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 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

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 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.

Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/3.0/).

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

http://dx.doi.org/10.1016/j.ajem.2014.11.046
0735-6757/Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/3.0/).
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 10000 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

---

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 (MHHLF, 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 ≪ 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, MLHF 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

<table>
<thead>
<tr>
<th></th>
<th>Overall database</th>
<th>Study population</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MHLF</td>
<td>BMF</td>
</tr>
<tr>
<td>Population, n</td>
<td>757</td>
<td>242</td>
</tr>
<tr>
<td>Male, n (%)</td>
<td>562 (74%)</td>
<td>179 (74%)</td>
</tr>
<tr>
<td>Female, n (%)</td>
<td>195 (26%)</td>
<td>63 (26%)</td>
</tr>
<tr>
<td>Age, years, mean (SD)</td>
<td>38 (15)</td>
<td>47 (21)</td>
</tr>
<tr>
<td>Mechanism of injury</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Blunt, n (%)</td>
<td>664 (88%)</td>
<td>216 (89%)</td>
</tr>
<tr>
<td>Penetrating, n (%)</td>
<td>84 (11%)</td>
<td>26 (11%)</td>
</tr>
<tr>
<td>Hospital transfer, n (%)</td>
<td>0 (0%)</td>
<td>118 (49%)</td>
</tr>
<tr>
<td>Prehospital airway intubation, n (%)</td>
<td>165 (22%)</td>
<td>97 (40%)</td>
</tr>
<tr>
<td>ISS, median (IQR)b</td>
<td>17 (9-34)</td>
<td>17 (9-26)</td>
</tr>
<tr>
<td>Prehospital fluid volume, mL, median (IQR)d</td>
<td>300 (100-628)</td>
<td>100 (50-250)</td>
</tr>
<tr>
<td>24-h PRBC transfusion volumes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>≥ 1 unit, n (%)</td>
<td>153 (20%)</td>
<td>62 (26%)</td>
</tr>
<tr>
<td>≥ 9 units, n (%)</td>
<td>36 (5%)</td>
<td>11 (5%)</td>
</tr>
<tr>
<td>Overall mortality, n (%)</td>
<td>85 (11%)</td>
<td>28 (12%)</td>
</tr>
<tr>
<td>Died before admission to ED, n (%)</td>
<td>36 (42%)</td>
<td></td>
</tr>
<tr>
<td>Died after admission to ED, n (%)</td>
<td>49 (58%)</td>
<td>21 (75%)</td>
</tr>
</tbody>
</table>

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.
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.

![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 A1.](image)

### 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

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

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

### 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

<table>
<thead>
<tr>
<th>Feature set description (features)</th>
<th>24-h PRBC volume</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>≥ 1</td>
<td>≥ 5</td>
<td>≥ 9</td>
</tr>
<tr>
<td>Population (controls, cases)</td>
<td>402 (348, 54)</td>
<td>402 (377, 25)</td>
<td>402 (386, 16)</td>
<td></td>
</tr>
<tr>
<td>Core feature set (HR, RR, SBP, PP)</td>
<td>0.79 (0.70-0.85)</td>
<td>0.85 (0.73-0.92)</td>
<td>0.86 (0.73-0.94)</td>
<td></td>
</tr>
<tr>
<td>Core feature set + RSA (HR, RR, SBP, PP, RSA)</td>
<td>+0.00</td>
<td>+0.01</td>
<td>+0.02</td>
<td></td>
</tr>
<tr>
<td>Core feature set – HR (RR, SBP, PP)</td>
<td>+0.00</td>
<td>+0.01</td>
<td>+0.00</td>
<td></td>
</tr>
<tr>
<td>Core feature set – HR + RSA (RR, SBP, PP, RSA)</td>
<td>+0.00</td>
<td>+0.01</td>
<td>+0.02</td>
<td></td>
</tr>
<tr>
<td>Core feature set – RR (HR, SBP, PP)</td>
<td>−0.02</td>
<td>−0.03</td>
<td>−0.05</td>
<td></td>
</tr>
<tr>
<td>Core feature set – RR + RSA (HR, SBP, PP, RSA)</td>
<td>+0.00</td>
<td>−0.01</td>
<td>+0.00</td>
<td></td>
</tr>
<tr>
<td>Core feature set – (SBP, PP) (HR, RR)</td>
<td>−0.10 (b)</td>
<td>−0.07 (b)</td>
<td>−0.09</td>
<td></td>
</tr>
<tr>
<td>Core feature set – (SBP, PP) + RSA (HR, RR, RSA)</td>
<td>−0.06</td>
<td>−0.07 (b)</td>
<td>−0.05</td>
<td></td>
</tr>
</tbody>
</table>

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 1.0, greater than or equal to 5.0, and greater than or equal to 9.0 units of 24-hour PRBC vol. When we added RSA to the core feature set (SBP, RR, and HR), 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 AUCs were +0.00, +0.00, and +0.02 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 hypovolemic trauma, tachycardia, tachypnea, reduced SBP, and reduced RR (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 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

<table>
<thead>
<tr>
<th>Features</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>SBP (mmHg)</td>
<td>150 (120-150)</td>
<td>150 (120-150)</td>
<td>150 (120-150)</td>
<td>150 (120-150)</td>
<td>150 (120-150)</td>
<td>150 (120-150)</td>
<td>150 (120-150)</td>
</tr>
<tr>
<td>PP (mmHg)</td>
<td>100 (75-125)</td>
<td>100 (75-125)</td>
<td>100 (75-125)</td>
<td>100 (75-125)</td>
<td>100 (75-125)</td>
<td>100 (75-125)</td>
<td>100 (75-125)</td>
</tr>
<tr>
<td>SDNN (ms)</td>
<td>25 (20-30)</td>
<td>25 (20-30)</td>
<td>25 (20-30)</td>
<td>25 (20-30)</td>
<td>25 (20-30)</td>
<td>25 (20-30)</td>
<td>25 (20-30)</td>
</tr>
</tbody>
</table>

References


Reisner AT, Chen L, McKenna TM, Reifman J. Automatically-computed prehospital severity scores are equivalent to scores based on medical documentation. J Trauma 2006;65:915–23.


