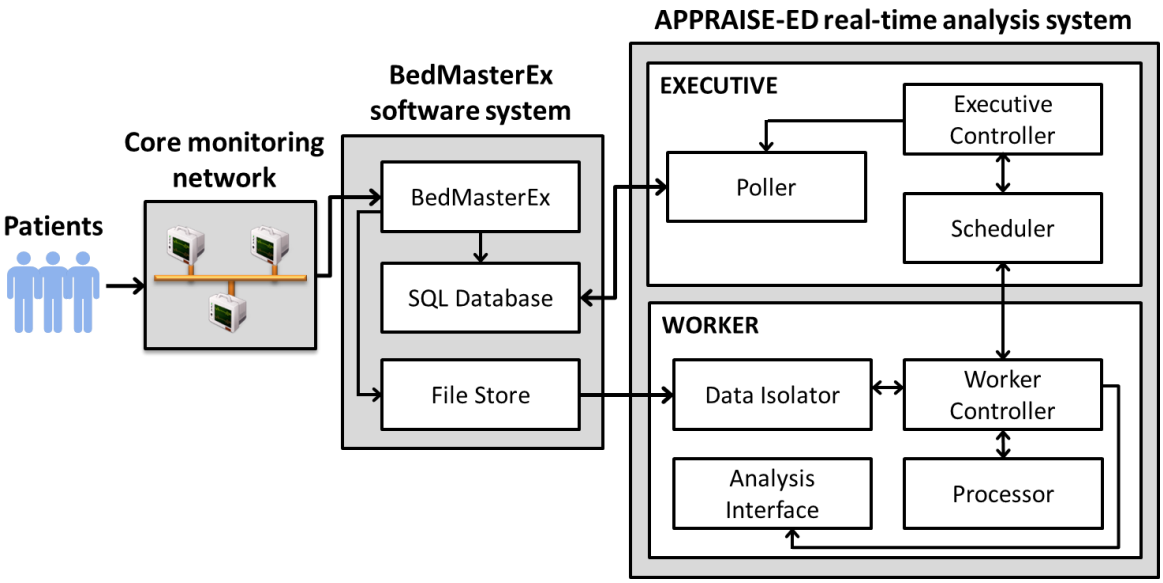



Figure 1. Illustration of the major components that comprise the proposed platform for the real-time acquisition and processing of physiological signals in the emergency department. APPRAISE-ED: Automated Processing of the Physiological Registry for Assessment of Injury Severity in the Emergency Department.

performance issues. We carefully examined two aspects of the system: first, reliability of data communication between the core monitoring network and the novel analysis platform, and second, the reliability of the real-time analysis.

## II. Methods

### A. Description

Fig. 1 illustrates the complete system, consisting of three major components: the core monitoring network, which is a system of 16 Solar patient monitors (General Electric [GE], Milwaukee, WI); the proprietary BM software system; and our APPRAISE-ED system.

BM is hosted on a dedicated personal computer (PC) and, as per the manufacturer's specifications, collects physiological data from the GE patient monitors and saves it to binary (STP) files archived on the Windows file system. In parallel, BM archives associated data in an SQL Server (Microsoft Corp., Redmond, WA) database [4] archives associated data, including patient identifying information (i.e., protected health information [PHI]), monitor status, and indices of how data are stored within each STP file. The BM system standardizes waveform and vital-sign data frequencies to 240 and 1 Hz, respectively, and provides a common name for signals, thus offering a uniform means for access and analysis of collected data at the expense of averaging or missing higher frequency sensor data.

The APPRAISE-ED software runs on a dedicated server. The software consists of an executive module responsible for managing overall APPRAISE-ED system functionality. Also, for each monitor/bay under analysis, it creates an instance of a dedicated worker module.

The executive module includes three key sub-modules. The executive controller sub-module directs the tasks to be performed by the other sub-modules. The poller sub-module determines which monitor/bays are in active use by querying the BM SQL Server database. In addition, as new physiological data are continuously accumulated through time, the poller sub-module determines the timing information about when each BM STP file was updated, as well as the location of the new data within the updated STP file. The poller's queries are performed every 5 s (query intervals are configurable), and this information is passed on to the scheduler sub-module.

The scheduler sub-module creates instances of the worker modules for each monitor/bay in active use. The scheduler also coordinates the transfer of newly acquired physiological data to each worker module, depending on its estimate of the time required to perform each analysis.

It is the individual worker modules which, as directed by the scheduler, are responsible for actually extracting physiological data from BM and then performing analyses. The worker controller sub-module in each worker directs the operations to be performed by each of the sub-modules. Each worker's data isolator sub-module locates and extracts the newest physiological data from the STP file. We rely on a software tool, Stp2Xml, available from the BM vendor, to transform the STP file into an intermediate data format. The data isolator sub-module then reads this intermediate XML file and deletes PHI from the extracted data. Finally, the data isolator stores the physiological data in a new portable binary file format (HDF: Hierarchical data format [5]).

Each worker's processor sub-module performs additional data handling and monitoring, such as tracking update times and verifying that new physiological data were obtained. The processor also checks the physiological data for indications that the patient was off-monitor and/or that a new patient may have been swapped into that monitor/bay. We found that, in a busy ED, there were often cases where patients were being removed from the monitor/bay without being electronically discharged, or patients were swapped without any update of the patient identity within the GE monitoring network. Accordingly, when no physiological data are detected for 5 min (configurable) the active session is

terminated. If data re-appear after that, a new session handled by a new instance of the worker module is started.

Each worker's analysis interface sub-module communicates with a MATLAB process (The Mathworks Inc., Natick, MA) to perform an analysis of isolated data via the MATLAB application programming interface (API). Algorithms are used to determine the reliability of waveform (e.g., electrocardiogram) and vital-sign data (e.g., heart rate) [6-8]. A primary focus of our research is early identification of patients with hemorrhage, and we have investigated a methodology involving multivariate classification [9] with the sequential probability ratio test [10] (the latter is an established technique for identifying abnormal patterns in a series of repeated measurements). The analysis interface has the ability to simultaneously run multiple instances of the analytic algorithms on the data from a given session at the same time, so that results from different algorithms can be compared. We constructed an analysis viewer to allow for real-time viewing of analysis results from a remote, networked location. At the end of each session, the HDF files and the analysis results are saved on the storage server for post-hoc review.

In prior work, we examined the impact of BM on the function of the core GE monitoring network and its constituent monitors [3]. For the current project, it was a priority that any newer functionality should not pose any additional risk to the same core monitoring network. Accordingly, the core GE monitoring network remains an isolated network without direct connection to the internet or the APPRAISE-ED software, except indirectly through the BM host PC, which serves as a bridge between the two networks. The APPRAISE-ED software resides on a separate server which communicates (read-only) with the BM host PC through select open ports. Queries from the APPRAISE-ED software to the SQL Server database (on the BM host PC) are designed to be low impact, i.e., minimum number of queries. Other aspects of query design were to limit the size of returned data corresponding to a given query, and to minimize the number of connections to the database. The core GE monitoring network thus remains a closed network, with its data travelling first to the BM host PC and then, through highly restricted ports and read-only access, to the APPRAISE-ED server and storage systems. In order to minimize any exposure to malware, neither the BM host PC nor the APPRAISE-ED analysis server is used for any other purpose than the aforementioned data processing.

### B. Testing

Overall, testing was similar to that employed by Khitrov et al. [2], although the current system adds the complexity of up to 16 bays/monitors to analyze at a time, rather than a single transport monitor. We tested the APPRAISE-ED software component by component by using a unit-testing approach that exercised each module and sub-module. Next, we tested integrated system function during real-time operation by implementing a "simulated" ED consisting of three networked GE monitors, a virtual BM installation bridging the GE network and the general laboratory network, and a virtual server installation hosting the APPRAISE-ED software. Patients were simulated using Netech MiniSim 1000 (Netech Corp, Farmingdale, NY) patient simulators.

Real-time functionality, including BM data extraction, algorithm processing, and result archiving, were compared to offline analysis of the same raw data (sourced from the BM archive). We confirmed that the software met all the aforementioned functional specifications (see *Description* above).

We conducted an exploration of several potential failure scenarios. We attempted to operate the GE Solar monitors in unusual fashions (e.g., turning monitors off in the middle of a session; swapping patients without formally discharging the initial patient within the GE network, etc.). Also, we simulated network failure scenarios during an ongoing session, such as interrupting APPRAISE-ED access to the SQL Server by blocking the SQL Server port, and blocking access to the STP file location by unmapping the shared drive on the PC running BM.

After successful laboratory testing, the system was tested in clinical use, in the Massachusetts General Hospital's ED, where we compared the system's resultant HDF files and real-time analysis results to offline analysis of data sourced directly from the BM archive. We also reviewed Windows Performance Monitor to assess the function of the BM host PC during clinical use.

## III. Results

Here, we summarize the main results and present several notable findings. During laboratory testing, we confirmed that the software met all the aforementioned functional specifications. The system was able to begin and end analysis sessions when new patients were placed on or removed from the monitor. The system was able to process simultaneous patient sessions as intended. During simulated network communication interruptions, the system appropriately logged the events and recovered from those interruptions as designed.

We performed clinical testing on weekdays, mostly during morning hrs, over a span of 39 days. Over a total of 230 hrs (2200 patient session hrs) of real-time ED operation, we did not observe network communication errors between the components of the overall system. Real-time analysis was executed as per our design, on up to 16 monitors/bays simultaneously, without any operational errors.

Based on the data within the HDF files (both vital-sign numeric as well as waveform data), we ascertained that the data passed from BM in real time matched the source BM record, with one exception. We uncovered one trivial but consistent discrepancy in the data values for the pulse oximetry (SpO2) waveform that were passed to the real-time system. Specifically, the first three samples of the waveform (representing the first 0.0013 ms) at the beginning of each 60-s segment received by the real-time system did not match the source BM data. Based on our internal analysis and in discussion with the BM vendor [11], we learned that the Stp2Xml data export tool applied a moving-window average after extracting SpO2 waveform data excerpts from the BM STP file. This moving-window average distorted the data at the beginning of the excerpt where it lacked preceding SpO2

waveform for proper averaging. We addressed this issue by using the Stp2Xml tool to transform the previous SpO2 waveform excerpt along with the current segment so all data within the BM extraction window would be available.

When we compared real-time analysis in the ED (the results of which were available within the HDF data archive) versus offline retrospective analysis of the same source data (as archived by the BM system), there was convincing agreement. The mean standard error between these methods was 0.00. There were no episodes of unusual operation of the BM host PC, according to the logs of the Windows Performance Monitor.

## IV. Discussion

We have successfully developed, validated, and deployed the APPRAISE-ED system for prospectively testing real-time decision-support algorithms during clinical operations. The goal of this report is to highlight the technical issues that we confronted. Details of these issues may be valuable to other software developers or to those interested in procuring a vendor solution for similar functionality, which is often a six-figure investment. These issues may not be readily apparent to clinicians, administrators, and researchers interested in acquiring this functionality.

First, we felt it was important to carefully consider the integrity of the core GE monitoring network. In prior work, we assessed whether the BM archiving system could alter the functionality of the core GE monitoring network [3]. By design, the newly added functionality did not interact with the core monitoring network, but used the BM host PC as the indirect communication bridge through which real-time physiological data were obtained. Interactions between the APPRAISE-ED server and the BM host PC were kept to a minimum (i.e., read-only access, query frequency minimized, query date returns minimized, and working within elevated levels of BM host PC security).

Second, we felt it was important to consider the reliability of the communication between the software components. In our system design, we added functionality to log interruptions and gracefully recover from such interruptions automatically. We also identified at least one condition in which the data passed from the BM system in real time were not exactly the same as the data actually archived by BM for retrospective analysis. Although the differences were trivial, this issue—the integrity of data communicated for real-time analysis—is not trivial. It will be essential to consider this issue for any and all interoperable systems if such healthcare decision-support functionality becomes normative in the future.

Third, we felt it was important to carefully consider the validity of the real-time paradigm. Our paradigm involved "quasi" real-time processing, where there were brief but non-zero delays in the frequency of checking for new real-time data, then additional brief delays in processing those data. (We felt that delays of 2 min or less were acceptable when seeking to identify a physiological condition that is unlikely to progress substantially in that time frame.) Overall, we validated that this integrated system was able to perform as intended, within a 2-min analysis latency. Obviously, if such real-time analyses become normative in healthcare, it is important to consider the "worst case scenario" in terms of latency for any decision support upon which tomorrow's clinicians grow to depend. An important corollary is that the latency and overall performance are a function of the computational complexity of the algorithms; a system that performs suitably with one set of algorithms may not perform well with a different set of algorithms.

Lastly, although largely out of the scope of the current report, it is important to consider the "meta-data" of the system. For instance, is there any risk of clock error, or of associating data with the wrong patient? An expanded discussion of these issues can be found in [12].

## 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] *Medical Device Plug-and-Play (MD PnP) Interoperability Program* [Online]. Available: http://www.mdpnp.org/uploads/Oct12_MD_PnP_White_Paper.pdf [accessed March 7 2014].

[2] M. Y. Khitrov, M. Rutishauser, K. Montgomery, A. T. Reisner, and J. Reifman, "A platform for testing and comparing of real-time decision-support algorithms in mobile environments," *Conf Proc IEEE Eng Med Biol Soc,* vol. 2009, pp. 3417-20, 2009.

[3] S. Liddle, L. Grover, R. Zhang, M. Khitrov, J. C. Brown, J. P. Cobb, J. Goldman, J. Chou, D. Yagoda, B. Westover, and A. T. Reisner, "Safety evaluation of a Medical Device Data System," *Conf Proc IEEE Eng Med Biol Soc,* vol. 2012, pp. 5899-902, 2012.

[4] *BedMasterEx Operators Manual*, Excel Medical Electronics Inc., Jupiter, FL, 2010.

[5] *The HDF Group. Hierarchical data format version 5, 2000-2014.* [Online]. Available: http://www.hdfgroup.org/HDF5 [accessed March 7 2014].

[6] 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.

[7] 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. 915-23, Oct 2008.

[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. 309-20, May-Jun 2006.

[9] 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, 2008.

[10] 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.

[11] J. B. Smith, Department of Defense Biotechnology High Performance Computing Software Applications Institute, Fort Detrick, MD, private communication, July 16 2013.

[12] 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, pp. 392-402, 2013.

American Journal of Emergency Medicine 33 (2015) 254–261

Contents lists available at ScienceDirect

American Journal of Emergency Medicine

journal homepage: www.elsevier.com/locate/ajem

Original Contribution

# Is heart rate variability better than routine vital signs for prehospital identification of major hemorrhage?☆,☆☆

Shwetha Edla, PhD [a], Andrew T. Reisner, MD [b], Jianbo Liu, PhD [a], Victor A. Convertino, PhD [c], Robert Carter III, PhD [c], Jaques Reifman, PhD [a,*]

[a] Department of Defense Biotechnology High Performance Computing Software Applications Institute, Telemedicine and Advanced Technology Research Center, US Army Medical Research and Materiel Command, Fort Detrick, MD 21702, USA
[b] Department of Emergency Medicine, Massachusetts General Hospital, Boston, MA 02114, USA
[c] US Army Institute of Surgical Research, Fort Sam Houston, TX 78234, USA

ARTICLE INFO

*Article history:*
Received 19 August 2014
Received in revised form 2 November 2014
Accepted 24 November 2014

ABSTRACT

*Objective:* During initial assessment of trauma patients, metrics of heart rate variability (HRV) have been associated with high-risk clinical conditions. Yet, despite numerous studies, the potential of HRV to improve clinical outcomes remains unclear. Our objective was to evaluate whether HRV metrics provide additional diagnostic information, beyond routine vital signs, for making a specific clinical assessment: identification of hemorrhaging patients who receive packed red blood cell (PRBC) transfusion.

*Methods:* Adult prehospital trauma patients were analyzed retrospectively, excluding those who lacked a complete set of reliable vital signs and a clean electrocardiogram for computation of HRV metrics. We also excluded patients who did not survive to admission. The primary outcome was hemorrhagic injury plus different PRBC transfusion volumes. We performed multivariate regression analysis using HRV metrics and routine vital signs to test the hypothesis that HRV metrics could improve the diagnosis of hemorrhagic injury plus PRBC transfusion vs routine vital signs alone.

*Results:* As univariate predictors, HRV metrics in a data set of 402 subjects had comparable areas under receiver operating characteristic curves compared with routine vital signs. In multivariate regression models containing routine vital signs, HRV parameters were significant ($P < .05$) but yielded areas under receiver operating characteristic curves with minimal, nonsignificant improvements ($+0.00$ to $+0.05$).

*Conclusions:* A novel diagnostic test should improve diagnostic thinking and allow for better decision making in a significant fraction of cases. Our findings do not support that HRV metrics add value over routine vital signs in terms of prehospital identification of hemorrhaging patients who receive PRBC transfusion.



## 1. Introduction

A series of investigations have suggested that measures of heart rate variability (HRV) offer a promising capability for the identification of trauma patients who require life-saving interventions (LSIs), which are time-sensitive clinical interventions, such as packed red blood cell (PRBC) transfusion, endotracheal intubation, and operative interventions. Heart rate variability, which can be measured via routine electrocardiography, represents the beat-to-beat fluctuations in the R-R intervals (RRIs) of the electrocardiogram (ECG), revealing the state of the patient's autonomic nervous system. A wide range of different HRV metrics have been investigated [1], including frequency domain metrics [2-7], time domain metrics [2,3,5-11], and complexity metrics [2-4,6,8,10,12].

In trauma patients, it is clear that, on average, those patients who subsequently require an LSI have reduced HRV during prehospital and emergency department (ED) monitoring [4,6,8,12]. There are also significant differences in HRV group averages between trauma patients with and without traumatic brain injury [7,11] and between survivors vs fatalities [2,3,5,7]. Moreover, diagnostic test characteristics have been encouraging, with 80% sensitivity and 75% specificity reported in patients who require surgical intervention in the operating room [9] and 86% sensitivity with 74% specificity reported in patients who

☆ Reprints: N/A.

☆☆ Conflicts of interest and sources of funding: None of the authors have any conflicts of interest to disclose. This work was supported by the US 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 US Army Medical Research and Materiel Command, Fort Detrick, MD, USA. The study sponsors did not have any role in the study design, data collection, analysis and interpretation of data, report writing, or the decision to submit the article for publication.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 US Army or of the US Department of Defense. This article has been approved for public release with unlimited distribution.

* Corresponding author at: Department of Defense Biotechnology High Performance Computing Software Applications Institute, Telemedicine and Advanced Technology Research Center, US Army Medical Research and Materiel Command, ATTN: MCMR-TT, 504 Scott St, Fort Detrick, MD 21702, USA. Tel.: +1 301 619 7915; fax: +1 301 619 1983.
*E-mail address:* jaques.reifman.civ@mail.mil (J. Reifman).

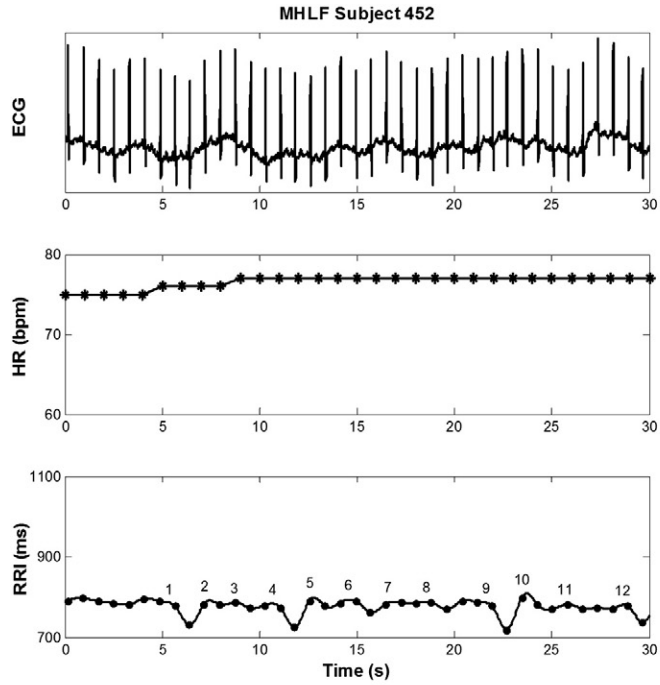
http://dx.doi.org/10.1016/j.ajem.2014.11.046

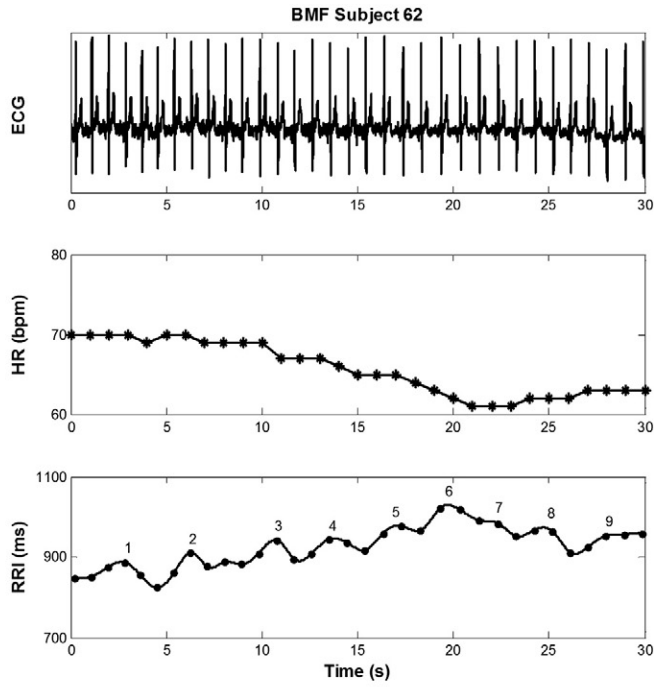
require any LSI [10]. However, these findings are tempered by several other reports, which suggest that, for that subset of trauma patients with normal vital signs, HRV metrics have a low sensitivity (16%) for LSI prediction [6], and their diagnostic potential is reduced by notable intersubject variability as well as intrasubject temporal variability [13].

To date, HRV monitoring has not become routine practice, although PubMed lists more than 10 000 citations relevant to HRV from over 3 decades, spanning a diversity of potential clinical applications. This suggests that there may be some barrier (eg, economic, regulatory, educational, etc) that is hampering the dissemination of a potentially useful technology. Alternatively, it may be that the aforementioned research studies have been suboptimal in terms of answering precisely how (or if) HRV can improve patient care. Many of the published reports about HRV offer intriguing associations but do not provide explicit comparisons vs the routine clinical data used in standard decision making. For instance, if HRV is to be used in deciding whether a trauma patient requires trauma center care, it may be elucidating to compare it against standard criteria for trauma center transport [14]. Likewise, if HRV is to be used for diagnosing traumatic brain injury, it could be compared against standard criteria for neuroimaging after head injury, for example, the Canadian head computed tomography rule [15].

To better understand the value of HRV for decision making, we decided to focus on the identification of trauma patients with major hemorrhage who receive PRBC transfusion because exsanguination is a leading cause of death in both civilian [16] and military [17] trauma populations, whereas many hemorrhagic deaths can be prevented with time-sensitive interventions such as surgery and optimal resuscitation [18,19]. In theory, a reliable and simple diagnostic indicator of which patients require such interventions could enhance the quality and efficiency of clinical decision making, leading to optimal patient outcomes. Fig. 1 illustrates 2 cases in which the patients' vital signs are similar, but HRV metrics indicate whether or not the patients are suffering life-threatening hemorrhage.

To this end, we conducted a multivariate analysis, using routine vital signs as the comparator, to test the hypothesis that HRV metrics can improve the identification of patients with major hemorrhage. By focusing




| Subject characteristics | Normal range |
|---|---|
| HRV metrics | |
| RSA (cpm): 23 | 19-26 |
| SDNN (ms): 37.15 | 23-53 |
| SampEn: 0.48 | 0.9-1.4 |
| Vital signs | |
| HR (bpm): 77 | 77-105 |
| SBP (mmHg): 100 | 120-147 |
| Injury description: Grade II liver laceration, Grade IV spleen laceration | |
| 24-hr PRBC volume: 13 | |

| Subject characteristics | Normal range |
|---|---|
| HRV metrics | |
| RSA (cpm): 18.68 | 19-26 |
| SDNN (ms): 51.88 | 23-53 |
| SampEn: 1.51 | 0.9-1.4 |
| Vital signs | |
| HR (bpm): 66 | 77-105 |
| SBP (mmHg): 99 | 120-147 |
| Injury description: Superficial spleen laceration, thoracic cavity injury with hemo- and pneumo-thorax | |
| 24-hr PRBC volume: 0 | |

**Fig. 1.** The 2 cases—30-second excerpts of ECG, HR, and RRI waveforms from 2 different subjects—are selected examples where HRV metrics, but not routine vital signs, can differentiate between patients with (left) and without (right) hemorrhagic injuries requiring substantial 24-hour PRBC transfusion. For each subject, the RRI waveform is illustrated, along with each cycle of sinus arrhythmia that was identified by computer algorithm (each cycle indicated by numerals above the RRI waveform); see text for more details about computation of HRV metrics. The "normal ranges" listed in the tables above represent the interquartile range for subjects who did not receive any 24-hour PRBC transfusion.

on a specific clinical condition of clear importance, that is, substantial hemorrhage after injury, and quantitatively comparing HRV metrics vs routine vital signs as diagnostic tests, it may be possible to better understand if and how HRV metrics may be used to improve trauma patient management.

