Algorithms to qualify respiratory data collected during the transport of trauma patients

Liangyou Chen¹, Thomas McKenna¹, Andrew Reisner¹,² and Jaques Reifman¹

¹ Bioinformatics Cell, US Army Medical Research and Materiel Command, ATTN: MCMR-ZB-T, Building 363, Miller Drive, Ft. Detrick, MD 21702, USA
² Department of Emergency Medicine, Massachusetts General Hospital, MGH, 55 Fruit Street, Clinics 116, Boston, MA 02114, USA

E-mail: lchen@bioanalysis.org, tmckenna@bioanalysis.org, andrewtreisner@yahoo.com and jaques.reifman@us.army.mil

Received 12 April 2006, accepted for publication 1 June 2006
Published 27 June 2006
Online at stacks.iop.org/PM/27/797

Abstract

We developed a quality indexing system to numerically qualify respiratory data collected by vital-sign monitors in order to support reliable post-hoc mining of respiratory data. Each monitor-provided (reference) respiratory rate (RRR) is evaluated, second-by-second, to quantify the reliability of the rate with a quality index (QI R). The quality index is calculated from: (1) a breath identification algorithm that identifies breaths of ‘typical’ sizes and recalculates the respiratory rate (RRC); (2) an evaluation of the respiratory waveform quality (QIW) by assessing waveform ambiguities as they impact the calculation of respiratory rates and (3) decision rules that assign a QIR based on RRR, RRC and QIW. RRC, QIW and QIR were compared to rates and quality indices independently determined by human experts, with the human measures used as the ‘gold standard’, for 163 randomly chosen 15 s respiratory waveform samples from our database. The RRC more closely matches the rates determined by human evaluation of the waveforms than does the RRR (difference of 3.2 ± 4.6 breaths min⁻¹ versus 14.3 ± 19.3 breaths min⁻¹, mean ± STD, p < 0.05). Higher QIW is found to be associated with smaller differences between calculated and human-evaluated rates (average differences of 1.7 and 8.1 breaths min⁻¹ for the best and worst QIW, respectively). Establishment of QIW and QIR, which ranges from 0 for the worst-quality data to 3 for the best, provides a succinct quantitative measure that allows for automatic and systematic selection of respiratory waveforms and rates based on their data quality.
1. Introduction

Respiratory waveform data collected from trauma patients during transport from the scene of the accident to a care facility are noisy, leading to difficult post-hoc trauma data mining. This paper presents methodologies and algorithms to automatically and systematically qualify respiratory waveforms and their derived rates, thereby permitting the formulation of a numerical score or quality index (QI) for the data that allow investigators to select high-quality data for mining and development of decision-support algorithms. This outcome is accomplished through assessment of waveform quality, recalculation of respiratory rates from the waveforms and comparison of the calculated respiratory rates with the vital-sign monitor-derived rates.

The respiratory rate is a fundamental vital sign. It is associated with the modulation of cardiac vagal outflow and cardiac vagal tone through respiratory sinus arrhythmia (Berntson et al. 1993, Yasuma and Hayano 2004). Acute respiratory distress syndrome and serious sleep-related breathing disorder have deleterious effects on the cardiovascular system and are associated with heart failure, artery diseases and lung injuries (Javaheri 2003, Kallet et al. 2006, Obenza Nishime et al. 2000). Moreover, patterns of respiration, such as alternating respiration (Thomas et al. 2004) and Cheyne–Stokes respiration (Lorenz and Ito 1978), have diagnostic value (Isakov et al. 1979, Bounhoure et al. 2005). Thus, clinicians suggest the continuous monitoring of respiratory status to identify events related to life-threatening respiratory, cardiac and neuromuscular diseases (Burdett-Smith 1997, Folke et al. 2003, Mellies et al. 2005) and to employ patterns and rate of respiration to evaluate respiratory characteristics during mechanical ventilation (Yang and Yang 2002, Lucangelo et al. 2005), indexing of cardiac parasympathetic control (Grossman et al. 2003) and estimation of trauma severity (Frankema et al. 2005).

The respiratory rate is usually derived from respiratory waveforms continuously calculated by vital-sign-monitor devices (Brouillette et al. 1987, Protocol® Systems Inc. 1998). Respiratory waveform signals are generated through sensor electrodes externally attached to patients as a means of non-invasive monitoring. Techniques, such as impedance pneumography and inductive plethysmography, are used to measure the movement of the rib cage and the abdominal compartment during respiratory cycles and generate the respiratory waveforms (Brouillette et al. 1987, Protocol® Systems Inc. 1998, Adams et al. 1993). Respiratory waveforms may also be derived from electrocardiogram (ECG) waveforms (Moody et al. 1986, De Chazal et al. 2003), if the ECG sampling rate is sufficiently high to allow accurate measurement of the QRS complex (i.e., the shape of the ECG wave). Respiratory waveform has been used for clinical detection of apnea and ventilation (Weese-Mayer et al. 2000, Que et al. 2002); however, its primary use is the derivation of respiratory rates through identification of breaths encoded in the waveform. Many vital-sign monitors can provide respiratory waveforms and rates in a digital data format that can be stored for post-hoc analysis.

Despite the broad utility of respiratory data, methods for determination of signal quality are not widely available, even though the respiratory signal quality itself may have clinical utility (Alametsa et al. 2005). Unfortunately, the low-frequency respiratory signal is subject to
Figure 1. Low-quality respiratory waveforms affect the calculation of respiratory rates. Examples include: (a) uncertain breaths at time intervals 686–687 and 695–696; (b) breaths immersed in quick fluctuations at time interval 40–45; (c) noise at time intervals 917–918 and 925–926 and (d) spikes at time intervals 622–627 and 635–636.

movement artefact and erroneous placement of sensor electrodes on the body (Cohen et al 1997, Ernst et al 1999, Khambete et al 2000), leading to respiratory waveforms with low signal-to-noise ratio and ambiguities between breaths and signal fluctuations. Such ambiguities can be seen in the waveforms illustrated in figure 1, where relatively low-amplitude breaths, waveform signal fluctuations, noise and spikes can all generate ambiguous breaths. Ambiguities in breath determination greatly affect the calculation of the respiratory rate. For example, based on a 15 s window, missing only one breath from a 20 breaths min$^{-1}$ respiratory waveform will result in a 20% difference between the calculated and actual respiratory rates. This causes a substantial difficulty for accurate respiratory rate analysis. Moreover, respiratory rates calculated by different methods can differ significantly in the presence of ambiguous breaths, and the lack of a common standard for breath definition leads to subjective respiratory rate evaluation, even by experienced clinicians (Adams et al 1993, Weese-Mayer et al 2000, Liu et al 2004, Burns 2003).

