Development and Validation of a Portable Platform for Deploying Decision-Support Algorithms in Prehospital Settings

A. T. Reisner¹,²; M. Y. Khitrov¹; L. Chen¹; A. Blood²; K. Wilkins³; W. Doyle³; S. Wilcox²,³; T. Denison³; J. Reifman¹

¹Biotechnology High Performance Computing Software Applications Institute (BHSAl), Telemedicine and Advanced Technology Research Center (TATRC), U.S. Army Medical Research and Materiel Command (USAMRMC), Fort Detrick, MD, USA;
²Massachusetts General Hospital Department of Emergency Medicine, Boston, MA, USA;
³Boston MedFlight, Bedford, MA, USA

Keywords
Decision-support algorithms, prehospital care, device connectivity, vital-sign data, combat casualty care

Summary

Background: Advanced decision-support capabilities for prehospital trauma care may prove effective at improving patient care. Such functionality would be possible if an analysis platform were connected to a transport vital-signs monitor. In practice, there are technical challenges to implementing such a system. Not only must each individual component be reliable, but, in addition, the connectivity between components must be reliable.

Objective: We describe the development, validation, and deployment of the Automated Processing of Physiologic Registry for Assessment of Injury Severity (APPRAISE) platform, intended to serve as a test bed to help evaluate the performance of decision-support algorithms in a prehospital environment.

Methods: We describe the hardware selected and the software implemented, and the procedures used for laboratory and field testing.

Results: The APPRAISE platform met performance goals in both laboratory testing (using a vital-sign data simulator) and initial field testing. After its field testing, the platform has been in use on Boston MedFlight air ambulances since February of 2010.

Conclusion: These experiences may prove informative to other technology developers and to healthcare stakeholders seeking to invest in connected electronic systems for prehospital as well as in-hospital use. Our experiences illustrate two sets of important questions: are the individual components reliable (e.g., physical integrity, power, core functionality, and end-user interaction) and is the connectivity between components reliable (e.g., communication protocols and the metadata necessary for data interpretation)? While all potential operational issues cannot be fully anticipated and eliminated during development, thoughtful design and phased testing steps can reduce, if not eliminate, technical surprises.
1. Introduction

In this report, we describe the development, validation, and deployment of a platform intended to serve as a test bed to help evaluate the performance of decision-support algorithms in a prehospital environment. The Automated Processing of Physiologic Registry for Assessment of Injury Severity (APPRAISE) platform consists of a ruggedized computer that is connected to a portable vital-sign monitor, software that communicates with the monitor and acquires patient vital signs, and algorithms that analyze the vital signs for any indication of potentially dangerous medical conditions. Technical requirements for its development included reliable communication with the vital-sign monitor, adequate data storage and processing capabilities, and a form factor that is acceptable to the end users. Our experience in tackling these challenges may be relevant to broader efforts in deploying automated decision-support functionality in prehospital environments. In addition, the experience is relevant to broader hospital-wide efforts to deliver automated decision support via electronic device interconnectivity [1].

The United States (U.S.) armed forces have long been interested in advanced decision-support capabilities for prehospital care of trauma casualties. The motivation is to assist caregivers on a battlefield, who may be inexperienced or overwhelmed by multiple casualties and hostile threats. The ideal system would automatically process medical data in real time and accurately identify the state of the casualty, the appropriate course of management, and the patient's priority for evacuation [2]. Multiple studies dating back to the Vietnam War through the conflicts in Iraq and Afghanistan have supported the notion that soldiers' lives could be saved with superior battlefield identification of wounds that are life threatening, yet treatable with timely intervention [3, 4].

More generally, it has been suggested that reliable electronic communication between analysis engines and medical devices (and other sources of clinical data) will make it possible "to improve patient safety, treatment efficacy, and workflow efficiency, reducing medical errors and healthcare costs to the benefit of patients throughout the continuum of care—from the home, to out-of-hospital transport, and to clinical areas as diverse as the operating room, intensive care unit, and general hospital ward" [5]. Our decision-support algorithms, intended to discriminate between trauma patients with and without life-threatening hemorrhage through analysis of vital-sign patterns [6, 7], are one example of the type of decision-support functionality that could become common throughout healthcare, if there were easy, reliable solutions to electronic connectivity between medical devices and analysis platforms. However, achieving such connectivity poses technological challenges, which we confronted in this project. Our experiences may prove informative to other technology developers and healthcare stakeholders seeking to invest in connected electronic systems for prehospital as well as in-hospital use.