## 2. Materials and methods

### 2.1. Clinical data collection

We examined 2 pooled datasets, the first originally collected on board Memorial Hermann Life Flight (MHLF, Houston, TX) air ambulances [5,20] between August 2001 and April 2004 and the second from Boston Medflight (BMF, Bedford, MA) air ambulances between February 2010 and December 2012 with institutional review board approval. Routine vital sign and ECG data sourced from Propaq 206 patient monitors (Welch-Allyn, Beaverton, OR) were acquired from adult (age ≥18 years) trauma patients en route to level 1 trauma centers and ultimately archived in our database. Additional clinical data, including demographics, injury descriptions, prehospital interventions, hospital treatments, etc., were obtained via retrospective chart review.

We studied all subjects with at least 1 reliable measurement of each investigational metric, allowing for a meaningful comparison of the investigational metrics (see below for definition of measurement reliability). Subjects who died before hospital admission were excluded because it was difficult to determine what volume of blood transfusion they would have received within 24 hours (or, in some cases, whether they were truly bleeding or not). For the primary analysis, we excluded patients who received PRBC transfusion but lacked explicitly hemorrhagic injuries, that is, no documented solid organ injury, no thoracic or abdominal hematoma, and no vascular injury requiring a procedure for hemostasis. (We reexamined these patients in a secondary sensitivity analysis to determine whether the major findings of the primary analysis were different for the excluded population.)

### 2.2. Routine vital signs

We studied the average of reliable vital signs (heart rate [HR], respiratory rate [RR], systolic blood pressure [SBP], and pulse pressure [PP = SBP − diastolic blood pressure]) measured up to the 15th minute of each subject's prehospital data record. The reliability of each vital sign was determined using automated computer algorithms [21-23]. The HR reliability algorithm [23] evaluated whether the ECG waveform was clean, whether the heart rhythm was regular, and whether the Propaq HR was close in value to the algorithm's independent computation of HR. The RR reliability algorithm [21] evaluated whether the impedance pneumogram waveform was clean, whether the breaths were regular, and whether the Propaq RR was close in value to the algorithm's independent computation of RR. The blood pressure reliability algorithm [22] evaluated whether the relationship between SBP, mean arterial pressure, and diastolic blood pressure was normative and whether the HR measured from the oscillometric cuff was close to the HR measured by the ECG.

### 2.3. Heart rate variability metrics

We studied the average value of 3 reliable HRV metrics (SD of the RRIs in the ECG signal [SDNN], sample entropy [SampEn], and rate of sinus arrhythmia [RSA]) measured up to the 15th minute of each subject's prehospital data record. SDNN [8-11] and SampEn [2-4,6,8,10,12] have been investigated in recent reports, whereas RSA offered encouraging performance in prior exploratory analysis.

To compute SDNN, we upsampled each ECG segment to 2000 Hz by cubic spline interpolation and identified the location of each R-wave using an HR estimation algorithm [23]. We computed the difference between successive R-waves, which established the RRI time series. Fig. 1 shows examples of these RRI time series. For every second of ECG recorded, we computed SDNN: the SD of RRIs from the preceding 5 minutes. The computed SDNN was considered reliable when the preceding 5 minutes of ECG waveforms were at least 80% clean and reliable, per the ECG waveform reliability algorithm [23]. We used a 5-minute window for SDNN calculation in accordance with consensus guidelines [24]. When the change in RRI from one beat to the next was too large or too small (as per the quantitative criteria of Malik et al [25]), that beat was considered aberrant. R-R intervals from the interval immediately before or immediately after the aberrant beat were excluded whenever SDNN was computed.

To compute SampEn, which is a measure of similarity within the RRI time series, we used the PhysioTools software "sampen.m" [26], which implements the method of Richman and Moorman [27]. Sample entropy is the probability that, if an RRI time series has a repeated "similar" pattern of data points of length $m$ (where $m \ll N$), then the similarity will also persist when the length of data points is extended to $m + 1$. Similarity is defined mathematically, that is, when any 2 sequences of data points have the same data point values in the same order within some tolerance $r$. A detailed explanation of this calculation can be found in the online PhysioTools tutorial [28]. In this work, for every second of ECG recorded, we computed SampEn from the preceding 200 ECG beats (equivalent to $N = 201$), using $r = 0.20$ times the SD of the RRI series, and $m = 2$. The computed SampEn values were considered reliable only if all the 200 ECG beats were reliable (as per the ECG waveform reliability algorithm [23]) and without any aberrant beats (defined above). The values of $N$, $m$, and $r$ were selected in accordance with several recent reports evaluating SampEn in trauma patients [3,10,12].

Lastly, we computed RSA, which is the frequency of oscillation of the HR (HR typically varies in a rhythmic fashion, often synchronized to the rate of respiration [29], although sometimes faster [30] or slower [31] than respiration). Fig. 1 shows examples of the oscillatory RSA. For computational purposes, we treated the RRI time series as a form of respiratory waveform [29] and applied our previously developed RR measurement and reliability algorithms [21] to compute RSA for every second and to determine whether the waveform was reliable or not.

### 2.4. Univariate analysis

We analyzed the association between each investigational metric (HR, RR, SBP, PP, SDNN, SampEn, and RSA) vs PRBC transfusion received over 24 hours. Specifically, we computed the area under the receiver operating characteristic curve (ROC AUC) for each investigational metric as a predictor of different 24-hour PRBC volumes (24-hour PRBC vol: ≥1, ≥5, and ≥9 units). We compared each of the HRV metrics (SDNN, SampEn, and RSA) vs routine vital signs (HR, RR, SBP, and PP), testing whether there were any differences per DeLong's test [32] with a significance threshold of $P < .05$.

### 2.5. Multivariate analysis

We conducted multivariate logistic regression analysis using the "glmfit" routine in MATLAB version 7 (The Mathworks, Inc, Natick, MA). First, we evaluated a baseline multivariate model containing all routine vital signs (core feature set: HR, RR, SBP, and PP) and compared this model vs other models that included an HRV metric and/or lacked one of the routine vital signs. For each model, we determined which input parameters were statistically significant, and we computed ROC AUCs for the same outcomes as the univariate analysis (ie, ≥1, ≥5, and ≥9 units of 24-hour PRBC vol).

### 2.6. Net reclassification improvement

We tested whether the HRV metrics were associated with a significant net reclassification improvement (NRI), using the statistical method of Pencina et al [33]. First, we computed the probability of hemorrhage given pairs of logistic regression models (baseline model with different combinations of routine vital signs vs a model with the

same set of vital signs plus an HRV metric) for the same outcomes as the univariate analysis (ie, ≥1, ≥5, and ≥9 units of 24-hour PRBC vol). For each subject, we assessed which model gave an "improved classification" (defined as a higher probability of hemorrhage in hemorrhage patients or a lower probability of hemorrhage in control patients). Then we evaluated whether one model was significantly different from the other using the z-test (the null hypothesis was that each model had an equal likelihood of improved classification).

### 2.7. Sensitivity analysis

Many subjects were excluded from the primary analyses because they lacked a complete set of reliable investigational metrics within their prehospital physiological data. To check whether this led to notable selection bias, we repeated the univariate analysis on a broader set of subjects to determine whether the univariate findings were sensitive to the exclusion criteria. For each investigational metric, we identified all subjects with at least 1 reliable value (subjects who did not necessarily have a complete set of reliable investigational metrics). We then computed the univariate ROC AUC of each metric for these larger populations to predict 24-hour PRBC vol (ie, ≥1, ≥5, and ≥9 units of 24-hour PRBC vol). However, we could not perform paired comparisons of these ROC AUCs because each result arose from somewhat different subject subsets.

In addition, we repeated the multivariate analysis for our 2 populations, MHLF and BMF. For each, we computed the ROC AUC for the core feature set (HR, RR, SBP, and PP) with and without the investigational HRV metrics: SDNN, SampEn, and RSA.

We also repeated the univariate and multivariate analyses for an alternative outcome, namely, subjects who received 24-hour PRBC vol greater than or equal to 1, greater than or equal to 5, and greater than or equal to 9 units and did not necessarily have explicitly hemorrhagic injuries.

## 3. Results

We had a total of 999 patients in the overall database (subjects with at least 1 routine vital sign from the Propaq 206 monitors), from which 402 patients composed the primary study population. We excluded the following:

1. 43 patients in whom the presence and extent of hemorrhagic injury was unknowable because of death during transport or before being admitted to the hospital,
2. 90 patients who received PRBC transfusion without an explicitly hemorrhagic injury (these 90 patients were reincluded and analyzed in the sensitivity analysis), and
3. 464 patients in whom a paired comparison could not be performed because the patients lacked a complete set of all vital signs during their initial 15 minutes of transport (these 464 patients were reincluded and analyzed in the sensitivity analysis).

Table 1 shows the overall database and the study population characteristics. Most of the differences between the overall database and the study population were minor, except for a slightly higher overall mortality rate in the overall database.

Fig. 2 displays the distributions of all investigational metrics (also see Table A.1). Table 2 reports the univariate ROC AUCs of the basic vital signs and investigational HRV metrics for the identification of 24-hour PRBC vol greater than or equal to 1, greater than or equal to 5, and greater than or equal to 9 units. We observed that both the HRV metrics (ROC AUCs, 0.60-0.79) and routine vital signs (ROC AUCs, 0.65-0.79) had statistically significant discriminatory power, but none of the 3 HRV metrics were significantly superior to any of the routine vital signs.

Of the investigational HRV metrics, RSA yielded the highest univariate ROC AUCs. Table 3 shows the multivariate analysis testing whether RSA provided significant independent information above and beyond routine vital signs. When added to multivariate logistic regression models that included the core feature set (and subsets of the core feature set), RSA was found to be a significant, independent predictor of 24-hour PRBC transfusion. However, the resultant improvements in ROC AUCs when RSA was added to the core feature set, and its subsets were minor, and neither improvements in ROC AUCs nor NRIs were statistically significant.

We also performed the same multivariate analysis using SDNN and SampEn (with the same feature sets listed in the first column of Table 3, but using SDNN and SampEn in place of RSA). We found that SDNN was a significant, independent predictor of PRBC transfusions only in the model that consisted of RR, SBP, PP, and SDNN. Similarly, SampEn was a significant, independent predictor of PRBC transfusion only in the model that consisted of RR, SBP, PP, and SampEn. When

**Table 1**
Characteristics of the overall database and the study population

| | Overall database | | Study population | |
|---|---|---|---|---|
| | MHLF | BMF | MHLF | BMF |
| Population, n | 757 | 242 | 273 | 129 |
| Male, n (%) | 562 (74%) | 179 (74%) | 207 (76%) | 97 (75%) |
| Female, n (%) | 195 (26%) | 63 (26%) | 66 (24%) | 32 (25%) |
| Age, years, mean (SD)[a] | 38 (15) | 47 (21) | 37 (14) | 43 (19) |
| Mechanism of injury | | | | |
| Blunt, n (%) | 664 (88%) | 216 (89%) | 238 (87%) | 118 (92%) |
| Penetrating, n (%) | 84 (11%) | 26 (11%) | 30 (11%) | 11 (9%) |
| Hospital transfer, n (%) | 0 (0%) | 118 (49%) | 0 (0%) | 65 (50%) |
| Prehospital airway intubation, n (%) | 165 (22%) | 97 (40%) | 52 (19%) | 51 (40%) |
| ISS, median (IQR)[b] | 17 (9-34) | 17 (9-26) | 13 (8-34) | 17 (9-26) |
| Prehospital GCS, median (IQR)[c] | 15 (12-15) | 15 (6-15) | 15 (13-15) | 15 (5-15) |
| Prehospital fluid volume, mL, median (IQR)[d] | 300 (100-628) | 100 (50-250) | 300 (100-600) | 100 (50-200) |
| 24-h PRBC transfusion volumes | | | | |
| 24-h PRBC vol ≥1 unit, n (%) | 153 (20%) | 62 (26%) | 38 (14%) | 16 (12%) |
| 24-h PRBC vol ≥9 units, n (%) | 36 (5%) | 11 (5%) | 11 (4%) | 5 (4%) |
| Overall mortality, n (%) | 85 (11%) | 28 (12%) | 12 (4%) | 9 (7%) |
| Died before admission to ED, n (%) | 36 (42%) | 7 (25%) | 0 (0%) | 0 (0%) |
| Died after admission to ED, n (%) | 49 (58%) | 21 (75%) | 12 (100%) | 9 (100%) |

The overall database consists of all subjects who had at least 1 available routine vital sign from the Propaq 206 monitor. See text for details about the study population. Abbreviations: *GCS*, Glasgow Coma Scale; *IQR*, interquartile range; *ISS*, injury severity score.

[a] No age information available for 6 patients in the overall database and 3 patients in the study population.

[b] No injury severity score information available for 186 patients in the overall database and 64 patients in the study population.

[c] No Glasgow Coma Scale information available for 75 patients in the overall database and 29 patients in the study population.

[d] No prehospital fluid volume information available for 36 patients in the overall database and 17 patients in the study population.

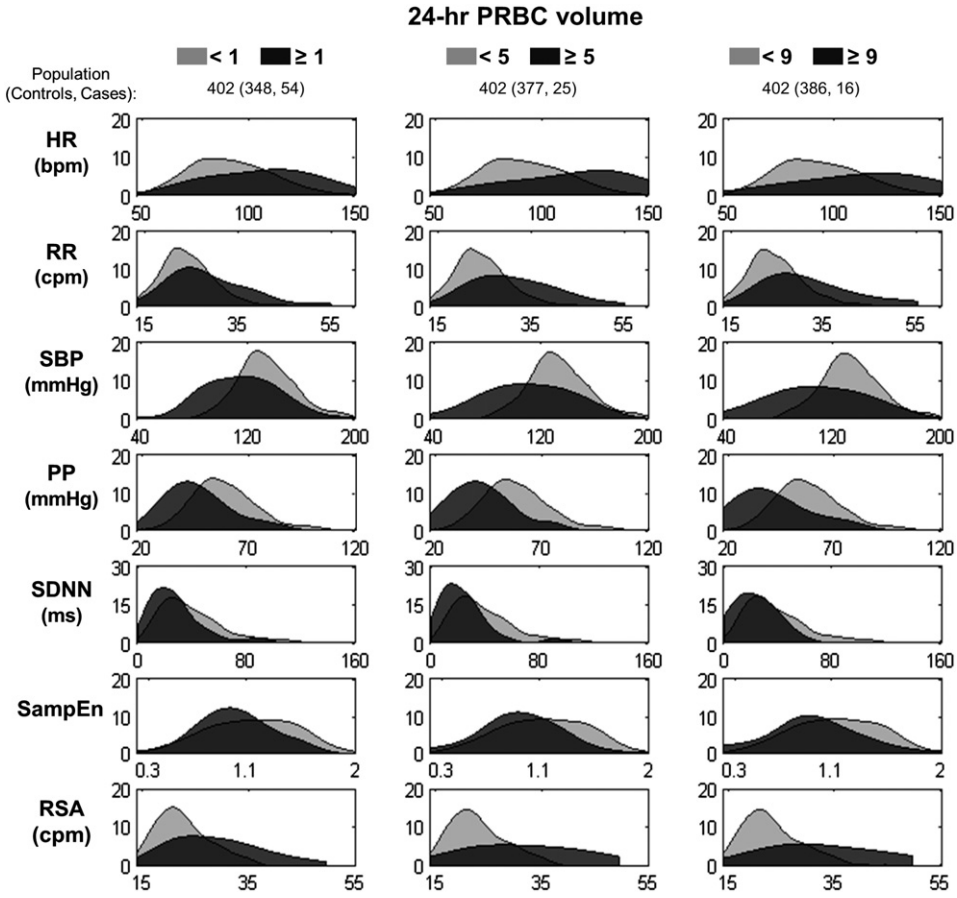



**Fig. 2.** Distributions of routine vital signs and HRV metrics for trauma patients grouped by different 24-hour PRBC volumes. The medians and interquartile ranges of the distributions are provided in Table A.1.

SDNN and SampEn were separately added to the multivariate models together with the core feature set (or subsets of the core feature set), the resultant improvements in ROC AUCs were minor (+0.03 or lesser), and neither improvements in ROC AUCs nor NRIs were statistically significant.

**Table 2**
Areas under the receiver operating characteristic curves and 95% confidence intervals of routine vital signs and HRV metrics (univariate performance) for predicting 24-hour PRBC volume

| | 24-h PRBC volume | | |
|---|---|---|---|
| Features | ≥1 | ≥5 | ≥9 |
| Population (controls, cases) | 402 (348, 54) | 402 (377, 25) | 402 (386, 16) |
| Routine vital signs | | | |
| HR | 0.68 (0.59-0.76) | 0.74 (0.59-0.84) | 0.72 (0.53-0.85) |
| RR | 0.65 (0.56-0.73) | 0.74 (0.63-0.83) | 0.73 (0.53-0.84) |
| SBP | 0.70 (0.61-0.78) | 0.72 (0.58-0.82) | 0.73 (0.55-0.86) |
| PP | 0.74 (0.65-0.81) | 0.79 (0.68-0.88) | 0.79 (0.61-0.90) |
| HRV metrics | | | |
| SDNN | 0.67 (0.59-0.75) | 0.72 (0.61-0.82) | 0.71 (0.57-0.82) |
| SampEn | 0.60 (0.53-0.68)[a] | 0.63 (0.52-0.73)[a] | 0.62 (0.46-0.75) |
| RSA | 0.72 (0.64-0.79) | 0.76 (0.62-0.85) | 0.79 (0.64-0.89) |

The 3 HRV metrics are compared to each of the routine vital signs for significant differences ($P < .05$) using DeLong's test.

[a] Area under the receiver operating characteristic curves is significantly different from the PP ROC AUC.

**Table 3**
Areas under the receiver operating characteristic curves and 95% confidence intervals of the multivariate logistic regression models consisting of different combinations of routine vital signs and rate of sinus arrhythmia for predicting 24-hour PRBC volume

| | 24-h PRBC volume | | |
|---|---|---|---|
| Feature set description (features) | ≥1 | ≥5 | ≥9 |
| Population (controls, cases) | 402 (348, 54) | 402 (377, 25) | 402 (386, 16) |
| Core feature set (HR[a], RR[a], SBP, PP) | **0.79 (0.70-0.85)** | **0.85 (0.73-0.92)** | **0.86 (0.73-0.94)** |
| Core feature set + RSA (HR, RR[a], SBP, PP, RSA[a]) | +0.00 | +0.01 | +0.02 |
| Core feature set − HR (RR[a], SBP, PP[a]) | +0.00 | +0.01 | +0.00 |
| Core feature set − HR + RSA (RR[a], SBP, PP, RSA[a]) | +0.00 | +0.01 | +0.02 |
| Core feature set − RR (HR[a], SBP, PP) | −0.02 | −0.03 | −0.05 |
| Core feature set − RR + RSA (HR, SBP, PP, RSA[a]) | +0.00 | −0.01 | +0.00 |
| Core feature set − (SBP, PP) (HR[a], RR[a]) | −0.10[b] | −0.07[b] | −0.09 |
| Core feature set − (SBP, PP) + RSA (HR, RR[a], RSA[a]) | −0.06 | −0.07[b] | −0.05 |

The bold numbers in the first row show the performance of the core feature set in terms of ROC AUCs and the 95% confidence intervals. The subsequent rows represent the relative change in ROC AUC with respect to that of the core feature set.
Note: Findings for the SDNN and SampEn were similar; see Results section for details.

[a] The coefficient of the corresponding feature is significantly different from zero ($P < .05$) in at least 1 of the models for predicting 24-hour PRBC volume greater than or equal to 1, greater than or equal to 5, or greater than or equal to 9 units.

[b] Area under the receiver operating characteristic curves is significantly different ($P < .05$) from the core feature set ROC AUC by DeLong's test.

### 3.1. Sensitivity analysis

We performed several sensitivity analyses, to test whether our exclusion criteria affected our findings. Here, we summarize the findings (detailed results are provided in Appendix A under *Sensitivity analysis*).

We repeated the univariate analysis on a broader set of subjects (subjects who did not necessarily have a complete set of reliable vital signs and HRV metrics). Compared with the primary results (ie, Table 2), there were neither any significant changes nor notable trends.

We repeated the primary multivariate analysis for 2 subpopulations (MHLF vs BMF) for greater than or equal to 1, greater than or equal to 5, and greater than or equal to 9 units of 24-hour PRBC vol. When we added RSA to the core feature set (SBP, PP, HR, and RR), respective increases in ROC AUCs were +0.00, +0.00, and +0.02 in MHLF and +0.01, +0.05, and +0.03 in BMF. When we added SDNN to the core feature set, respective increases in ROC AUCs were +0.00, +0.01, and +0.00 in MHLF and +0.00, +0.01, and +0.01 in BMF. When we added SampEn to the core feature set, respective increases in ROC AUC were +0.00, +0.00, and +0.00 in MHLF and +0.00, +0.03, and +0.00 in BMF. Overall, increases in ROC AUCs were very similar in MHLF vs BMF.

We also repeated the primary analysis with an alternative outcome definition: subjects who received PRBC transfusions whether they had explicitly hemorrhagic injuries. As in the primary analysis (ie, Table 3), improvements in the ROC AUCs were minimal after adding RSA, SampEn, or SDNN to the core feature set or its subsets (ROC AUC improvements were +0.02 or less).

## 4. Discussion

After a life-threatening injury, some trauma casualties may temporarily evidence normal vital signs, belying the severity of their condition. This motivated the substantial interest in HRV metrics as indexes of cardiovascular stability for trauma patients, to better distinguish between patients who require time-sensitive interventions vs those with less acute conditions.

Our analysis of prehospital vital signs demonstrated that before hospital arrival, many patients with substantial bleeding (defined by a large 24-hour PRBC vol) had abnormal vital signs consistent with hypovolemia: tachycardia, tachypnea, reduced SBP, and reduced PP (ie, reduced stroke volume). Multivariate analysis allowed for very good separation between patients with and without substantial bleeding (ie, ROC AUC, 0.86 in Table 3). Heart rate variability metrics of autonomic tone were also significantly different from controls in many patients with substantial hemorrhage. However, when combined with routine vital signs, HRV added negligible additional discriminatory value (see Table 3). This finding may indicate that discriminatory changes in HRV and changes in standard vital signs develop at similar stages during progressive hemorrhage.

In theory, there should be compensatory changes in the autonomic system during the very earliest stages of the response to serious injury. Indeed, population averages of HRV indexes have been shown to correlate with central blood volume loss in animal hemorrhage experiments [34] and hypovolemia in human lower body negative-pressure studies [13,35-37]. Clinically, significant group differences in HRV metrics have been reported between trauma patients who require LSIs and those who do not [4,6,8,12] and between survivors and fatalities [2,3,5,7]. In terms of discriminatory power, our own findings suggest that HRV metrics are comparable to routine vital signs, in terms of their possible utility for identifying substantial bleeding.

At the same time, there are physiological reasons why HRV metrics might not add discriminatory value above and beyond routine vital signs. First of all, vital signs include HR, which alone provides some basic measure of the autonomic system; that is, tachycardia represents sympathetic activation, whereas bradycardia represents parasympathetic dominance. Although HRV metrics represent a more nuanced quantification of the sympathetic and parasympathetic states, it is worth noting that the autonomic system is highly sensitive to physiologic stimuli. For instance, performing mental arithmetic has been shown to alter HRV metrics [38]. In theory, such sensitivity to disparate stimuli might confound the association between HRV and hemorrhage. Previous studies suggest that the complexity of interindividual and intraindividual variability in autonomic compensatory responses weakens the association between HRV metrics and blood loss and weakens their potential diagnostic value [13].