Our ultimate goal is to develop a new generation of automated triage, diagnostic and prognostic algorithms that can offer decision support to caregivers during the prehospital care of trauma casualties. Automatic qualification of respiratory data is a first step towards this goal. We consider respiratory data collected in a dynamic, field environment as of suspect quality, as movement artefact and unstable sensor attachment are common. Under such conditions, a primary objective, prior to post-hoc data mining, is to evaluate respiratory data for quality and, thereby, allow investigators to choose what data are suitable for analysis.
Figure 2. The respiratory data qualification algorithm is composed of three components: breath identification, waveform qualification and decision logic.

2. Methods

Our respiratory data qualification method calculates a quality index QI for each sampled reference respiratory rate, RR, and its associated waveform from which the rate is computed. The calculated respiratory rate quality index, QIR, and respiratory waveform quality index, QIW, range from 0 to 3 to reflect the quality of the respiratory rate and waveform from worst to best, respectively. This method, depicted in figure 2, consists of three components: (1) breath identification—to independently recalculate respiratory rates RR from the monitor-provided respiratory waveforms; (2) waveform qualification—to assess the waveform quality QI and (3) decision logic—to assess a respiratory rate quality index QIR based on the waveform quality QIW and the level of similarity between RR and RR.

Our method does not depend on a specific vital-sign monitor. Inputs to the method include a reference respiratory rate and its associated respiratory waveform segment, and the outputs are the quality indices QIR and QIW. Method parameters can be easily adjusted for respiratory waveforms sampled at different sampling rates. Before detailing the method, we first introduce the data set used to develop and validate the qualification algorithm.

2.1. The trauma data set

Complete vital-sign records for 711 trauma injured patients (of which 81 died) are warehoused in our system for management and analysis of time-series physiological data termed the physiology analysis system (PAS) (Reifman et al. 2004). These data were collected from patients during transport by Life Flight helicopter service from the site of injury to the Memorial Hermann Hospital, an urban regional trauma center at Houston, Texas (Convertino and Holcomb 2003). Vital-sign data were collected during flight with a Propaq Encore® 206EL vital-sign monitor (Protocol® Systems Inc. 1998); the data were then exported onto a personal digital assistant flashcard, uploaded to a local database (Trauma Vitals 2004) and then exported into PAS. These patients vary in age, gender, race and severity of injury. The lack of structure and potential for movement artefacts in the prehospital setting generally results in poor-quality signals corrupted by noise. Furthermore, the variety of patients’ demographics and mechanisms of injury increase the difficulty in identifying common patterns in the respiratory waveforms.
The Propaq monitor collects respiratory waveforms by impedance pneumography from the same leads used to collect electrocardiogram data. The monitor determines which signals are artefacts and which signals result from actual respiratory efforts. The respiratory waveform is collected at 22.7 Hz (equivalent to one sample every 44 ms) and it is used to compute reference respiratory rates $RR_R$, which are provided at 1.0 Hz intervals (Protocol\textsuperscript{©} Systems Inc. 1998).

### 2.2. Calculation of respiratory rate from respiratory waveform

Each calculated respiratory rate at time $t$, $RR_C(t)$, is based on a 15 s waveform window ending at $t$. For reasonable respiratory rates of 20–60 breaths min$^{-1}$, this window contains 5–15 breaths. Shorter windows may not contain sufficient breaths for reliable estimation of respiratory rates. Conversely, longer windows may not reflect an accurate respiratory rate at the end of the window due to potential averaging effects. A 15 s window balances these two concerns. After calculating $RR_C(t)$, the window is advanced by 1 s and the new window is used to calculate $RR_C(t + 1)$. This continues to the end of the waveform.

Our respiratory rate computation is based on the estimation of a respiratory waveform ‘baseline.’ A baseline is a moving average of the waveform and is calculated by sequentially averaging the amplitude of overlapping waveform segments defined by a moving window in time and thus forming a trace of averaged values for the waveform. By using windows of different lengths, we attain different baselines or trends of the waveform. The relative amplitude of the original waveform above a chosen baseline is used to identify a breath, which is then used to calculate the respiratory rate.

An explicit assumption in our methodology is that only breaths of ‘typical sizes’ are real respiratory efforts, and deviance from a typical-size breath leads to poor data quality. Under this assumption, our algorithm identifies typical breaths in a waveform segment by dynamically adjusting its breath identification criteria based on the initially identified breaths, and calculates $RR_C$ and $QIW$ from the typical breaths. Before detailing this method, we first introduce some concepts.

#### 2.2.1. Concepts

Let a time series of length $T$ be represented by a vector $x$, whose value at position $t$, $t = 1, 2, \ldots, T$, is denoted as $x(t)$. A time window of size $p$ at position $t$, denoted as $\varpi_p t$, is a vector of consecutive $p$ integers where $t$ is located at the centre position when $p$ is odd, and one position to the right of the centre position when $p$ is even. For example, $(4, 5, 6)$ is a time window of size 3 at position 5, and $(7, 8)$ is a time window of size 2 at position 8. Mathematically, the elements of $\varpi_p t$ range consecutively from $t - \lfloor p/2 \rfloor$ to $t - \lfloor p/2 \rfloor + p - 1$, where $\lfloor \rfloor$ denotes the floor function, which takes the integer value of a floating variable.

Given a time series $x$ of length $T$ and a time window $\varpi_p t$ of size $p$, a $p$-baseline moving average of $x$, $b^p_x(t)$, is a time series whose elements are calculated as

$$b^p_x(t) = \frac{1}{p} \sum_{t' \in \varpi_p t} x(t'), \quad \text{for} \quad t = \lfloor p/2 \rfloor + 1, \lfloor p/2 \rfloor + 2, \ldots, \lfloor p/2 \rfloor + T - p + 1,$$

where at the extremities, $t = 1$ and $t = T$, $b^p_x(t)$ takes the average value of the first two samples and last two samples, respectively, of the time series $x$, and for the other positions near the extremities but undefined by (1), i.e., from 2 to $\lfloor p/2 \rfloor$ and from $\lfloor p/2 \rfloor + T - p + 2$ to $T - 1$, $b^p_x(t)$ is computed according to (1) based on the largest possible reduced-size window.

