

A Robust Fusion Model for Estimating Respiratory Rate from Photoplethysmography and Electrocardiography

Drew A. Birrenkott*, Marco A.F. Pimentel, Peter J. Watkinson, and David A. Clifton

Abstract—Objective: Respiratory rate (RR) estimation algorithms based on the photoplethysmogram (PPG) and electrocardiogram (ECG) lack clinical robustness. This is because the PPG and ECG respiratory modulations are dependent on patient physiology, regardless of general signal quality. The present work describes an RR estimation algorithm using respiratory quality indices (RQIs) which assess the presence or absence of the PPG- and ECG-derived respiratory modulations. **Methods:** Six respiratory waveforms are derived from the amplitude modulation, frequency modulation, and baseline wander of the PPG and ECG. The respiratory quality of each modulation is assessed using RQIs based on the FFT, autoregression, and autocorrelation. The individual RQIs are fused to obtain a single RQI per modulation per time window. Based on a tunable threshold, the RQIs are used to discard poor modulations and weight the remaining modulations to provide a single RR estimation per time window. **Results:** The proposed method was tested on two independent data sets and found that using a conservative threshold, the mean absolute error (MAE) was 0.71 ± 0.89 and 3.12 ± 4.39 brpm while discarding only 1.3% and 23.2% of all time windows, for each data set, respectively. **Conclusion:** These errors are either better than or comparable to current methods, and the number of windows discarded is far lower demonstrating improved robustness. **Significance:** This work describes a novel pre-processing algorithm that can be implemented in conjunction with other RR estimation techniques to improve robustness by specifically considering the quality of the respiratory information.

Index Terms—Respiratory rate (RR), photoplethysmography (PPG), electrocardiography (ECG), respiratory quality index (RQI), signal quality, signal fusion.

The work of D.A. Birrenkott was supported by the Rhodes Trust. The work of M.A.F. Pimentel was supported by a Health Innovation Challenge Fund from the Wellcome Trust and the Department of Health. The work of P.J. Watkinson was supported by the NIHR Biomedical Research Centre, Oxford. The work of D.A. Clifton was supported by the Centre of Excellence in Personalised Healthcare funded by the Wellcome Trust and EPSRC under Grant WT 088877/Z/09/Z, the Royal Academy of Engineering, and Balliol College, Oxford. Asterisk indicates corresponding author.

*D.A. Birrenkott, M.A.F. Pimentel, and D.A. Clifton are with the Institute of Biomedical Engineering, Department of Engineering Science, University of Oxford, Oxford OX3 7DQ, UK (e-mail: drew.birrenkott; marco.pimentel; david.clifton@eng.ox.ac.uk).

P.J. Watkinson is with the Oxford University Hospitals NHS Trust, Oxford OX3 9DU, UK (e-mail: peter.watkinson@ndcn.ox.ac.uk).

Manuscript received May 8, 2017; revised November 19, 2017; accepted November 24, 2017. Copyright (c) 2017 IEEE. Personal use of this material is permitted. However, permission to use this material for any other purposes must be obtained from the IEEE by sending an email to pubpermissions@ieee.org.

I. INTRODUCTION

ABNORMAL respiratory rate (RR) is one of the earliest and most prevalent indicators of in-hospital catastrophic deterioration. Abnormal RR preempts adverse events such as cardiac and respiratory arrest, systemic inflammatory response syndrome (SIRS), renal failure, transfer to the intensive care unit (ICU), and in-hospital mortality [1]–[5]. The normal range for RR is 8–20 breaths per minute (brpm) for adults [6]. In a study of abnormal RR, it was found that 54% of all cardiac arrest patients in an internal medicine unit had $RR > 27$ brpm during at least one recording in the three days prior to cardiac arrest [2]. Elevated RR was found to be a more reliable indicator of potential arrest than elevated heart rate or systolic blood pressure [2].

Despite the importance of measuring RR in identifying patient deterioration, RR has historically been the least monitored and recorded vital sign [5], [7], [8]. One of the primary causes for this is that, unlike other vital signs, the measurement of RR is performed manually in most settings because there is no widely available automated device to measure RR with suitable clinical robustness [4]–[7].

The importance of RR and the lack of adequate RR monitoring suggests that the implementation of a clinically robust, automated, and continuous respiratory monitoring method would be of great clinical significance. While automated RR estimation methods, such as capnography, impedance plethysmography, and flow thermography, are used in ICUs and high dependency units (HCUs), these methods are too invasive, cumbersome, and costly for widespread use in general wards, rehabilitation centers, and for in-home use [9].

However, continued advances in photoplethysmography (PPG) and electrocardiography (ECG) have led to their widespread use both in and out of the hospital [10]–[12]. In particular, there is substantial interest in incorporating reflectance PPG technology into smart sensors and wearable devices, such as consumer-grade smart watches [13]. Currently, PPG and ECG are primarily used to estimate RR; however, respiratory activity modulates both of these waveforms [14].