In terms of specific HRV metrics, we studied 2, which have been the focus of other trauma reports: SDNN [8-11] and SampEn [2-4,6,8,10,12]. We also studied RSA, which we previously found offered encouraging performance. Ectopic beats, transient events (ie, nonstationary signal), motion artifacts, and length of data acquisition are technical factors that can affect these HRV calculations [6,37], and we used previously validated algorithms [23] to exclude unreliable segments of ECG (ie, either noisy or with ectopic beats). Note that we did not study frequency domain metrics, which are less robust to some of the aforementioned factors affecting HRV calculations and are likely impractical for trauma patient monitoring [24,34,39].

It is worth noting that time averaging of HRV and vital signs, as we did in our analysis, reduced the effects of temporal variability and, therefore, may have increased overall diagnostic performance [40]. Time averaging likely represents a "best case" for vital signs and HRV metrics because, in practice, clinicians do not use time-averaged parameters, and episodic fluctuations can result in misleading vital sign patterns [41]. As a point of comparison, Zarzaur et al [42] reported that a single isolated measurement of SBP and HR (ie, the Shock Index) yielded an ROC AUC of 0.78 for predicting greater than or equal to 4 units of blood in 48 hours. Moreover, there is room for improvement: at the 90% sensitivity operating point of our receiver operating characteristic curve for vital signs alone (multivariate model using HR, RR, SBP, and PP), specificity was only 40%.

In terms of limitations, it is possible that HRV may be valuable for other clinical applications or that our findings may not be generalizable to alternative HRV metrics (beyond those studied in this report). However, our study design, whereby HRV metrics were directly compared to routine clinical data for assessing diagnostic thinking efficacy, remains relevant, with the potential to enhance future HRV investigations. Second, HRV metrics can be affected by disparate factors [38], and it is possible that another data set may offer significantly different findings. However, inconsistent findings in different data sets, due to HRV's established sensitivity to confounding effects, would be another reason for caution about HRV in trauma care.

There are 2 primary implications of this research. First, from a clinical standpoint, our findings do not support that HRV metrics add value over routine vital signs, in terms of prehospital identification of substantial bleeding. Given a multivariate regression model, the HRV metrics added negligible diagnostic value. Moreover, clinicians are unlikely to weigh the information from HRV as carefully as this multivariate model, and there is some theoretical risk to having incorrect decision making, that is, some clinicians might be overreliant on HRV metrics rather than routine vital signs.

The second implication relates to research methodology. By way of background, Pearl [43] described a 7-tier hierarchical approach to evaluating diagnostic testing. The type of analysis in the current report—directly comparing HRV to routine vital signs—corresponds to Pearl's third tier "diagnostic thinking efficacy," which includes the "percentage of cases in which the final diagnosis changed after testing." What is notable among HRV clinical investigation is a scarcity of comparisons against standard criteria for decision making, for example, standard criteria for trauma center transport [14] or standard criteria for neuroimaging after head injury [15]. Arguably, there would be a better understanding of the appropriate role of HRV in clinical medicine if a larger proportion of the 10 000 HRV citations currently listed by PubMed focused on Pearl's third or higher tiers of evaluation.

## 5. Conclusions

We investigated whether HRV was useful for the identification of trauma patients who require blood transfusion. Heart rate variability metrics were comparable to routine vital signs in univariate analysis. However, in multivariate analysis, HRV metrics did not significantly improve diagnostic performance. Our findings do not support that HRV would improve today's standard care for this clinical application.

## Appendix A. Sensitivity analysis

We repeated the univariate analysis on a broader set of subjects (subjects who did not necessarily have a complete set of reliable vital signs and HRV metrics). Table A.2 shows the results. The ROC AUCs in this secondary population were similar to the primary analysis in terms of the relative performance of the HRV metrics vs routine vital signs. There were neither any significant changes nor notable trends. All ROC AUCs for this secondary analysis were within the 95% confidence intervals of the primary analysis (see Table 2).

We also repeated the primary analysis with an alternative outcome definition: subjects who received PRBC transfusions whether they had explicitly hemorrhagic injuries. As before, we excluded the subjects who died during transport and those who did not have reliable investigational metrics. The findings were similar to the primary analysis, with all the ROC AUCs within the 95% confidence intervals of the primary analysis with the following exceptions: the ROC AUC corresponding to RR for the prediction of 24-hour PRBC vol greater than or equal to 1 unit was 0.54; the ROC AUC corresponding to RSA for predicting 24-hour PRBC vol greater than or equal to 1 unit was 0.63. As in the primary analysis, none of the HRV metrics were significantly better as univariate predictors of 24-hour PRBC transfusions than routine vital signs (with one exception: RSA was significantly superior to RR for the prediction of 24-hour PRBC vol ≥1 unit, but not for ≥5 or ≥9 units).

As in the primary multivariate analysis, RSA was significant in all multivariate models that included the core feature set (and subsets of the core feature set) for predicting 24-hour PRBC transfusion. However, SDNN was not a significant predictor of PRBC transfusions in any of the multivariate models. SampEn was significant when included in the model that consisted of RR, SBP, and PP as in the primary analysis and in the model that consisted of the core feature set, unlike the primary analysis. Regardless, improvements in the aforementioned ROC AUCs were minimal after adding RSA, SDNN, or SampEn to the core feature set or its subsets (ROC AUC improvement was +0.03 or less).

**Table A.2**
Areas under the receiver operating characteristic curves and 95% confidence intervals of routine vital signs and heart rate variability metrics for predicting 24-hour packed red blood cell volume

| | 24-h PRBC volume | | |
|---|---|---|---|
| Features | ≥1 | ≥5 | ≥9 |
| Population (controls, cases) | 797 (698, 99) | 797 (749, 48) | 797 (765, 32) |
| HR | 0.66 (0.60-0.72) | 0.71 (0.61-0.79) | 0.70 (0.58-0.80) |
| Population (controls, cases) | 508 (439, 69) | 508 (474, 34) | 508 (484, 24) |
| RR | 0.61 (0.53-0.69) | 0.66 (0.55-0.75) | 0.65 (0.52-0.76) |
| Population (controls, cases) | 837 (736, 101) | 837 (788, 49) | 837 (808, 29) |
| SBP | 0.75 (0.69-0.80) | 0.78 (0.70-0.84) | 0.76 (0.65-0.84) |
| Population (controls, cases) | 837 (736, 101) | 837 (788, 49) | 837 (808, 29) |
| PP | 0.76 (0.70-0.81) | 0.81 (0.73-0.87) | 0.79 (0.68-0.87) |
| Population (controls, cases) | 563 (488, 75) | 563 (528, 35) | 563 (540, 23) |
| SDNN | 0.66 (0.59-0.73) | 0.67 (0.56-0.76) | 0.61 (0.47-0.74) |
| Population (controls, cases) | 522 (453, 69) | 522 (490, 32) | 522 (502, 20) |
| SampEn | 0.58 (0.51-0.65) | 0.62 (0.52-0.72) | 0.65 (0.50-0.77) |
| Population (controls, cases) | 668 (579, 89) | 668 (626, 42) | 668 (641, 27) |
| RSA | 0.72 (0.65-0.77) | 0.75 (0.66-0.82) | 0.78 (0.68-0.86) |

Shown above are the univariate performance results for a secondary sensitivity analysis using less restrictive inclusion criteria: subjects with at least 1 reliable value for each investigational metric. Unlike the primary study population, this population did not necessarily have a full set of every reliable investigational metric. Abbreviations: *HR*, heart rate; *PP*, pulse pressure = systolic blood pressure (SBP) − diastolic blood pressure; *PRBC*, packed red blood cells; RR, respiratory rate; *RSA*, rate of sinus arrhythmia; *SampEn*, sample entropy; *SDNN*, SD of the normal R-R intervals in the electrocardiogram signal.

## References


[1] Ryan ML, Thorson CM, Otero CA, Vu T, Proctor KG. Clinical applications of heart rate variability in the triage and assessment of traumatically injured patients. Anesthesiol Res Pract 2011;2011:416590–7.
[2] Batchinsky AI, Cancio LC, Salinas J, Kuusela T, Cooke WH, Wang JJ, et al. Prehospital loss of R-to-R interval complexity is associated with mortality in trauma patients. J Trauma 2007;63:512–8.
[3] Batchinsky AI, Salinas J, Kuusela T, Necsoiu C, Jones J, Cancio LC. Rapid prediction of trauma patient survival by analysis of heart rate complexity: impact of reducing data set size. Shock 2009;32:565–71.
[4] Cancio LC, Batchinsky AI, Salinas J, Kuusela T, Convertino VA, Wade CE, et al. Heart-rate complexity for prediction of prehospital lifesaving interventions in trauma patients. J Trauma 2008;65:813–9.
[5] Cooke WH, Salinas J, Convertino VA, Ludwig DA, Hinds D, Duke JH, et al. Heart rate variability and its association with mortality in prehospital trauma patients. J Trauma 2006;60:363–70.
[6] Rickards CA, Ryan KL, Ludwig DA, Convertino VA. Is heart period variability associated with the administration of lifesaving interventions in individual prehospital trauma patients with normal standard vital signs? Crit Care Med 2010;38:1666–73.


**Table A.1**
Median and interquartile range of routine vital signs, heart rate variability metrics, injury severity score, and Glasgow Coma Scale for trauma patients grouped by different 24-hour packed red blood cell transfusion volumes

| | 24-h PRBC volume | | | | | |
|---|---|---|---|---|---|---|
| Features | <1 | ≥1 | <5 | ≥5 | <9 | ≥9 |
| HR (bpm) | 90 (77-105) | 108 (85-125) | 90 (77-105) | 117 (88-135) | 91 (77-106) | 116 (88-132) |
| RR (cpm) | 24 (20-27) | 26 (23-35) | 24 (20-28) | 30 (24-37) | 24 (20-28) | 30 (24-37) |
| SBP (mmHg) | 132 (120-147) | 115 (92-134) | 131 (119-146) | 106 (85-134) | 131 (119-146) | 106 (82-134) |
| PP (mmHg) | 57 (48-67) | 44 (34-55) | 56 (47-67) | 40 (31-50) | 56 (47-66) | 39 (27-51) |
| SDNN (ms) | 34 (23-53) | 25 (14-33) | 33 (22-53) | 21 (11-30) | 33 (22-53) | 22 (10-33) |
| SampEn | 1.2 (0.9-1.4) | 1.0 (0.8-1.2) | 1.1 (0.9-1.4) | 1.0 (0.8-1.2) | 1.1 (0.9-1.4) | 0.9 (0.7-1.3) |
| RSA (cpm) | 22 (19-26) | 29 (22-34) | 22 (19-27) | 32 (23-40) | 22 (19-27) | 32 (25-41) |
| ISS[a] | 13 (8-25) | 27 (17-43) | 14 (9-26) | 29 (21-52) | 14 (9-26) | 41 (28-66) |
| GCS[b] | 15 (12-15) | 15 (6-15) | 15 (12-15) | 15 (8-15) | 15 (12-15) | 15 (4-15) |

Population (controls, cases) for 24-hour PRBC volume greater than or equal to 1, 5, and 9 units: 402 (348, 54), 402 (377, 25), 402 (386, 16), respectively. Abbreviations: *GCS*, Glasgow Coma Scale; *HR*, heart rate; *ISS*, injury severity score; *PP*, pulse pressure = systolic blood pressure (SBP) − diastolic blood pressure; PRBC, packed red blood cells; *RR*, respiratory rate; *RSA*, rate of sinus arrhythmia; *SampEn*, sample entropy; *SDNN*, SD of the normal R-R intervals in the electrocardiogram signal.

[a] No ISS information available for 64 of the 402 patients in the study population.
[b] No GCS information available for 29 of the 402 patients in the study population.

[7] Ryan ML, Ogilvie MP, Pereira BM, Gomez-Rodriguez JC, Manning RJ, Vargas PA, et al. Heart rate variability is an independent predictor of morbidity and mortality in hemodynamically stable trauma patients. J Trauma 2011;70:1371–80.
[8] Cancio LC, Batchinsky AI, Baker WL, Necsoiu C, Salinas J, Goldberger AL, et al. Combat casualties undergoing lifesaving interventions have decreased heart rate complexity at multiple time scales. J Crit Care 2013;28:1093–8.
[9] King DR, Ogilvie MP, Pereira BM, Chang Y, Manning RJ, Conner JA, et al. Heart rate variability as a triage tool in patients with trauma during prehospital helicopter transport. J Trauma 2009;67:436–40.
[10] Mejaddam AY, Birkhan OA, Sideris AC, Van der Wilden GM, Imam AM, Hwabejire JO, et al. Real-time heart rate entropy predicts the need for lifesaving interventions in trauma activation patients. J Trauma Acute Care Surg 2013;75:607–12.
[11] Proctor KG, Atapattu SA, Duncan RC. Heart rate variability index in trauma patients. J Trauma 2007;63:33–43.
[12] Peev MP, Naraghi L, Chang Y, Demoya M, Fagenholz P, Yeh D, et al. Real-time sample entropy predicts life-saving interventions after the Boston Marathon bombing. J Crit Care 2013;28:1109.e1-4.
[13] Ryan KL, Rickards CA, Ludwig DA, Convertino VA. Tracking central hypovolemia with ECG in humans: cautions for the use of heart period variability in patient monitoring. Shock 2010;33:583–9.
[14] Sasser SM, Hunt RC, Faul M, Sugerman D, Pearson WS, Dulski T, et al. Guidelines for field triage of injured patients: recommendations of the National Expert Panel on Field Triage, 2011. MMWR Recomm Rep 2012;61:1–20.
[15] Stiell IG, Wells GA, Vandemheen K, Clement C, Lesiuk H, Laupacis A, et al. The Canadian CT Head Rule for patients with minor head injury. Lancet 2001;357:1391–6.
[16] Sauaia A, Moore FA, Moore EE, Moser KS, Brennan R, Read RA, et al. Epidemiology of trauma deaths: a reassessment. J Trauma 1995;38:185–93.
[17] Eastridge BJ, Mabry RL, Seguin P, Cantrell J, Tops T, Uribe 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):S431–7.
[18] Cotton BA, Reddy N, Hatch QM, LeFebvre E, Wade CE, Kozar RA, et al. Damage control resuscitation is associated with a reduction in resuscitation volumes and improvement in survival in 390 damage control laparotomy patients. Ann Surg 2011; 254:598–605.
[19] Riskin DJ, Tsai TC, Riskin L, Hernandez-Boussard T, Purtill M, Maggio PM, et al. Massive transfusion protocols: the role of aggressive resuscitation versus product ratio in mortality reduction. J Am Coll Surg 2009;209:198–205.
[20] Holcomb JB, Salinas J, McManus JM, Miller CC, Cooke WH, Convertino VA. Manual vital signs reliably predict need for life-saving interventions in trauma patients. J Trauma 2005;59:821–8.
[21] Chen L, McKenna T, Reisner A, Reifman J. Algorithms to qualify respiratory data collected during the transport of trauma patients. Physiol Meas 2006;27:797–816.
[22] Reisner AT, Chen L, McKenna TM, Reifman J. Automatically-computed prehospital severity scores are equivalent to scores based on medic documentation. J Trauma 2008;65:915–23.
[23] Yu C, Liu Z, McKenna T, Reisner AT, Reifman J. A method for automatic identification of reliable heart rates calculated from ECG and PPG waveforms. J Am Med Inform Assoc 2006;13:309–20.
[24] Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology. Heart rate variability: standards of measurement, physiological interpretation and clinical use. Circulation 1996;93:1043–65.
[25] Malik M, Cripps T, Farrell T, Camm AJ. Prognostic value of heart rate variability after myocardial infarction. A comparison of different data-processing methods. Med Biol Eng Comput 1989;27:603–11.
[26] MATLAB Code for Calculating Sample Entropy. Available at: http://www.physionet.org/physiotools/sampen/matlab/1.1-1/. [Accessed: March 27, 2014].
[27] Richman JS, Moorman JR. Physiological time-series analysis using approximate entropy and sample entropy. Am J Physiol Heart Circ Physiol 2000;278:H2039–49.
[28] Costa M, Goldberger AL, Peng C-K. Multiscale Entropy Analysis (MSE). Available at: http://physionet.org/physiotools/mse/tutorial/tutorial.pdf. [Accessed: March 27, 2014].
[29] Hirsch JA, Bishop B. Respiratory sinus arrhythmia in humans: how breathing pattern modulates heart rate. Am J Physiol Heart Circ Physiol 1981;241:H620–9.
[30] Bernardi L, Salvucci F, Suardi R, Solda PL, Calciati A, Perlini S, et al. Evidence for an intrinsic mechanism regulating heart rate variability in the transplanted and the intact heart during submaximal dynamic exercise? Cardiovasc Res 1990;24:969–81.
[31] Berger RD, Saul JP, Cohen RJ. Transfer function analysis of autonomic regulation. I. Canine atrial rate response. Am J Physiol Heart Circ Physiol 1989;256:H142–52.
[32] DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics 1988;44:837–45.
[33] Pencina MJ, D'Agostino Sr RB, D'Agostino Jr RB, Vasan RS. Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond. Stat Med 2008;27:157–72.
[34] Batchinsky AI, Cooke WH, Kuusela T, Cancio LC. Loss of complexity characterizes the heart rate response to experimental hemorrhagic shock in swine. Crit Care Med 2007;35:519–25.
[35] Convertino VA, Ryan KL, Rickards CA, Salinas J, McManus JG, Cooke WH, et al. Physiological and medical monitoring for en route care of combat casualties. J Trauma 2008;64(4 Suppl):S342–53.
[36] Cooke WH, Rickards CA, Ryan KL, Convertino VA. Autonomic compensation to simulated hemorrhage monitored with heart period variability. Crit Care Med 2008;36: 1892–9.
[37] Hinojosa-Laborde C, Rickards CA, Ryan KL, Convertino VA. Heart rate variability during simulated hemorrhage with lower body negative pressure in high and low tolerant subjects. Front Physiol 2011;85:1–9.
[38] Bernardi L, Wdowczyk-Szulc J, Valenti C, Castoldi S, Passino C, Spadacini G, et al. Effects of controlled breathing, mental activity and mental stress with or without verbalization on heart rate variability. J Am Coll Cardiol 2000;35:1462–9.
[39] Sethuraman G, Ryan KL, Rickards CA, Convertino VA. Ectopy in trauma patients: cautions for use of heart period variability in medical monitoring. Aviat Space Environ Med 2010;81:125–9.
[40] Chen L, Reisner AT, Gribok A, Reifman J. Exploration of prehospital vital sign trends for the prediction of trauma outcomes. Prehosp Emerg Care 2009;13:286–94.
[41] Chen L, Reisner AT, Chen X, Gribok A, Reifman J. Are standard diagnostic test characteristics sufficient for the assessment of continual patient monitoring? Med Decis Making 2013;33:225–34.
[42] Zarzaur BL, Croce MA, Fischer PE, Magnotti LJ, Fabian TC. New vitals after injury: shock index for the young and age x shock index for the old. J Surg Res 2008;147: 229–36.
[43] Pearl WS. A hierarchical outcomes approach to test assessment. Ann Emerg Med 1999;33:77–84.

**SHOCK,** Vol. 43, No. 5, pp. 429–436, 2015

# AUTOMATED ANALYSIS OF VITAL SIGNS TO IDENTIFY PATIENTS WITH SUBSTANTIAL BLEEDING BEFORE HOSPITAL ARRIVAL: A FEASIBILITY STUDY

**Jianbo Liu,* Maxim Y. Khitrov,* Jonathan D. Gates,† Stephen R. Odom,‡ Joaquim M. Havens,† Marc A. de Moya,§ Kevin Wilkins,‖ Suzanne K. Wedel,‖ Erin O. Kittell,¶ Jaques Reifman,* and Andrew T. Reisner¶**

**Department of Defense Biotechnology High Performance Computing Software Applications Institute, Telemedicine and Advanced Technology Research Center, US Army Medical Research and Materiel Command, Fort Detrick, Maryland; and †Department of Surgery, Division of Trauma, Burns and Surgical Critical Care, Brigham and Women's Hospital, Boston, Massachusetts; ‡Department of Surgery, Division of Acute Care Surgery, Trauma, and Surgical Critical Care, Beth Israel Deaconess Medical Center, Boston, Massachusetts; and §Department of Surgery, Division of Trauma, Emergency Surgery and Surgical Critical Care, Massachusetts General Hospital, Boston, Massachusetts; ‖Boston MedFlight, Bedford, Massachusetts; and ¶Department of Emergency Medicine, Massachusetts General Hospital, Boston, Massachusetts*

Received 4 Nov 2014; first review completed 24 Nov 2014; accepted in final form 22 Dec 2014

ABSTRACT—Trauma outcomes are improved by protocols for substantial bleeding, typically activated after physician evaluation at a hospital. Previous analysis suggested that prehospital vital signs contained patterns indicating the presence or absence of substantial bleeding. In an observational study of adults (aged ≥18 years) transported to level I trauma centers by helicopter, we investigated the diagnostic performance of the Automated Processing of the Physiological Registry for Assessment of Injury Severity (APPRAISE) system, a computational platform for real-time analysis of vital signs, for identification of substantial bleeding in trauma patients with explicitly hemorrhagic injuries. We studied 209 subjects prospectively and 646 retrospectively. In our multivariate analysis, prospective performance was not significantly different from retrospective. The APPRAISE system was 76% sensitive for 24-h packed red blood cells of 9 or more units (95% confidence interval, 59% – 89%) and significantly more sensitive ($P < 0.05$) than any prehospital Shock Index of 1.4 or higher; sensitivity, 59%; initial systolic blood pressure (SBP) less than 110 mmHg, 50%; and any prehospital SBP less than 90 mmHg, 50%. The APPRAISE specificity for 24-h packed red blood cells of 0 units was 87% (88% for any Shock Index ≥1.4, 88% for initial SBP <110 mmHg, and 90% for any prehospital SBP <90 mmHg). Median APPRAISE hemorrhage notification time was 20 min before arrival at the trauma center. In conclusion, APPRAISE identified bleeding before trauma center arrival. *En route*, this capability could allow medics to focus on direct patient care rather than the monitor and, via advance radio notification, could expedite hospital interventions for patients with substantial blood loss.

KEYWORDS—Trauma, hemorrhage, massive transfusion, decision-support systems, prehospital emergency care

## INTRODUCTION

### *Background*

Hemorrhage is recognized as the leading treatable cause of death after injury (1). Improved outcomes in trauma patients have been shown when trauma centers apply specific protocols for patients with substantial bleeding (2, 3). These protocols encompass damage-control resuscitation, including aggressive measures to avoid coagulopathy (via permissive hypotension that slows blood loss, adequate restoration of coagulation factors via transfusion, and minimization of hypothermia), which is important because trauma-induced coagulopathy affects between 24% and 56% of critically injured patients (4). For these patients, massive transfusion of packed red blood cells (PRBCs), that is, 10 or more units in 24 h (5), is often necessary. Damage-control resuscitation is paired with damage-control surgery, the operative strategy of prioritizing early surgical control of bleeding, while sparing noncritical surgical repairs that are undertaken only after the patient has sufficiently recovered.

Although management protocols for patients with substantial bleeding are associated with mortality benefits (3, 6), there are no widely accepted criteria for their initiation. Holcomb and Gumbert (2) commented that, in the report by Cotton et al. (3), activation had been subjective after a surgeon's evaluation of the patient. Riskin et al. (6) reported that the Stanford Protocol was activated subjectively "at the discretion of the attending physician." Several clinical scores to predict whether trauma patients will require massive transfusion have been developed, including the McLaughlin score, the Trauma Associated Severe Hemorrhage (TASH) score, and the Assessment of Blood Consumption (ABC) score (5). These scores are based on vital sign data; mechanism of injury or anatomic details (TASH score and ABC score); and abdominal sonography

Address reprint requests to Jaques Reifman, PhD, US Department of Defense Biotechnology High Performance Computing Software Applications Institute, Telemedicine and Advanced Technology Research Center, US Army Medical Research and Materiel Command, ATTN: MCMR-TT, 504 Scott St., Fort Detrick, MD 21702. E-mail: jaques.reifman.civ@mail.mil.

This work was supported by the US 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 US Army Medical Research and Materiel Command, Fort Detrick, Maryland. 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.

Supplemental digital content is available for this article. Direct URL citation appears in the printed text and is provided in the HTML and PDF versions of this article on the journal's Web site (www.shockjournal.com).

DOI: 10.1097/SHK.0000000000000328

(ABC score), laboratory testing (McLaughlin score), or both (TASH score).

