
A COMPARISON OF VITAL SIGNS CHARTED BY NURSES WITH AUTOMATED ACQUIRED VALUES USING WAVEFORM QUALITY INDICES

Monica Sapo², Shaozhi Wu^{1,4}, Shadnaz Asgari¹,
Norma McNair, MSN, RN³, Farzad Buxey²,
Neil Martin, MD⁵ and Xiao Hu, PhD¹

Sapo M, Wu S, Asgari S, McNair N, Buxey F, Martin N, Hu X. A comparison of vital signs charted by nurses with automated acquired values using waveform quality indices.

J Clin Monit Comput 2009

ABSTRACT. Objective. (1) To investigate if there exist any discrepancies between the values of vital signs charted by nurses and those recorded by bedside monitors for a group of patients admitted for neurocritical care. (2) To investigate possible interpretations of discrepancies by exploring information in the alarm messages and the raw waveform data from monitors. **Methods.** Each charted vital sign value was paired with a corresponding value from data collected by an archival program of bedside monitors such that the automatically archived data preceded the charted data and had minimal time lag to the charted value. Next, the absolute differences between the paired values were taken as the discrepancy between charted and automatically-collected data. Archived alarm messages were searched for technical alarms of sensor/lead failure types. Additionally, 7-min waveform data around the place of large discrepancy were analyzed using signal abnormality indices (SAI) for quantifying the quality of recorded signals. **Results.** About 31,145 pairs of systolic blood pressure (BP-S) and 67,097 pairs of SpO₂ were investigated. Seven and a half percent of systolic blood pressure pairs had a discrepancy greater than 20 mmHg and less than one percent of the SpO₂ pairs had a discrepancy greater than 10. We could not find any technical alarms from the monitors that could explain the large difference. However, SAI calculated for the waveforms associated with this group of cases was significantly larger than the SAI values calculated for the control waveform data of the same patients with small discrepancies. **Conclusion.** Charted vital signs reflect in large the raw data as reported by bedside monitors. Poor signal quality could partially explain the existence of cases of large discrepancies.

KEY WORDS. clinical decision support, patient monitor, nurse charting, signal abnormality index.

Monica Sapo and Shaozhi Wu made an equal contribution in writing this paper.

From the ¹Neural Systems and Dynamics Lab, Department of Neurosurgery, David Geffen School of Medicine, University of California, NPI 18-240, Los Angeles, CA 90095, USA; ²Neuroinformatics Service and Support, Department of Neurosurgery, David Geffen School of Medicine, University of California, Los Angeles, CA, USA; ³Department of Nursing, Ronald Reagan UCLA Medical Center, Los Angeles, CA, USA; ⁴School of Computer Science and Engineering, University of Electronic Science and Technology of China, 610054 Chengdu, China; ⁵Department of Neurosurgery, Ronald Reagan UCLA Medical Center, Los Angeles, CA, USA.

Received 24 April 2009. Accepted for publication 7 July 2009.

Address correspondence to X. Hu, Neural Systems and Dynamics Lab, Department of Neurosurgery, David Geffen School of Medicine, University of California, NPI 18-240, Los Angeles, CA 90095, USA.

E-mail: xhu@mednet.ucla.edu

INTRODUCTION

The practice of utilizing electronic medical records (EMR) in clinical decision support is increasing. Integrating information extracted from vital signs including blood pressure, heart rate, respiratory rate, etc. is an important way of realizing a data-driven clinical decision support system. Considering that the common practice of capturing these data into EMR is through frequent and extensive nurse charting, the fidelity of vital signs thus captured is therefore a fundamental question to be investigated before developing vital signs-based decision support algorithms.

A handful of articles have been written investigating the accuracy and higher resolution of the data captured using automatic documentation systems [1–6]. For these analyses, data were simultaneously collected using an automated data archival program and manually charted nurse documentation from either paper records or through a computer-charting application. The quality of the data extracted from the automated monitoring systems was assessed by measuring the percentage of time each manually charted parameter matched the corresponding raw values [2, 7–10]. In several studies, obvious artifacts were excluded prior to comparison [3, 4].

Alternatively, never before have these studies utilized the waveforms and alarms from the bedside monitoring systems to provide quantitative interpretations of discrepancies that do exist between charting data and the source data. Artifacts have been identified through the study of analog waveforms, but these results were commonly excluded and have never been the focus of a study [4]. The aim of this paper was to identify and discuss the source of the discrepancies found through analysis of waveforms and alarms extracted from a bedside monitor archival system.

METHODS AND MATERIALS

Subjects and data acquisition

During an 11 month period, continuously acquired vital signs of 384 consecutive adult patients admitted into the neurosurgical intensive care unit (ICU) at UCLA Medical Center were studied. A waiver of patient consent was granted by the institutional internal review board for this retrospective data analysis that does not involve patient information. No exclusion criteria were applied. Our focus was primarily on two vitals including blood oxygen saturation (SpO₂) and systolic arterial blood pressure (ABP-S). All SpO₂ and ABP-S measurements were recorded using two clinical information systems (CIS) simultaneously. The nursing staff regularly enters vital signs into a nurse charting system, Essentris™, at the end of each hour, and more frequently depending on the severity of the patient's condition. Concurrently, these vital signs are automatically captured every 15 min by the BedMasterEx™ program that was configured to automatically archive raw data for all 24 beds in the neurosurgical ICU. Data from each source are regularly extracted into well organized databases using SQL-based extraction–transformation–loading (ETL) tools developed in house.