The respiratory waveform can be extracted from the PPG and ECG via multiple modulations, three of the most common are amplitude modulation (AM), frequency modulation (FM), and baseline wander (BW) (Table I, Fig. S.1) [15]. However, the presence of each of these modulations is patient-specific; may appear and disappear for an individual through time; and has been shown to vary based on factors such as pre-existing

TABLE I
SPECIFIC PHYSIOLOGIC RATIONALE FOR EXTRACTION OF EACH RR MODULATION IN BOTH PPG AND ECG [18].

Signal Feature	PPG Mechanism	ECG Mechanism
Amplitude Modulation (AM)	Decrease in Cardiac Output	Thoracic Impedance & Cardiac Rotation
Frequency Modulation (FM)	Respiratory Sinus Arrhythmia (RSA)	Respiratory Sinus Arrhythmia (RSA)
Baseline Wander (BW)	Intrathoracic Pressure Variation	Thoracic Impedance & Cardiac Rotation

health conditions, cardiopulmonary system function, gender, age, and body position [16], [17].

Multiple methods for deriving and estimating RR from PPG and ECG have been described based on digital filtering, wavelets, autoregression (AR), principal component analysis (PCA), and Gaussian processes. Furthermore, a number of methods for combining the estimates from multiple modulations have been developed. For a review of these techniques see [9]. The current state-of-the-art in estimating RR from cardiac waveforms was developed by Karlen *et al.* (2013) [19]. In this method, multiple RR waveforms are extracted from PPG signals via the AM, FM, and BW modulations and fused using the author's "SmartFusion" algorithm. A shortfall of this methodology is that even though PPG or ECG is non-artifactual, quality respiratory modulations may not be present.

In such algorithms, if the PPG or ECG is deemed to be non-artifactual, it is passed to an RR-estimation module whether or not the respiratory modulations carry meaningful physiological information. One recent study used Hjorth descriptors to calculate a "spectral purity metric" on the respiratory waveform after it was extracted from the original signal. This metric was incorporated as a weighting factor into a Kalman filter-based respiratory fusion method [17]. While similar to the widespread implementation of signal quality indices (SQIs), this approach is different because it specifically calculates respiratory modulation robustness. To our knowledge, this is the first example of an explicit quality index for extracted respiratory modulations. Our current work expands this concept by developing novel respiratory quality indices (RQIs) and fusing them into a single robust RQI that can be used to facilitate improved fusion of respiratory estimates from multiple sensors and modulations.

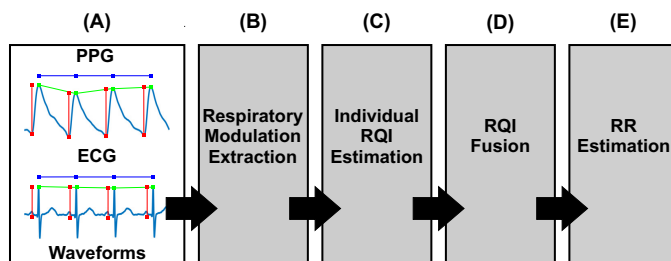


Fig. 1. Outline of the proposed RR estimation method. (A) PPG and ECG waveforms and the respiratory modulations (AM, FM, BW) shown as features of these waveforms. (B) extraction of respiratory waveforms via aforementioned respiratory modulations. (C) estimation of a set of individual (RQIs) for each modulation. (D) fusion of individual RQIs obtained for each modulation into a single RQI. (E) estimation of a single RR for a given time window through RQI-based estimation method.

II. METHODS

The processing sequence for using RQIs in RR estimation from a fusion of PPG and ECG respiratory modulations is shown in Fig. 1. The algorithm consists of four steps: respiratory modulation extraction (Fig. 1B), individual RQI calculation (Fig. 1C), individual RQI fusion into a single fused RQI per modulation (Fig. 1D), and combination of modulation RR estimates using RQIs into a single RR estimate per time window (Fig. 1E). The following sections will describe the data sets used to train and validate the algorithm, the data pre-processing procedure, and each step of the proposed algorithm.

A. Data Sets

We consider two independent, publicly-available data sets: CapnoBase [20] and MIMIC II [21], [22]. Both data sets contain simultaneous PPG and ECG waveforms and an RR reference which is either capnography (CapnoBase) or IP (MIMIC II). The CapnoBase data set represents a younger subject population than the MIMIC II data set, but the vital sign measures for both data sets are comparable (Table S.1).

The CapnoBase data set, \mathcal{D}_c , consisted of the same 42 subjects (29 pediatric and 13 adult) described in [19] (Table S.1). One high-quality, eight-minute segment of data, taken during elective surgery or routine anesthesia, was available for each subject. Each eight-minute segment was partitioned into 15 non-overlapping windows of 32s duration. The CapnoBase reference values, $R_{ref,i}$ for the i^{th} window, were obtained via manual annotations of the capnography waveform, which was performed by the authors of that study. For our analysis, the data set was randomly divided into $\mathcal{D}_{c,1}$, a training-and-validation set (32 subjects), and $\mathcal{D}_{c,2}$, a test set (10 subjects).

The MIMIC II data set, \mathcal{D}_m , consisted of the same 53 subjects as described in [23] and is available at the author's website¹ (Table S.1). One eight-minute segment was extracted between the $t = 60$ and $t = 68$ minutes of the recording and partitioned into 32s windows following [23]. The onset at $t = 60$ minutes was selected to minimize signal artifacts arising at the start of monitoring.