A $p$-baseline has some useful properties: (1) it reflects the trend, or trace of averaged values, for each position of the original waveform without any time lag, and (2) it is smoother than the original waveform and the smoothness is controlled by the window size, i.e., the larger the window size, the smoother the resulting baseline. Figures 3(a) and (b) show a 7-baseline
Figure 3. Illustration of baselines. (a) A 7-baseline $b_7^x$ follows the original respiratory waveform $x$ very closely. (b) A 21-baseline $b_{21}^x$ is much flatter than a $b_7^x$, causing lower amplitude waveform segments to fall below the baseline. The asterisk (*) at time $t_i$ illustrates the dependence of breath identification on the window size of the baseline.

and a 21-baseline, respectively, for time series $x$. Clearly, a smaller window generates a baseline that more closely follows the original series, and a larger one generates a baseline that moves further apart from, but is smoother than, the original series. A $p$-baseline median filter is similar to the $p$-baseline moving average in (1) but with the median value calculated instead of the average value.

Given a respiratory waveform $x$ and its $p$-baseline $b_p^x$, a respiratory breath is defined by three consecutive intersections of $x$ and $b_p^x$, where for any position $t$ between the first two intersecting points the inequality

$$x(t) - b_p^x(t) > 0$$

(2)

is satisfied, defining an upper respiratory cycle relative to the $p$-baseline. For positions $t$ between the last two intersecting points, (2) is not satisfied, defining a lower respiratory cycle relative to $b_p^x$ (figure 4). The breath width, $w$, is defined by the distance between the first and third intersecting points, or equivalently by the total length of the upper and lower cycles. An upper cycle or a lower cycle alone defines a half breath; the height of a half breath, $h_U$ and $h_L$, is defined as the maximum value of $|x(t) - b_p^x(t)|$ for any $t$ in the corresponding cycle and the area of a half breath, $a_U$ and $a_L$, is defined as the sum of $|x(t) - b_p^x(t)|$ over all $t$ in the corresponding cycle (figure 4). Standalone half breaths exist when a lower cycle is present at the front end of the waveform, or an upper cycle is present at the back end of the waveform.

Importantly, a respiratory breath is defined relative to a selected baseline and changing a baseline redefines the start and end positions of breath cycles or the presence of a cycle. An example can be seen in figures 3(a) and (b), where the 7-baseline in figure 3(a) identifies an upper cycle around $t_i$ marked by an asterisk (*), while the 21-baseline in figure 3(b) does not. Our algorithm optimizes the selection of the ‘best’ baseline by estimating a typical breath width from the waveform and using it as the best window size. The reason for such choice is presented in the appendix.

2.2.2. Breath identification algorithm. The breath identification algorithm consists of four steps: (1) a preliminary smoothing process to remove noise and sharp spikes; (2) determination
Figure 4. A respiratory breath is defined by three consecutive intersecting points (marked by open circles) of a waveform $x$ and its $p$-baseline $b_p$. A breath consists of two half breaths, an upper cycle $U$ and a lower cycle $L$, which are characterized by their corresponding heights $h_U$ and $h_L$, and areas $a_U$ and $a_L$. A breath is also characterized by its width $w$.

of a representative breath width using a fixed $p$-baseline; (3) identification of potential breaths using an adjusted baseline and (4) analysis of breaths statistics using their width, height and area information and, if significant noise is present, further denoising the waveform and reidentifying potential breaths. The first step is performed once for the entire waveform, and the subsequent steps are performed for each 15 s overlapping waveform segment shifted by 1 s. Figure 5 together with the following four steps provides a detailed description of the algorithm.

Step 1. Preliminary smoothing. The original respiratory waveform $x$ is first denoised with a 7-baseline median filter followed by a 6-baseline average filter to generate a smoothed waveform $x_d$. This removes noise and sharp spikes whose widths are narrower than 7 samples or approximately 0.3 s. Figure 5(a) shows an example of the original waveform $x$ and its corresponding smoothed waveform $x_d$ in figure 6(b). After the preliminary smoothing, each overlapping 15 s waveform segment, corresponding to a reference respiratory rate $RR_R$ at the end of the segment, is provided as the starting point of the initial breath identification in step 2.

Step 2. Determination of a representative breath width. Breath cycles relative to an arbitrarily selected 102-baseline ($p = 102$) are identified from the provided waveform (figure 5(c)). And the computed median width $\tilde{w}$ of the breaths, corresponding in this case to 26 data samples, is used as an initial estimate of a representative breath width for the segment. The algorithm is not sensitive to the initial selection of $p$ as long as it is larger than the mean breath width so that low-frequency noise is not considered as breaths (refer to the appendix).

Step 3. Identification of potential breaths. Potential breaths are identified based on a $p$-baseline, with $p = \tilde{w}$, from step 2 (figure 5(d)). This assumes that a modified baseline, based on the initially determined breath width, is the best choice of $p$ for the identification of actual breaths. Figure 5(d) shows that the selection of $p = \tilde{w} = 26$ identifies three additional breaths (marked by arrows) not identified using the initial 102-baseline.

Step 4. Analysis of respiratory breath statistics and further waveform denoising. The set of widths $W = \{w_1, w_2, \ldots, w_n\}$, areas $A = \{a_1, a_2, \ldots, a_n\}$ and heights $H = \{h_1, h_2, \ldots, h_n\}$
of the $n$ respiratory breaths identified in step 3 over the waveform window are computed and used to define effective breaths. The $n$th respiratory breath is denoted as an effective breath if it meets the following conditions:

$$
(w_n > 7) \text{ and } (h_n > \text{sort}_{4/5}(H)/4) \text{ and } (a_n > \text{sort}_{4/5}(A)/4) \text{ and } (a_n > a_0),
$$

where $\text{sort}_{4/5}(Z)$ denotes the value of the element of set $Z$ at the 80th percentile position after $Z$ has been ordered in an ascending order, and $a_0$ denotes a lower limit threshold for the area of a respiratory breath. Here, $a_0$ is set as the product of the narrowest possible breath width (7 samples, or 0.3 s) and 1/20 of the standard deviation of the denoised waveform $x_d$ computed in step 1.
Breaths not satisfying (3) are defined as pseudo breaths. If pseudo breaths are present among the identified breaths, they are considered to be noise detected as breaths. Then the waveform is further denoised with a moving average filter of window size equal to 1/2 of the initial breath width identified in step 2, and steps 2–4 are repeated to identify effective breaths from the denoised waveform. Otherwise, the identified breaths are considered to be actual breaths and used for calculating respiratory rates. We do not repeat steps 2–4 more than twice because additional denoising could transform the waveform to the point of affecting actual respiratory breaths. Respiratory breaths not satisfying (3) at the end of this iterative process are not used to estimate respiratory rates.