Discover discrepancy

For this study, patient data were matched from each source by parameter type and time. Since vitals are not expected to be charted at the precise time of automatically archived data, each charted nursing vital sign value was compared with the most recent data point captured by BedMaster within a 15-min time interval prior to charting time. The data was filtered in such a way that the automatically archived data preceded the charted data and had minimal time lag to the charted value. A list of all corresponding values was assembled for each parameter and then analyzed further. For ABP-S, any value below 20 or greater than 250 was eliminated as erroneous data for both sources.

The absolute differences between the paired values were calculated for each record and designated as the discrepancy between charted and automatically-collected data. For SpO₂, any discrepancy greater than 10 was considered “extreme”. Additionally, for ABP-S, any discrepancy greater than 20 mmHg was considered “extreme”.

Interpret discrepancy

We hypothesize that the discrepancy is most likely due to the noise in the raw signals, which was screened and ignored by the nurse documenting process. We next sought for evidence of this using two means. Archived alarm messages from the monitor were first investigated for each of the “extreme” values on the basis that modern patient monitors usually have certain built-in noise recognition ability and that they are capable of detecting noise and artifacts due to technical faults such as sensor detachment [11]. Each parameter has a designated set of alarm codes which were used to filter the dataset and identify technical alarms including sensor and lead failure. Each alarm message is comprised of a start and end time as well as a brief description of the error that has occurred. Any vitals recorded within the start and end times of the alarm messages were noted.

Additionally, 7-min waveform data around the time instant of large discrepancy were collected for the ten patients with the most “extreme” data points for each parameter. For comparison purposes, we also collected 7-min waveform data around a control instance where little to no discrepancy existed. These waveform data were further processed to quantify their quality using a recently proposed signal abnormality index (SAI) [12] for ABP waveform and a new algorithm (to be described in next Section) for the SpO₂ signal. With these quantitative measures of signal quality, we can test the hypothesis that

the signal has a better quality at control regions as compared to those around the point with large discrepancy between charted and automatically-archived data.

Signal quality quantification

Original signal abnormality index (OSAI) for arterial blood pressure signal

OSAI algorithm was proposed by Sun [12] to flag abnormal ABP pulse waveforms. Flowchart of the algorithm is shown in Figure 1. First, a pulse detection algorithm [13] is used to mark the onset of each pulse. The following signal features are then obtained from each consecutive pulse: heart rate (HR), systolic blood pressure (P_s), diastolic blood pressure (P_d), mean blood pressure (P_m), pulse amplitude (P_p), change of systolic pressure ($\text{delta}P_s = P_s[k] - P_s[k-1]$), change of diastolic pressure ($\text{delta}P_d = P_d[k] - P_d[k-1]$), and change of heart rate ($\text{delta}T = \text{HR}[k] - \text{HR}[k-1]$). Each of these features is then judged by a corresponding abnormality criterion listed in Table 1 so that a score of 1 will be assigned to the pulse if any of the criteria is satisfied otherwise a score of 0 will be assigned. To obtain a cumulative score (cSAI) for a segment of ABP signal, an average of scores of individual pulses is obtained that will be between 0 and 1. cSAI can be considered as quantitative measure of the signal quality. A large cSAI indicates a segment of ABP pulse of poor quality.

Extension of signal abnormality index for a general pulsatile signal

Criteria listed in Table 1 are physiologically motivated. Therefore, one could imagine that a new set of criteria would have to be defined if the algorithm is to be applied to a pulsatile signal other than ABP. As an alternative approach, we have recently proposed a signal subspace decomposition approach to determine whether a pulse is valid [14]. This algorithm was originally conceived for

Table 1. The criteria for flagging an ABP beat

Feature	Abnormality criteria
P_s	$P_s > 300$ mmHg
P_d	$P_d < 20$ mmHg
P_m	$P_m < 30$ or $P_m > 200$ mmHg
Hr	$\text{Hr} < 20$ or $\text{Hr} > 200$ bpm
P_p	$P_p < 20$ mmHg
$\text{delta}P_s$	$P_s[k] - P_s[k-1] \mid P_s \mid > 20$ mmHg
$\text{delta}P_d$	$P_d[k] - P_d[k-1] \mid P_d \mid > 20$ mmHg
$\text{delta}T$	$T[k] - T[k-1] \mid T \mid > 2/3$ s

determining the validity of intracranial pressure pulse as a component of a computerized algorithm of intracranial pressure pulse analysis [15]. Here we extend its usage as a signal quality index of general pulsatile signals such as SpO₂. We call it subspace decomposition signal abnormality index (SDSAI). A comparison of this subspace based algorithm and the physiological-criteria based SAI approach on the same pulsatile signals can be found in Asgari [14]. The flowchart of SDSAI is shown in Figure 2.