B. Pre-processing Procedure

Data pre-processing consisted of: data windowing, RR reference calculation, and peak and trough detection.

First, data from each eight-minute recording was partitioned into 15 non-overlapping windows of 32s duration. This resulted in $N = 480$ windows in $\mathcal{D}_{c,1}$, $N = 150$ in $\mathcal{D}_{c,2}$ and $N = 795$ in \mathcal{D}_m . Previous work shows that signal stationarity can be assumed for RR signals for a 32s duration [17].

¹<http://www.robots.ox.ac.uk/~davidc>

Next, the reference RR value (“gold standard”), $R_{ref,i}$, was obtained for each window in \mathcal{D}_c using the annotations of the peaks and troughs as unique labels. For \mathcal{D}_m , $R_{ref,i}$ was obtained from manual annotations of breaths using the IP waveforms by two of the authors of the current work. Using the two sets of independent labels for both \mathcal{D}_c and \mathcal{D}_m , two independent RR estimates were obtained by averaging the time between annotations in each window. $R_{ref,i}$ is the average of the two RR estimates. However, when the difference between the estimates was ≥ 2 brpm, the window was discarded. This resulted in 470 (97.9%) windows being retained in $\mathcal{D}_{c,1}$, 150 (100%) in $\mathcal{D}_{c,2}$, and 737 (92.7%) in \mathcal{D}_m .

Finally, the peaks and troughs from the PPG and ECG waveforms were determined. The PPG waveform was filtered using a 0.05 - 4.00 Hz bandpass filter prior to peak detection. PPG peaks were detected using the peak delineator defined in [24] and ECG peaks were detected using the search-back algorithm in [25]. Detected peaks were corrected to be the local maximum ± 0.2 s around each detected peak. Troughs were detected as the local minimum 0.2s before a peak.

C. RR Estimation

The PPG and ECG AM, FM, and BW modulations for each window were extracted from the peaks and troughs (Fig. S.1). The peaks and troughs were represented as a series of pairs, where $(t_{pk}, y_{pk})_{n \dots N}$ represents the time ($t_{pk,n}$) and amplitude ($y_{pk,n}$) of the n^{th} peak and $(t_{tr}, y_{tr})_{n \dots N}$ represents the time ($t_{tr,n}$) and amplitude ($y_{tr,n}$) of the n^{th} trough.

- AM is the magnitude of the signal from the peak to the prior trough: $y_{AM,n} = (t_{pk,n}, y_{pk,n} - y_{tr,n})_{n \dots N}$
- FM is the change in the time interval between subsequent peaks: $y_{FM,n} = (t_{pk,n}, t_{pk,n+1} - t_{pk,n})_{n \dots N}$
- BW is the envelope of the signal: $y_{BW,n} = (t_{pk,n}, y_{pk,n})_{n \dots N}$

The resulting six (three PPG and three ECG) modulations were filtered using a 5th-order Butterworth bandpass filter between 0.083 Hz and 1.000 Hz (representing 5 to 60 brpm) and downsampled to 4 Hz. For each modulation, AM_{PPG} , FM_{PPG} , BW_{PPG} , AM_{ECG} , FM_{ECG} , and BW_{ECG} , an RR estimate was obtained using the ARSpec algorithm [26]. These estimates were used to validate our proposed method. For all six modulations the estimates are denoted $R_{M,1} \dots R_{M,6}$, respectively.

D. RQI Calculation

Three RQIs, $Q_{S,k}$ ($k = 1 \dots 3$) were defined based on the Fourier transform ($Q_{S,1}$), autocorrelation ($Q_{S,2}$), and autoregression ($Q_{S,3}$). Two other RQIs based on the Hjorth complexity [27] and Cosine correlation [28] were defined; however, these RQIs lacked robustness. Each RQI was applied to the respiratory modulations described in the previous section.

1) *FFT RQI*: $Q_{S,1}$ used the power spectrum ($X_{i=1 \dots 64}$) of each 4 Hz, 32s time-series, as estimated via the 128-point FFT. Each component of the FFT covered a frequency band of $\frac{1}{32}$ Hz. We defined $Q_{S,1}$ as follows:

$$Q_{S,1} = \frac{\hat{X}}{\sum_j X_j} \quad (1)$$

where:

$$\hat{X} = \max \left\{ \sum_{k=z-1}^z X_k \middle| (z-1), (z) \in [2 \ 32] \right\} \quad (2)$$

and where $j \in [2 \ 32]$. The FFT components with indices [2 32] corresponded to the range of plausible respiratory frequencies [0.083 1.000] Hz. $Q_{S,1}$ defined the ratio of (the maximum power in any 2-component sub-range, corresponding to a window range of 0.031 Hz, of these respiratory frequencies) to (the total power in the range of respiratory frequencies).