For example, from the three additional breaths in figure 5(d), the first two are pseudo breaths while the third one, at 557, is identified as an effective breath. This is achieved by further denoising the 15 s waveform segment, reestablishing a representative breath width ($p = 28$) with a 102-baseline through step 2 (figure 5(e)), and identifying effective breaths, shown in figure 5(f), through step 3. Effective breaths satisfying (3) are found through this iterative process and used for estimating respiratory rates.

2.2.3. Respiratory rate calculation. Once respiratory breaths have been determined, the algorithm attempts to identify ‘invalid’ waveform data before calculating the respiratory rate.
RR_C. Two conditions are assumed to invalidate a data point. The first relates to data points that are part of a flat line and the second relates to points close to a spike. Three or more consecutive data points with the same value constitute a flat line. Given a waveform x and its denoised waveform x_d, a spike is said to exist at position t_spike if the absolute difference between x(t_spike) and x_d(t_spike) is larger than the standard deviation of the difference between x_d and its 102-baseline b_x, i.e., |x(t_spike) − x_d(t_spike)| > std(x_d − b_x). Spikes and their neighbouring six data points (three from each side) are considered as invalid data.

The validity ratio V(t) of a 15 s waveform segment ending at time t is taken as the ratio of the number of valid data points v in the segment to the total number of data points in the segment. For a sampling rate r, the validity ratio V(t) at time t is calculated as

\[
V(t) = \frac{v}{15/r},
\]

where V(t) ranges from 0.0 to 1.0.

Given V(t), the calculated respiratory rate RR_C(t) (breaths min\(^{-1}\)) at time t is defined as

\[
RR_C(t) = \begin{cases} 
M \times 60/(t - t_1), & V(t) > 0.5 \\
0, & V(t) \leq 0.5,
\end{cases}
\]

where M is the number of effective breaths found in step 4 of the breath identification algorithm and t_1 is the start time of the first effective breath, i.e., the position of the first crossing between the waveform and its baseline for the first effective breath. We calculate RR_C only when the validity ratio V is greater than 0.5.

2.3. Respiratory waveform qualification

We calculate the respiratory waveform quality index QI_W to indicate the quality of the 15 s respiratory waveform segment used for the calculation of the corresponding respiratory rate. QI_W can be placed in a database along with the reference respiratory rates and used to search for waveforms of a given quality, with the assumption that their derived respiratory rates can be trusted to the extent of the quality of the underlying waveform. The computation of QI_W relies on the identification of ambiguous breaths.

Based on a set of J candidate baselines, an ambiguous breath is defined as a breath that is observed with a certain baseline but is no longer present when a different baseline is chosen. For example, the breath marked with an asterisk (*) at t_i in figure 3 is an ambiguous breath.

The existence of an ambiguous breath depends on the choices of candidate baselines used to assess the breath. For QI_W purposes, we use a set of J = 34 baselines whose window sizes p = 10, 20, . . . , 340 are selected to be shorter than the waveform window. For a given waveform x, we identify the number of effective breaths M using each baseline b_x^p. The number of distinct values of M over the set of J baseline windows is denoted as \( \kappa \), which is associated with the number of ambiguous breaths. For example, if 12 breaths are identified by a 10-baseline, and 14 breaths are identified by all other baselines, then the distinct values of M are 12 and 14, thus \( \kappa = 2 \).

We may now define a regularity ratio \( G_t \) of waveform x as

\[
G_t = 1 - \frac{(\kappa - 1)}{J}.
\]

The regularity ratio \( G_t \) decreases from 1 to 1/J as \( \kappa \) increases from 1 to J. A larger \( G_t \) reflects a more regular waveform and the existence of a smaller number of ambiguous breaths.
Given a 15 s waveform segment ending at time $t$ with a regularity ratio $G_r$ defined by (6) and a validity ratio $V(t)$ defined by (4), the waveform quality $Q_W$ at time $t$ is defined as

$$Q_W(t) = G_r^2 \times V(t).$$

(7)

$Q_W$ ranges from 0.0 to 1.0, with 1.0 representing the best possible quality, and provides the means to empirically estimate the quality index $Q_{IW}$ at each time $t$ as

$$Q_{IW}(t) = \begin{cases} 
3, & 0.8 < Q_W(t) \leq 1.0; \\
2, & 0.7 < Q_W(t) \leq 0.8; \\
1, & 0.6 < Q_W(t) \leq 0.7; \\
0, & Q_W(t) \leq 0.6. 
\end{cases}$$

(8)

The purpose of $Q_{IW}$ assignment is to facilitate automatic assignment of waveform quality using simple quality levels from 0 to 3, corresponding to waveform quality ranging from low to high. Based on (7) and (8), the best-quality index level $Q_{IW} = 3$ requires a squared regularity ratio $G_r^2 > 0.8$, or equivalently, an ambiguous breath count $\kappa \leq 4$, and a validity ratio $V(t) > 0.8$.

### 2.4. Respiratory rate qualification

A second quality index, $Q_{IR}$, for the reference respiratory rate $RRR$, is calculated based on the quality $Q_{IW}$ of the corresponding waveform segment and the similarity between $RRC$ and $RRR$ according to the following rules:

$$Q_{IR}(t) = \begin{cases} 
3, & \{Q_{IW}(t) = 3\} \text{ AND } \{|RRC(t) - RRR(t)| < 10\}; \\
2, & \{Q_{IW}(t) = 2\} \text{ AND } \{|RRC(t) - RRR(t)| < 5\}; \\
1, & \{Q_{IW}(t) = 3\} \text{ OR } \{|RRC(t) - RRR(t)| < 10\}; \\
0, & \text{otherwise.} 
\end{cases}$$

(9)

The best quality $Q_{IR} = 3$ means that $RRR$ is very reliable since it matches the calculated $RRC$ and the waveform from which they are derived is deemed to have good quality, i.e., a $Q_{IW}$ of 2 or 3. The lower quality rates $Q_{IR} < 3$ exhibit more relaxed requirements. The thresholds used in (9) to define $Q_{IR}$ are empirically determined based on manual examination of the waveform quality and corresponding $Q_{IW}$ levels, and expert suggestions of permissible differences between $RRC$ and $RRR$ at each $Q_{IW}$ level.

