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).

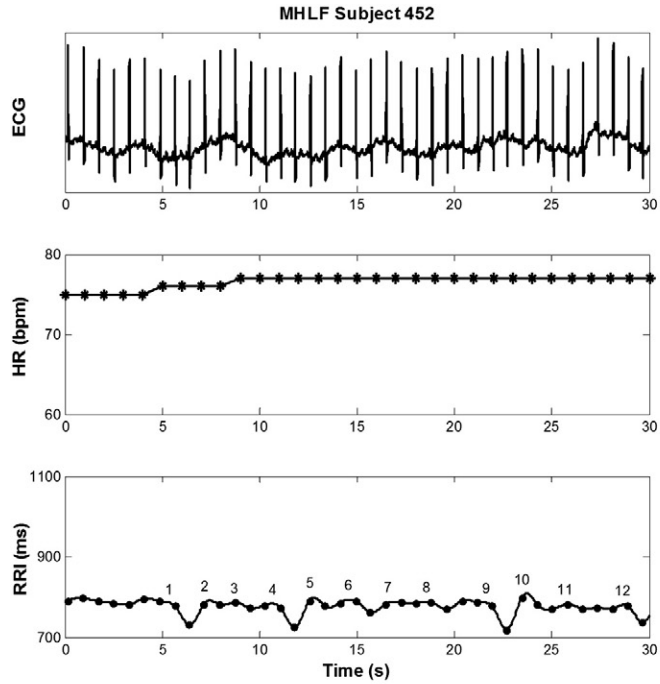
http://dx.doi.org/10.1016/j.ajem.2014.11.046

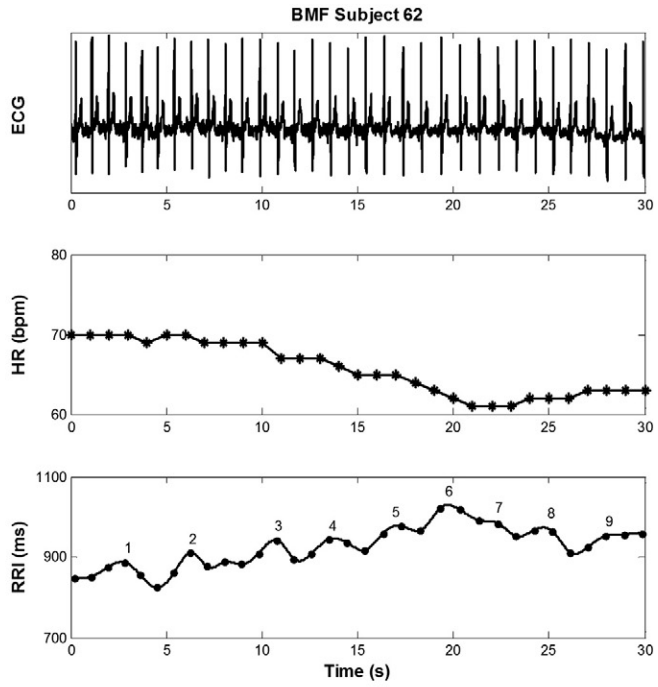
require any LSI [10]. However, these findings are tempered by several other reports, which suggest that, for that subset of trauma patients with normal vital signs, HRV metrics have a low sensitivity (16%) for LSI prediction [6], and their diagnostic potential is reduced by notable intersubject variability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

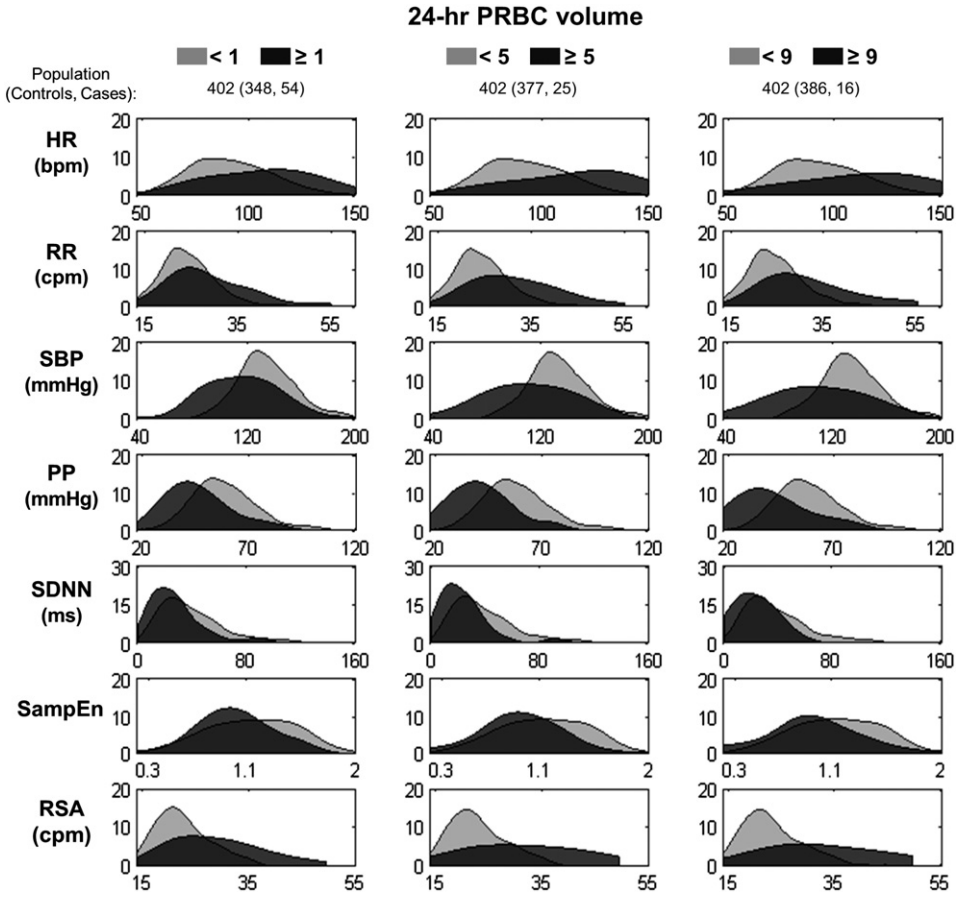


Fig. 2. Distributions of routine vital signs and HRV metrics for trauma patients grouped by different 24-hour PRBC volumes. The medians and interquartile ranges of the distributions are provided in Table A.1.

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.