2) *Autocorrelation RQI*: $Q_{S,2}$ used the autocorrelation of the respiratory waveform for every sample lag within the physiologically relevant respiratory range ($k = 4 \dots 48$) where each lag denotes a phase shift of $\frac{4}{k}$ Hz.

$$Q_{S,2} = \max_{k=4 \dots 48} \left\{ \frac{\frac{1}{N-1} \sum_{n=1}^{N-k} (x(n) - \bar{x}) * (x(n+k) - \bar{x})}{c_0} \right\} \quad (3)$$

where c_0 is the signal variance, $N = 128$ is the length of the signal, \bar{x} is the signal mean, and k is the signal lag.

3) *Autoregression RQI*: $Q_{S,3}$ fitted an AR process to the $N = 128$ values for each window using the Yule-Walker method [29]. AR processes of order $M = 1 \dots 30$ were considered, with the optimal order M selected using Akaike’s Information Criterion [29]. Using the optimal AR process, the power spectral estimate was estimated at $i = 481$ evenly-spaced points ($X_{i=1 \dots 481}$) from 0.0 to 2.0 Hz (the Nyquist frequency of the time-series) with each X_i representing a frequency band of $\frac{2}{480}$. $Q_{S,3}$ was calculated as:

$$Q_{S,3} = \frac{\hat{X}}{\sum_j X_j} \quad (4)$$

where:

$$\hat{X} = \max \left\{ \sum_{k=z-4}^{z+4} X_k \middle| (z-4), (z+4) \in [21 \ 241] \right\} \quad (5)$$

and where $j \in [21 \ 241]$. The 481-point spectrum had components with indices [21 241] corresponding to the range of plausible respiratory frequencies [0.83 1.00] Hz. $Q_{S,3}$ was the ratio of (the maximum power in any 9-component sub-range, corresponding to a window range of 0.03 Hz, of these respiratory frequencies) to (the total power in the range of respiratory frequencies).

E. RQI Fusion

Two supervised methods: linear regression (LR) and support vector regression (SVR) and one unsupervised method: principal component analysis (PCA) were considered to fuse $Q_{S,k}$ ($k = 1 \dots 3$) into a single, fused RQI. Each model was trained using $\mathcal{D}_{c,1}$ where data was not differentiated based on modulation nor waveform type. Model performance was investigated using $\mathcal{D}_{c,2}$ and \mathcal{D}_m .

RQI values, for all data sets, were standardized (a zero-mean and unit-variance normalization) using the mean and variance of $\mathcal{D}_{c,1}$. We defined an error term for each window, $E = |R_{M,j} - R_{ref}|$ where R_{ref} is the value of the reference RR for the window and $R_{M,j}$ is the estimated RR for a given modulation as described in section II-C.

Formally, each RQI-fusion candidate method \tilde{Q}_F is defined $\tilde{Q} : \mathcal{X} \mapsto \mathcal{Y}$, where $\mathcal{X} \in [0, 1]^3$ and $\mathcal{Y} \in \mathbb{R}$, and where the output y is a regression onto E . That is, the RQI-fusion methods take as input the 3-dimensional vector of RQI values for (one modulation of) a window and attempts to estimate the error E for that window. The error term E represents how erroneous the value of RR provided by the standard implementation of an existing algorithm (used to produce $R_{M,j}$), is in comparison to the “gold standard” for that window (R_{ref}). Of note is that E depends on the modulation RR estimation method. We chose the ARSpec algorithm [26] due to its reliability and availability. The method performs similarly to other techniques including the FFT-based method implemented in [19] as shown in supplementary Table S.2.

1) *Linear Regression*: The LR model was trained using 5-fold cross-validation (CV) where $\mathcal{D}_{c,1}$ was subdivided into 5 equal sub-samples to create 5 independent training-and-validation data sets. The individual RQIs, $Q_{S,k}$ were used as triplet observations for each modulation and window to determine the LR regularization parameter, λ from $\lambda = 0 \dots 16.38$, where the models were trained independently on each training set and where the optimal regularization parameters were determined by minimizing the average mean squared error (MSE) of the independent LR model on the held-out validation sets. Using the optimal λ , the LR model was trained using the complete $\mathcal{D}_{c,1}$ data set. The resulting model was applied to $\mathcal{D}_{c,2}$ and \mathcal{D}_m resulting in a single fused RQI, $\tilde{Q}_{F,1}$, for each modulation of each window in those data sets.

2) *Support Vector Regression*: The SVR fusion model was trained using the radial basis function, $\exp(-\gamma\|\mathbf{x} - \mathbf{x}'\|^2)$, to transform the individual RQI values, $Q_{S,k}$, and allow for the observation of non-linear relationships between $Q_{S,k}$ and E . The SVR model was trained using the same 5-fold CV training and validation sets as the LR method described in section II-E1. The SVR has two hyperparameters: the penalty parameter, C , and the kernel bandwidth, γ . The optimal value for these was found using grid search, where $C = 2^a$ for $a \in [-5, -4, \dots, 15]$ and $\gamma = 2^b$ for $b \in [-15, -14, \dots, 5]$. The grid search was conducted independently for each CV fold; the optimal values for C and γ were determined by minimizing the sum of MSEs, averaged over all CV folds. Using a_{opt} and b_{opt} (the best values averaged over all CV folds), a further fine grid search was conducted for $a \in [a_{opt} - 0.75, a_{opt} - 0.50, \dots, a_{opt} + 0.75]$ and $b \in [b_{opt} - 0.75, b_{opt} - 0.50, \dots, b_{opt} + 0.75]$ using increments of 0.25. The hyperparameters were the refined values from the fine grid search, C^* , and γ^* . Using these hyperparameters, the SVR model was trained using the individual RQIs from $\mathcal{D}_{c,1}$. The SVR model was applied to the RQIs from $\mathcal{D}_{c,2}$ and \mathcal{D}_m to obtain the SVR RQIs ($\tilde{Q}_{F,2}$).