Our group was interested in whether it would be feasible to identify patients with substantial bleeding before arrival at the hospital, without relying on testing or expertise that is normally hospital based, such as sonography or laboratory testing. Growing evidence shows that assessment of multiple vital signs together may be more effective than univariate approaches for detecting hemorrhagic hypovolemia (7, 8). In addition, there have been encouraging reports of computational techniques (9, 10) to account for the fact that high-acuity trauma patients demonstrate complex temporal fluctuations in their prehospital vital signs (11–13), and to identify unreliable vital signs (14), because spurious measurements are so common (15–18).

In this report, we evaluate the hypothesis that it is feasible to identify patients with substantial 24-h PRBC transfusion requirements by automated analysis of prehospital vital signs. To test this prospectively, a specialized real-time computing platform was developed and deployed into an active prehospital operation (19). If it is feasible to identify patients with substantial bleeding by automated analysis of prehospital vital signs, there might be improved *en route* care as well as in-hospital care. *En route*, caregivers could focus more on patient care rather than split attention with reexamining and reevaluating the vital sign monitor. The automated system could notify the caregivers when the vital signs were statistically consistent with bleeding and display an on-screen checklist of expected responses. The receiving hospital could be provided with advance radio notification, offering a head start for careful preparation of a patient with major hemorrhage, for example, prewarming of the patient's bay (to prevent hypothermia), preparation of fresh frozen plasma for immediate transfusion, and mobilization of surgical assets (for early surgical intervention).

## METHODS

### Setting and study population

We examined a convenience sample of adult (aged ≥18 years) trauma patients transported by air emergency medical service (EMS) to participating level I trauma centers. With institutional review board approval, we collected a prospective data set from Boston MedFlight (BMF, Bedford, Mass) and compared the findings with an archival data set originally collected from Memorial Hermann Life Flight (MHLF, Houston, Tex) by Cooke et al. (20) and Holcomb et al. (21). In both data sets, we analyzed all subjects with at least one recorded non-zero systolic blood pressure (SBP). Patients who died before hospital admission (e.g., in the emergency department) were excluded from analysis, because resuscitation was often terminated before large-volume PRBC transfusion could be completed, regardless of whether or not the patient had significant hypovolemia.

Our primary study outcome was 24-h PRBC transfusion volume in patients with hemorrhagic injury, defined as a documented hemorrhagic injury that unequivocally caused some loss of blood (laceration or fracture of a solid organ; documented hematoma within the thorax, peritoneum, retroperitoneum, or pelvis; vascular injury that required operative repair; or limb amputation) and PRBC transfusion within 24 h. Patients who received PRBCs but lacked a documented hemorrhagic injury were excluded from the primary analysis because, in the absence of an explicitly hemorrhagic injury, it was challenging to determine whether the transfusion was clinically indicated. Whether the patient had documented hemorrhagic injury was determined by automated text search, searching for injuries that met the aforementioned criteria (records were also jointly reviewed by two investigators, J.L. and A.T.R., who confirmed that the automated text search had not omitted any applicable hemorrhagic injuries nor included nonhemorrhagic injuries).

The excluded patients who lacked explicitly hemorrhagic injuries were reincluded and analyzed in a sensitivity analysis (see Appendix, Supplemental Digital Content 1, at http://links.lww.com/SHK/A267).

### Vital sign data processing

For the prospective cohort, we deployed the APPRAISE (Automated Processing of the Physiological Registry for Assessment of Injury Severity [19]) system 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, Oreg) networked to the GoBook ultracompact ruggedized personal computer (General Dynamics Itronix, Sunrise, Fla) running analytic algorithms developed for this research project (19). As a practical matter, this meant that all vital sign data processing and analyses for BMF were done automatically and in real time.

The following routine vital signs were monitored by the Propaq 206 monitor: heart rate (HR), respiratory rate (RR), oscillometric SBP, and pulse pressure ([PP], the difference between SBP and diastolic blood pressure). The APPRAISE software created an electronic record of the Propaq data, analyzed the vital sign data in real time using algorithms described below and 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).

The retrospective data originally had been collected onboard 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 (21). Subsequently, those data were uploaded to our data warehousing system (22) and analyzed offline.

We analyzed the prospective and the retrospective Propaq 206 data using the exact same computational methodology. First, the automated algorithms identified and excluded unreliable vital sign measurements (Fig. 1). The reliability

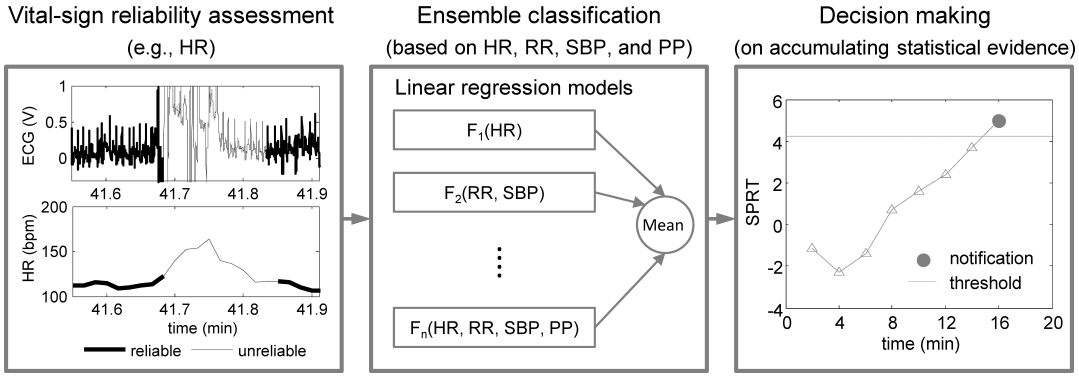



FIG. 1. **Analytic methodology for hemorrhage identification**. In the first step (left panel), algorithms were used to identify, and exclude, unreliable vital signs. In the second step (middle panel), ensemble classification was applied, which consisted of a set of different linear regression models, $F_1$, $F_2$, ⋯, $F_n$, 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 was evaluated by the SPRT, a statistical test of whether or not measurements repeated across time are consistent with a control distribution or with a different (e.g., hemorrhagic patient) distribution. bpm—beats per minute; ECG—electrocardiography; V—volt.

algorithms for HR and RR involved analysis of the electrocardiography (ECG) and impedance pneumography waveforms. This allowed us to discriminate between a clean source signal versus an unreliable segment caused by signal artifacts (23, 24). The SBP and PP reliability algorithms assessed signal quality by analyzing the relationship between systolic, diastolic, and mean arterial pressures and by comparing HR measured by ECG versus HR measured by oscillometry (25). These automated algorithms, which have been shown to agree with human experts' opinions (23, 24), can significantly increase the diagnostic value of vital signs by removing spurious measurements (25, 26).

The second step of real-time analysis involved an ensemble classifier, which is a set of multivariate regression models whose numerical outputs were averaged to yield the final output (Fig. 1). We trained the multivariate regression models (i.e., set the weights for the input variables) for a binary outcome as per Chen et al. (27), using the initial 15 min of vital sign data from each MHLF subject. For the model training, the binary outcome was whether patients received 1 or more PRBCs for an unambiguous hemorrhagic injury or not. This model training yielded a classifier that, on the basis of the input vital signs (HR, RR, SBP, and PP), quantified whether the pattern was similar to the population with hemorrhage (output closer to 1) or to the nonhemorrhagic control population (output closer to 0). The ensemble classifier was originally developed for use at a single time point, for example, on 15 min of prehospital data collection, for prediction of 24-h PRBC more than 0, and it was cross-validated using 50%/50% training/testing (27). There were no significant differences ($P > 0.05$) when the receiver operating characteristic area under the curve (ROC AUC) of 10-fold cross-validation was compared with the ROC AUC for 100%/100% training/testing ($\Delta$ROC AUC $\pm$ 0.01). Compared with routine multivariate regression, an ensemble classifier can provide two advantages. First, the ensemble can still classify patients even when a complete set of reliable vital signs is unavailable. Second, it can offer performance that is more consistent from one data set to the next (27, 28).

Every 2 min, this analysis was repeated. For the prospective trial, this occurred in real time. For the retrospective analysis, we reapplied the algorithms at every 2-min mark of the patient's electronic record, simulating real-time application. 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 up to the time of analysis. (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.) The rationale for analyzing data reaching back to the start of the mission arose from previous analysis suggesting that prehospital vital signs contained enormous variability—likely caused by pain, medications, or other transient stimuli—and that time averaging was an effective method to remove some of the confounding data perturbations and achieve superior diagnostic performance (9).

The third and final step of real-time analysis involved the Wald sequential probability ratio test (SPRT) for determining whether to issue a "hemorrhage notification" on the basis of accumulated evidence from the ensemble classifier outputs (Fig. 1). The SPRT (29) is a useful statistical technique for determining 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). Thresholds for the SPRT were set as per Chen et al. (10), where the SPRT was shown to reduce false alarms at the expense of some alarm latency.

### Clinical outcomes

For the BMF data set, 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). The data were archived electronically using REDCap (30). We obtained injury severity scores from each hospital's trauma registry. For the MHLF data set, a chart review was conducted by the original study authors (21).

### Statistical analysis

Data were summarized using mean with standard deviation or median with interquartiles for continuous variables, and frequency with percentage for categorical variables. We computed the proportion of patients who received a hemorrhage notification as a function of the number of units of PRBCs that each patient received during the initial 24 h in the hospital ("24-h PRBC volume"). For comparison, we also computed the proportion of patients with other hemodynamic abnormalities: initial SBP less than 110 mmHg, any prehospital SBP less than 90 mmHg, or any prehospital Shock Index (SI) of 1.4 or higher (where SI = HR/SBP). The threshold for SI was chosen based on the findings in Mutschler et al. (31). We tested for significant differences between those proportions using McNemar's test.

For BMF patients, MHLF patients, and the pooled data set, we developed logistical regression models to quantify the likelihood of a patient receiving a hemorrhage notification as a function of 24-h PRBC volume. We also tested whether the likelihood of a patient receiving a hemorrhage notification differed between the two populations (BMF and MHLF) controlling for the 24-h PRBC volume. Finally, to investigate other factors that may have influenced whether a patient received a hemorrhage notification or not, we applied multivariate logistical regression to a set of parameters quantifying potential sources of variability: age, mechanism of trauma, prehospital factors (elapsed time since injury, volume of resuscitation, endotracheal intubation, duration of transport), and anatomy of the injuries based on the trauma registry Abbreviated Injury Scale scores. Two-sided values of $P < 0.05$ were considered as statistically significant.

## RESULTS

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-h PRBC transfusion while lacking explicitly hemorrhagic injuries (MHLF, 64; BMF, 25); these patients were examined in the sensitivity analysis (see Appendix, Supplemental Digital Content 1, at http://links.lww.com/SHK/A267). Table 1 describes the primary study population.

### Correlation of basic vital signs and 24-h PRBC volume

In the MHLF data set, SBP, PP, HR, and RR were significantly ($P < 0.001$) correlated with 24-h PRBC transfusion volume: $\rho = -0.32$, $\rho = -0.36$, $\rho = +0.24$, and $\rho = +0.24$, respectively.

In the BMF data set, SBP ($P < 0.001$) and PP ($P < 0.01$) were significantly correlated with 24-h PRBC transfusion volume: $\rho = -0.30$ and $\rho = -0.23$, respectively. Heart rate showed a nonsignificant trend ($P = 0.051$), with $\rho = +0.14$, whereas RR was not significantly correlated.

### Diagnostic test characteristics

Table 2 shows the relationship between the incidence of APPRAISE hemorrhage notification and 24-h PRBC transfusion volume. With increasing 24-h PRBC transfusion volume,

TABLE 1. **Study population characteristics**

| | Memorial Hermann Life Flight | Boston MedFlight |
|---|---|---|
| Population, n | 646 | 209 |
| Sex, male, n (%) | 479 (74) | 155 (74) |
| Age, mean (SD), years | 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) |
| Prehospital blood transfusion, n (%) | 0 (0) | 15 (7) |
| 24-h PRBC volume >0 unit, n (%) | 75 (12) | 31 (15) |
| 24-h PRBC volume ≥3 units, n (%) | 57 (9) | 18 (9) |
| 24-h PRBC volume ≥10 units, n (%) | 22 (3) | 8 (4) |
| Survival to discharge, n (%) | 608 (94) | 191 (91) |

GCS—Glasgow Coma Scale; IQR—interquartile range; ISS—Injury Severity Score; SD—standard deviation.

Table 2. **Relationship between prehospital APPRAISE hemorrhage notification versus 24-h PRBC transfusion volume**

| | 24-h PRBC volume (units) | | | | |
|---|---|---|---|---|---|
| | 0 | 1 – 2 | 3 – 8 | ≥9 | Total |
| **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) | |
| **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) | |
| **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) | |

the proportion of APPRAISE-positive subjects exhibited an increasing trend in both the MHLF and BMF studies. In the pooled data set (MHLF and BMF), sensitivities for 24-h PRBC transfusion volume of 9 or more units for APPRAISE notification, SI of 1.4 or higher, initial SBP less than 110 mmHg, and any hypotension (SBP <90 mmHg) were 76% (59% – 89%), 59% (41% – 75%), 50% (32% – 68%), and 50% (32% – 68%), respectively, and we found that the sensitivity of APPRAISE notification was significantly higher than SI of 1.4 or higher ($P$ = 0.014), initial SBP less than 110 mmHg ($P$ = 0.007), and any hypotension, that is, SBP less than 90 mmHg ($P$ = 0.007). The sensitivities of APPRAISE notification for 24-h PRBC transfusion volume of 9 or more units were similar for the MHLF versus BMF data sets: 76% (55% – 91%) and 78% (40% – 97%), respectively.

In the pooled data set (MHLF and BMF), specificities for 24-h PRBC transfusion volume of 0 units (i.e., no blood transfusion at all) for APPRAISE notification, SI of 1.4 or higher, initial SBP less than 110 mmHg, and any hypotension (SBP <90 mmHg) were 87% (85% – 89%), 88% (85% – 90%), 88% (86% – 91%), and 90% (88% – 92%), respectively, and we found that the specificity of APPRAISE was not significantly different from initial SBP less than 110 mmHg or any prehospital SI of 1.4 or higher. Compared with any prehospital SBP less than 90 mmHg, APPRAISE notification showed a significantly lower specificity ($P$ < 0.05), although the absolute magnitude of the difference was 3%. The specificities of APPRAISE notification for 24-h PRBC transfusion volume of 0 units were 86% (83% – 89%) for the MHLF data set and 90% (85% – 94%) for the BMF data set. In the pooled data set, negative predictive values (24-h PRBC transfusion volume = 0 units vs. ≥1 unit) for APPRAISE notification, SI of 1.4 or higher, initial SBP less than 110 mmHg, and any hypotension (SBP <90 mmHg) were similar: 94% (92% – 96%), 92% (90% – 94%), 92% (90% – 94%), and 92% (90% – 94%), respectively.

Incidentally, there were three subjects who received prehospital needle decompression, and all received hemorrhage notifications during transport (24-h PRBC volumes for these subjects were 0, 4, and >20, respectively).

### *Timelines*

Figure 2 illustrates prehospital timelines for all subjects with 24-h PRBC volume of 9 or more units, showing the timing of blood pressure measurements, of APPRAISE hemorrhage notifications, and episodes of hypotension (SBP <90 mmHg). The median notification time after the start time of transport was 6 min (interquartiles 4 – 16) for MHLF and 10 min for BMF (interquartiles 8 – 40). The median notification time before arrival at the hospital was 17 min for MHLF and 52 min for BMF, and the difference was largely caused by shorter transport times for MHLF (the median transport time for subjects with 24-h PRBC volume ≥9 units was 28 min [interquartiles 24 – 36] for MHLF and 65 min [interquartiles 35 – 78] for BMF). Combining the two populations, APPRAISE notification occurred in the first half of the transportation in 73% of the cases.

Nine subjects returned to APPRAISE-negative status after a hemorrhage notification: six MHLF subjects who were actually false positive (i.e., 24-hr PRBC = 0) and three BMF subjects who were true positive (i.e., 24-hr PRBC ≥ 1) and received prehospital PRBC transfusion.

### *Multivariate logistic regression*

Using logistic regression to model the likelihood of APPRAISE hemorrhage notification as a function of 24-h PRBC transfusion volume further demonstrated that the results were similar in both data sets (Fig. 3). Each PRBC unit transfused was associated with a 43% (95% CI, 30 – 57%) increase in the odds of APPRAISE hemorrhage notification for MHLF and a 44% (95% CI, 24 – 67%) increase for BMF. The odds ratio of APPRAISE notification per unit of PRBC transfused was not significantly different between the two data sets (i.e., BMF versus MHLF) when fitting a regression model to the pooled data set ($P$ = 0.635). However, there was a nonsignificant trend toward a lower overall likelihood of hemorrhage notification in the BMF data set when compared with the MHLF data set ($P$ = 0.053), including a lower likelihood of hemorrhage notifications in patients without bleeding (24-h PRBC = 0) and with substantial bleeding (24-h PRBC ≥9), which is apparent in the offset between the two regression curves (Fig. 3). Note that the specificities and sensitivities extracted from Figure 3 are slightly different from those reported in Table 2 because of the nature of the regression fit.

We investigated the factors associated with whether subjects received an APPRAISE hemorrhage notification; see univariate and multivariate logistic regression analyses in Table 3. In multivariate analysis, four independent factors were significant predictors of whether the patient received an APPRAISE notification: increasing 24-h PRBC volume, increasing severity

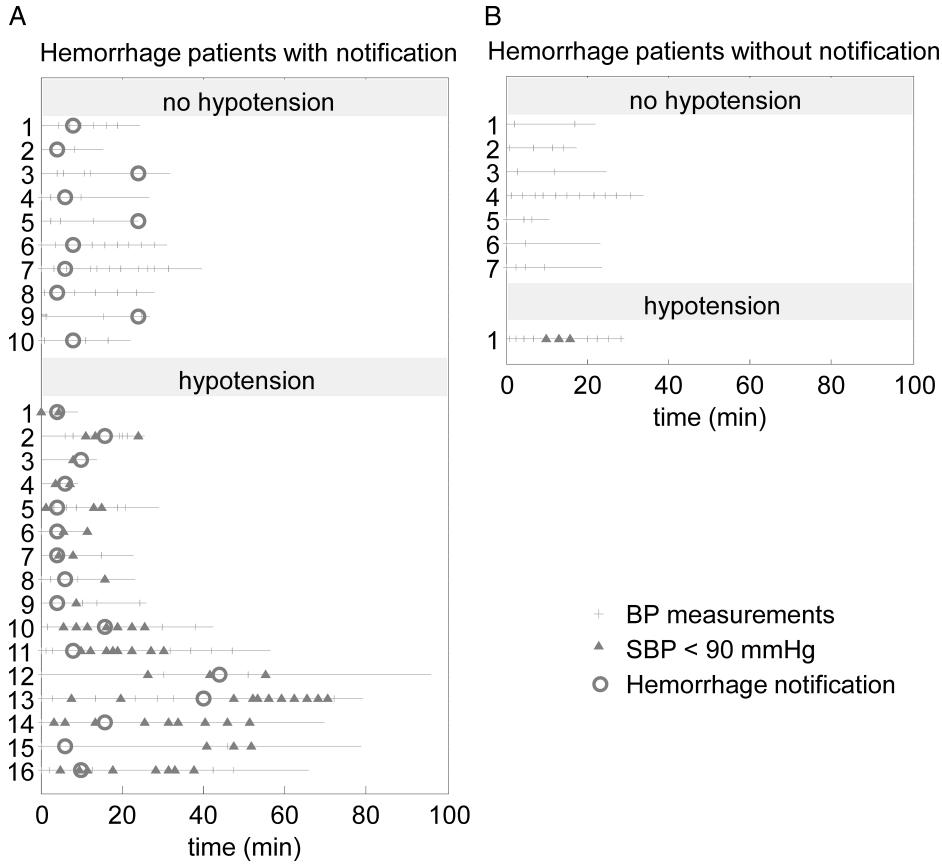



FIG. 2. **Timelines for patients with substantial bleeding (i.e., 24-h PRBC volume ≥9 units) indicating time of hypotensive episodes (SBP <90 mmHg) and hemorrhage notification during prehospital transport.** (A) Bleeding patients who received an APPRAISE hemorrhage notification (i.e., true positives) and (B) bleeding patients who did not receive a notification (i.e., false negatives).

of chest injury, longer flight duration, and younger age. Neither abdominal, nor head, nor extremity injury severity had a significant association with false-negative alarms. Prehospital PRBC transfusion was only found in the BMF cohort, and those patients had a significantly increased risk of APPRAISE hemorrhage notification.

## DISCUSSION

This investigation demonstrated that there was a strong association between 24-h PRBC transfusion volume and abnormal prehospital vital signs, and that the majority of patients with large transfusion requirements could be distinguished from other trauma patients using techniques for time series and multivariate analysis. The automated APPRAISE system required neither oversight nor input by the flight crew; it operated wholly autonomously, only requiring that the flight crew use their Propaq transport monitor as per standard procedure. The performance of the APPRAISE algorithms for early identification of patients with 24-h PRBC of 9 or more units was quite similar in actual prospective real-time use (the BMF data set) versus simulated real-time use (the MHLF data set).

### *Potential benefits of prehospital identification of substantial bleeding*

Automated functionality that reliably provides a notification whenever important patterns develop would permit the caregiver to focus much more on the patient (e.g., better pain control, better management of retching patients who could aspirate, etc.) and not constantly split attention between the patient and the travel monitor. Consistent fully automated detection of hypovolemic vital signs may be most clinically valuable if the EMS caregiver is inexperienced, fatigued, or distracted.

With reliable notification that a bleeding patient is about to arrive, the receiving facility could prepare for hemorrhage-specific management. Today's typical practice involves a trauma team evaluation—postarrival—before deciding whether to activate protocols for substantial bleeding (2). At best, this adds a small delay to care and, in some cases, resultant delays can be substantial. In one report describing the benefits of an institutional protocol for substantial bleeding, interventions such as transfusion of fresh-frozen plasma were not initiated for several hours in many cases (6). By analogy, the common practice of activating the cardiac catheterization team when the prehospital ECG shows ST-elevation myocardial infarction in a patient with chest pain illustrates the potential value of readying the hospital for an exsanguinating patient based on a simple objective prehospital indicator: by initiating in-hospital preparations based on prehospital notification, the time delay to percutaneous coronary intervention can be reduced (32). Of note, cardiologists still conduct expert evaluations before undertaking catheterization, and prehospital notification does not remove clinical authority from hospital caregivers.

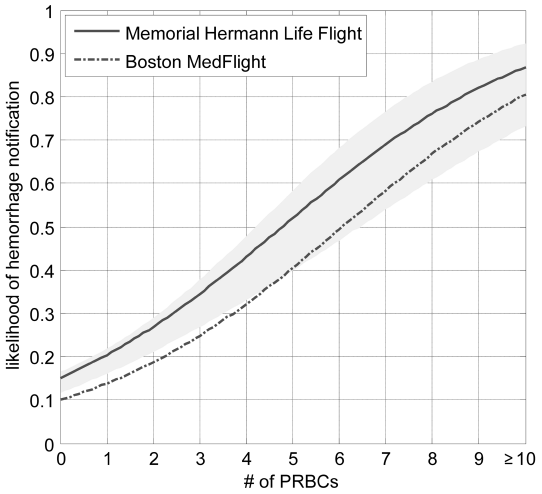



Fig. 3. **Modeling the rate of APPRAISE hemorrhage notification using logistic regression**. The slopes of the BMF and MHLF curves were the same: each PRBC unit transfused was associated with a 43% (95% CI, 30% – 57%) increase in the odds of MHLF hemorrhage notification and also a 44% (95% CI, 24% – 67%) increase for BMF hemorrhage notification. The offset between MHLF and BMF was not statistically significant: further inclusion of a population parameter (1 for MHLF and 0 for BMF) into the regression model for the combined (BMF and MHLF) data set did not yield a statistically significant coefficient for the population parameter. Shaded areas are the 95% confidence intervals for the combined population regression model. The confidence interval becomes wider as a result of a smaller patient population with larger 24-h PRBC volumes. Patients who received 10 or more units of PRBCs were combined into a single category.

It may be most clinically valuable if, rather than a simple alert or notification, the automated system were to display an on-screen list of bulleted action items to remind the EMS caregiver of each and every expected action item for trauma patients with abnormal circulation, for example, check for compressible hemorrhage, check for tension pneumothorax (as noted, the APPRAISE system generated a hemorrhage notification for all three subjects with documented prehospital needle decompression), hold fluids unless SBP was less than 90 mmHg, keep patient warm, and so on. Note that protocol compliance is an underlying challenge throughout health care (33), and checklists are a valuable tool to improve protocol compliance (34–36).