Subspace decomposition signal abnormality index needs a collection of valid pulses, which is termed the reference pulse library. A pulse normalization algorithm is then applied to the pulses in the library so that all pulses have the same length and unit standard deviation. After this normalization, the pulse library can be represented as a matrix R . Singular value decomposition (SVD) is then applied to R such that $R = U \times S \times V$ where the columns of orthogonal matrix U are organized based on the descending order of their corresponding singular values. Based on the theory of SVD, columns of U are further separated into a signal and a noise subspaces U_1 and U_2 , where U_1 consists of the first k columns and U_2 contains the rest of columns. k is hence called the dimension of the signal subspace. The quality of a SpO₂ pulse x is quantified

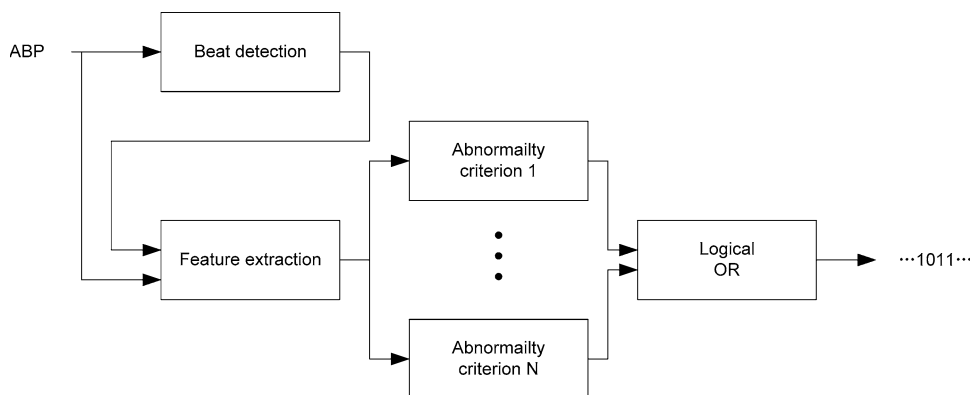


Fig. 1. OSAI block diagram. Input is a sequence of arterial blood pressure pulses. Output is a logic value (no flag = 0, flag = 1) for each pulse.

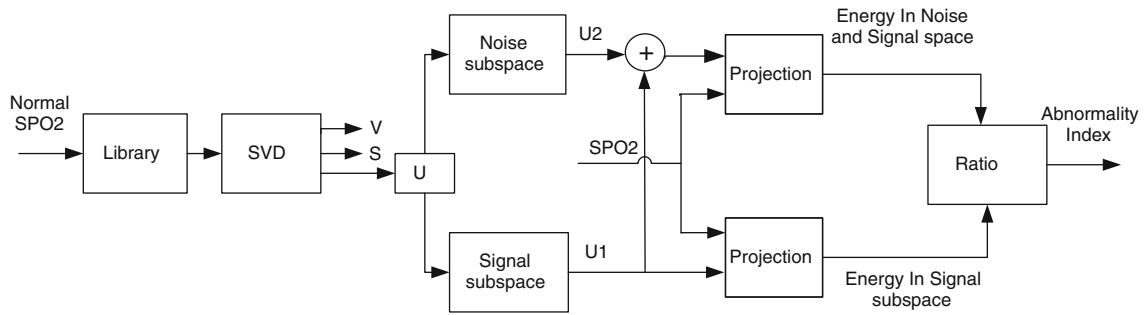


Fig. 2. SDSAI calculation flow chart. The key step in the figure is to project each pulse to the noise and signal subspaces that are constructed from using the singular value decomposition of a library of SPO₂ pulses.

by using the ratio between the energy of x 's projection on the signal subspace and energy of the original pulse as $SDSAI = 1 - \frac{\|x_s\|^2}{\|x\|^2}$ where $x_s = U_1 U_1^T x$ and $\|x\|^2$ is the square of the 2-norm of a vector. Based on this definition, a large SDSAI corresponds to low signal quality.

In the present work, we built a reference library comprised of 230 valid SpO₂ pulses, sampled at 400 Hz, which were manually selected from 29 patients. To apply the SDSAI approach, the lengths of the pulses in the library were all resized to 393 samples, that is, the 90th percentile of the lengths of all pulses in the library. The optimal dimension of signal subspace k was determined as the “knee” point of the singular value spectrum. For this dataset, $k = 9$ was found to be optimal.

RESULTS

At least 1 h of physiological vital sign data was collected using BedMaster for 384 ICU admissions, which included 302 unique patients. Table 2 shows characteristics of the ICU admissions. About 38,022 h of BedMaster vital sign data were collected, during which time over 100,000 manual charting instances were recorded. A total number of 31,145 pairs of systolic blood pressure and 67,097 pairs of SpO₂ were investigated.