2. Case Report

2.1 System Description

The goal of the project was to reliably connect an analysis platform to a standard transport vital-sign monitor to enable near-real-time decision support during prehospital operations. For vital-sign monitoring, we selected the Propaq 206 Encore monitor (Welch Allyn, Skaneateles Falls, NY [8]) due to its use on Boston MedFlight air ambulances, the site of our first deployment. This monitor outputs standard vital signs, such as heart rate (HR), blood O2 saturation, etc., at a frequency of 1 Hz. The electrocardiogram (ECG), photoplethysmogram (PPG), and impedance pneumogram (IP) comprise the waveforms, reported at 182, 91, and 23 Hz, respectively.

For the analysis platform hardware, we selected the GoBook MR-1 ruggedized personal computer (PC) (General Dynamics Itronix, Sunrise, FL [9]), which was securely attached to the Propaq. The GoBook MR-1 meets U.S. military standards MIL-STD-810F and IPS4 (protection against dust and splashing water) for operation in hazardous environments. It is connected to the Propaq via an RJ-12 to DE-9 RS-232 serial cable. Wireless communication is not supported by the Propaq 206 (and is prohibited by U.S. Federal Aviation Agency regulations). The Propaq transmits vital-sign data to the PC in discrete packets at regular time intervals, per the proprietary Protocol Systems, Inc.
Communication Protocol (PSI/CP). ▶ Figure 1 shows both hardware components and the serial connection cable in their disassembled state.

Software components of the APPRAISE platform reside on the GoBook. These include the communication controller, the analysis controller, MATLAB (MathWorks, Natick, MA [10]), and all investigational decision-support algorithms (▶ Figure 2). The communication controller establishes and maintains the serial connection whenever the Propaq is turned on. A new analysis session is created when the controller receives a HR value between 10 and 350 beats/min, which was our heuristic for detecting new patients. Sessions are terminated when no such value has been received for 5 min. To allow for post hoc assessment of the APPRAISE system's operation, the communication controller records every raw data packet received from the Propaq, along with a high-resolution timestamp, to a single file on the hard drive. A session that lasts for 1 h requires ~5 MB of disk space, meaning that the platform is capable of storing almost 6 mo of uninterrupted vital-sign data on a 20 GB data partition.

A key goal was to facilitate real-world testing of investigational algorithms implemented in or callable from MATLAB, a popular language for computational research and development. An analysis controller program, started at the beginning of each session, uses the MATLAB Engine Application Programming Interface to push all available vital-sign data into the MATLAB process memory and to execute the main analysis function at regular time intervals. That function then calls all installed algorithms in a pre-defined sequence, giving them access to the most recent data.

One primary challenge in this process is the conversion of PSI/CP packets (which contain data for multiple vital signs at different sampling frequencies) into neat vectors for each vital sign, synchronized on a common timeline. The conversion takes into account the start time, sampling rate, and duration for each variable and uses this information to automatically detect and fix problems with data synchronization that may result from corrupt, lost, or delayed packets. This presentation of vital signs as simple constant-frequency vectors is an abstraction layer that aims to hide some of the complexity associated with data acquisition. Algorithm developers do not have to worry about how data are transferred from the monitor to the running MATLAB process, which makes their code simpler and allows future versions of APPRAISE to support new hardware without having to modify any analysis components.

Finally, it is essential that the system possesses sufficient computational capability to execute the analysis algorithms. The analysis controller provides a configuration option that is used to limit the actual analysis rate anywhere from a few seconds to several minutes, because re-running the algorithms for every single datum is computationally expensive and rarely useful. When the analysis rate is set to a specific interval, vital-sign data are buffered in memory until the delay timeout expires. The buffered data are then pushed out into MATLAB and the next analysis iteration is triggered. This sequence continues until the end of the current session, at which point the decision-support algorithms are notified of the impending termination and are given a last chance to produce their output.