The clinical benefits of this system are speculative because we did not assess clinical impact in the current investigation (an institutional review board–related matter; see Methods). Yet, it seems reasonable to move toward bedside computing for certain tasks, such as statistical analyses that can quantify whether a sequence of vital signs is abnormal, and thereby permit caregivers to focus on quality bedside care.

### *Physiological interpretation of the findings*

At a rudimentary level, this study suggests that patients with massive 24-h blood transfusion requirements demonstrate hypovolemic physiology before hospital arrival. This intuitive finding is consistent with other prediction rules for massive transfusion where hypotension and tachycardia are established predictive factors for massive transfusion (5).

Unlike the other massive transfusion prediction rules, the APPRAISE system only involves vital sign data analyzed during prehospital transport. The APPRAISE system uses well-known statistical techniques, such as time averaging and the SPRT, for analyzing data that fluctuate through time, and it detects hemorrhage by considering the temporal accumulation of evidence. The system does *not* seek to identify trends through time (e.g., downward drifts in SBP), which may seem counterintuitive, but it has been clearly demonstrated that prehospital vital signs fluctuate substantially frequently without obvious overt directional trends (10, 12–14).

In addition to time series techniques, another common sense principle incorporated in the APPRAISE system was multivariate analysis. Like several prediction rules for massive transfusion (5), the APPRAISE system's algorithms used the independent diagnostic information from more than one vital sign. This is consistent with recent reports that the SI (the ratio of HR to SBP) is a valuable diagnostic tool for identification of hemorrhage (7, 8). The APPRAISE system identifies hypovolemia by a combination of low SBP, low PP, high HR, and high RR. A minority of the massive transfusion patients were not detected by the APPRAISE system; those generally lacked hypotension (Fig. 2B), suggesting that they were not substantially hypovolemic during transport.

There were also APPRAISE hemorrhage notifications in patients who did *not* require massive transfusion. These patients were likely hypovolemic during transport yet without the ongoing blood losses that necessitate massive transfusion (of note, among patients who never needed any PRBCs, those who

Table 3. **Factors associated with APPRAISE hemorrhage notification**

| | Univariate | | Multivariate | |
|---|---|---|---|---|
| Factor | Odds ratio | 95% CI | Odds ratio | 95% CI |
| 24-h PRBC transfusion volume (per 1 unit) | 1.43‡ | (1.32 – 1.55) | 1.4‡ | (1.29 – 1.52) |
| **Demographics** | | | | |
| Age (per 10 years) | 0.91 | (0.82 – 1.02) | 0.87* | (0.77 – 0.99) |
| **Prehospital course** | | | | |
| Endotracheal intubation (y/n) | 2.22‡ | (1.53 – 3.20) | | |
| IVF (per 500 mL) | 1.51‡ | (1.27 – 1.79) | | |
| Time to begin transport (per 10 min) | 1.05 | (0.99 – 1.12) | | |
| Duration of transport (per 10 min) | 1.18† | (1.05 – 1.32) | 1.17* | (1.02 – 1.35) |
| **Injury mechanism** | | | | |
| **Blunt trauma (y/n)** | 1.28 | (0.70 – 2.33) | | |
| **Penetrating trauma (y/n)** | 0.88 | (0.48 – 1.62) | | |
| **Injury description** | | | | |
| Head AIS ≥3 (y/n) | 1.03 | (0.67 – 1.58) | | |
| Abdomen or pelvis AIS ≥ 3 (y/n) | 4.41‡ | (2.73 – 7.12) | | |
| Extremity, not pelvis AIS ≥3 (y/n) | 1.26 | (0.80 – 1.98) | | |
| Thorax AIS ≥3 (y/n) | 3.73‡ | (2.51 – 5.53) | 2.58‡ | (1.64 – 4.04) |

AIS—Abbreviated Injury Scale; CI—confidence interval; IVF—intravenous fluids; (y/n)—binary variables.
Odds ratio significantly different from 1.0: *$P < 0.05$, †$P < 0.01$, ‡$P < 0.001$.

received an APPRAISE hemorrhage notification had significantly higher average injury severity scores). Whether the alert would offer clinical value in this population without massive transfusion requirements is an open question. As discussed above, the APPRAISE system would not obviate the need for clinical assessments by prehospital and receiving facility personnel. Rather, the system is a tool for optimizing vital sign information, offering automated consistent notification when the patterns suggest hypovolemia, and these patterns are strongly associated with subsequent blood transfusion requirements.

### *Limitations*

The study outcome, hemorrhage severity, was quantified by each patient's 24-h PRBC volume. However, the quantity of PRBCs that a patient actually receives is a function of multiple factors, including the speed and effectiveness of surgical hemorrhage control, and some subjective clinical decision making. The generalizability of the findings, that is, the notification incidence versus 24-h PRBC volume, and their applicability to guiding initial resuscitation may have limitations. Yet, the notable consistency (Fig. 3) between the MHLF and BMF results during aeromedical transport to one and three distinct trauma centers, respectively, suggests that such confounding factors can average out across different trauma systems, yielding consistent relationships between prehospital notification incidence and hemorrhage severity.

The prospective BMF arm of this study was sufficient to demonstrate that the real-time system can perform encouragingly well (seven of nine massive transfusion BMF subjects received a real-time prehospital notification). However, the BMF data set was too small to directly compare test characteristics of the APPRAISE notification versus hypotension or SI.

## CONCLUSIONS

We conclude that real-time multivariate time series analysis of vital signs is a feasible means of identifying prehospital trauma patients with substantial bleeding, and that prospective investigation of the clinical value of this automated methodology is justified.

## REFERENCES


1. Kauvar DS, Lefering R, Wade CE: Impact of hemorrhage on trauma outcome: an overview of epidemiology, clinical presentations, and therapeutic considerations. *J Trauma* 60(6 suppl):S3–S11, 2006.
2. Holcomb JB, Gumbert S: Potential value of protocols in substantially bleeding trauma patients. *Curr Opin Anaesthesiol* 26(2):215–220, 2013.
3. Cotton BA, Reddy N, Hatch QM, LeFebvre E, Wade CE, Kozar RA, Gill BS, Albarado R, McNutt MK, Holcomb JB: Damage control resuscitation is associated with a reduction in resuscitation volumes and improvement in survival in 390 damage control laparotomy patients. *Ann Surg* 254(4):598–605, 2011.
4. Curry N, Davis PW: What's new in resuscitation strategies for the patient with multiple trauma? *Injury* 43(7):1021–1028, 2012.
5. Elmer J, Wilcox SR, Raja AS: Massive transfusion in traumatic shock. *J Emerg Med* 44(4):829–838, 2013.
6. Riskin DJ, Tsai TC, Riskin L, Hernandez-Boussard T, Purtill M, Maggio PM, Spain DA, Brundage SI: Massive transfusion protocols: the role of aggressive resuscitation versus product ratio in mortality reduction. *J Am Coll Surg* 209(2):198–205, 2009.
7. King RW, Plewa MC, Buderer NM, Knotts FB: Shock index as a marker for significant injury in trauma patients. *Acad Emerg Med* 3(11):1041–1045, 1996.
8. Sohn CH, Kim WY, Kim SR, Seo DW, Ryoo SM, Lee YS, Lee JH, Oh BJ, Won HS, Shim JY, et al.: An increase in initial shock index is associated with the requirement for massive transfusion in emergency department patients with primary postpartum hemorrhage. *Shock* 40(2):101–105, 2013.
9. Chen L, Reisner AT, Gribok A, Reifman J: Exploration of prehospital vital sign trends for the prediction of trauma outcomes. *Prehosp Emerg Care* 13(3): 286–294, 2009.
10. Chen L, Reisner AT, Chen X, Gribok A, Reifman J: Are standard diagnostic test characteristics sufficient for the assessment of continual patient monitoring? *Med Decis Making* 33(2):225–234, 2013.
11. Morris JA Jr, Auerbach PS, Marshall GA, Bluth RF, Johnson LG, Trunkey DD: The Trauma Score as a triage tool in the prehospital setting. *JAMA* 256(10): 1319–1325, 1986.
12. Shapiro NI, Kociszewski C, Harrison T, Chang Y, Wedel SK, Thomas SH: Isolated prehospital hypotension after traumatic injuries: a predictor of mortality? *J Emerg Med* 25(2):175–179, 2003.
13. Lipsky AM, Gausche-Hill M, Henneman PL, Loffredo AJ, Eckhardt PB, Cryer HG, de Virgilio C, Klein SL, Bongard FS, Lewis RJ: Prehospital hypotension is a predictor of the need for an emergent, therapeutic operation in trauma patients with normal systolic blood pressure in the emergency department. *J Trauma* 61(5):1228–1233, 2006.
14. 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* 27(5):533.e1–533.e10, 2012.
15. Edmonds ZV, Mower WR, Lovato LM, Lomeli R: The reliability of vital sign measurements. *Ann Emerg Med* 39(3):233–237, 2002.
16. Jones DW, Appel LJ, Sheps SG, Roccella EJ, Lenfant C: Measuring blood pressure accurately: new and persistent challenges. *JAMA* 289(8):1027–1030, 2003.
17. Lovett PB, Buchwald JM, Stürmann K, Bijur P: The vexatious vital: neither clinical measurements by nurses nor an electronic monitor provides accurate measurements of respiratory rate in triage. *Ann Emerg Med* 45(1): 68–76, 2005.
18. Friesdorf W, Konichezky S, Gross-Alltag F, Fattroth A, Schwilk B: Data quality of bedside monitoring in an intensive care unit. *Int J Clin Monit Comput* 11(2):123–128, 1994.
19. Reisner AT, Khitrov MY, Chen L, Blood A, Wilkins K, Doyle W, Wilcox S, Denison T, Reifman J: Development and validation of a portable platform for deploying decision-support algorithms in prehospital settings. *Appl Clin Inform* 4(3):392–402, 2013.
20. Cooke WH, Salinas J, Convertino VA, Ludwig DA, Hinds D, Duke JH, Moore FA, Holcomb JB: Heart rate variability and its association with mortality in prehospital trauma patients. *J Trauma* 60(2):363–370; discussion 370, 2006.
21. Holcomb JB, Salinas J, McManus JM, Miller CC, Cooke WH, Convertino VA: Manual vital signs reliably predict need for life-saving interventions in trauma patients. *J Trauma* 59(4):821–828, 2005.
22. McKenna TM, Bawa G, Kumar K, Reifman J: The physiology analysis system: an integrated approach for warehousing, management and analysis of time-series physiology data. *Comput Methods Programs Biomed* 86(1):62–72, 2007.
23. Yu C, Liu Z, McKenna T, Reisner AT, Reifman J: A method for automatic identification of reliable heart rates calculated from ECG and PPG waveforms. *J Am Med Inform Assoc* 13(3):309–320, 2006.
24. Chen L, McKenna T, Reisner A, Reifman J: Algorithms to qualify respiratory data collected during the transport of trauma patients. *Physiol Meas* 27(9): 797–816, 2006.
25. Reisner AT, Chen L, McKenna TM, Reifman J: Automatically-computed prehospital severity scores are equivalent to scores based on medic documentation. *J Trauma* 65(4):915–923, 2008.
26. Chen L, Reisner AT, Gribok A, McKenna TM, Reifman J: Can we improve the clinical utility of respiratory rate as a monitored vital sign? *Shock* 31(6): 574–580, 2009.
27. Chen L, McKenna TM, Reisner AT, Gribok A, Reifman J: Decision tool for the early diagnosis of trauma patient hypovolemia. *J Biomed Inform* 41(3): 469–478, 2008.
28. Dietterich TG: Ensemble methods in machine learning. In *Multiple Classifier Systems*. Vol. 1857. Berlin, Germany: Springer, pp 1–15, 2000.
29. Wald A: Sequential tests of statistical hypotheses. *Ann Math Stat* 16(2): 117–186, 1945.
30. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG: Research electronic data capture (REDCap)—a metadata-driven methodology and workflow process for providing translational research informatics support. *J Biomed Inform* 42(2):377–381, 2009.
31. Mutschler M, Nienaber U, Munzberg M, Wolfl C, Schoechl H, Paffrath T, Bouillon B, Maegele M, The TraumaRegister, DGU: The Shock Index revisited—a fast guide to transfusion requirement? A retrospective analysis on 21,853 patients derived from the TraumaRegister DGU®. *Crit Care* 17(4):R172, 2013.
32. Bradley EH, Nallamothu BK, Curtis JP, Webster TR, Magid DJ, Granger CB, Moscucci M, Krumholz HM: Summary of evidence regarding hospital strategies

to reduce door-to-balloon times for patients with ST-segment elevation myocardial infarction undergoing primary percutaneous coronary intervention. *Crit Pathw Cardiol* 6(3):91–97, 2007.
33. Cabana MD, Rand CS, Powe NR, Wu AW, Wilson MH, Abboud PA, Rubin HR: Why don't physicians follow clinical practice guidelines? A framework for improvement. *JAMA* 282(15):1458–1465, 1999.
34. Spector JM, Lashoher A, Agrawal P, Lemer C, Dziekan G, Bahl R, Mathai M, Merialdi M, Berry W, Gawande AA: Designing the WHO Safe Childbirth Checklist program to improve quality of care at childbirth. *Int J Gynaecol Obstet* 122(2):164–168, 2013.
35. Arriaga AF, Bader AM, Wong JM, Lipsitz SR, Berry WR, Ziewacz JE, Hepner DL, Boorman DJ, Pozner CN, Smink DS, et al. Simulation-based trial of surgical-crisis checklists. *N Engl J Med* 368(3):246–253, 2013.
36. Pronovost P: Interventions to decrease catheter-related bloodstream infections in the ICU: the Keystone Intensive Care Unit Project. *Am J Infect Control* 36(10):S171.e1–S171.e5, 2008.

# 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)$^{\dagger}$** | **97 (87–128)$^{\dagger\dagger}$** | **120 (92–136)$^{\dagger\dagger\dagger}$** |
| | 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)$^{\dagger\dagger}$** | **28 (24–35)$^{\dagger\dagger}$** |
| | 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)$^{\dagger\dagger}$** | **106 (94–117)$^{\dagger\dagger\dagger}$** | **112 (87–125)$^{\dagger\dagger\dagger}$** |
| | 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)$^{\dagger\dagger}$** | **44 (34–48)$^{\dagger\dagger\dagger}$** | **34 (28–49)$^{\dagger\dagger\dagger}$** |
| | 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.

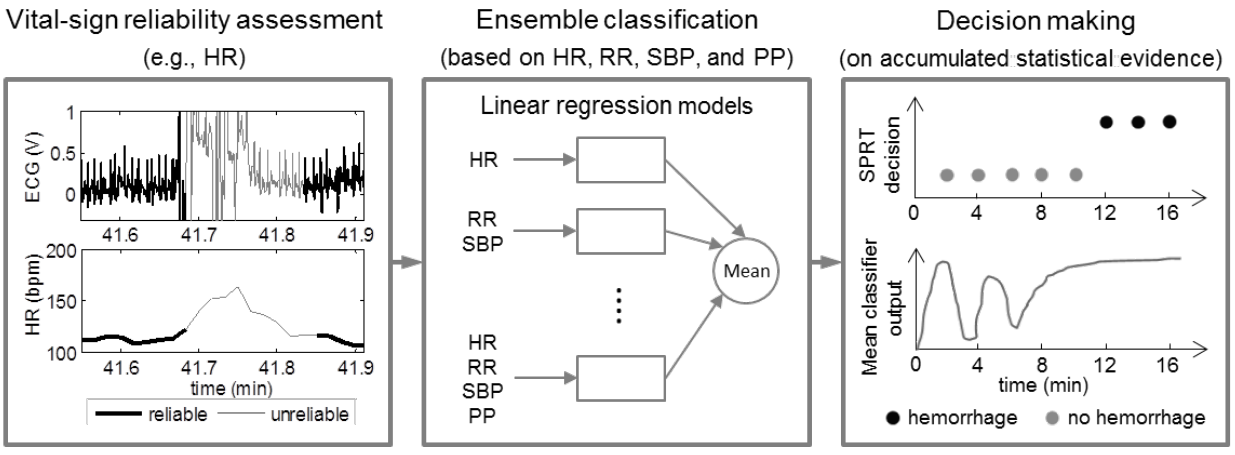



**Figure 2: Analytic methodology for hemorrhage identification. In the first step (left panel), algorithms were used to identify, and exclude, unreliable vital signs. In the second step (middle panel), ensemble classification was applied, which consisted of a set of different linear regression models, 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 $\theta_0$ and $\theta_1$ 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 $A$<br>2. Window size $L$ | $0 < A < 1$<br>$L$ = 2, 15, 60 min |
| Sequential probability ratio testing (SPRT) | 1. Upper limit $A$<br>2. Lower limit $B$<br>3. Window size $L$ | $-2.2 < A < 6.9$<br>$-6.9 < B < 2.2$<br>$L$ = 2, 15, 60 min |
| Risk-adjusted SPRT (RASPRT) | 1. Upper limit $A$<br>2. Lower limit $B$<br>3. Window size $L$ | $-2.2 < A < 6.9$<br>$-6.9 < B < 2.2$<br>$L$ = 2, 15, 60 min |
| Cumulative sum (CUSUM) | 1. Upper limit $A$<br>2. Offset $w$<br>3. Window size $L$ | $0 < A < 1$<br>$0 < w < 1$<br>$L$ = 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%).

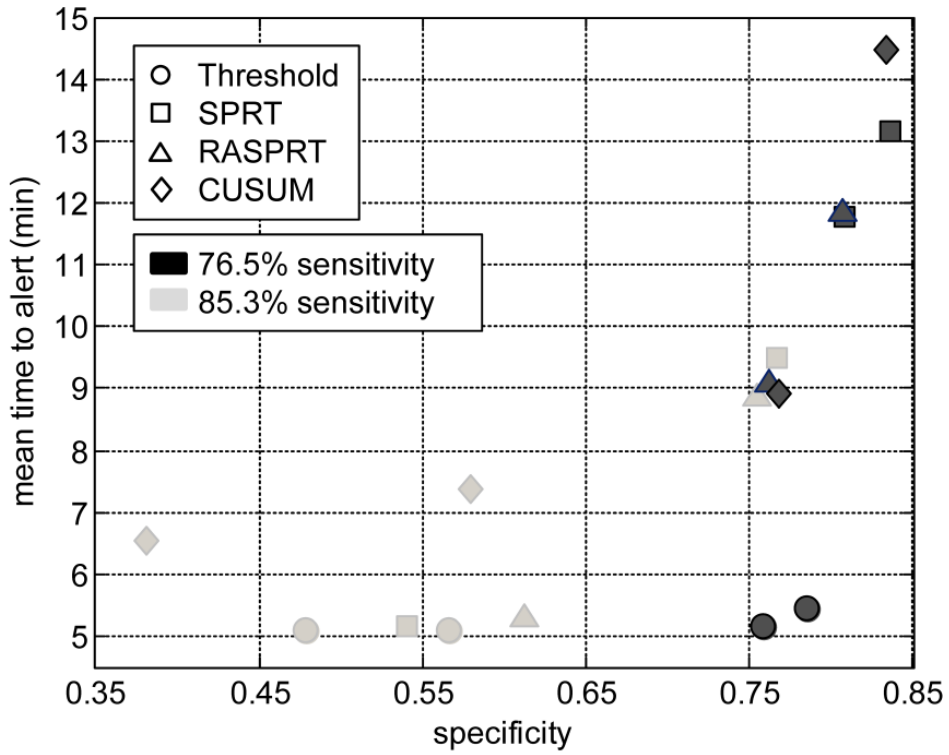



**Figure 3: The trade-off between mean time to alert and specificity at fixed sensitivity levels of 76.5% and 85.3%. The four investigational strategies yielded a spectrum of results varying in sensitivity, specificity, and time to alert (depending on the setting of parameter values; see Table 4). Above, 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.

[15] Nwawka OK, Kabutey NK, Locke CM, Castro-Aragon I, Kim D. Ultrasound-guided needle localization to aid foreign body removal in pediatric patients. J Foot Ankle Surg 2014;53(1):67–70. http://dx.doi.org/10.1053/j.jfas.2013.09.006 [Epub 2013 Nov 13].
[16] Borgohain B, Borgohain N, Handique A, Gogoi PJ. Case report and brief review of literature on sonographic detection of accidentally implanted wooden foreign body causing persistent sinus. Crit Ultrasound J 2012;4(1):10. http://dx.doi.org/10.1186/2036-7902-4-10.
[17] Jecković M, Anupindi SA, Barbir SB, Lovrenski J. Is ultrasound useful in detection and follow-up of gastric foreign bodies in children? Clin Imaging 2013;37(6):1043–7.
[18] Tahmasebi M, Zareizadeh H, Motamedfar A. Accuracy of ultrasonography in detecting radiolucent soft-tissue foreign bodies. Indian J Radiol Imaging 2014;24(2):196–200.

## Utility of shock index calculation in hemorrhagic trauma

We read with great interest the article by Edla et al [1] comparing heart rate variability (HRV) metrics vs routine vital signs as diagnostic tests to improve trauma patient management focusing on the identification of trauma patients with major hemorrhage. They conducted a multivariate analysis using routine vital signs (heart rate, respiratory rate, systolic blood pressure, and pulse pressure) as the comparator to test the hypothesis that HRV metrics can improve the identification of patients with major hemorrhage. However, when combined with routine vital signs, HRV added negligible additional discriminatory value. The authors addressed a very important question as far as the most substantial clinical problem facing physicians being the identification of hemorrhagic trauma. In prehospital setting, current trauma triage relies on abnormal physiological criteria to determine the patient's mode of transport, priority of treatment, destination for treatment, and need for possible life-saving interventions.

We would like to go further into the debate and speculate that calculation of the shock index (SI) may be more useful for caregivers than isolated measurements of systolic blood pressure (SBP) and heart rate (HR) in the compensatory phase of shock. The *SI* is defined as the ratio of HR to SBP. This easily calculable score in the field has been demonstrated to be a pragmatic and useful guide for diagnosing acute hypovolemia in the presence of normal HR and blood pressure. Shock index has been shown to correlate with other indices of end-organ perfusion such as central venous oxygen saturation and arterial lactic acid concentration [2]. Compared with HR or SBP alone, SI has been suggested to be a better measure of hemodynamic stability [3]. Rady et al [4] evaluated a SI cutoff point of 0.9 in a cohort of 275 adult patients presenting to an emergency department with stable vital signs. The authors found that a SI greater than 0.9 was associated with an illness that was treated immediately, admission to the hospital, and intensive therapy on admission. A given set of vital signs may on initial interpretation appear unalarming, but calculation of SI added additional perspective that could influence clinical decisions [5].

To conclude, we would like to know if the authors, maybe based on a retrospective analysis of the data set of 402 subjects, could test the usefulness of SI (with a cutoff value of 0.9) in initial assessment of patients with ongoing exsanguinations?

Marilyn Franchin, MD*
Daniel Jost, MD
Hugues Lefort, MD
Stephane Travers, MD
Jean-Pierre Tourtier, MD
*Emergency Medical Department, Fire Brigade of Paris, Paris, France*
*Corresponding author at: Emergency Medical Department
Fire Brigade of Paris. 1, place Jules Renard F-75017 Paris, France
*E-mail address* marilyn.franchin@pompiersparis.fr

http://dx.doi.org/10.1016/j.ajem.2015.04.001

## References

[1] Edla S, Reisner AT, Liu J, et al. Is heart rate variability better than routine vital signs for pre-hospital identification of major hemorrhage? Am J Emerg Med 2015;33(2):254–61.
[2] Rady MY. The role of central venous oximetry, lactic acid concentration and shock index in the evaluation of clinical shock: a review. Resuscitation 1992;24:55–60.
[3] Birkhahn RH, Gaeta TJ, Terry D, Bove JJ, Tloczkowski J. Shock index in diagnosing early acute hypovolemia. Am J Emerg Med 1994;23:323–6.
[4] Rady MY, Smithline HA, Blake H, Nowak R, Rivers E. A comparison of the shock index and conventional vital signs to identify acute, critical illness in the emergency department. Ann Emerg Med 1994;24:685–90.
[5] Vandromme MJ, Griffin RL, Kerby JD, et al. Identifying risk for massive transfusion in the relatively normotensive patient: utility of the prehospital shock index. J Trauma 2011;70:384–8.