Figure 3 shows the histograms of values of the absolute discrepancy of ABP-S and SpO₂ between the nursing chart and monitor data. 7.5% of systolic blood pressure pairs had a discrepancy greater than 20 mmHg, of which only two instances were associated with alarm messages. In addition, only 0.4% of SpO₂ pairs had a discrepancy greater than 10, of which none was associated with alarm messages. We present four representative examples comparing a 24-h time series of ABP-S (Panels a and b) and SPO₂ (Panels c and d) from the nursing chart and the monitors in Figure 4. It is observed that the raw data, as recorded by monitors, often show wide swings in

Table 2. Patient characteristics by physiological parameter

Patient characteristic	ABP-S ($n = 123$)	SpO ₂ ($n = 304$)
Age, year, mean + SD	49.0 ± 16	51.4 ± 19
Gender		
Female (%)	42.28%	41.45%
Male (%)	57.72%	58.55%
Hospital length of stay, days, median (IQR)	12 (6–30)	9 (5–21)
ICU length of stay, days, median (IQR)	5 (1–14)	2 (1–7)

physiological parameters which are missing from the less frequently charted data. This is considered as a smoothing effect of nurse charting on the raw data.

Figure 5 gives an example of using the SAI algorithm to flag noisy ABP pulses. It can be seen that seven pulses in the data segment shown in the figure were marked as noise based on the amplitude criterion in Table 1. Figure 6 compares the histogram of the cSAI calculated for the 1-min around the time points of large chart-monitor discrepancy and a control minute of data that did not have large discrepancy. It can be observed that cSAI is highly skewed toward lower scores for the control (normal) group with 30% more cases in the lowest bracket. Figure 7 displays a representative example of calculating the SDSAI score for noisy SpO₂ pulses. It can be seen that the algorithm is highly sensitive to the change of the quality of SpO₂ pulses. By studying the histograms of the SpO₂ SDSAI scores for the two groups in Figure 8, we observed the similar pattern as in cSAI scores of ABP pulses. SDSAI for the control group is biased toward the lower spectrum of SDSAI histogram indicating a generally better quality of SpO₂ pulses. However, this difference is less than what is observed in ABP pulses.

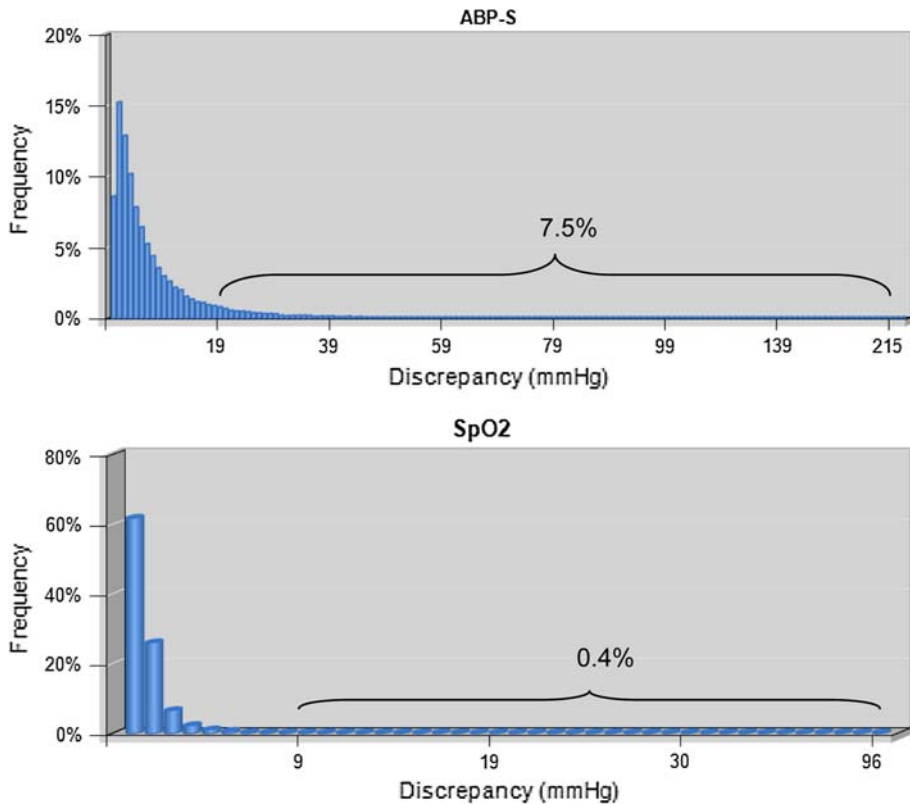


Fig. 3. Histogram of the quantified difference between each manually charted and automated record of systolic ABP and SpO₂. (a) Results for systolic arterial blood pressure (ABP-S). Differences greater than 20 mmHg were considered ‘discrepancies’ that comprises 7.5% of all BP records. (b) Results for percentage oxygen saturation in the blood (SpO₂). Differences greater than 10 were considered ‘discrepancies’ that include 0.4% of all SpO₂ records.