### 3. Results

We evaluate the performance of our quality indexing system through: (1) demonstration of the responses of the respiratory rate quality index with real respiratory data; (2) comparison of the respiratory rate qualification against human expert evaluation; (3) evaluation of the PAS data set using the algorithm-generated quality indices and (4) examples of using qualified respiratory rates for data mining.

### 3.1. Demonstration of respiratory rate quality index

We start by demonstrating the performance of our algorithm through the analysis of an example. Figure 6 shows a case where a respiratory waveform is corrupted by spikes and noise. In intervals (a) and (d), the respiratory waveform quality is good, $Q_{IW} \geq 2$ (panels I and II), the reference respiratory rates closely match the calculated respiratory rates within
Table 1. Comparison of human expert versus computer algorithm assignment of respiratory waveform quality indices.

<table>
<thead>
<tr>
<th>Human-rated QIHW</th>
<th>Algorithm-rated QIW</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0 1 2 3</td>
</tr>
<tr>
<td>1</td>
<td>2 8 16 1</td>
</tr>
<tr>
<td>2</td>
<td>0 10 24 21</td>
</tr>
<tr>
<td>3</td>
<td>0 1 9 51</td>
</tr>
</tbody>
</table>

5 breaths min\(^{-1}\) (panel III) and the best quality index QIR = 3 is assigned (panel IV). In interval (b), the waveform is corrupted by spikes, yielding a QIW ≤ 1, but the reference respiratory rates match the calculated respiratory rates within 10 breaths min\(^{-1}\), so a quality index QIR = 1 is assigned. In interval (c), the waveform is affected by noise, yielding a QIW < 3 for the interval except for one time point in the middle where QIW = 3. The difference between the reference and the calculated respiratory rates is ≥10 breaths min\(^{-1}\), so a quality index QIR = 0 is assigned for the points in interval (c), except for that one time point where QIR = 1. The surprisingly high respiratory rate RR\(_R\) in this interval may potentially be due to two reasons: the increased intensity of spikes preceding interval (c), or data processing and manipulation errors. Regardless of the reason, RR\(_R\) provided at this time interval is unreliable and should not be used.

3.2. Comparison of respiratory rate qualification algorithm against human evaluation

We compared our methodology against 200 samples reviewed by two human experts. Each sample consisted of a randomly selected reference respiratory rate RR\(_R\) and its associated 15 s respiratory waveform segment. The human experts were asked to visually inspect and evaluate the quality of the respiratory waveform QIHW, calculate a respiratory rate RR\(_H\) from the waveform and use the decision rules in (9) to form a quality index QIHR for each data sample. The experts determined QIHW through visual inspection according to the following rules:

A respiratory waveform is ranked: (i) excellent, or QIHR = 3, if all breaths are clearly identifiable; (ii) good, or QIHR = 2, if some breaths are ambiguous, but the ambiguous breaths occupy less than one-third of the total waveform in length; (iii) poor, or QIHR = 1, if one or more breaths can be identified, but the identifiable breaths occupy less than one-third of the total waveform in length and (iv) bad, or QIHR = 0, if the above rules are not met or the waveform is corrupted by spikes or flat lines. (Due to ambiguities, rules (ii) and (iii) are arbitrated by the experts and are assigned the most likely level.)

Out of the 200 samples, 163 (or 82%) were assigned the same QIHW by the two experts. For these consensus samples, the two expert-computed rates were averaged and used as the consensus RR\(_H\) for each sample. We use these consensus QIHW samples and the corresponding RR\(_H\) values as the ‘gold standard’ to evaluate our algorithm-calculated RR\(_C\), QIW and QIR, as follows.

3.2.1. Evaluation of respiratory waveform quality index QIW. Table 1 compares the algorithm-calculated QIW against the human-rated QIHW for the 163 samples. QIW matches
Algorithms to qualify respiratory data collected during the transport of trauma patients

Table 2. Comparison of human-calculated (RRH) versus reference (RRR) and algorithm-calculated (RRC) respiratory rates as a function of waveform quality (QIW).

<table>
<thead>
<tr>
<th>QIW</th>
<th>Samples where a RRH is determined</th>
<th>D1:</th>
<th>D2:</th>
<th>Paired Student’s t-test between D1 and D2 (p-value)</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>73</td>
<td>1.7 ± 2.9</td>
<td>8.3 ± 12.4</td>
<td>1.7e−5</td>
</tr>
<tr>
<td>2</td>
<td>49</td>
<td>4.3 ± 5.5</td>
<td>19.8 ± 23.2</td>
<td>5.8e−5</td>
</tr>
<tr>
<td>1</td>
<td>23</td>
<td>5.5 ± 5.6</td>
<td>21.9 ± 24.2</td>
<td>5.8e−3</td>
</tr>
<tr>
<td>0</td>
<td>18</td>
<td>8.1 ± 0.0</td>
<td>26.5 ± 3.5</td>
<td>8.5e−2</td>
</tr>
<tr>
<td>All</td>
<td>163</td>
<td>3.2 ± 4.6</td>
<td>14.3 ± 19.3</td>
<td>1.4e−10</td>
</tr>
</tbody>
</table>

Table 3. Comparison of human expert versus algorithm assignment of respiratory rate quality indices QIHR and QIR.

<table>
<thead>
<tr>
<th>Human-rated QIHR</th>
<th>Algorithm-calculated QIR</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>44 22 3 3</td>
</tr>
<tr>
<td>1</td>
<td>3 8 0 4</td>
</tr>
<tr>
<td>2</td>
<td>0 3 2 6</td>
</tr>
<tr>
<td>3</td>
<td>0 4 2 59</td>
</tr>
</tbody>
</table>

QIHW in 99 (61%) cases (diagonal entries in the table) and differs in 64 (39%) cases. In 161 (99%) cases, QIW differs from QIHW within one level (entries once removed from the diagonal). QIW and QIHW are correlated with each other (p < 0.05, Pearson correlation test value 0.8). These results indicate that QIW generally reflects the human evaluation of the waveform quality.