3) *Principal Component Analysis*: The PCA method obtains the principal components (PCs) of a training set, and has no hyperparameters—therefore, CV is not required. We thus calculated the PC decomposition of the individual RQIs in $\mathcal{D}_{c,1}$. The resulting decomposition was applied to $\mathcal{D}_{c,2}$ and \mathcal{D}_m and the PCA RQI ($\tilde{Q}_{F,3}$) was defined of the first PC.

4) *Fusion RQI Bounding*: The RQI fusion models were trained to estimate the RR error E such that low E maps to a low value for \tilde{Q}_F ; however, the minimum and maximum

values for \tilde{Q}_F were not bounded and the individual RQIs were mapped such that high values for Q_S mapped to low values of E . Therefore, for comparison purposes, all values for \tilde{Q}_F were inverted and bounded from 0 to 1 using a logistic transformation as described in section S.4. The transformed RQI fusion values are labeled as Q_F .

F. RR Estimation

After completing the analysis described above, we produced fused RQI values, $Q_{F,i}$, with $i = 1 \dots 3$ representing the three RQI fusion candidate methods (LR, SVR, PCA). We have fused RQIs for each of the six modulations (j), and hereafter use $Q_{F,i,j}$ to represent the fused RQI for a window of the j^{th} modulation type using the i^{th} fusion method. For a given window, we also have an RR estimate, $R_{M,j}$ for each of the $j = 6$ modulations. We now seek to use the fused RQI, $Q_{F,i,j}$, and the RR estimates, $R_{M,j}$ to produce a maximally robust RR estimate for a given window. The current state-of-the-art for fusing RR estimates from multiple modulations into a single estimate is to average all values, $R_{M,j}$, for a time window. Additional methods to improve the accuracy of the estimates at the expense of discarding data as unusable can also be applied, one commonly-used example of which is “SmartFusion” [19].

We hypothesized that using our RQIs as a measure of confidence in the accuracy of an RR estimate for a given modulation, j , RR estimation accuracy could be improved. We therefore defined a new RR-estimation algorithm using a tunable accuracy-control threshold where:

$$R_{F,i} = \frac{\sum_{j \in s} Q_{F,i,j} R_{M,j}}{\sum_{j \in s} Q_{F,i,j}} \quad (6)$$

where $i = 1 \dots 3$ defines the fusion model, $j = 1 \dots 6$ defines the modulation type, $Q_{F,i,j}$ is the fused RQI, $R_{M,j}$ is the modulation-specific RR estimate, and where s is the subset of modulations j for which the fused RQI exceeds T :

$$s = j \mid Q_{F,i,j} \geq T \quad (7)$$

In the above, T is a tunable threshold parameter that defines a restricted subset s of modulations for which our fused RQI suggests data of significantly high quality. As the value of T is increased, the size of subset s decreases, as more of the $j = 6$ modulations are deemed to be of insufficient quality.

G. Evaluation Methods

The ability of our $k = 3$ individual (unfused) RQIs, $Q_{S,k}$, and our $i = 3$ fused RQIs, $Q_{F,i}$, to discriminate high-from low-quality respiratory segments was assessed using two methods: (i) Spearman’s rank-order correlation ρ , between Q and E , and (ii) by observing the change in the mean value of error E . For this latter analysis, we defined a metric:

$$E_{50} = \frac{\sum_N E}{N} - \frac{\sum_m E}{m} \quad (8)$$

where N represents the complete super set of N windows and $j = 6$ modulations (therefore comprising $6N$ values), and where M represents the indices of those $\frac{6N}{2}$ windows and modulations with values of Q greater than the median value

of Q . This metric E_{50} therefore takes greater positive values when Q may be used to distinguish windows with high error E from windows with low error E .

We compared the performance of $Q_{S,k}$ and $Q_{F,i}$, to a baseline control and an “ideal” control. The baseline control was a commonly-used SQI based on the F1 Score, which assesses the waveform quality prior to extraction of respiratory modulations [25]. The “ideal” control is an artificially-generated quality index produced with knowledge of E (and therefore of the gold standard, R_{ref}). For this metric, Spearman’s correlation between the metric scalar and error E is $\rho = -1.00$, and it therefore acts as the “ideal” RQI. That is, a good RQI will have high negative correlation with E , and hence the “ideal” is straightforwardly set to a scalar value varying such that $\rho = -1.00$ (it perfectly “estimates” E , because it is based on a negative scalar multiple of E).