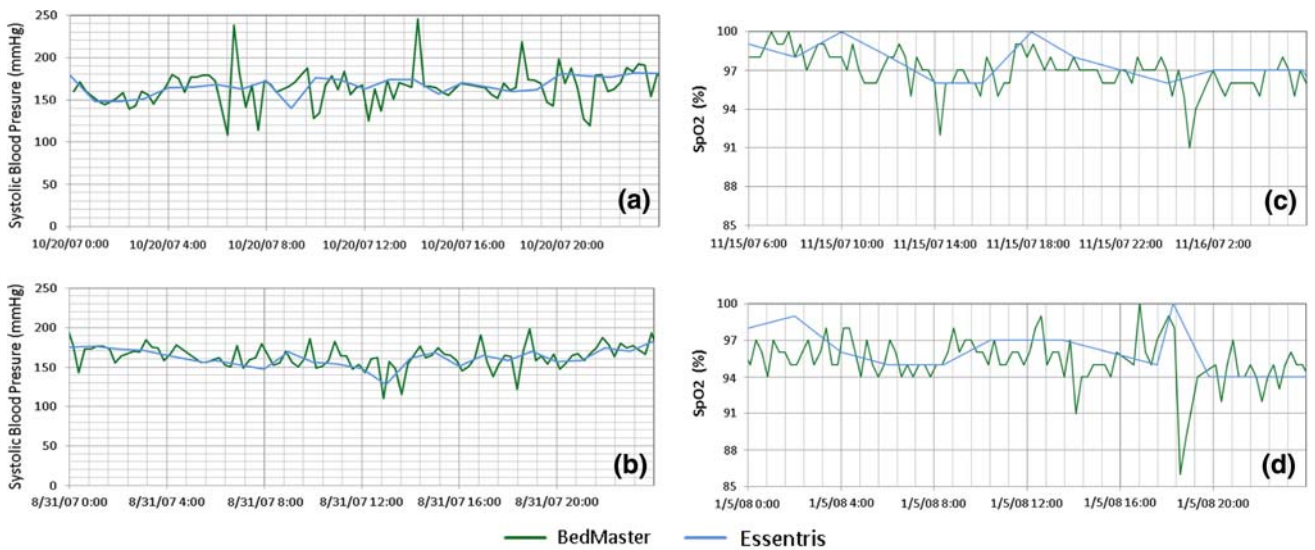


Fig. 4. Four examples of raw patient data collected for a 24 h time period for four distinct patients comparing manually charted vitals (blue) and automated data (green). These examples show that the charting process has a smoothing effect on the raw data.