3.2.2. Evaluation of calculated respiratory rate RRc. We also tested how the waveform quality QIW relates to the accuracy of the algorithm-calculated respiratory rates compared to expert-calculated rates. We illustrate in table 2 the mean and the standard deviation (STD) of the absolute differences between RRc and RRH as a function of QIW for the 143 samples where the experts were able to determine an RRH. These results show that the higher the QIW, the closer the match between the human and the algorithm-calculated respiratory rates. For the best-quality waveforms, QIW = 3, there is a 1.7 ± 2.9 (mean ± STD) breaths min⁻¹ difference, and for the worst-quality waveforms, QIW = 0, the difference is more than 8 breaths min⁻¹. The difference between the reference and human rate estimations |RRR - RRH| follows a similar trend but with a much larger mean and STD. The average |RRc - RRq| is significantly (p < 0.05, Student’s t test) smaller than the average |RRR - RRH| (3.2 ± 4.6 breaths min⁻¹ versus 14.3 ± 19.3 breaths min⁻¹).

3.2.3. Evaluation of respiratory rate quality index. Table 3 compares the algorithm-calculated QIR against the human-rated QIHR for the 163 previously described samples. QIR matches QIHR in 113 (69%) cases and differs in 50 cases. In 149 (91%) cases, QIR differs from QIHR within one level. QIR and QIHR are correlated with each other (Pearson test 0.83, p < 0.05). Note that there are three samples where the human-rated QIHR = 0, while the algorithm-inferred QIR = 3. These samples contain ambiguous breaths that result in the
human-calculated respiratory rates \( R_{Rh} \) differing from the reference rates \( R_{Rr} \) by more than 10 breaths min\(^{-1}\), but the algorithm-calculated rates \( R_{Rc} \) match \( R_{Rr} \) within 5 breaths min\(^{-1}\). The algorithm trusts \( R_{Rc} \) and \( R_{Rr} \) in these cases and assigns a high \( QI_{R} \) for the reference respiratory rates following the first decision rule in (9).

3.3. Evaluation of the PAS dataset

We also evaluated the quality indices \( QI_W \) and \( QI_R \) for the 711 patient records in PAS, consisting of a total of about 280 h of time-series data per physiologic variable. The proportions of \( QI_W \) levels of 0, 1, 2 and 3 for the respiratory waveform data are 45%, 10%, 22% and 23%, respectively. As would be expected from these proportions, the distribution of \( QI_R \) levels of 0, 1, 2 and 3 is similarly shifted towards the lower end, to yield 59%, 14%, 4% and 23%, respectively. This result suggests caution when using field-collected respiratory data for post-hoc data analysis.

3.4. Examples of using qualified respiratory rates for data mining

We have used the qualified respiratory rates in PAS to calculateprehospital severity scores, including the revised trauma score (RTS) (Champion et al 1989) and the prehospital index (PHI) (Koehler et al 1986), to predict mortality and the requirement for blood. Given a medic’s evaluation of the cognitive status of a patient, termed the Glasgow Coma Scale (Udekwu et al 2004), and a number of physiology variables, including respiratory rate, heart rate and blood pressure, RTS and PHI scores are calculated from a linear combination of these input variables. By considering the quality of the respiratory rates based on \( QI_R \), we facultatively selected respiratory rates from different time intervals and consequently calculated different RTS and PHI scores for the same patient. The differently calculated scores were evaluated for sensitivity and specificity of outcome prediction by constructing receiver operation curves (ROC). A greater area under the ROC curve (AUC) suggests a better prediction performance. Statistical significance of the differences between paired AUCs was tested using the ROCKIT software (ROCKIT 2006, Dorfman and Alf 1969).

For mortality prediction using RTS scores from a population of 187 patients (of which 15 died), the AUCs are 0.88 and 0.83 for the best- and the worst-quality data, respectively, and the difference is statistically significant at a \( p < 0.25 \) level. For blood requirement (40 of the 187 patients received blood) the AUCs are 0.73 and 0.58, respectively, for the best- and the worst-quality data, with \( p < 0.001 \). Predictions from the PHI scores show a similar profile. These results suggest that higher quality respiratory data are more predictive than lower quality data, and highlight the usefulness of the proposed method.

Importantly, we find that the respiratory rate quality index itself is significantly associated with patient death outcome. That is, patients with high-quality respiratory data have a greater chance of survival than patients expressing low-quality respiratory data (figure 7). This finding indicates that the quality of the data itself may have clinical value.

4. Discussions

Despite their low frequency and noisy characteristics, respiratory waveforms can be used to independently calculate respiratory rates to validate the rates provided by life-sign monitors. We calculate the respiratory rate based on the principle of counting typical breaths, which has been adopted by many applications. Weese-Mayer et al (2000) defined typical breaths for end-tidal \( CO_2 \) and nasal/oral thermister measurements as a deflection \( \geq 75\% \) (for end-tidal
CO₂ measurement) or ≥25% (for thermister measurement) of the mean excursion observed during a quiet period preceding the evaluation. The application of such a protocol depends on an observed quiet period prior to data collection and, thus, is not suitable for unstructured noisy environments and continuous breath calculations. Chen et al. (2005) detected respiratory rhythm during sleep by applying a wavelet transformation and a zero-crossing point detection algorithm to the respiratory waveform. Their assumption is based on a relatively stable condition during sleep, which may not be suitable for trauma patients in a dynamic field environment.

Compared to the above algorithms, our method for breath detection is more systematic and adaptable to changes in waveform amplitude. We discovered that the major difficulty for breath identification comes from the existence of ambiguous breaths that, based on differently chosen baselines, may or may not be counted as actual breaths. Our iterative procedure is able to estimate typical breaths from a 15 s waveform segment at run time based on an adjustable baseline. Thus, this technique is insensible to waveform signal fluctuations and is adaptable to patients under dynamic transportation environment. Furthermore, our algorithm is computationally cheaper than algorithms using wavelet transformation without loss of accuracy. Indeed, the calculated respiratory rates provide a statistically significant closer match to rates determined by human experts than those provided by a life-sign monitor. There is only a 3.2 ± 4.6 breaths min⁻¹ average difference between the algorithm-calculated and the human-evaluated respiratory rates. This suggests that the algorithm-calculated rates are trustworthy. The significantly greater difference of 14.3 ± 19.3 breaths min⁻¹ between the monitor-provided and human-evaluated rates for the same sample set may be due to different methodologies and assumptions in the monitor’s algorithm and the criteria used for human evaluation of the waveforms. Nevertheless, the difference between the two algorithms provides yet another comparative metric to highlight the improved accuracy of the proposed method.