III. RESULTS

A. Performance of Individual RQIs

Table II shows that every individual RQI, $Q_{S,k}$, discriminated high- from low-quality respiratory modulations better than an SQI for all three data sets. The value for E_{50} for all $Q_{S,k}$ in $\mathcal{D}_{c,1}$ and \mathcal{D}_m exceeds 1 indicating that when 50% of the respiratory modulations are discarded, the average error in the RR estimate for all modulations decreases by at least 1 bpm (this is not seen in $\mathcal{D}_{c,2}$, however, the magnitude of E_{50} for the ideal control is also significantly smaller in $\mathcal{D}_{c,2}$ compared to other data sets). This performance can be compared to the SQI control in which the E_{50} never exceeds 1 and the “ideal” control where the optimal value for E_{50} varies significantly by data set but ranges from 1.38 to 4.54 bpm.

While all three individual RQIs have similar performance for E_{50} and ρ compared to E , two aggregate measures, the individual RQIs do not perform identically. This can be observed by looking at the rank correlation among the individual RQIs, $Q_{S,k}$, in which the values between RQIs range from 0.70 to 0.87 amongst all data sets (Table S.3). While these correlations are expected to be strong, as the individual RQIs should be similarly large for the highest quality respiratory modulations, they value of the correlations is low enough to suggest that for the borderline cases, those respiratory modulations which may or may not contain respiratory information, the RQIs perform differently, as is expected since each RQI has been defined to look at different potential features of the respiratory waveform (as is later described in section IV-A).

Fig. 2 shows the results of the analysis for held-out data sets $\mathcal{D}_{c,2}$ and \mathcal{D}_m (results for the training set $\mathcal{D}_{c,1}$ appear in supplementary Fig. S.2). In this analysis, we examined the effect of discarding a modulation when an RQI value $Q < V$, for some threshold V . If $V = 0$, no data is discarded corresponding to 100% data retention, shown leftmost on the horizontal axis. As V is increased, the number of data that are retained decreases, corresponding to moving rightward along the horizontal axis. Fig. 2 therefore presents the MAE of increasingly more stringent subsets of data for an increasing value of V where each subset contains only the windows and modulations where $Q > V$. The results further demonstrate

TABLE II
 E_{50} AND SPEARMAN’S CORRELATION ρ FOR $Q_{S,k}$ ($k = 1 \dots 3$), $Q_{F,i}$ ($i = 1 \dots 3$), AND CONTROLS.

	$\mathcal{D}_{c,1}$		$\mathcal{D}_{c,2}$		\mathcal{D}_m	
	E_{50}	ρ	E_{50}	ρ	E_{50}	ρ
$Q_{S,1}$	1.92±0.07	-0.50	0.99±0.08	-0.44	1.49±0.06	-0.37
$Q_{S,2}$	1.79±0.07	-0.52	0.86±0.09	-0.54	1.47±0.06	-0.38
$Q_{S,3}$	1.77±0.07	-0.48	1.09±0.07	-0.50	1.26±0.06	-0.33
$Q_{F,1}$	2.03±0.07	-0.52	1.04±0.08	-0.49	1.52±0.07	-0.39
$Q_{F,2}$	2.35±0.07	-0.70	0.97±0.08	-0.42	-1.30±0.07	0.32
$Q_{F,3}$	1.88±0.07	-0.53	0.97±0.08	-0.53	1.41±0.07	-0.38
SQI	0.57±0.08	-0.04	0.06±0.10	0.03	0.29±0.06	-0.11
Control	3.07±0.05	-1.00	1.38±0.06	-1.00	4.54±0.04	-1.00

the positive association of high individual RQI values with low RR estimation error for all $Q_{S,k}$. The figure shows that the “ideal” method results in monotonically-decreasing MAE and that the SQI alone, without an RQI (as is current practice reported in the literature), does not reliably reduce MAE.

B. Determination of Model for RQI Fusion

Performance of the individual RQIs show that no single $Q_{S,k}$ is optimal for every data set. The value of E_{50} for $Q_{S,1}$ is highest in $\mathcal{D}_{c,1}$ and \mathcal{D}_m and is highest for $Q_{S,3}$ for $\mathcal{D}_{c,2}$ (Table II). This motivates our proposed RQI-fusion methods $Q_{F,i}$, the results for which are shown in Table II and discussed below. Values for the hyperparameters of the fusion methods are included in (Figs. S.3, S.4, and S.5).

Fusion model $Q_{F,1}$ provided the best general performance across all three data sets and will be the focus of the remainder of the work described. While the magnitude of E_{50} and ρ were highest for $Q_{F,2}$ in $\mathcal{D}_{c,1}$, the performance of $Q_{F,2}$ for the held-out data sets $\mathcal{D}_{c,1}$ and \mathcal{D}_m did not outperform $Q_{F,1}$, $Q_{F,3}$, or even the conventional SQI, indicating its lack of ability to generalize with previously unseen data.