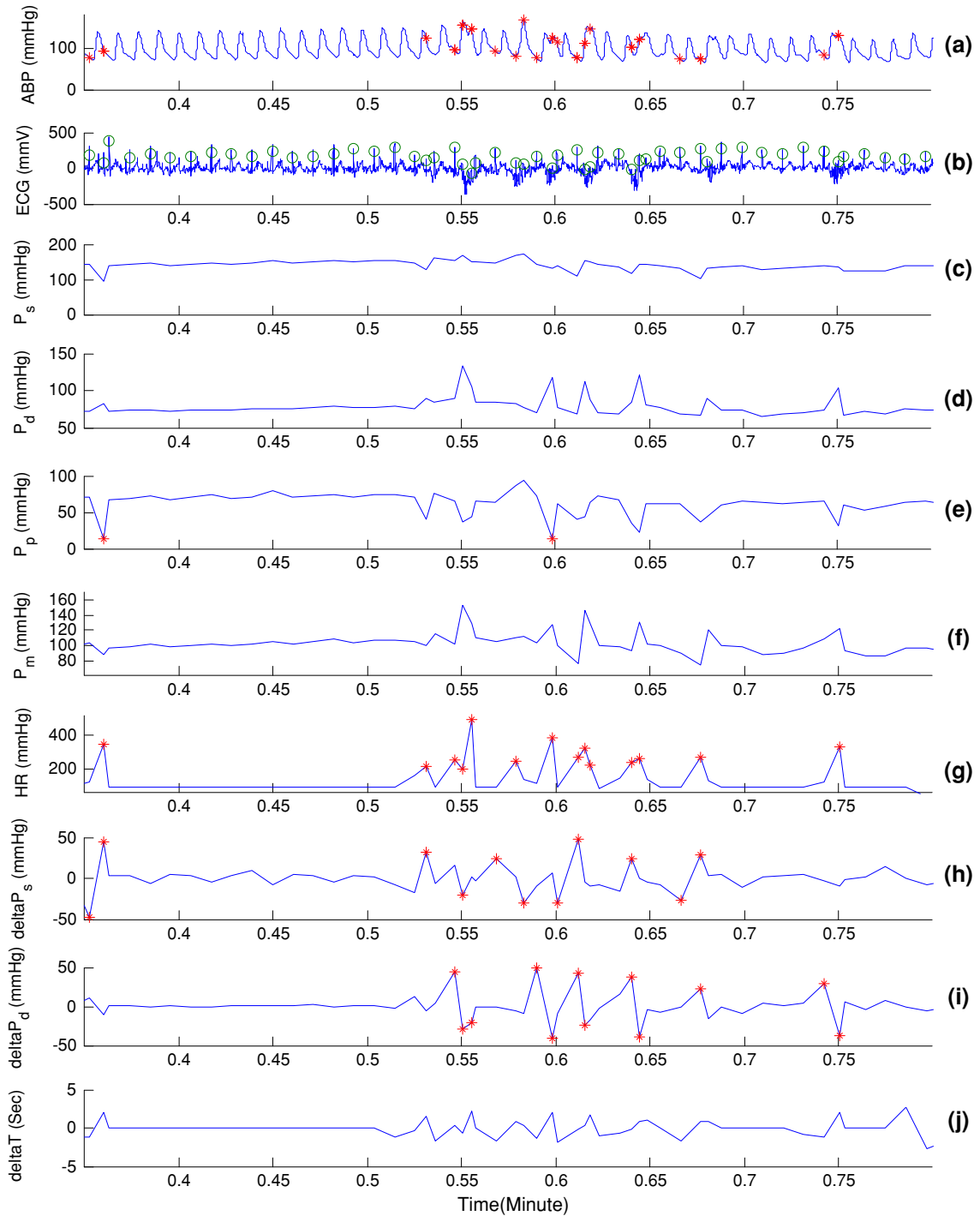


Fig. 5. Demonstration of cSAI scoring results for an arbitrary ABP waveform that contains low quality pulses. Each noisy pulses are marked with a red ‘*’. In addition, the corresponding criteria that triggered the recognition of these noisy pulses are also marked.

Table 3 lists the mean and standard deviation of the two signal abnormality indices for the 7 min of data around the time instant of extreme discrepancy and those of their corresponding control 1-min segment. Cases are marked

with “*” when the difference of the mean signal abnormality indices are significantly ($P < 0.05$) different, by a two-sample t -test, between the extreme and the control groups.

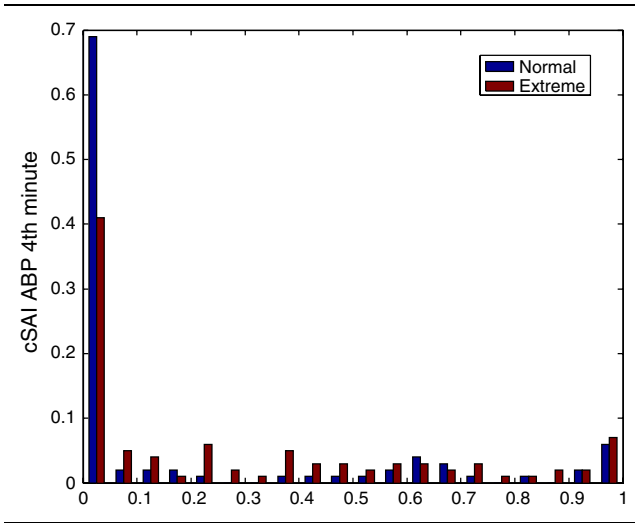


Fig. 6. Comparing the histogram of cSAI of systolic ABP values that were automatically recorded around the point of large monitor-chart discrepancies and that of systolic ABP values recorded around a point of minimal discrepancies.

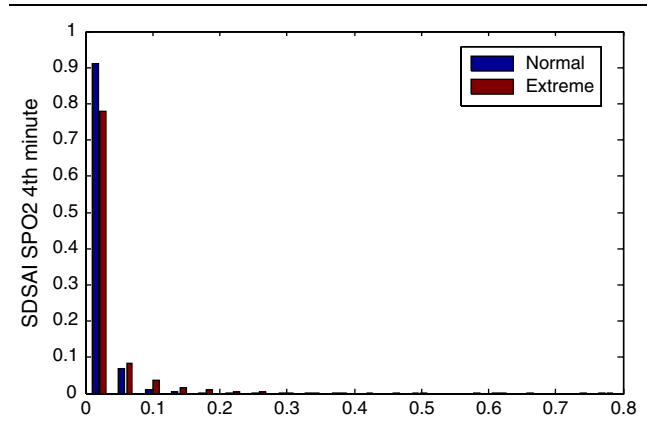


Fig. 8. Comparing the histogram of SDSAI of SpO_2 values that were automatically recorded around the point of large monitor-chart discrepancies and that of SpO_2 values recorded around a point of minimal discrepancies.

DISCUSSION

Based on our comparison between nurse charting and archived data from the monitors, the “phenomenon of smoothing” as recognized in 1991 [16] does indeed exist. Despite that vital signs captured by nurses have an overall agreement with automatically archived data, nursing data have fewer extreme samples resulting in a larger discrepancy between them when the raw data are of extreme values. This trend was first noted in a study by

Block [17], which retrospectively analyzed abnormal automated recordings of noninvasive heart rate and blood pressure. He found that while the majority of the readings were within normal ranges, 3.6% of systolic pressure readings, 13.25% of diastolic pressure readings and 4.25% of heart rate readings were recorded outside normal ranges. He concluded that fluctuations in heart rate and blood pressure are common and unexpectedly “out of range” values subject to averaging or dismissal are often accurate. Later, Reich et al. [9] did a comparison study analyzing 81 pairs of handwritten and computer-generated anesthesia records using a matched sample design. The handwritten records confirmed the “smoothing effect” previously