## In reply to "Utility of shock index calculation in hemorrhagic trauma"☆

*To the Editor,*

We wish to thank the correspondents for their interest and comments regarding our report [1]. We agree that multivariate vital-sign analysis is a powerful tool. The Shock Index (SI), which scales the heart rate (HR) to the systolic blood pressure (SBP), is attractive because it can be computed mentally at the bedside. At least in theory, by examining multiple vital signs, one may better distinguish abnormal vital signs due to psychological distress (typically tachycardia with hypertension) vs blood loss and shock (relative tachycardia with normal or reduced blood pressure). In addition to the reports cited by the correspondents, SI has been studied in trauma registries of more than 16 000 [2] and 21 000 [3] patients, demonstrating that blood transfusion requirement and mortality are associated with increasing SI.

To address the question posed by the correspondents, we computed the areas under receiver operating characteristic curves (ROC AUCs) for SI using the same data set of 402 subjects from Edla et al [1]. We used that report's methodology for excluding unreliable vital signs and analyzed the average vital-sign values from each subject's initial 15 minutes of physiological data. The ROC AUCs for SI were 0.76, 0.80, and 0.81 for predicting 24-hour red blood cell transfusion greater than or equal to 1, 5, and 9 units, respectively. These ROC AUCs for SI trend higher than the ROC AUCs for HR and SBP (available in Table 2 from Edla et al [1]), although the differences were not statistically significant. The sensitivity and specificity of SI greater than 0.9 as a predictor of massive transfusion (defined as 24-hour red blood cell transfusion ≥ 9 units) were 63% and 83%, respectively, using the 15-minute average of SBP and HR.

One challenge of SI is that its value changed minute by minute because the patient's HR fluctuated. Many patients developed SI greater than 0.9 at least at some time point during early trauma care. In a separate analysis of 855[1] subjects during prehospital transport [4], we found that 57% of the patients with *no* significant bleeding nonetheless demonstrated SI greater than 0.9, at least transiently. We found that SI greater than 1.4 was a more practical cutoff, with a false-positive rate of only 12% in patients without bleeding; and it was sensitive to 59% of massive transfusion patients. (For comparison, note that SBP < 90 mmHg had a false-positive rate of 10% in patients without bleeding; and it was sensitive to 50% of massive transfusion patients.)

At the bedside, clinicians should consider computing SI using a time-averaged value of HR and SBP from a multiminute observation interval to reduce false alarms [5]. There are also statistical techniques that can

☆ Conflicts of interest: None of the authors have any conflicts of interest to disclose.

[1] The 402 subjects from Edla et al [1] comprise the subset of this larger data set of 855 subjects [4] with a full set of reliable vital signs and at least 5 minutes of reliable electrocardiogram waveform data for heart rate variability analysis.

objectively distinguish transient vs clinically meaningful vital-sign abnormalities in trauma patients and that have been shown to be significantly superior to SI alone, but these techniques require specialized bedside computing capabilities [4].

## Acknowledgments

This work was supported by the US Department of Defense Medical Research and Development Program (grant D10_I_AR_J6_773) and by the Combat Casualty Care Research Area Directorate of the US Army Medical Research and Materiel Command, Fort Detrick, MD, USA. The study sponsors did not have any role in the study design, data collection, analysis and interpretation of data, report writing, or decision to submit the article for publication.

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 US Army or of the US Department of Defense. This correspondence has been approved for public release with unlimited distribution.

Shwetha Edla, PhD
*Department of Defense Biotechnology High Performance Computing Software Applications Institute, Telemedicine and Advanced Technology Research Center, US Army Medical Research and Materiel Command, Fort Detrick, MD 21702, USA*

Andrew T. Reisner, MD
*Department of Emergency Medicine, Massachusetts General Hospital Boston, MA 02114, USA*

Jianbo Liu, PhD
*Department of Defense Biotechnology High Performance Computing Software Applications Institute, Telemedicine and Advanced Technology Research Center, US Army Medical Research and Materiel Command, Fort Detrick, MD 21702, USA*

Victor A. Convertino, PhD
Robert Carter III, PhD
*US Army Institute of Surgical Research, Fort Sam Houston, TX 78234, USA*

Jaques Reifman, PhD
*Department of Defense Biotechnology High Performance Computing Software Applications Institute, Telemedicine and Advanced Technology Research Center, US Army Medical Research and Materiel Command, Fort Detrick, MD 21702, USA*
Corresponding author. Department of Defense Biotechnology High Performance Computing Software Applications Institute, Telemedicine and Advanced Technology Research Center, US Army Medical Research and Materiel Command, ATTN: MCMR-TT, 504 Scott St, Fort Detrick, MD 21702, USA. Tel.: +1 301 619 7915; fax: +1 301 619 1983
*E-mail address:* jaques.reifman.civ@mail.mil

http://dx.doi.org/10.1016/j.ajem.2015.04.002

## References

[1] Edla S, Reisner AT, Liu J, Convertino VA, Carter III R, Reifman J. Is heart rate variability better than routine vital signs for prehospital identification of major hemorrhage? Am J Emerg Med 2015;33:254–61.
[2] Zarzaur BL, Croce MA, Fischer PE, Magnotti LJ, Fabian TC. New vitals after injury: shock index for the young and age × shock index for the old. J Surg Res 2008;147:229–36.
[3] Mutschler M, Nienaber U, Munzberg M, Wolfl C, Schoechl H, Paffrath T, et al. The Shock Index revisited—a fast guide to transfusion requirement? A retrospective analysis on 21,853 patients derived from the TraumaRegister DGU. Crit Care 2013;17:R172.
[4] Liu J, Khitrov MY, Gates JD, Odom SR, Havens JM, de Moya MA, et al. Automated analysis of vital signs to identify patients with substantial bleeding prior to hospital arrival: a feasibility study. Shock 2015. http://dx.doi.org/10.1097/SHK.0000000000000328.
[5] Chen L, Reisner AT, Gribok A, Reifman J. Exploration of prehospital vital sign trends for the prediction of trauma outcomes. Prehosp Emerg Care 2009;13:286–94.

## Peritoneal dialysis and potassium: pains and gains in the ED

*To the Editor,*

The article by Roseman et al [1] is indeed interesting, as the authors had brought peritoneal dialysis (PD) back to frontline and as an option for patients with severe hyperkalemia in resource-limited emergency department. However, few aspects of this article need contemplation based on our experiences with regard to potassium clearance [2]. It is well known that potassium clearance achieved by PD is markedly lower than hemodialysis.

Clearance of potassium averages approximately 17 mmol/min for intermittent PD and approximately 7 mmol/min for continuous ambulatory peritoneal dialysis (CAPD). Interestingly, higher potassium clearance (24 mmol/min) is obtained during the first hour than that of the remaining period due increased release of potassium from the cells that line the peritoneal cavity. Peritoneal dialysis patients have normal or low plasma potassium probably because of greater shift of this ion into intracellular compartment, which is facilitated by initial low pH and/or by the hyperosmolality of the instilled dialysate, which does not contain potassium [3].

Thus, patients on PD in general have high intracellular potassium content, more so those on CAPD. This process is also further enhanced due to the continuous glucose absorption from the dialysis solutions and the subsequent stimulation of intracellular uptake of potassium, mediated by insulin. However, potassium entry into peritoneal epithelium declines as patients on PD started developing peritoneal sclerosis. This intracellular overload is not only difficult to correct but also makes them susceptible for hyperkalemia easily [4]. After removal of potassium from extracellular compartment by dialysis, there will be a rebound as the intracellular potassium moves to extracellular compartment. This continues till the total body potassium is depleted. Hence, to solve these problems, there is a need for a long and sustained dialysis using a 2-L CAPD exchange 4 times per day with potassium-free dialysate [5]. We have also noticed normalization of plasma potassium levels and steady state of plasma potassium of 5 mmol/L in our cases [2]. One can estimate the potassium removal close to 33 to 35 mmol/d to avoid hyperkalemic rebound in the postdialytic period.

Peritoneal dialysis offers a unique and timely opportunity for the emergency physician to rescue; however, the limitations of potassium exchange and noninfectious complications of PD have to be kept in the mind, and the alternatives have to be discussed with patients and caregivers before preparing them for PD.

Subramanian Senthilkumaran, MD
*Department of Emergency & Critical Care Medicine, Sri Gokulam Hospital Salem, Tamil Nadu, India*
Corresponding author: Department of Emergency & Critical Care Medicine Sri Gokulam hospital, Salem, Tamil Nadu, India
Tel.: +91 9994634444, +91 9846349444 (Mobile)
*E-mail address:* maniansenthil@yahoo.co.in

Florence Benita, MBBS
*Department of Emergency Medicine*
*Kokilaben Dhirubhai Ambani Hospital and Medical Research Institute Mumbai, India*

Gore B. Sandeep, MBBS
*Department of Accident & Emergency Medicine, Fortis Hospital Mumbai, India*

# THE GEORGE WASHINGTON UNIVERSITY

## Faculty Positions-Emergency Medicine

The George Washington University Medical Faculty Associates, an independent non-profit academic clinical practice group affiliated with The George Washington University, is seeking full-time academic Emergency Medicine physicians. The Department of Emergency Medicine (**http://smhs.gwu.edu/emed/**) provides staffing for the emergency units of George Washington University Hospital, United Medical Center, the Walter Reed National Military Medical Center, and the Washington DC Veterans Administration Medical Center. The Department's educational programs include a four-year residency and ten fellowship programs.

Responsibilities include providing clinical and consultative service; teaching Fellows, Residents, and Medical Students; and maintaining an active research program. These non-tenure track appointments will be made at a rank (Instructor/Assistant/Associate/Full Professor) and salary commensurate with experience.

**Basic Qualifications: Applicants must** be ABEM or AOBEM certified, or have completed an ACGME or AOA certified Emergency Medicine residency, **and** be eligible for licensure in the **District of Columbia, at the time of appointment**.

**Application Procedure:** Complete the online faculty application at **http://www.gwu.jobs/postings/56800** and upload a CV and cover letter. Review of applications will be ongoing beginning November 30, 2018 and will continue until positions are filled. Only complete applications will be considered. Employment offers are contingent on the satisfactory outcome of a standard background screening. Questions about these positions may be directed to Department Chair, Robert Shesser M.D., at rshesser@mfa.gwu.edu.

Academic Emergency Medicine
Official Journal of the Society for Academic Emergency Medicine

BRIEF REPORT

# Muscle Oxygen Saturation Improves Diagnostic Association Between Initial Vital Signs and Major Hemorrhage: A Prospective Observational Study

Andrew T. Reisner, MD, Shwetha Edla, PhD, Jianbo Liu, PhD, John T. Rubin, Jill E. Thorsen, Erin Kittell, MSN, Jason B. Smith, Daniel D. Yeh, MD, and Jaques Reifman, PhD

## Abstract

***Objectives:*** During initial assessment of trauma patients, vital signs do not identify all patients with life-threatening hemorrhage. We hypothesized that a novel vital sign, muscle oxygen saturation ($SmO_2$), could provide independent diagnostic information beyond routine vital signs for identification of hemorrhaging patients who require packed red blood cell (RBC) transfusion.

***Methods:*** This was an observational study of adult trauma patients treated at a Level I trauma center. Study staff placed the CareGuide 1100 tissue oximeter (Reflectance Medical Inc., Westborough, MA), and we analyzed average values of $SmO_2$, systolic blood pressure (sBP), pulse pressure (PP), and heart rate (HR) during 10 minutes of early emergency department evaluation. We excluded subjects without a full set of vital signs during the observation interval. The study outcome was hemorrhagic injury and RBC transfusion $\geq 3$ units in 24 hours (24-hr RBC $\geq 3$). To test the hypothesis that $SmO_2$ added independent information beyond routine vital signs, we developed one logistic regression model with HR, sBP, and PP and one with $SmO_2$ in addition to HR, sBP, and PP and compared their areas under receiver operating characteristic curves (ROC AUCs) using DeLong's test.

***Results:*** We enrolled 487 subjects; 23 received 24-hr RBC $\geq 3$. Compared to the model without $SmO_2$, the regression model with $SmO_2$ had a significantly increased ROC AUC for the prediction of $\geq 3$ units of 24-hr RBC volume, 0.85 (95% confidence interval [CI], 0.75–0.91) versus 0.77 (95% CI, 0.66–0.86; $p < 0.05$ per DeLong's test). Results were similar for ROC AUCs predicting patients ($n = 11$) receiving 24-hr RBC $\geq 9$.

***Conclusions:*** $SmO_2$ significantly improved the diagnostic association between initial vital signs and hemorrhagic injury with blood transfusion. This parameter may enhance the early identification of patients who require blood products for life-threatening hemorrhage.

ACADEMIC EMERGENCY MEDICINE 2016;23:353–357 © 2016 by the Society for Academic Emergency Medicine

From the Department of Emergency Medicine, Massachusetts General Hospital (ATR, JTR, JET, EK), Boston, MA; the Department of Defense Biotechnology High Performance Computing Software Applications Institute, Telemedicine and Advanced Technology Research Center, U.S. Army Medical Research & Materiel Command (SE, JL, JBS, JR), Fort Detrick, MD; and the Department of Surgery, Massachusetts General Hospital (DDY), Boston, MA. Mr. Smith is currently with the National Institute of Mental Health Data Archive, OMNITEC Solutions, Inc., Bethesda, MD.

Presented at the 19th Annual New England Society for Academic Emergency Medicine Regional Conference (NERDS), Newton, MA, April 1, 2015; and the Society of Academic Emergency Medicine Annual Meeting, San Diego, CA, May 12–15, 2015.

This work was supported by the Combat Casualty Care Research Area Directorate of the U.S. Army Medical Research and Materiel Command, Fort Detrick, MD. The study sponsors did not have any role in the study design, data collection, analysis and interpretation of data, report writing, or decision to submit the article for publication. 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.

Received July 15, 2015; revision received October 7, 2015; accepted October 9, 2015.

The authors have no relevant financial information or potential conflicts to disclose.

Supervising Editor: Shahriar Zahtabchi, MD.

Address for correspondence and reprints: Jaques Reifman, PhD; e-mail: jaques.reifman.civ@mail.mil.



ISSN 1069-6563
PII ISSN 1069-6563583





doi: 10.1111/acem.12899

To help clinicians determine which trauma patients have life-threatening hemorrhage, several clinical scores have been developed for predicting the need for massive blood transfusion.[1] However, these scores cannot be computed immediately upon emergency department (ED) arrival because they require either blood testing or imaging results. An alternative methodology that could be used upon initial evaluation (or even during prehospital transport) for detecting significant risk of exsanguination could be useful. Near-infrared spectrometry (NIRS), which has shown to correlate with high acuity and poor outcomes in trauma patients,[2–4] has been studied specifically as a triage tool in one study, where it was found to predict the need for blood transfusion in combat casualties who lacked early hypotension.[5] To further investigate whether NIRS offers potential value as a triage tool in the preliminary assessment of trauma patients upon ED arrival (i.e., prior to the availability of any diagnostic testing except for routine vital signs), we undertook an investigation, testing the hypothesis that NIRS tissue oxygen monitoring improves the early identification of patients with major hemorrhage compared with initial vital signs alone.

## MATERIALS AND METHODS

### Study Setting and Population

We received protocol approval from the institutional review board (IRB), including a waiver of informed consent as per 45 CFR § 46.116(d). We studied a convenience sample of trauma patients ≥ 18 years of age evaluated in the ED of a Level I trauma center. A priori exclusion criteria were: 1) transfer from another hospital if prior workup already ruled out hemorrhagic injury; 2) no suitable NIRS sensor placement site overlying the deltoid or thigh due to either tattoos, visible skin injury, gross blood, visible rash, clothing, request of treating clinician, or evident hirsutism (when patients had visible body hair, we did not attempt to place the oximeter because subsequent removal of the adhesive from hairy skin was expected to be painful, and the subjects had not provided consent for any painful procedure); 3) per manufacturer's recommendation, estimated body mass index < 19 or > 40 kg/m$^2$; 4) minor trauma, e.g., fall from standing to flat ground; and 5) failure to record muscle oxygen saturation ($SmO_2$), heart rate (HR), and blood pressure (BP) within a matching 10-minute interval during the patient's initial evaluation.

### Measurements

$SmO_2$ was measured using the CareGuide 1100 tissue oximeter (Reflectance Medical, Inc., Westborough, MA) placed by dedicated study staff on skin overlying the deltoid or thigh. The sensor remained in place for a minimum of 3 minutes. The CareGuide sensor measures $SmO_2$ using principles that are similar to other NIRS oximeters, while incorporating proprietary technology that is designed to eliminate spectral inference from skin pigmentation and fat.[6]

Vital signs were measured as per clinical routine using Solar patient monitors (General Electric, Milwaukee, WI). In most ED bays, data were electronically archived using BedMasterEx software (Excel Medical, Jupiter, FL). Unreliable vital sign data were identified and excluded using validated software algorithms.[7] For ED bays that lacked the BedMasterEx system, we relied on the vital signs documented by ED nurses and corroborated by vital signs simultaneously documented by dedicated study staff.

### Outcome

We studied the prediction of patients with hemorrhagic injuries and the receipt of ≥3 units of packed red blood cells (RBCs) in the first 24 hours. (Hemorrhagic injury was defined as any of the following: laceration or fracture of a solid organ; documented hematoma within the thorax, peritoneum, retroperitoneum, or pelvis; vascular injury that required operative repair or angioembolization; or limb amputation.) The secondary outcome was receipt of ≥9 units of RBCs in patients with hemorrhagic injury. Injuries, injury severity score, Glasgow coma scale, and operative interventions were obtained from the medical record (clinical documentation, radiology reports, and operative reports) and the trauma registry. Data were archived electronically using REDCap.[8] We determined whether or not the patient had documented hemorrhagic injury by automated text search, searching for injuries that met the aforementioned criteria (all records were also jointly reviewed by two investigators to confirm that the automated text search had not omitted any applicable hemorrhagic injuries, nor included nonhemorrhagic injuries).

Patients who received RBCs but lacked a documented hemorrhagic injury were excluded from analysis, because of unresolvable uncertainty about whether the RBC transfusion was clinically indicated in the absence of explicitly hemorrhagic injuries.

### Data Analysis

We computed the mean values of $SmO_2$, HR, systolic BP (sBP), pulse pressure (PP = sBP – diastolic BP), and the shock index (SI = HR/sBP) measurements from a 10-minute window starting upon the first simultaneous occurrence of a full set of HR, BP, and $SmO_2$ values.

We applied DeLong's test to the areas under receiver operating characteristic curves (ROC AUCs) from two logistic regression models, the first using only routine vital signs (HR, sBP, PP) and the second adding the investigational metric (HR, sBP, PP, $SmO_2$). The null hypothesis was that $SmO_2$ did not provide additional diagnostic information compared with using routine vital signs alone.

## RESULTS

Between June 2012 and October 2014, we enrolled 487 subjects. Figure 1 shows the enrollment flowchart. Subjects were predominantly male (68%), mechanism of injury was predominantly blunt (90%), and median age was 47 years (interquartile range [IQR] = 31–64). Median injury severity score was 16 (IQR = 9–25) and mortality was 3%. There were 23 patients who received ≥3 units of RBCs within 24 hours and 11 who received ≥9 units.

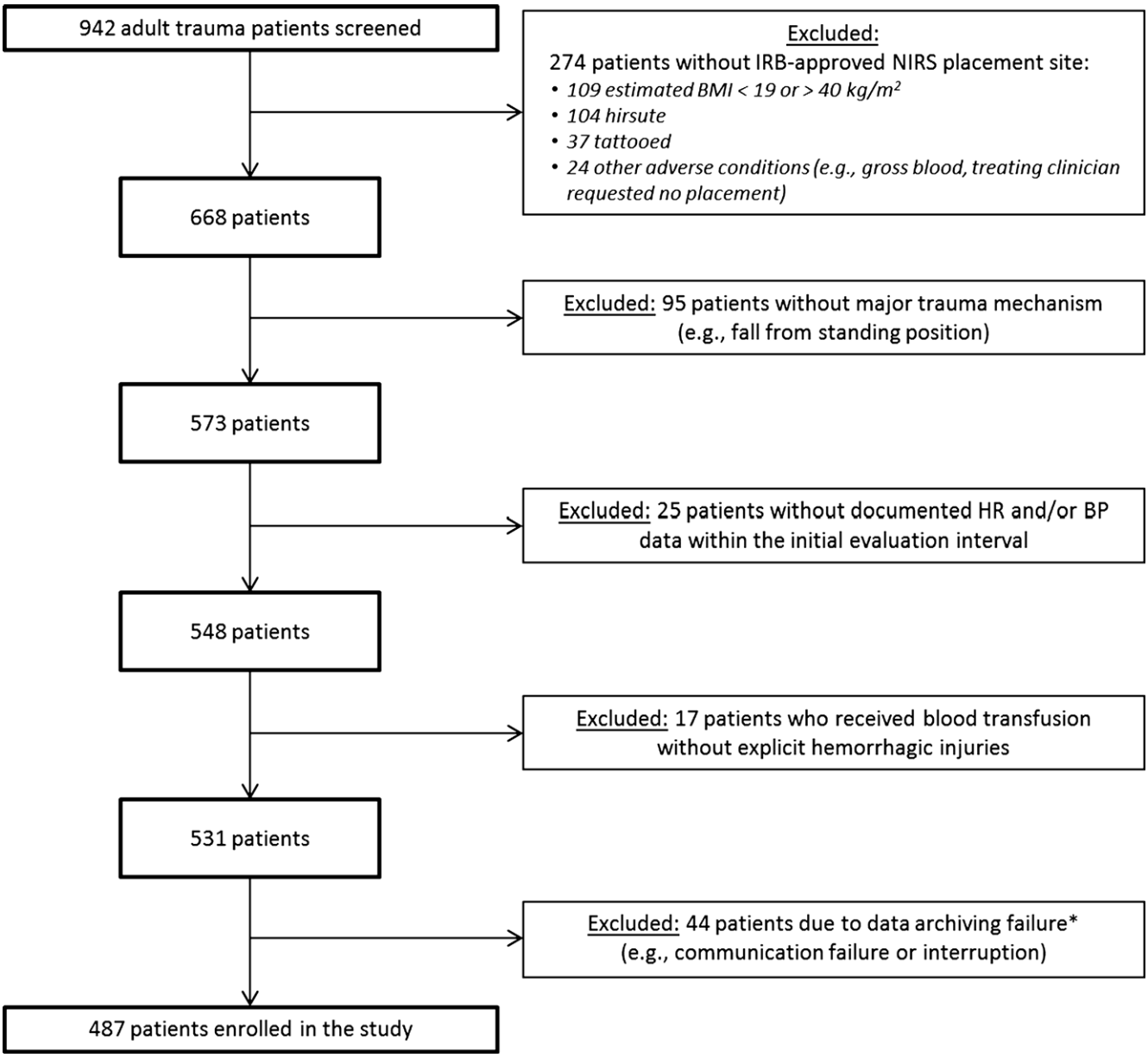



**Figure 1.** Flowchart of subject enrollment. *Data archiving failures involved the archiving system (ruggedized GoBook personal computer connected to the CareGuide $SmO_2$ sensor) that we assembled for this investigation. For the final 16 months of the investigation, we reinforced the electronic and mechanical connectivity, and had only one additional subject with data archiving failure in that time interval. BMI = body mass index; BP = blood pressure; HR = heart rate; IRB = institutional review board; NIRS = near-infrared spectrometry.

$SmO_2$, BP, and HR were collected early in the subjects' clinical courses: the median time elapsed between the ED admission time and the onset of the initial evaluation time window (i.e., the 10-minute window used for analysis) was 3.65 minutes (IQR = 1.85–6.22 minutes). Median values for vital signs and $SmO_2$ are presented in Table 1.

In terms of diagnostic association between hemorrhagic injury with 24-hour RBC transfusion volume ≥ 3 units:

- The multivariate regression model using HR, sBP, and PP *alone* yielded a ROC AUC of 0.77 (95% confidence interval [CI], 0.66–0.86), which was similar to the ROC AUC for SI (Table 1).
- The multivariate regression model using HR, sBP, and PP *plus* $SmO_2$ yielded a ROC AUC of 0.85 (95% CI, 0.75–0.91).

Per DeLong's test, these ROC AUCs were significantly different ($p < 0.05$).