2.2 Laboratory Validation

For the initial validation and deployment of APPRAISE, the system was configured to execute our algorithms for calculating the reliability of vital-sign data [11–13] and identifying hemorrhage patients based on patterns in the vital signs [6, 7]. Preliminary laboratory testing allowed us to carefully validate the connectivity and operation of all system components. For this testing, we developed a Propaq emulator, a computer program that perfectly duplicated the way Propaq transmits sensor data. The emulator read archived physiological data from a plain-text file and "re-played" the data as if being output in real time from a monitored patient. A separate computer, containing the emulator and its input data, replaced the Propaq shown in ▶ Figure 2.

We selected a convenience sample of 20 patients from our existing archive of trauma vital-sign data to be used as the emulator input. Included in this set of patients were those with the longest records, the highest fraction of noisy signals (as measured by automated data reliability algorithms [11–13]), and cases with and without a full set of available vital signs. The vital-sign data for each of the 20 patients were transmitted by the emulator with a 1-h delay between patients. This process allowed us to test not only the data processing components but also the session start/stop mechan-
ism. For each session, we compared the recorded vital-sign values with the original emulator input. We also compared analysis results from the real-time analysis of streaming data versus independent offline analysis.

When we compared the source archive data versus the data recorded by APPRAISE, we found only trivial differences for ECG and IP waveforms (<0.5%), consistent with quantization errors due to the rounding of raw signal values into 12-bit representation during the emulator encoding process. More importantly, when we compared the vital-sign reliability and classifier outputs generated in real time by APPRAISE versus the precomputed results, we found no differences. We concluded that the connectivity and operation was acceptable, and we continued to field validation.

2.3 Field Validation

With local and U.S. Army Institutional Review Board (IRB) approval to study trauma patients, the platform was deployed on a Boston MedFlight helicopter on July 29, 2009. We compared the date/time of data archives automatically created by the APPRAISE system versus MedFlight’s log of missions. Figure 3 shows the prehospital timelines of the first 38 MedFlight patients. For most of the cases, APPRAISE began recording and analyzing vital signs a few minutes after the medics’ log indicated they had reached the patient. Monitoring with the Propaq typically continued throughout the flight and was terminated after the medics’ log indicated they had landed at the receiving hospital. However, records for subjects 5, 29, 37, and 38 were incomplete due to a premature system shut-down, likely because of insufficient battery power. The log files of the communication controller were consistent with this explanation, documenting an abrupt termination without any preceding errors.

Next, we compared the hand-recorded vital signs documented by the medics versus those archived by the APPRAISE system for 33 of 38 charts corresponding to eligible patients transported to a trauma center participating in our study (we did not have IRB approval to access the medical records of the other 5 patients). We defined a “perfect match” when vital signs were identical (+/-0%) within a 10-min window. For the median subject, the APPRAISE data archive contained a perfect match for 100% of all medic-charted HR, systolic blood pressure, and diastolic blood pressure values (means: 98%, 89%, and 91%, respectively). Overall, the majority of medic-charted vital signs of each subject matched the APPRAISE archive. Figure 4 shows several examples of archived data from APPRAISE (continuous line) versus hand-recorded vital signs by medics (solid circles) during the transport of three consecutive patients. Aside from those missions that terminated early due to loss of power, we did not identify any issues related to data communication or archiving.

Finally, we found that our analysis algorithms produced their outputs every 2 min, as configured, indicating that the system had sufficient computational capability for typical usage. The records had a mean duration of 26.5 min (SD 17.2 min). With the analysis routines configured to repeat every 2 min, we found that the average analysis time was 12.1 sec (SD 2.3 sec), with a maximum duration of 18 sec.

3. Discussion

There has been recent interest in the development of a new generation of decision-support, alarm, and automated control systems, which is made possible by connecting an analysis platform to one or more external sources of electronic clinical data [5]. Our own group is working to apply multivariate analysis and time-series analysis to pre-hospital vital signs (e.g., Ref. [7]) so that patients at high-risk of bleeding to death after trauma can be identified before arriving at a hospital, allowing for the earliest initiation of time-sensitive management protocols that have been shown to improve outcomes in trauma patients [14]. In addition, we are aware of a number of active commercial and research efforts to achieve related functionality, i.e., automated decision support driven by analysis of physiological data (e.g., Refs. [15–20]).