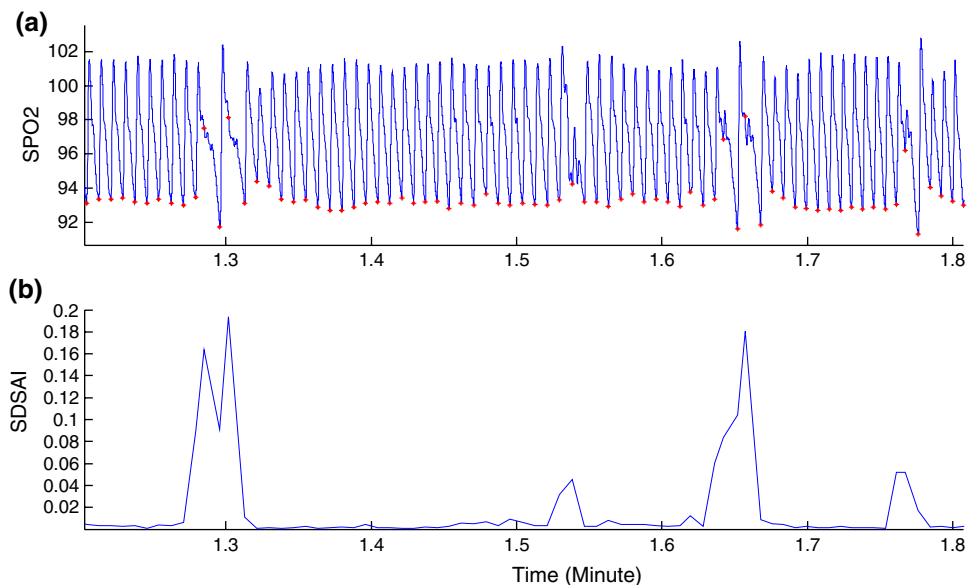


Fig. 7. Illustration SDSAI scoring process for a SpO_2 signal that contains noisy pulses. Higher SDSAI corresponds to noisy pulses.

Table 3. The mean and standard deviation of ABP cSAI and that of SPO2 SDSAI

Time point relative to time of large discrepancy (min)	Mean ABP cSAI		Mean SPO2 SDSAI	
	Normal	Extreme	Normal	Extreme
-3	0.2136 ± 0.3399	0.2873 ± 0.3673	0.0169 ± 0.0331 ^a	0.0219 ± 0.0403
-2	0.1973 ± 0.3313 ^a	0.3066 ± 0.3843	0.0164 ± 0.0303 ^a	0.0215 ± 0.0383
-1	0.1869 ± 0.3266 ^a	0.3271 ± 0.3815	0.0160 ± 0.0266 ^a	0.0257 ± 0.0467
0	0.1812 ± 0.3198 ^a	0.2933 ± 0.3361	0.0164 ± 0.0293 ^a	0.0248 ± 0.0481
1	0.1797 ± 0.3065 ^a	0.2798 ± 0.3153	0.0172 ± 0.0341 ^a	0.0217 ± 0.0413
2	0.1814 ± 0.3116 ^a	0.2205 ± 0.3080	0.0155 ± 0.0250 ^a	0.0225 ± 0.0425
3	0.2223 ± 0.3376	0.2135 ± 0.3135	0.0166 ± 0.0266 ^a	0.0220 ± 0.0477

^aIndicates the significant cases as detected by a two-sample *t*-test.

described by Block. The results consisted of lower peak and higher trough values for both systolic and diastolic blood pressure, in addition, extreme values were recorded more frequently in computerized records than in the handwritten records.

We furthered previous studies by searching the most possible reasons for the “smoothing effect”. This was attempted by co-registering alarm data and raw waveform data with the sampled vital signs. However, alarms did not provide much explanation for these discrepancies, probably for the reason that the source of erroneous raw data was due to inadequate/improper measurement conditions, which can be avoided by nurses at the time of charting. Nurses evaluate these vital signs based on what is on the monitor, previous trends of the patient and in response to alarms. Such evaluation process is a critical component in determining if data should be charted. For example, if a nurse is drawing blood from an arterial line, the alarm will sound as the ABP will alarm as “high”. However, the nurse is aware that she/he is drawing blood and would not include that alarm in any data that is gathered. Should the computer gather this data, the nurse would evaluate it and determine if it was correct or not. If the data is not correct, the nurse could then correct the information. On the other hand, artifact-related alarms, which were studied in the present work, would have only been triggered for technical failures including lead failure or sensor disconnections.