Because $Q_{F,1}$ is a fusion method derived using $Q_{S,1 \dots 3}$, it does not outperform the best individual RQI for each data set; however, it does outperform the remaining two individual RQIs for each data set. Because the best individual RQI for each data set is not the same and because each individual RQI is important for a full understanding of the respiratory waveform, $Q_{F,1}$ represents a more robust RQI measure than any single individual RQI alone when no *a priori* information on the data set, subject, or available respiratory modulations is known. This is further demonstrated in section S.8.

C. Improving RR Estimation via RQIs

The previous two sections demonstrate the capacity of the RQIs to serve as a measure of the confidence that a given respiratory modulation contains physiologically-useful respiratory information. This is only clinically relevant if an accurate RR estimation can be obtained for a given window. We now examine the performance of the respiratory modulation fusion method described in section II-F.

We compare the results from our proposed RR-estimation method, RQIFusion, to (i) the average of value of $R_{M,j}$ (the RR estimates obtained as described in section II-C); (ii) the existing SmartFusion method; and (iii) a modified version of

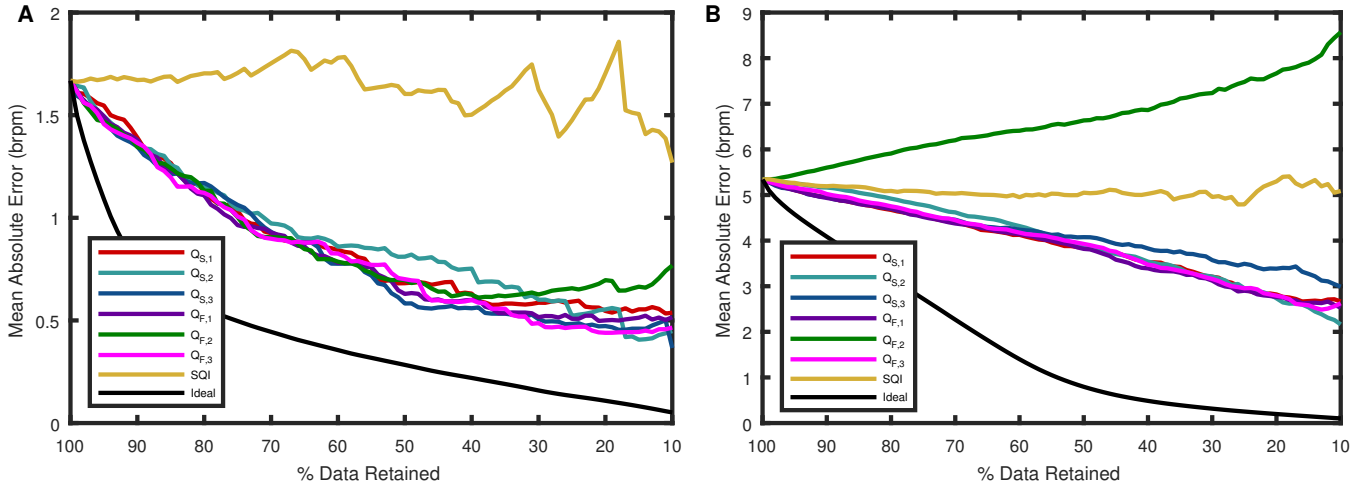


Fig. 2. MAE of $Q_{S,k}$, $Q_{F,i}$, and controls on (A) $\mathcal{D}_{c,2}$ and (B) \mathcal{D}_m , as windows are discarded based on comparison of RQIs to a threshold V . Note that the horizontal axis shows the percentage of data retained, which decreases as the value of V increases.

SmartFusion. In the above, (ii) and (iii) take the average of RR estimates for each available respiratory modulation and discard any window of data for which the standard deviation of those RR estimates is ≥ 4 brpm. SmartFusion as described in [19] uses only PPG-based RR estimates ($R_{M,1} \dots R_{M,3}$), while the SmartFusion modification uses all six estimates ($R_{M,1} \dots R_{M,6}$) as a fair comparison to RQIFusion.

RQIFusion uses the threshold parameter, $T \in [0, 1]$, which is the threshold applied to $Q_{F,1,j}$ to remove modulations from subsequent RR estimation (cf. section II-F). A trade-off exists between the data retention and the error in RR estimates (as seen in Fig. 2). This trade-off can be seen in $\mathcal{D}_{c,1}$ in Fig. 3, which shows MAE for RR estimates vs. data discarded on a window-by-window basis. The results in Fig. 3 show data discard in terms of the number of whole windows for which no RR estimate could be produced—this only occurs for our proposed method when all six of the modulations for a given window have $Q_{F,1,j} < T$. RQIFusion is able to produce an RR estimate for a window if at least one modulations has an RQI greater than T . This is a key advantage of RQIFusion compared to existing methods; the latter will only report an RR estimate for a window if the s.d. of estimates for all modulations is sufficiently small.²

Fig. 3 shows that all three comparator methods have substantially different rates of data-discard, while the proposed method allows selection of the data-discard rate (by varying threshold T). RQIFusion provides RR estimates that have lower MAE than the comparator methods.