There is a limited body of work describing algorithms that estimate the quality of respiratory data. Alametsa et al (2005) developed algorithms to detect respiratory waveform spikes for the evaluation of the severity of respiratory disturbance during sleep. Their algorithm assessed large relative increases in the amplitude of the waveform, which does not depend on the absolute waveform amplitude. Our respiratory qualification algorithm includes a spike
detection algorithm, which may be used for similar purposes. In addition, we also measure the presence of noise, irregularity and flat signals in the respiratory waveform, and, therefore, can provide a more systematic measure of the evaluation of signal quality and patients’ respiratory efforts. The subsequently calculated respiratory waveform quality index \( Q_{IW} \), ranging from 0 to 3, provides a quantitative measure of the quality of respiratory waveforms.

The decision to use respiratory waveform quality index \( Q_{IW} \) as a feature in constructing the corresponding respiratory rate quality index \( Q_{IR} \) is supported by observations from clinical data. We found that the higher the \( Q_{IW} \), the closer the match between the human- and the algorithm-calculated respiratory rates. When the waveform quality index attains its highest value, \( Q_{IW} = 3 \), there is only a \( 1.7 \pm 2.9 \) breaths min\(^{-1} \) average difference between the algorithm-calculated and the human-evaluated respiratory rates. This result provides a level of confidence for selecting respiratory data based on the waveform quality.

The respiratory rate quality index \( Q_{IR} \), presented simply as a numerical score ranging from 0 to 3, allows investigators to select high-quality respiratory rates for data mining. The best quality index \( Q_{IR} = 3 \) is characterized by good-quality respiratory waveforms and closely matches between the algorithm-calculated and the monitor-provided reference respiratory rates. We found that only 23% of the respiratory time-series data archived in PAS have the best quality \( Q_{IR} = 3 \), while 59% have the worst quality \( Q_{IR} = 0 \), confirming (and quantifying) that field-collected respiratory data are especially noisy. Our respiratory data qualification method provides a reliable, systematic and automatic method to curate respiratory data collected in unstructured field environments.

Qualified data provide a more reliable platform for data mining activities, such as the development of trauma scoring methods for the prognosis of trauma patients, than do unqualified data. It is likely that real-time qualification and analysis of continuous respiratory information will improve sensitivity and specificity parameters for the detection of pathophysiological states that threaten trauma patients. Confirmation of such an improvement in the evaluation of revised trauma scores presented here further validates the usefulness of our respiratory data qualification approach. The significant association found between respiratory data quality and physiological outcome suggests that caution should be exercised in the selection of respiratory data for the development of algorithms that predict some clinical outcome. Our data qualification methodology represents a first step to manage noisy respiratory data, which is an essential element for the future discovery of robust and clinically useful prognostic algorithms.

5. Conclusions

We developed an automated, quantitative respiratory data quality indexing system for time-series respiratory data. Elements of this system are a method to count breaths and, therefore, respiratory rate; an algorithm to qualify ambiguous breaths and, therefore, qualify the respiratory waveform and a set of decision-logic rules that combine these elements to systematically qualify the respiratory rates. A respiratory rate quality index ranging from 0 to 3 allows rapid searches and extraction of qualified respiratory rates for data mining by providing a succinct, point-by-point indication of their reliability.

Acknowledgments

We are grateful to the University of Texas Health Sciences Center, COL John Holcomb, and Dr Jose Salinas of the US Army Institute of Surgical Research, Fort Sam Houston,
Algorithms to qualify respiratory data collected during the transport of trauma patients

San Antonio, Texas, for providing access to the vital signs (Trauma) database. We thank Dr Jingyu Liu for her participation in the expert evaluation of our data and Gagandeep Bawa and the Physiology Analysis System team for providing a data analysis platform for the trauma patient data. The work presented here was supported by the Combat Casualty Care and the Military Operational Medicine Directorates of the US Army Medical Research and Materiel Command, Fort Detrick, Maryland.

Appendix. Optimal selection of baseline window size for breath identification

Another feature of interest of a respiratory waveform is the peak of a breath. Let \( x \) be a respiratory time series and \( k \) an integer. Then, \( x(t_{\text{peak}}) \) at position \( t_{\text{peak}} \) is defined as a peak of a breath within a time window \( \sigma_{t_{\text{peak}}}^k \) of length \( k \) if, for any \( t \in \sigma_{t_{\text{peak}}}^k \) and \( t \neq t_{\text{peak}} \), \( x(t) < x(t_{\text{peak}}) \). The longest window length \( k^* \) for which the definition of a peak holds is defined as the width of the peak. (To simplify the notation, henceforth \( k \) is used to denote the longest window length \( k^* \).) Moreover, as illustrated in figure A1, the height \( h_{t_{\text{peak}}}^k \) of a peak with width \( k \) is defined as

\[
h_{t_{\text{peak}}}^k = \max_{t \in \sigma_{t_{\text{peak}}}^k} [x(t_{\text{peak}}) - x(t)]. \tag{A.1}
\]

We introduce two theorems that together provide insight regarding the appropriate selection of the relative lengths of the width \( k \) of a peak and the window size \( p \) of a baseline moving average that optimizes the identification of actual respiratory breaths.

**Theorem 1.** Let \( x \) be a respiratory waveform, \( b_p^x \) its \( p \)-baseline, and \( x(t_{\text{peak}}) \) a peak of a respiratory breath at position \( t_{\text{peak}} \) with width \( k \). Then, if \( p \leq k \), \( t_{\text{peak}} \) must be located in the upper cycle of a breath.