Repeating the same analysis for the alternative RBC cutoff, i.e., ≥9 units of RBCs within 24 hours, we found similar results: the regression model that included $SmO_2$ in addition to HR, sBP, and PP yielded an increased ROC AUC (0.89; 95% CI, 0.76–0.95) that was

Table 1
Association of Initial Vital Signs and $SmO_2$ With 24-Hour RBC Transfusion Volume ≥ 3 Units

| | Median (IQR) for Patients With 24-Hour RBC Volume < 3 Units ($n$ = 464) | Median (IQR) for Patients With 24-Hour RBC Volume ≥ 3 Units ($n$ = 23) | ROC AUCs (95% CIs) |
|---|---|---|---|
| HR (beats/min) | 83 (70–96) | 99 (82–119) | 0.70 (0.56–0.81) |
| sBP (mm Hg) | 141 (127–158) | 136 (96–147) | 0.62 (0.47–0.75) |
| PP (mm Hg) | 61 (50–73) | 48 (33–63) | 0.68 (0.54–0.80) |
| SI ([beats/min]/[mm Hg]) | 0.59 (0.49–0.70) | 0.80 (0.63–0.94) | 0.75 (0.61–0.85) |
| $SmO_2$ (%) | 67 (62–72) | 61 (50–64) | 0.76 (0.65–0.84) |

HR = heart rate; IQR = interquartile range; PP = pulse pressure (= sBP – diastolic blood pressure); RBC = packed red blood cells; ROC AUC = area under the receiver operating characteristic curve; sBP = systolic blood pressure; SI = shock index (= HR/sBP); $SmO_2$ = muscle oxygen saturation.

significantly greater ($p < 0.05$) than the ROC AUC (0.77; 95% CI, 0.61–0.87) of the regression model with HR, sBP, and PP alone.

## DISCUSSION

In this investigation, we found that $SmO_2$ added significant discriminatory information beyond the initial values of HR and BP. The significantly higher ROC AUC implies that using $SmO_2$ offered a higher combination of sensitivity and specificity than could be achieved using only routine vital signs.

Why would low $SmO_2$ indicate a patient with life-threatening hemorrhage, if routine vital signs are not patently abnormal? Presumably, such a patient would be physiologically compensating for blood loss with marked peripheral vasoconstriction, maintaining BP during blood loss, but at the expense of peripheral tissue hypoperfusion (and a resultant low $SmO_2$). We note that some hemorrhagic patients with low $SmO_2$ also had simultaneous hypotension, while others did not.

Our findings suggest that $SmO_2$ can enhance initial vital signs, but cannot replace them: some hemorrhagic patients had hypotension but without low $SmO_2$. We speculate that this reflects individual variability in vasoconstriction during blood loss:[9] Patients with maximum vasoconstriction can maintain BP at the expense of peripheral perfusion/$SmO_2$, while those patients with minimal vasoconstriction can maintain peripheral perfusion/$SmO_2$ while experiencing earlier hypotension. (As well, there were hemorrhagic patients with neither hypotension nor reduced $SmO_2$; possibly these patients had not yet suffered significant blood loss or had baseline hypertension and/or bradycardia that masked their progression into hypovolemic physiology.)

The current investigation provides evidence that NIRS oximetry can provide additional information for early detection of hemorrhage than initial routine vital signs alone. These findings are consistent with a prior report[5] evaluating NIRS oximetry as a triage tool, which likewise found that it could predict the need for blood transfusion in patients who lacked early hypotension. The NIRS sensor is relatively easy to place, attaching to the patient's skin via an adhesive sleeve. A simple measure that can improve the early identification of patients with major hemorrhage may be particularly useful when caregivers are novice, distracted, or fatigued.

## LIMITATIONS

A substantial number of subjects were excluded for lacking a suitable IRB-approved NIRS placement site, and our outcome, blood transfusion, involved subjective clinical decision-making; both factors limit the generalizability of the findings. This investigation only analyzed the initial ED assessment, and our findings do not address whether or not better information than initial vital signs alone translates into better clinical judgments and patient outcomes. We did not evaluate whether there is value of NIRS oximetry in later phases of trauma care when additional sources of diagnostic data are available, i.e., vital sign trends through time, lab results, such as lactate and base deficit, and imaging. Sources of diagnostic error for the CareGuide sensor, whether it is more reliable than other NIRS oximeters and how to interpret temporal trends in $SmO_2$, were not investigated. We note that the State of Minnesota has added tissue spectroscopy to their official guidelines for "Tier-One Trauma Team Activation Criteria."[10] Our findings corroborate the usefulness of this metric for early identification of major hemorrhage, but also highlight remaining questions about relying on this technology for clinical decision-making.

## CONCLUSIONS

We compared muscle oxygen saturation to heart rate, sBP, and pulse pressure alone in the early ED evaluation of trauma patients and found that use of muscle oxygen saturation significantly improved the diagnostic association between vital signs and hemorrhagic injury requiring blood transfusion. The results offered prima facie evidence that near-infrared spectrometry might provide a tool for the early ED identification of patients with life-threatening hemorrhage.

## References

1. Brockamp T, Nienaber U, Mutschler M, et al. Predicting on-going hemorrhage and transfusion requirement after severe trauma: a validation of

six scoring systems and algorithms on the TraumaRegister DGU. Crit Care 2012;16:R129.
2. Cohn SM, Nathens AB, Moore FA, et al. Tissue oxygen saturation predicts the development of organ dysfunction during traumatic shock resuscitation. J Trauma 2007;62:44–54.
3. Moore FA, Nelson T, McKinley BA, et al. Massive transfusion in trauma patients: tissue hemoglobin oxygen saturation predicts poor outcome. J Trauma 2008;64:1010–23.
4. Sagraves SG, Newell MA, Bard MR, et al. Tissue oxygenation monitoring in the field: a new EMS vital sign. J Trauma 2009;67:441–3.
5. Beekley AC, Martin MJ, Nelson T, et al. Continuous noninvasive tissue oximetry in the early evaluation of the combat casualty: a prospective study. J Trauma 2010;69(Suppl 1):S14–25.
6. Yang Y, Soyemi O, Scott PJ, et al. Quantitative measurement of muscle oxygen saturation without influence from skin and fat using continuous-wave near infrared spectroscopy. Opt Express 2007;15:13715–30.
7. Reisner AT, Chen L, McKenna TM, Reifman J. Automatically-computed prehospital severity scores are equivalent to scores based on medic documentation. J Trauma 2008;65:915–23.
8. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)–a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform 2009; 42:377–81.
9. Convertino VA, Rickards CA, Ryan KL. Autonomic mechanisms associated with heart rate and vasoconstrictor reserves. Clin Auton Res 2012;22:123–30.
10. Minnesota Department of Health. Level 3 Trauma Hospital Criteria. Available at: http://www.health.state.mn.us/traumasystem/hospresources/criteria_level3.pdf. Accessed Oct 6, 2015.

Injury, Int. J. Care Injured 49 (2018) 1654–1660

Contents lists available at ScienceDirect

Injury

journal homepage: www.elsevier.com/locate/injury

# Tachycardic and non-tachycardic responses in trauma patients with haemorrhagic injuries

Andrew T. Reisner[a], Shwetha Edla[b], Jianbo Liu[b], Jiankun Liu[b], Maxim Y. Khitrov[b], Jaques Reifman[b,*]

[a] Department of Emergency Medicine, Massachusetts General Hospital, Boston, MA, USA
[b] 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, USA

ARTICLE INFO

*Article history:*
Accepted 29 April 2018

*Keywords:*
ATLS
Blood loss
Heart rate
Tachycardia
Trauma
Vital signs

ABSTRACT

*Background:* Analyses of large databases have demonstrated that the association between heart rate (HR) and blood loss is weaker than what is taught by Advanced Trauma Life Support training. However, those studies had limited ability to generate a more descriptive paradigm, because they only examined a single HR value per patient.
*Methods:* In a comparative, retrospective analysis, we studied the temporal characteristics of HR through time in adult trauma patients with haemorrhage, based on documented injuries and transfusion of ≥3 units of red blood cells (RBCs). We analysed archived vital-sign data of up to 60 min during either pre-hospital or emergency department care.
*Results:* We identified 133 trauma patients who met the inclusion criteria for major haemorrhage and 1640 control patients without haemorrhage. There were 55 haemorrhage patients with a normal median HR and 78 with tachycardia. Median ΔHR was −0.8 and +0.7 bpm per 10 min, respectively. Median time to documented hypotension was 8 and 5 min, respectively. RBCs were not significantly different; median volumes were 6 (IQR: 4–13) and 10 units (IQR: 5–16), respectively. Time-to-hypotension and mortality were not significantly different. Tachycardic patients were significantly younger ($P < 0.05$). Only 10 patients with normal HR developed transient/temporary tachycardia, and only 11 tachycardic patients developed a transient/temporary normal HR.
*Conclusions:* The current analysis suggests that some trauma patients with haemorrhage are continuously tachycardic while others have a normal HR. For both cohorts, hypotension typically develops within 30 min, without any consistent temporal increases or trends in HR.

Published by Elsevier Ltd.

## Introduction

Multiple reports have demonstrated that the current Advanced Trauma Life Support (ATLS) training course is inaccurate regarding vital-sign changes in trauma patients with haemorrhage [1–4]. Analyses of large datasets have demonstrated that the association between heart rate (HR) and blood loss is weaker than what is taught by ATLS [2,3]. Studying nearly 200,000 trauma patients in a trauma registry, Guly et al. [2] reported that "[w]ith increasing estimated blood loss there is a trend to increasing HR and a reduction in systolic blood pressure (SBP), but not to the degree suggested by the ATLS classification of shock." Studying over 35,000 trauma patients, Mutschler et al. [3] concluded that "[t]his study indicates that the ATLS classification of hypovolaemic shock does not seem to reflect clinical reality accurately."

If it has been established that ATLS is not accurate in describing HR changes during haemorrhage, an alternative paradigm describing HR patterns in trauma patients has not emerged. In part, this is because the aforementioned large registry studies only examined a single HR value per patient, whereas in reality, HR is continuously monitored during trauma patient management. By studying only single HR values per patient, it cannot be determined how often tachycardia develops as haemorrhage progresses. As well, it cannot be determined whether the weak association between HR and haemorrhage was *i*) because HR varied substantially in individual trauma patients (i.e. large intra-subject variability), and/or *ii*) because HR

* Corresponding author at: Department of Defense Biotechnology High Performance Computing Software Applications Institute, Telemedicine and Advanced Technology Research Center, U.S. Army Medical Research and Materiel Command, ATTN: MCMR–TT, 504 Scott Street, Fort Detrick, MD 21702, USA.
*E-mail address:* jaques.reifman.civ@mail.mil (J. Reifman).

https://doi.org/10.1016/j.injury.2018.04.032
0020-1383/Published by Elsevier Ltd.

responses varied substantially between patients (i.e. large inter-subject variability).

Having a better understanding of the temporal characteristics of HR through time in trauma patients with haemorrhage could contribute to a more accurate and useful alternative to ATLS. Accordingly, we analysed an archived dataset of continual vital signs in trauma patients, seeking to characterise the HR patterns recorded through time. We evaluated the extent to which trauma patients demonstrated tachycardia over time, and whether there were salient clinical differences between patients who demonstrated different types of HR responses.

## Materials and methods

### *Study design, setting, population, and outcome*

This was a comparative study carried out by a secondary analysis of three pooled datasets. We studied adult trauma patients with haemorrhagic injuries during initial care (either during pre-hospital transport or upon arrival in the emergency department). Dataset 1 was originally collected aboard air ambulances between February 2010 and December 2012 [5], Dataset 2 from an emergency department between June 2012 and December 2014 [6], and Dataset 3 during air transport between August 2001 and April 2004 [7,8]. All datasets were collected with the approval of local institutional review boards.

For our outcome, haemorrhagic injury, we used the following criteria: documented haemorrhagic injuries, and transfusion of three or more units of red blood cells within 24 h (24-h RBCs). Explicitly documented haemorrhagic injuries were identified by chart review, and defined as solid organ injuries, thoracic or abdominal haematomas noted in imaging or operative reports, vascular injuries that required a procedure for haemostasis, or limb amputations.

For Dataset 1, eligible patients were identified by querying the air ambulance administrative database for adult trauma transports. Next, we queried the receiving hospital's electronic medical records to identify the subset who received at least three units of 24-h RBCs. This review was conducted by either a physician or nurse practitioner with clinical experience in trauma care, and who was blinded to subjects' physiological data. These data were collected and managed using REDCap electronic data capture tools [9]. Abstractors were first trained using training cases from Dataset 3. Next, the abstractors' adjudications about whether or not the subject had a haemorrhagic injury were confirmed by running an automated text-search through the trauma registry database, to independently corroborate that the subject had at least one of a list of haemorrhagic injuries. Cohen's K between the data abstractor adjudication and automated text search results was 0.67. All discrepancies were subsequently resolved by two-investigator adjudication.

For Dataset 2, eligible patients were first identified by electronically querying the source hospital's trauma registry for adult trauma patients. The remainder of the subject selection methodology, in terms of 24-h RBC volume and presence of haemorrhagic injury, was the same as that used for Dataset 1.

Data collection for Dataset 3 was conducted under a protocol that yielded an inventory of injuries and 24-h RBCs in a convenience sample of high-acuity trauma patients [7]. The methodology for determining presence of haemorrhagic injury was the same as that used for Dataset 1.

### *Study measurements*

To collect HR and blood pressure (BP) data during real-time care, data streaming from patients' vital-sign monitors were electronically recorded via software solutions [8,10,11]. The electronic recording system used for Dataset 1 was an *ad hoc* software system described by Reisner et al. [10]. The recording system used for Dataset 2 was the BedMasterEx system (Excel Medical, Jupiter FL). The recording system used for Dataset 3 was another *ad hoc* system described by Cooke et al. [7].

From these recordings, we analysed vital-sign data of up to 60 min in duration, beginning with the first recorded non-zero vital sign. We studied HR from intervals with high-quality electrocardiograms (ECGs), as determined by the consensus of an automated algorithm (which has been shown to be more conservative than human expert evaluation [12]) and a human adjudicator. When there was disagreement, a second human adjudicator evaluated the reliability of the data segment.

For Datasets 1 and 2, study staff performed retrospective chart review to extract additional clinical data, including demographics, injury descriptions, clinical interventions, and mortality, using the methodology detailed above. All of these data were compared with an electronic report from the hospital's independent trauma registry, and discrepancies were resolved by two-investigator adjudication. Clinical data abstraction for Dataset 3 was conducted in accord with a previous study [7].

### *Data analysis*

By convention, tachycardia is defined as a HR of 100 bpm or greater. We examined whether 100 bpm was a clinically valid cut-off to discriminate between patients with and without haemorrhage, and calculated the diagnostic testing characteristics of tachycardia and the associated receiver operating characteristic (ROC) curve [13]. To investigate whether patients with haemorrhage demonstrated tachycardia at variable time intervals, we calculated how often those with a normal HR developed transient/temporary tachycardia (at least 5 min of tachycardia within any 10-min time window), and how often those with tachycardia developed a transient/temporary normal HR (at least 5 min of normal HR within any 10-min time window). We also performed a sensitivity analysis to investigate whether our findings were sensitive to the definition of clinical haemorrhage, by computing ROC curves for predicting a set of secondary outcomes: 24-h RBCs $\geq$1, $\geq$3, $\geq$5, $\geq$7, and $\geq$10 units, regardless of documented injuries. In addition to the aforementioned analyses using median HR, we developed a logistic regression model using median HR for estimating the probability of haemorrhage, and tested its goodness-of-fit using the Hosmer-Lemeshow test.

We compared the haemodynamic and clinical characteristics of haemorrhage patients with a normal HR to those of haemorrhage patients with tachycardia. Variability in HR was quantified by the root mean square (RMS) around the mean of each patient's HR time series, while slope of HR as a function of time was computed using linear regression. We computed the BP characteristics of both cohorts, including the incidence of measured hypotension and the time elapsed until hypotension was first measured. Hypotension was defined as an SBP of less than 90 mmHg or a mean arterial pressure (MAP) of less than 70 mmHg. We also computed the pulse pressure (SBP – diastolic BP) and the Shock Index (SI = HR/SBP) for both cohorts. We compared clinical characteristics, including demographics, injury descriptions, clinical interventions, and mortality. We performed analyses in MATLAB version 9.0 (The MathWorks, Inc., Natick, MA). Data distributions were compared using the Wilcoxon rank-sum test for continuous variables and categorical variables using Fisher's exact test. We used a threshold for statistical significance of $P < 0.05$.

Finally, we studied the change in HR in the subset of haemorrhage patients who developed new onset hypotension. New onset hypotension was defined as follows: *i*) at least one non-hypotensive

BP measured within the 10 min *prior to* the first recording of hypotension; and *ii*) at least one subsequent hypotensive BP. We compared the HR and BP characteristics of tachycardic and non-tachycardic haemorrhage patients before and upon the onset of hypotension, and we also compared clinical characteristics. For this subgroup analysis, we only included HR data measured contemporaneously with each BP measurement, i.e. within a 2-min window, to preserve the relationship between HR and BP.

## Results

The overall characteristics of the study population are shown in Table 1. From the three combined datasets we identified 142 patients who met our criteria for haemorrhagic injury (6.5%), nine of which were subsequently excluded for insufficient ECG reliability (high acuity patients with very short ECG recordings). There were 1640 control patients who survived and received no RBC transfusions, 53 of which were excluded for insufficient ECG reliability.

We computed each patient's median ("patient-median") HR. For patients with haemorrhagic injury, the population median of patient-median HR (and interquartile range [IQR]) was 102 (87–126) bpm. For the control patients, the population median of patient-median HR was 87 (74–99) bpm. The area under the receiver operating characteristic curve (ROC AUC) of HR for distinguishing between patients with haemorrhagic injury and control patients was 0.71 (95% CI: 0.65–0.76), which is consistent with a diagnostic test of low-to-moderate accuracy [13]. In our sensitivity analysis, which investigated whether the diagnostic performance of HR varied depending on alternative definitions of haemorrhage, we found similar ROC AUCs for predicting alternative haemorrhage-related outcomes, i.e. 24-h RBCs of $\geq$1, $\geq$3, $\geq$5, $\geq$7, and $\geq$10 units (see Fig. 1). Regarding the suitability of a logistic regression model for estimating the probability of haemorrhage based on median HR, we found no evidence of poor calibration ($p > 0.05$, Hosmer-Lemeshow test). Using the conventional cut-off for tachycardia (HR $\geq$ 100 bpm), HR was 59% sensitive for haemorrhagic injury and 75% specific for the control patients.

Of the patients with haemorrhagic injury, 78 had tachycardia (59% of all trauma patients with haemorrhage) based on median HR. Overall, this cohort had a median vital-sign recording duration of 22 min (IQR: 12–31). Most subjects in this cohort were tachycardic throughout their recording: only eleven transiently/temporarily developed a normal HR (normal HR for at least 5 min within any 10-min interval). Ten who transiently/temporarily developed a normal HR had a median HR between 100 and 110 bpm, i.e. minimal degree of tachycardia.

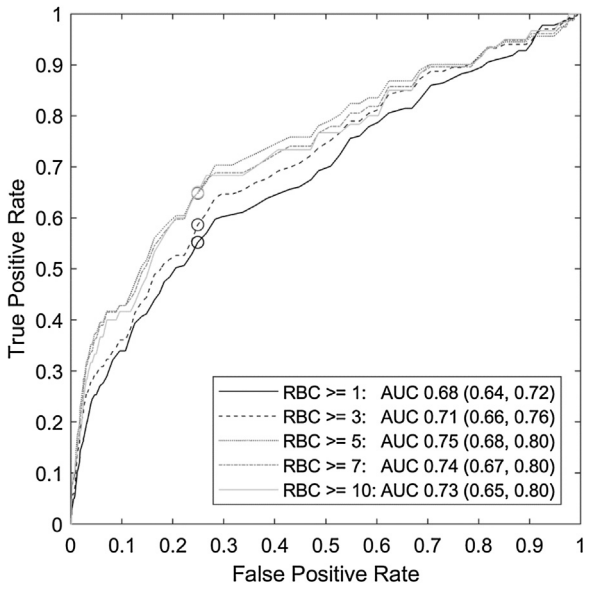



**Fig. 1.** Receiver operating characteristic (ROC) curves of the median heart rate (HR) for discriminating between different 24-h red blood cell (RBC) transfusion volumes (regardless of the patients' documented injuries), i.e. 24-h RBCs $\geq$1, $\geq$3, $\geq$5, $\geq$7, and $\geq$10 units. *Legend*: area under the ROC curve (AUC) with the 95% confidence interval in parentheses. *Circles* indicate a HR of 100 bpm, which is the conventional cut-off for tachycardia.

Based on median HR, 55 had a normal HR (41% of all trauma patients with haemorrhage). Overall, this cohort had a median vital-sign recording duration of 25 min (IQR: 15–36). Most subjects in this cohort had a normal HR throughout their recording: only ten subjects transiently/temporarily developed tachycardia (tachycardia for at least 5 min within any 10-min interval). Eight who transiently/temporarily developed tachycardia had a median HR between 90 and 100 bpm, i.e. at the upper extent of normal HR.

Overall, 25% of the patients *without* haemorrhage were tachycardic, based on their median HR. Patients without haemorrhage tended to be either persistently tachycardic or persistently non-tachycardic—only 17% ever changed from one state to the other even transiently/temporarily (i.e. for at least 5 cumulative minutes within any 10-min interval).

**Table 1**
Characteristics of trauma population datasets.

| Clinical characteristics | Dataset 1 (n = 209) | Dataset 2 (n = 1161) | Dataset 3 (n = 646) | Pooled dataset (n = 2016) |
|---|---|---|---|---|
| Setting | Pre-hospital | Emergency Dept | Pre-hospital | |
| Sex, male, n (%) | 155 (74) | 813 (70) | 479 (74) | 1447 (72) |
| Age, mean (SD), years | 45 (20) | 50 (21) | 38 (15) | 46 (19) |
| Blunt, n (%) | 188 (90) | 1008 (87) | 577 (89) | 1773 (88) |
| Penetrating, n (%) | 21 (10) | 144 (12) | 61 (9) | 226 (11) |
| ISS, median (IQR) | 16 (9–26) | 18 (10–26) | 16 (9–34) | 17 (9–29) |
| Inter-hospital transfer, n (%) | 103 (49) | 392 (34) | 0 (0) | 495 (25) |
| Glasgow Coma Scale Score (IQR) | 15 (8–15) | 15 (14–15) | 15 (13–15) | 15 (13–15) |
| 24-h RBC volume $\geq$1 unit, n (%) | 31 (15) | 153 (13) | 75 (12) | 259 (13) |
| 24-h RBC volume $\geq$3 unit, n (%) | 18 (9) | 75 (6) | 57 (9) | 150 (7) |
| 24-h RBC volume $\geq$10 unit, n (%) | 8 (4) | 24 (2) | 22 (3) | 54 (3) |
| Haemorrhage patient* | 18 (9) | 60 (15) | 55 (9) | 133 (7) |
| Survival to discharge, n (%) | 191 (91) | 1103 (95) | 608 (94) | 1902 (94) |

Abbreviations: IQR, interquartile range; ISS, Injury Severity Score; SD, standard deviation.

* Haemorrhage patient: Primary outcome was patients with at least one documented explicitly haemorrhagic injury (solid organ injuries, thoracic or abdominal haematomas, and/or vascular injuries requiring operative repair) and 24-h RBC volume $\geq$3 unit. Alternative definitions of haemorrhage were investigated in ancillary sensitivity analysis; see text for details.