The necessary technical requirements can be organized into two broad categories: the reliability of the system’s individual components and how these components are connected together. Our ex-
experiences illustrate the relevance of these issues and raise important questions that need to be asked during the development and acquisition of any similar systems.

3.1 Is Each Component of the System Reliable?

3.1.1 Physical integrity

The GoBook PCs were suitable for prehospital use. Initially, we encased the PCs in plastic cages, which were attached to the bottom of the Propaqs [19]. The weight and bulk of the cage were unpopular with the flight crew, so we mounted the PC, unprotected, to the top of the Propaq. Alone, the GoBook was sufficiently compact and light (~2 lb.) that the flight crew accepted its presence. Yet within 6 mo, the demands of the environment were evident (one cracked PC case, one bent ¼-in. steel-plate mounting plate, and one damaged serial port), even though the ruggedized PCs continued to function. As an alternative to the GoBook, tablet PCs may have appealing form factors and weights, but should be carefully evaluated for their resilience to physical damage, particularly to various input/output and power supply ports.

3.1.2 Power

During missions, the Propaq/GoBook ran on battery power. Between missions, we depended on a very busy flight crew to keep our PC charged (the helicopter has alternating current power outlets). To facilitate this, we tethered the power cords for the GoBook and the Propaq together, so that when the crew plugged in the Propaq it would be natural to plug in the GoBook at the same time. For portable computing applications, a re-charging plan is essential. Our solution was adequate for research purposes, but mission-critical functionality may require a fail-safe plan (e.g., training, audible alarms, back-up devices, etc.). In hospital, uninterruptable power supplies will be essential for key electronic components.

We also learned that we required a solution for automatically recovering from a system shutdown. Before we fielded the GoBook, we first field tested the Switchback PC (Black Diamond Advanced Technology, Tempe, AZ) [21], and we learned that its battery supply was sometimes (~5–10% of missions) insufficient, causing it to shut down before mission completion. Moreover, although the crew would plug in the PC after the completion of the mission, as of 2009, the Switchback could not be configured to automatically boot-up upon the application of external power. As a result, the recharged Switchback remained off during subsequent missions. For this reason, we switched to the GoBook PC, which we configured to automatically turn on anytime it was connected to alternating current power. This substantially reduced downtime, allowing the platform to collect and analyze 798 complete records out of 866 deployments (92%) as of November 2012. Of the 68 incomplete records, where the power ran out prior to normal session timeout, 49 (72%) contained at least 10 minutes of data.

3.1.3 Computational capability

The computational requirements for novel algorithms are important to consider. Our analysis platform accommodates algorithms of differing complexity via a configuration option to limit the actual analysis interval anywhere from a few seconds (for the simplest algorithms) to several minutes (for the most computationally demanding algorithms). In laboratory testing, we evaluated the analysis system under the computational load of the longest individual data records and determined that its performance was satisfactory. Of course, this performance is a function of the algorithm's complexity and it would be advisable to repeat such testing when substantial changes are made to the algorithm.

3.1.4 Accessories

An interconnected medical system can be affected by minor alterations. For example, our decision-support algorithms used the respiratory rate, among other parameters, as an indicator of hemorrhage (e.g., Ref. [22]). Yet, we discovered that, for a subset of MedFlight cases, the respiratory rate was not provided by the Propaq. After some investigation, we learned that the less-expensive ECG leads used on some of the helicopters did not support impedance pneumography. This illustrates how even minor accessories of an interconnected medical system can affect the system’s overall per-
formance and that after deployment, when any system component is altered, it will be important to consider if there are any ripple effects.

### 3.2 Is the Connectivity between Components Reliable?

#### 3.2.1 Device communication

Many medical devices use proprietary communication protocols, which pose challenges to reliable connectivity with non-source-vendor systems. There has been a movement to encourage open standards as a way of addressing some of these challenges (e.g., Ref. [5]). The current lack of protocol uniformity between manufacturers (and even between different device generations from the same manufacturer) makes it difficult to support “plug-and-play” connectivity. Different data formats and control logic lead to increased code complexity. On the legal side, documentation for these protocols is often protected by non-disclosure agreements, which prevents well-tested code from being shared in the form of reusable open-source libraries.