**Proof.** Let us assume that \( p \leq k \). From (1), we have

\[
b_p^x(t_{\text{peak}}) = \frac{1}{p} \sum_{t \in \sigma_{t_{\text{peak}}}^p} x(t). \tag{A.2}
\]
Because $p \leq k$, each $t \in \mathcal{P}_{t_{peak}}$ also belongs to $\mathcal{P}_{t_{peak}}$, and by the definition of a peak, $x(t) < x(t_{peak})$ for each $t \neq t_{peak}$, we have

$$\frac{1}{p} \sum_{t \in \mathcal{P}_{t_{peak}}} x(t) < x(t_{peak}). \quad (A.3)$$

Combining (A.2) and (A.3), we get

$$b_{p}^{x}(t_{peak}) < x(t_{peak}), \quad (A.4)$$

and

$$x(t_{peak}) - b_{p}^{x}(t_{peak}) > 0. \quad (A.5)$$

Hence, according to (A.5) and (2), $t_{peak}$ is located between the first two intersecting points of $x$ and its $p$-baseline $b_{p}^{x}$, therefore, it must be located in the upper cycle of a breath. □

**Theorem 2.** Let $x$ be a respiratory waveform, $b_{p}^{x}$ its $p$-baseline, and $x(t_{peak})$ a peak of a respiratory breath at position $t_{peak}$ with width $k$ and height $h_{t_{peak}}^{k}$. Then, if $p > k$ and $x(t) < x(t_{peak}) + (1/p) \cdot h_{t_{peak}}^{k}$ for all $t \in \mathcal{P}_{t_{peak}}, t_{peak}$ must be located in the upper cycle of a breath.

**Proof.** Let us assume that $p > k$ and

$$x(t) < x(t_{peak}) + (1/p) \cdot h_{t_{peak}}^{k}, \quad \text{for all } t \in \mathcal{P}_{t_{peak}}. \quad (A.6)$$

Let $t_{h} \in \mathcal{P}_{t_{peak}}$ be the location of the lowest amplitude of $x$ that defines the height $h_{t_{peak}}^{k}$ of $x(t_{peak})$, so that

$$x(t_{h}) = x(t_{peak}) - h_{t_{peak}}^{k}. \quad (A.7)$$

Because $p > k$, we also have $t_{h} \in \mathcal{P}_{t_{peak}}$. Hence, starting from the equality

$$\sum_{t \in \mathcal{P}_{t_{peak}}} x(t) = x(t_{h}) + \sum_{t \in \mathcal{P}_{t_{peak}}, t \neq t_{h}} x(t), \quad (A.8)$$

and applying (A.6) and (A.7), we obtain

$$\sum_{t \in \mathcal{P}_{t_{peak}}} x(t) < \left[ x(t_{peak}) - h_{t_{peak}}^{k} \right] + (p - 1) \cdot \left[ x(t_{peak}) + (1/p) \cdot h_{t_{peak}}^{k} \right],$$

$$\sum_{t \in \mathcal{P}_{t_{peak}}} x(t) < px(t_{peak}) - (1/p) \cdot h_{t_{peak}}^{k},$$

and as $h_{t_{peak}}^{k} > 0$, then

$$\sum_{t \in \mathcal{P}_{t_{peak}}} x(t) < px(t_{peak}). \quad (A.9)$$

Consequently, from (1) and (A.9), we have

$$b_{p}^{x}(t_{peak}) = \frac{1}{p} \sum_{t \in \mathcal{P}_{t_{peak}}} x(t) < \frac{1}{p} \left[ px(t_{peak}) \right], \quad (A.10)$$

$$b_{p}^{x}(t_{peak}) < x(t_{peak}).$$

Therefore,

$$x(t_{peak}) - b_{p}^{x}(t_{peak}) > 0, \quad (A.11)$$
and according to (A,11) and (2), $t_{\text{peak}}$ is located between the first two intersecting points of $x$ and its $p$-baseline; $b^*_p$, therefore, it must be located in the upper cycle of a breath.

A peak in a respiratory waveform $x$ is identified as a breath, relative to a $p$-baseline $b^*_p$, if the peak is located in the upper respiratory cycle, i.e., between the first two intercepts of $x$ and $b^*_p$. To that extent, theorem 1 states that if the selected $p$-baseline is smaller than or equal to the peak’s width $k$, i.e., $p \leq k$, then the peak will be identified as a breath. While such selection guarantees that a peak will be counted as a breath, it also has the potential detrimental effect of detecting low-frequency noise as respiratory breaths. This is where theorem 2 comes into play. It states that if $p > k$ selection will also identify a peak as a breath as long as the amplitude of $x(t)$, for all $t$ within a $p$-window of the peak, is smaller than the peak’s amplitude plus $1/p$ of the peak’s height. Such a selection guarantees that actual breaths will be identified as long as the variance of the amplitude of successive peaks within a $p$-window is kept within a threshold (of $1/p$ at the minimal peak height). As $p$ increases, the likelihood that this condition is met decreases. Therefore, it is important to maintain $p$ at about $k$ to ‘optimize’ breath identification.

Whether a peak in a waveform represents an actual breath or noise, however, is unknown a priori. Therefore, for this reason and computational simplicity, we approximate the selection of $k$ by the median of the breath widths $\tilde{w}$ over the waveform window based on an arbitrarily selected baseline and use $p = \tilde{w}$ as the starting point of the iterative breath identification algorithm described in the main text.

Disclaimer

The opinions or 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 the US Department of Defense. "This paper has been approved for public release; distribution is unlimited."

References

Isakov IuV, Krylova I V and Ioffe IuS 1979 Potentials of hyperbaric oxygenation in correcting central respiratory disorders in ischemic stroke Zh vopr Neurokhir Im N N Burdenko 3 22–8
Lorenz R and Ito A 1978 The definition of ‘Cheyne–Stokes rhythms’ Acta Neurochir. (Wien) 43 61–76
Lucangelo U, Bernabe F and Blanch L 2005 Respiratory mechanics derived from signals in the ventilation circuit Respir. Care 50 55–67
ROCKIT 2006 A program that compares differences between ROC curves, online at: http://www-radiology.uchicago.edu/kr/krL_ROC/software_index.htm
Trauma Vitals 2004 US Army Institute of Surgical Research http://traumavitals.tamu.edu/
Yang S C and Yang S P 2002 Effects of inspiratory flow waveforms on lung mechanisms, gas exchange and respiratory metabolism in COPD patients during mechanical ventilation Chest 122 2096–104
Yasuma F and Hayano J 2004 Respiratory sinus arrhythmia: why does the heartbeat synchronize with respiratory rhythm? Chest 125 683–90