We have some interesting observations from the analysis of waveform data. Both SDSAI (for SpO₂) and cSAI (for ABP pulse) in general indicate that signal quality was worse at situations of greater discrepancy. This is supportive of the fact that nurses can leverage the clinical context where the measurement is taken to filter out invalid values that could be associated, in an expected way, with low quality or noisy waveform. However, there were still many “extreme” cases where SDSAI and cSAI were within the

“normal” range indicating that signal quality was adequate. One possible reason for this observation could be that SDSAI and cSAI do not accurately reflect the quality of the waveform signals. However, we believe this may not be the primary reason because cSAI is a published approach and has been extensively validated. Similarly, we have extensively validated SDSAI approach before applying it in this study. Alternatively, one could argue that some unknown filtering operations were introduced during the nurse charting process. However, the exact characteristics of these filtering operations cannot be discerned from the present work. Despite that it is unclear whether these filtered extreme value carry any clinical value, our observation would caution the attempt to develop clinical decision-making algorithm based on nurse-charting data. Clinical decision systems would be more useful to capture truly extreme deviations from the normal physiological state if they were driven by data directly acquired from monitors and if effective noise filtering is in place. The motivation of developing algorithms including SDSAI and cSAI is exactly for such purpose.

Documentation of vital signs is an ongoing and time intensive task for nurses caring for critically ill patients. Patients in intensive care units often require hourly or more frequent documentation as nurses monitor responses to changes in medication dosing, nursing care procedures and other bedside activities. If reliable signal quality index can be developed and deployed at the spot of charting to eliminate erroneous data collection, valuable nursing time could then be used for other patient or family focused activities [17–19]. Such an effort needs further investigation.

This work is supported in part by a UCLA Clinical Translational Science Institute Scholar Award to XH.

REFERENCES

1. Ward NS. The accuracy of clinical information systems. *J Crit Care* 2004; 19: 221–225.
2. Vawdrey DK, Gardner RM, Evans RS, Orme JF Jr, Clemmer TP, Greenway L, Drews FA. Assessing data quality in manual entry of ventilator settings. *J Am Med Inform Assoc* 2007; 14: 295–303.
3. Friesdorf W, Konichezky S, Gross-Alltag F, Fattroth A, Schwilk B. Data quality of bedside monitoring in an intensive care unit. *Int J Clin Monit Comput* 1994; 11: 123–128.
4. Gregorini P. Comparison of four methods of automated recording of physiologic data at one minute intervals. *J Clin Monit* 1996; 12: 299–303.
5. Levy MM. Computers in the intensive care unit. *J Crit Care* 2004; 19: 199–200.
6. Gardner RM, Hawley WL, East TD, Oniki TA, Young HF. Real time data acquisition: recommendations for the Medical Information Bus (MIB). *Int J Clin Monit Comput* 1991; 8: 251–258.
7. Pringle M, Ward P, Chilvers C. Assessment of the completeness and accuracy of computer medical records in four practices committed to recording data on computer. *Br J Gen Pract* 1995; 45: 537–541.
8. Marasovic C, Kenney C, Elliott D, Sindhusake D. A comparison of nursing activities associated with manual and automated documentation in an Australian intensive care unit. *Comput Nurs* 1997; 15: 205–211.
9. Reich DL, Wood RK Jr, Mattar R, Krol M, Adams DC, Hossain S, Bodian CA. Arterial blood pressure and heart rate discrepancies between handwritten and computerized anesthesia records. *Anesth Analg* 2000; 91: 612–616.
10. Moret-Bonillo V, Mosqueira-Rey E, Alonso-Betanzos A. Information analysis and validation of intelligent monitoring systems in intensive care units. *IEEE Trans Inf Technol Biomed* 1997; 1: 87–99.
11. Mora FA, Passariello G, Carrault G, Le Pichon JP. Intelligent patient monitoring and management systems: a review. *IEEE Eng Med Biol Mag* 1993; 12: 23–33.
12. Sun JX, Reisner AT, Mark RG. A signal abnormality index for arterial blood pressure waveforms. *Comput Cardiol* 2006; 33: 13–16.
13. Hu X, Xu P, Lee DJ, Vespa P, Baldwin K, Bergsneider M. An algorithm for extracting intracranial pressure latency relative to electrocardiogram R wave. *Physiol Meas* 2008; 29: 459–471.
14. Asgari S, Xu P, Bergsneider M, Hu X. A subspace decomposition approach toward recognizing valid pulsatile signals. *Physiol Meas* 2009 (submitted).
15. Hu X, Xu P, Scalzo F, Vespa P, Bergsneider M. Morphological clustering and analysis of continuous intracranial pressure. *IEEE Trans Biomed Eng* 2009; 56: 696–705.
16. Asgari S, Bergsneider M, Hu X. Robust approach towards recognizing valid arterial blood pressure pulses. *IEEE Trans Inf Technol Biomed* 2009 (submitted).
17. Block FE Jr. Normal fluctuation of physiologic cardiovascular variables during anesthesia and the phenomenon of “smoothing”. *J Clin Monit* 1991; 7: 141–145.
18. East TD, Young WH, Gardner RM. Digital electronic communication between ICU ventilators and computers and printers. *Respir Care* 1992; 37: 1113–1123.
19. Donati A, Gabbanelli V, Pantanetti S, Carletti P, Principi T, Marini B, Nataloni S, Sambo G, Pelaia P. The impact of a clinical information system in an intensive care unit. *J Clin Monit Comput* 2008; 22: 31–36.