Those that undertake the challenge of extracting data from a specific medical device must begin by reimplementing the protocol from scratch, a process in which attention to detail is of the utmost importance. During the development of the APPRAISE system, we evaluated an initial implementation [19], which was created outside of our core research group. We noted that it was not completely reliable in reassembling and verifying the integrity of data packets sent over the serial connection and discovered possible data losses in its archiving mechanism, such as when an isolated corrupt packet caused subsequent valid packets to be discarded. This implementation was, therefore, never used in field operations. Accordingly, we undertook a complete rewrite of the software to 1) robustly detect and discard corrupt data packets; 2) correctly align vital signs of different frequencies in a way that preserves a common timeline, despite varying start times of individual vitals and missing data; and 3) accurately create a lossless archive of every packet sent and received during patient transport for post hoc forensic assessment of the platform and analysis routine operation. Incidentally, note that this simple system (one medical device connected to a PC) avoided another substantial risk of medical interconnectivity, which is the integrity of the overall network (e.g., Ref. [23]).

Interconnected medical systems may enable new capabilities, but they require careful consideration and testing. For those purchasing commercial analysis systems, it may be appropriate to question the vendor about the system's computational limits and reliability and survey other customers' experiences.

#### 3.2.2 Metadata accuracy

Metadata refers to the information necessary for a clinical informatics platform to analyze data from a peripheral medical device: timestamps, units of measurement, patient identifiers, etc. Incorrect, incomplete, or inconsistent metadata can lead to errors by the analysis engine, even if there is reliable communication of the primary electronic signals [24]. For example, if the analysis engine makes incorrect assumptions about the time-synchronization of the electronic signal, it may produce results that are completely meaningless. Such synchronization errors are more likely to occur if there is a lag between different signals, as we have described in a report about another data archiving platform [25].

Another example of an analysis error caused by incorrect metadata would be the association of a recorded signal with the wrong patient (such as inadvertently merging data from two consecutively monitored patients). Correct patient association must be carefully considered when conducting automated data analysis. The APPRAISE system relies on a simple heuristic to identify new sessions (patients), assuming that all patients transported by MedFlight will be separated by at least 5 min. To date, we have not witnessed any exceptions. In the hospital, however, where there are multiple patients and multiple devices sharing a network, the proper association of streaming electronic data with the correct patient can be more challenging. One report described an error in associating data with the correct patient due to the movement of patients from one location to another [26]. Broadly speaking, when attempting to analyze the data that are flowing through a connected system, it is important to anticipate potential metadata errors, such as mismatches in terms of time, units of measurement, and patient identity.
Conflict of Interest
There are no conflicts of interest reported by the authors.

Human Subjects Protections
We obtained IRB approval for data collection and analysis by the Massachusetts General Hospital and the U.S. Army Medical Research and Materiel Command Office of Research Protection, Fort Detrick, Maryland. The authors from the Biotechnology High Performance Computing Software Applications Institute received and analyzed deidentified data.

Acknowledgments
This work was sponsored by the U.S. Department of Defense Medical Research and Development Program (Grant # D10_I_AR_JG_773) and by the Combat Casualty Care Research Area Directorate of the U.S. Army Medical Research and Materiel Command, Fort Detrick, MD.
Fig. 1  Photograph of the Automated Processing of Physiologic Registry for Assessment of Injury Severity (AP-PRAISE) hardware components in the disassembled state. The GoBook MR-1 on the left is connected to the Propaq 206 on the right.

Fig. 2  Data path from the Propaq to outputs provided by the MATLAB decision-support algorithms.
Fig. 3 Transport timelines of the first 38 MedFlight patients demonstrate that the APPRAISE was operational during all 24 hours of the day. Dark gray bars represent the time during which the medics were with the patient. Black bars represent the period of helicopter flight. Light gray bars represent the time during which vital-sign data were recorded by APPRAISE.

Fig. 4 Comparison of archived and hand-written numeric data for three patients. Heart rate (HR) and blood pressure (BP) recorded by the medics (solid circles) are overlaid over the APPRAISE archives (continuous line).
References