**Table 2**
Comparison of clinical characteristics of tachycardic and non-tachycardic haemorrhage patients.

| | Tachycardic haemorrhage patients | | Non-tachycardic haemorrhage patients | |
|---|---|---|---|---|
| | All patients (n = 78) | Subset with hypotension onset (n = 12) | All patients (n = 55) | Subset with hypotension onset (n = 14) |
| *Clinical characteristics* | | | | |
| 24-h RBCs, median (IQR) | 10 (5–16) | 15 (7–30)* | 6 (4–13) | 7 (4–11)* |
| Age in years, median (IQR) | 32 (27–49)* | 34 (31–49) | 50 (36–62)* | 53 (36–71) |
| Intubation, n (%) | 42 (54%) | 8 (67%) | 22 (40%) | 8 (57%) |
| Pre-hospital patients, n (%) | 40 (51%) | 7 (58%) | 33 (60%) | 8 (57%) |
| Mortality, n (%) | 25 (32%) | 6 (50%) | 15 (27%) | 2 (14%) |
| Head AIS, median (IQR) | 0 (0–2) | 0 (0–4) | 0 (0–0) | 0 (0–0) |
| Abdomen AIS, median (IQR) | 3 (1–4) | 0 (0–4) | 2 (0–3) | 2 (0–3) |
| Extremity AIS, median (IQR) | 1 (0–3) | 3 (0–3) | 0 (0–3) | 0 (0–2) |
| Thorax AIS, median (IQR) | 3 (0–4) | 4 (2–4) | 3 (0–4) | 3 (0–4) |
| ISS, median (IQR) | 27 (18–43) | 26 (17–45) | 27 (18–41) | 22 (16–45) |
| *Vital-sign characteristics* | | | | |
| Duration of recording in min, median (IQR) | 22 (12–31) | 28 (22–42) | 25 (15–36) | 30 (26–38) |
| HR[a] in bpm, median (IQR) | 121 (109–136)[b] | 127 (115–137)[b] | 83 (73–91)[b] | 86 (79–89)[b] |
| Slope of HR in bpm/10 min, median (IQR) | +0.7 [(−2.9)–(+3.9)]* | +3.6 [(+0.7)–(+7.1)]* | −0.8 [(−4.4)–(+0.5)]* | −2.4 [(−3.7)–(−0.5)]* |
| RMS about mean HR in bpm, median (IQR) | 6 (3–9) | 8 (3–13) | 7 (5–11) | 6 (5–7) |
| SBP[a] in mmHg, median (IQR) | 110 (89–135) | 95 (82–108) | 104 (85–123) | 98 (93–115) |
| PP[a] in mmHg, median (IQR) | 44 (34–51) | 36 (31–44) | 44 (32–55) | 45 (34–57) |
| Time to hypotension in min, median (IQR) | 5 (1–16) | 13 (7–18) | 8 (2–14) | 11 (8–14) |
| Shock index in bpm/mmHg, median (IQR) | 1.08 (0.90–1.36)* | 1.34 (1.22–1.57)* | 0.83 (0.66–1.06)* | 0.88 (0.70–0.96)* |
| Incidence of hypotension, n (%) | 52 (67%) | 12 (100%) | 39 (71%) | 14 (100%) |

Tachycardic group: median HR ≥ 100 bpm; non-tachycardic group: median HR < 100 bpm.
Subset with hypotension onset: subjects with a systolic blood pressure of less than 90 mmHg or a mean arterial pressure of less than 70 mmHg.
Abbreviations: 24-h RBCs, red blood cell units transfused in 24 h; AIS, abbreviated injury scale; IQR, interquartile range; ISS, injury severity score; bpm, beats per minute; HR, heart rate; RMS, root mean square; MAP, mean arterial pressure; PP, pulse pressure = systolic blood pressure (SBP) − diastolic blood pressure.
[a] For vital signs HR, SBP, and PP, we computed the median value from each patient's record ("patient-median"); above, we report the population median of the patient-median values.
[b] Not tested for significant differences (because cohorts were defined by median HR).
* $P < 0.05$ via Wilcoxon rank-sum test comparing tachycardic group and non-tachycardic group.

Comparing the vital signs of the two haemorrhage cohorts (tachycardic and non-tachycardic patients with haemorrhage), both had similar BP characteristics (see Table 2) without significant differences in their average SBP, average PP, time of first recorded hypotension, and incidence of hypotension. The overall variability in the recorded HR (i.e. RMS around the mean) was not significantly different between the two cohorts. The differences in HR slope achieved statistical, but not clinical, significance; the median changes over 10 min were +0.7 and −0.8 bpm for the tachycardic and non-tachycardic cohorts, respectively. The SI was significantly higher in the tachycardic cohort. Fig. 2 illustrates the vital-sign patterns of several tachycardic and non-tachycardic patients who developed hypotension.

Fig. 3 shows the fraction of patients with haemorrhage who developed hypotension as a function of time. In both groups, after 30 min, the majority of patients had developed hypotension.

The clinical characteristics of the tachycardic and non-tachycardic cohorts were similar (Table 2). In terms of resuscitation, the volumes of 24-h RBCs for both cohorts were substantial but not significantly different; the corresponding median volumes were 10 units (IQR: 5–16) and 6 units (IQR: 4–13), respectively. Rates of mortality, endotracheal intubation, and injury severity coded by the abbreviated injury scale (AIS) were not significantly different. Only age was significantly different, although there was substantial overlap; the median ages were 32 (IQR: 27–49) and 50 (IQR: 36–62) years for the tachycardic and non-tachycardic cohorts, respectively.

We also studied the subset of patients who had documented onset of hypotension (i.e. hypotension but only after a non-hypotensive BP). We identified 26 subjects who met the inclusion criteria, 12 in the tachycardia cohort and 14 in the non-tachycardia cohort. HR did not change substantially upon the onset of hypotension (Table 3). Subjects in the tachycardic cohort had a median HR of 123 bpm (IQR: 107–132) in the 10 min prior to hypotension and a median HR of 127 bpm (IQR: 116–141) in the 10 min following the onset of hypotension. Subjects in the non-tachycardic cohort had a median HR of 87 bpm (IQR: 75–103) prior to hypotension and a median HR of 86 bpm (IQR: 78–89) in the 10 min upon the onset of hypotension. There was a statistically significant difference in the 24-h RBCs; tachycardic patients who developed hypotension received significantly more blood than did non-tachycardic patients who developed hypotension (Table 2).

## Discussion

Similar to prior analyses of large databases, we found that tachycardia was neither sensitive nor specific to haemorrhagic injury. The current analysis is novel in that we analysed HR measured continuously in the early evaluation of trauma patients (median durations of vital-sign recordings were 22 min and 25 min for tachycardic and non-tachycardic haemorrhage patients, respectively). We found that approximately half of trauma patients with haemorrhagic injury evidenced tachycardia and half did not. The former cohort demonstrated consistent tachycardia throughout their recording, with some fluctuation, but the tachycardia did not consistently increase over time, and there was no consistent change in HR upon the onset of hypotension. The latter cohort demonstrated normal HR throughout their recording, with some fluctuation, but the HR neither consistently increased over time nor showed any consistent change upon the onset of hypotension.

These findings add to our understanding about why prior analyses have found that the association between tachycardia and blood loss is "not to the degree suggested by the ATLS classification of shock" [1]. In general terms, there is a cohort of patients with haemorrhage who demonstrate tachycardia, but do not show any additional increase in HR, even through time, and even upon the

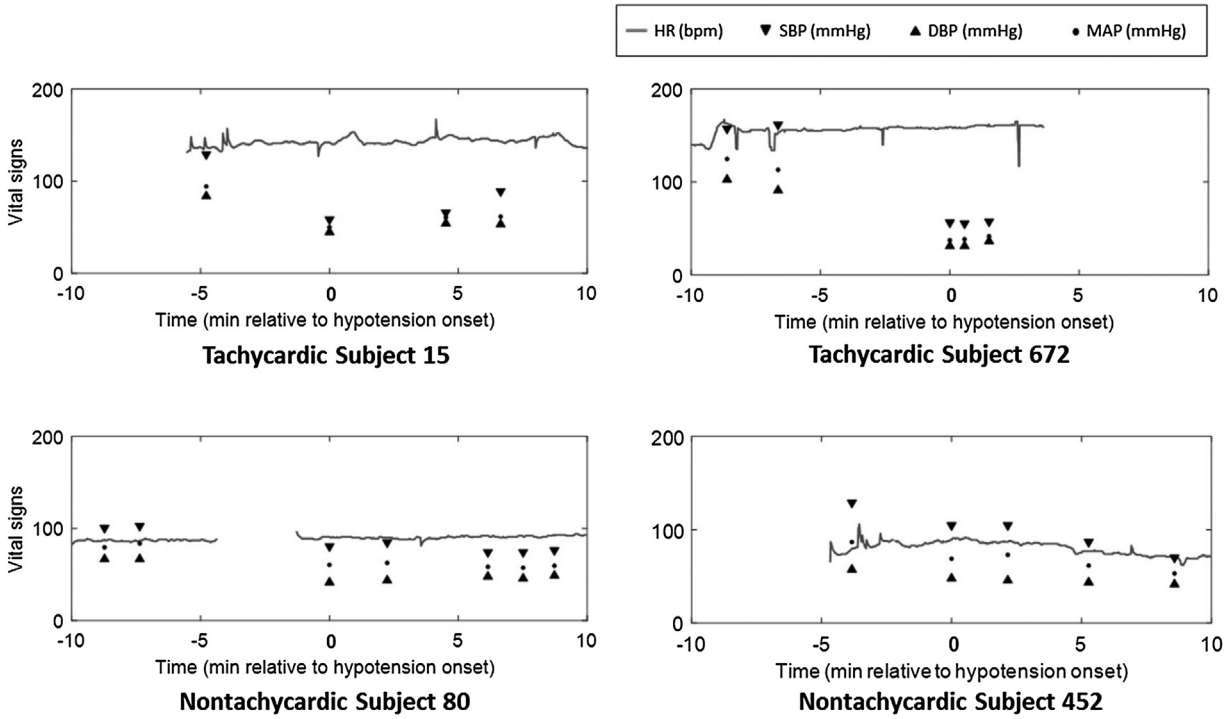



**Fig. 2.** Case examples of tachycardic and non-tachycardic subjects who developed hypotension. The panels above show vital signs of subjects in an analysis window that ranges from 10 min prior to 10 min after the onset of hypotension. Time $t = 0$ indicates the time of hypotension onset. Subjects in the tachycardic group had a pre-hypotension HR of at least 100 bpm, whereas those in the non-tachycardic group had a pre-hypotension HR of less than 100 bpm. *Top left:* 36-y old female, injuries included femoral fracture and arterial injury requiring angio-embolization, 24-h RBC volume = 6 units (hypotension onset after 6 min of vital-sign monitoring). *Top right:* 31-y old male, injuries included splenic laceration, 24-h RBC volume = 38 units (hypotension onset after 18 min of vital-sign monitoring). *Bottom left:* 42-y old male, injuries included liver laceration, 24-h RBC volume = 9 units (hypotension onset after 10 min of vital-sign monitoring). *Bottom right:* 76-y old female, injuries included splenic laceration, 24-h RBC volume = 13 units (hypotension onset after 5 min of vital-sign monitoring). Abbreviations: bpm, beats per minute; DBP, diastolic blood pressure; HR, heart rate; MAP, mean arterial pressure; RBC, red blood cell; SBP, systolic blood pressure. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

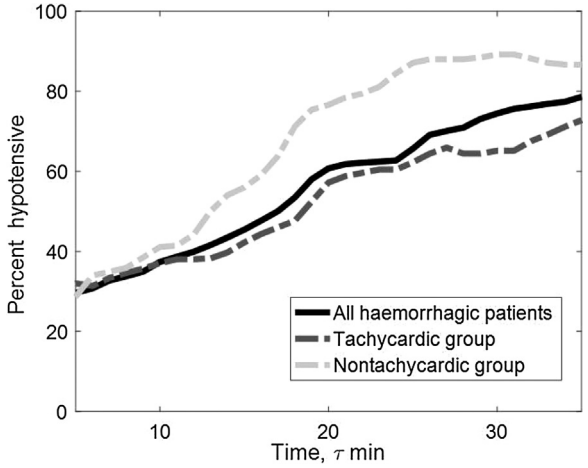



**Fig. 3.** Percentage of haemorrhage subjects (24-h red blood cell transfusion volume $\geq 3$ units with documented haemorrhagic injury) with hypotension. The solid black line represents the percentage of haemorrhage subjects with at least $\tau$ min of recorded data and who developed hypotension in the first $\tau$ min. The total number of patients decreased as a function of time, because the durations of pre-hospital and emergency department care were heterogeneous (of the 133 total haemorrhage subjects, 45 had at least 30 min of recorded data). For comparison, subjects were split into two groups: *i*) the tachycardic group, indicated by the dark gray dashed line, and *ii*) the non-tachycardic group, indicated by the light gray dashed line. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

onset of hypotension. Moreover, there is a separate cohort altogether that does not manifest tachycardia at all, even after the onset of hypotension. When we examined whether there were salient differences between these cohorts, we found few aside from age (tachycardic patients were significantly younger). Tachycardic or not, both generally developed hypotension within 30 min of vital-sign monitoring (Fig. 3). The two cohorts were similar in terms of other metrics of haemorrhage severity. There were no differences in the median time of the first recorded hypotension, overall rate of hypotension, 24-h RBCs, or mortality. AIS scores across all anatomic regions were also similar between the cohorts.

Regarding why non-tachycardic patients were significantly older, it is possible that this subpopulation has an attenuated cardiovascular control system that is less likely to mount a tachycardic response, because of either aging or medication, e.g. beta blockers. Yet, the age ranges for patients with and without tachycardia showed substantial overlap (IQRs of 27–49 years and 36–62 years, respectively; see Table 2). This indicates that age alone is not the sole determinant of a patient's haemorrhage response.

The determinants of HR during progressive haemorrhage have been investigated over decades of *in vivo* laboratory experimentation. Animal models demonstrate a basic paradigm in which progressive blood loss triggers tachycardia and vasoconstriction via activation of carotid and aortic baroreceptors [14] and also arterial chemoreceptors that are sensitive to local metabolic changes associated with hypovolaemia [15]. Afferent signals from these peripheral receptors are received by the cardiovascular center within the medulla oblongata, resulting in both sympathetic nervous signal activation and parasympathetic system inhibition, and, ultimately, increased pace of the heart's native pacemaker, the sino-atrial node. Moreover, a wide range of investigations has further demonstrated how these basic haemodynamic responses can be modified by a multitude of factors, including nociception [16], anaesthetics and analgesics [17,18], anxiety [19], gender [20], brain injury [21], cardiopulmonary baroreceptors [22], athletic

**Table 3**
Vital signs before and after the onset of hypotension (subset with hypotension onset).

| Vital sign | Median (IQR) of subjects | |
|---|---|---|
| | Tachycardic group (n = 12) | Non-tachycardic group (n = 14) |
| Pre-hypotension | | |
| HR (bpm) | 123 (107–132)[a] | 87 (75–103)[a] |
| SBP (mmHg) | 115 (101–148) | 126 (106–132) |
| PP (mmHg) | 46 (41–58) | 44 (35–72) |
| MAP (mmHg) | 88 (79–110) | 88 (78–102) |
| Onset of hypotension | | |
| HR (bpm) | 127 (116–141)[a] | 86 (78–89)[a] |
| SBP (mmHg) | 73 (67–81) | 86 (78–102) |
| PP (mmHg) | 23 (20–34)[*] | 41 (32–47)[*] |
| MAP (mmHg) | 57 (52–65) | 64 (56–68) |

Each patient's median HR, SBP, PP and MAP (patient-median) were computed for the 10-min interval before the onset of hypotension ("pre-hypotension") and the 10-min interval starting upon the onset of hypotension ("onset of hypotension"); above, we report the population median of the patient-median values. See text for additional details.
Abbreviations: bpm, beats per minute; HR, heart rate; IQR, interquartile range; MAP, mean arterial pressure; PP, pulse pressure = systolic blood pressure (SBP) − diastolic blood pressure.
[a] Not tested for significant differences (because cohorts were defined by median HR).
[*] $P < 0.05$ via Wilcoxon rank-sum test comparing tachycardic group and non-tachycardic group.

pre-conditioning [23], as well as chronic diabetes mellitus [24]. In many cases, there can be excitatory and inhibitory pathways activated at the same time, and the central nervous system integrates concordant and discordant afferent signals and generates the ultimate efferent output that drives HR [25]. We speculate that, during typical trauma patient management, there is heterogeneity in the determinants of HR, e.g. differing levels of pain, analgesia, haemorrhage, etc. This could explain why we observed heterogeneous HR responses in trauma patients.

The concept of categorizing patients based on above-average versus below-average sympathetic responses is consistent with a series of physiology reports conducted in a laboratory with healthy subjects, using lower body negative pressure (LBNP) to simulate progressive blood loss. These studies determined that the group of subjects with delayed onset of hypotension had relatively elevated HR and vasoconstriction, and denoted that cohort as "high tolerant" [26–29]. Our dataset corroborates the notion that both tachycardic and normal HR responses are common with progressive blood loss. Are patients with above-average sympathetic responses more tolerant of blood loss? Overall, there were no significant differences between tachycardic and non-tachycardic patients in terms of incidence of hypotension, 24-h RBCs, or mortality (Table 2)]. In contrast, in the smaller subset of haemorrhage patients whose onset of hypotension was recorded, tachycardic patients ended up with significantly larger volumes of 24-h RBCs (median, 15 units), suggesting that this subset may have been compensating for large blood volume losses prior to hypotension onset. Tachycardic or not, most patients with haemorrhage developed hypotension within 30 min of vital-sign monitoring (Fig. 3), with no statistically significant differences between the cohorts in terms of time of first recorded hypotension.

In terms of limitations of the current report, the vital-sign data used in this analysis were obtained during routine clinical care, and not during a carefully controlled laboratory investigation. Consequently, the measurement intervals were heterogeneous and recording durations were uneven, and the reliability of measurements may have been suboptimal. For HR, we were able to rely on ECGs to retrospectively identify unreliable measurements. For BP, we had no practical method of identifying unreliable non-invasive measurements. Excessive variability due to measurement errors, and the confounding effect of therapeutic interventions, such as volume administration or pain medication, might have masked differences between the cohorts (i.e. Type II statistical errors). Therefore, based on the current analysis, we cannot rule out subtle differences between tachycardic and non-tachycardic patients with haemorrhage. However, we note that the vital-sign data we analysed here are precisely those that a clinician must evaluate in providing treatment. Therefore, our comparative analysis would appear to be valid in terms of ruling in and ruling out significant cohort differences based on the actual vital-sign measurements that are evaluated by and acted upon by bedside clinicians.

As a second limitation, we did not have a feasible gold standard measurement of blood loss as a function of time. Therefore, the onset of hypotension may not always have indicated the progression of blood loss in some patients. However, we consider it likely that the development of frank hypotension was usually due to true hypovolaemia in this subject population with documented major haemorrhagic injuries, and who subsequently received three or more units of RBCs.

## Conclusions

In conclusion, trauma patients—both haemorrhagic and non-haemorrhagic—tend to fall into persistently tachycardic or persistently non-tachycardic groups during the first 30 min of monitoring. During initial assessment, it is reasonable to have an elevated concern for haemorrhage when tachycardia is present, keeping in mind the substantial limitation that tachycardia was only modestly specific (75%) and poorly sensitive (59%) for haemorrhage. Through time, there will be HR fluctuations, but diagnostically meaningful trends were not evident in the typical haemorrhage patient. Blood pressure should be carefully monitored, since hypotension was likely to manifest within 30 min in haemorrhage patients, and without any associated change in HR.

## Conflicts of interest

Andrew T. Reisner has received research funding from the Nihon-Kohden Corporation, Irvine, CA, to develop decision-support technology for sepsis patient management. Shwetha Edla, Jianbo Liu, Jiankun Liu, Maxim Khitrov, and Jaques Reifman report no conflict of interest.

## Sources of support

This work was supported by the Combat Casualty Care Program Area Directorate of the U.S. Army Medical Research and Materiel Command, Fort Detrick, MD, USA. The study sponsors did not have any role in the study design, data collection, analysis and interpretation of data, report writing, or in the decision to submit the article for publication.

## 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] Guly H.R., Bouamra O, Little R, Dark P, Coats T, Driscoll P, et al. Testing the validity of the ATLS classification of hypovolaemic shock. Resuscitation 2010;81:1142–7.
[2] Guly HR, Bouamra O, Spiers M, Dark P, Coats T, Lecky FE. Vital signs and estimated blood loss in patients with major trauma: testing the validity of the ATLS classification of hypovolaemic shock. Resuscitation 2011;82:556–9.

[3] Mutschler M, Nienaber U, Brockamp T, Wafaisade A, Wyen H, Peiniger S, et al. A critical reappraisal of the ATLS classification of hypovolaemic shock: does it really reflect clinical reality. Resuscitation 2013;84:309–13.
[4] Mutschler M, Paffrath T, Wolfl C, Probst C, Nienaber U, Schipper IB, et al. The ATLS classification of hypovolaemic shock: a well established teaching tool on the edge. Injury 2014;45:S35–8.
[5] Liu J, Khitrov MY, Gates JD, Odom SR, Havens JM, de Moya MA. Automated analysis of vital signs to identify patients with substantial bleeding prior to hospital arrival: a feasibility study. Shock 2015;43:429–36.
[6] Reisner AT, Edla S, Liu J, Rubin JT, Thorsen JE, Kittell E, et al. Muscle oxygen saturation improves diagnostic association between initial vital signs and major hemorrhage: a prospective observational study. Acad Emerg Med 2016;23:353–7.
[7] Cooke WH, Salinas J, Convertino VA, Ludwig DA, Hinds D, Duke JH. Heart rate variability and its association with mortality in prehospital trauma patients. J Trauma 2006;60:363–70.
[8] Holcomb JB, Salinas J, McManus JM, Miller CC, Cooke WH, Convertino VA. Manual vital signs reliably predict need for life-saving interventions in trauma patients. J Trauma 2005;59:821–8.
[9] Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap) – a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform 2009;42:377–81.
[10] Reisner AT, Khitrov MY, Chen L, Blood A, Wilkins K, Doyle W, et al. Development and validation of a portable platform for deploying decision-support algorithms in prehospital settings. Appl Clin Inform 2013;4:392–402.
[11] Smith JB, Reisner AT, Edla S, Liu J, Liddle S, Reifman J. A platform for real-time acquisition and analysis of physiological data in hospital emergency departments. Conf Proc IEEE Eng Med Biol Soc 2014;2014:2678–81.
[12] Yu C, Liu Z, McKenna T, Reisner AT, Reifman J. A method for automatic identification of reliable heart rates calculated from ECG and PPG waveforms. J Am Med Inform Assoc 2006;13:309–20.
[13] Akobeng AK. Understanding diagnostic tests 3: Receiver operating characteristic curves. Acta Paediatr 2007;96:644–7.
[14] Schadt JC. Experimental observations - animals. In: Secher NH, Pawelczyk JA, Ludbrook J, editors. Blood Loss and Shock. London: Edward Arnold; 1994.
[15] Landgren S, Neil E. Chemoreceptor impulse activity following haemorrhage. Acta Physiol Scand 1951;23:158–67.
[16] Bereiter DA, Benetti AP, Thrivikraman KV. Thermal nociception potentiates the release of ACTH and norepinephrine by blood loss. Am J Physiol 1990259: R1236–42.
[17] Zimpfer M, Manders WT, Barger AC, Vatner SF. Pentobarbital alters compensatory neural and humoral mechanisms in response to hemorrhage. Am J Physiol 1982;243:H713–21.
[18] Van Leeuwen AF, Evans RG, Ludbrook J. Effects of halothane, ketamine, propofol and alfentanil anaesthesia on circulatory control in rabbits. Clin Exp Pharmacol Physiol 1990;17:781–98.
[19] Schadt JC, Hasser EM. Hemodynamic effects of acute stressors in the conscious rabbit. Am J Physiol 1998;274:R814–21.
[20] Crofton JT, Share L. Sexual dimorphism in vasopressin and cardiovascular response to hemorrhage in the rat. Circ Res 1990;66:1345–53.
[21] McMahon CG, Kenny R, Bennett K, Kirkman E. Modification of acute cardiovascular homeostatic responses to hemorrhage following mild to moderate traumatic brain injury. Crit Care Med 2008;36:216–24.
[22] Oberg B, Thoren P. Increased activity in left ventricular receptors during hemorrhage or occlusion of caval veins in the cat. A possible cause of the vaso-vagal reaction. Acta Physiol Scand 1972;85:164–73.
[23] Ahlgren JK, Hayward LF. Role of lateral parabrachial opioid receptors in exercise-induced modulation of the hypotensive hemorrhage response in conscious male rats. Behav Brain Res 2012;226:404–10.
[24] Kastrup J, Norgaard T, Parving HH, Henriksen O, Lassen NA. Impaired autoregulation of blood flow in subcutaneous tissue of long-term type 1 (insulin-dependent) diabetic patients with microangiopathy: an index of arteriolar dysfunction. Diabetologia 1985;28:711–7.
[25] Wennergren G, Little R, Oberg B. Studies on the central integration of excitatory chemoreceptor influences and inhibitory baroreceptor and cardiac receptor influences. Acta Physiol Scand 1976;96:1–18.
[26] Convertino VA, Rickards CA, Ryan KL. Autonomic mechanisms associated with heart rate and vasoconstrictor reserves. Clin Auton Res 2012;22:123–30.
[27] Hinojosa-Laborde C, Rickards CA, Ryan KL, Convertino VA. Heart rate variability during simulated hemorrhage with lower body negative pressure in high and low tolerant subjects. Front Physiol 2011;2:85.
[28] Rickards CA, Ryan KL, Cooke WH, Convertino VA. Tolerance to central hypovolemia: the influence of oscillations in arterial pressure and cerebral blood velocity. J Appl Physiol 2011;111:1048–58.
[29] Sather TM, Goldwater DJ, Montgomery LD, Convertino VA. Cardiovascular dynamics associated with tolerance to lower body negative pressure. Aviat Space Environ Med 1986;57:413–9.