Table III shows a comparative summary of these results for $\mathcal{D}_{c,1}$, whereby RQIFusion is compared to each of these comparator methods either by (i) matching comparator method data-discard rate (ii) matching comparator method MAE (this matching is achieved by varying threshold T). For example, modified SmartFusion has an MAE of 1.00 with data-discard

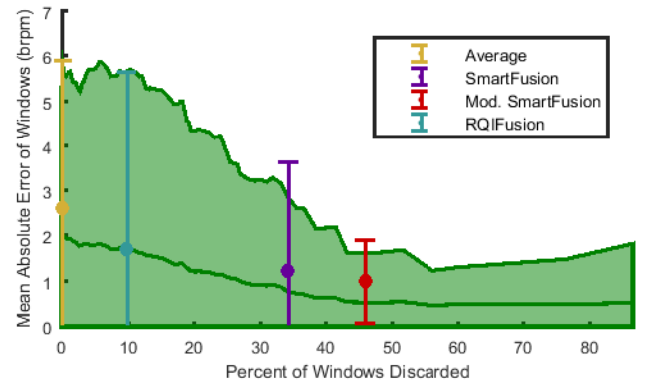


Fig. 3. MAE of RR estimates for windows from the proposed RQI-based estimation algorithm versus the percentage of windows discarded as threshold T is increased in $\mathcal{D}_{c,1}$. Results are also shown for the average RR estimate for a window (yellow), SmartFusion (purple), modified SmartFusion (red), and RQIFusion (cyan). Upper and lower bounds represent the standard deviation of MAE over all windows, shown plotted around the MAE.

rate of 46.0%. If RQIFusion matches this MAE of 1.00, it has a data-discard rate of 27%. Similarly, if RQIFusion matches the data-discard rate of 46%, it has an MAE of 0.53. In each instance, RQIFusion outperforms the comparator methods.³

For further analysis, the threshold value, T for RQIFusion has been set to 0.44. This value is the highest value for T in which no more than 10% of windows in $\mathcal{D}_{c,1}$ do not have an RR estimate (Fig. 3). However, T is tunable and further exemplar values of T are explored in section S.9.

Table IV shows the median error, MAE, and percentage of windows discarded for all of the RR-estimation algorithms considered for $\mathcal{D}_{c,2}$ and \mathcal{D}_m .⁴ Table IV shows that the median and MAE RR values obtained using RQIFusion are lower than the existing methods in all but one instance (SmartFusion

²We note that the presentation of results in Fig. 3 differs fundamentally from that in Fig. 2; the latter considers all windows \times modulations while the former considers the window-based results that correspond to use in practice, taking into account improved RR estimation within a window using RQIs for each of that window's modulations.

³It is not possible for the proposed method to match the MAE of the “average” comparator method because the MAE of the proposed method is < 2.61 for all values of T .

⁴Equivalent results for the training set, $\mathcal{D}_{c,1}$ and the additional exemplar values for T are provided as supplementary material, Table S.7.

TABLE III
EQUIVALENT PERCENTAGE OF WINDOWS DISCARDED (DIS. %) OR
EQUIVALENT MAE FOR RQI-BASED ESTIMATION ALGORITHM VS.
COMPARATOR METHODS IN $\mathcal{D}_{c,1}$.

Current Methods	MAE	Dis. %	RQI-based Estimation	
			Equiv. MAE (Fixed Dis. %)	Equiv. Dis. % (Fixed MAE)
Average	2.61	0.0%	1.98	
SmartFusion	1.23	34.3%	0.79	20.9%
Mod. SmartFusion	1.00	46.0%	0.53	27.0%

vs RQIFusion MAE in $\mathcal{D}_{c,2}$) and are significantly lower in a number of instances, particularly comparing the MAEs in \mathcal{D}_m from averaging (4.04 brpm) and SmartFusion (4.61 brpm) to RQIFusion (3.12 brpm). Furthermore, RQIFusion has a significantly lower percentage of windows discarded when compared to SmartFusion and modified SmartFusion in all instances. While the percentage of windows discarded from averaging compared to RQIFusion is not lower (as averaging discards no windows), the reduction in RR estimation error from RQIFusion far outperforms averaging in all instances.

These results support the conclusion that RQIFusion is effective at removing the lowest-quality modulations and, after removing them, the resulting RR estimate is more accurate than those obtained with current approaches. This is of particular importance for elderly populations, such as those in \mathcal{D}_m , where RR estimation is more difficult because the prevalence of certain RR modulations diminishes with age [16].

D. Performance of PPG and ECG as Individual Modalities

Additionally, we consider the performance of our method in the instance where the threshold T has been set to yield the equivalent number of windows discarded as SmartFusion in $\mathcal{D}_{c,2}$ and $\mathcal{D}_{m,2}$ when only modulations for PPG or ECG (i.e. only considering $Q_{F,1,j}$ and $R_{M,j}$ for $j = 1 \dots 3$ for PPG and $j = 1 \dots 3$ for ECG) (Table S.8).⁵ This analysis provides a direct comparison to the original SmartFusion method where only the three PPG modulations were considered for fusion. Our results show that under these constrained circumstances, our method still outperforms SmartFusion in every instance and suggests the applicability of our method even when only PPG or ECG is available.

E. RR Estimation for Individual Subjects

All previously presented results have considered how the proposed methods improve the performance of RR estimation for modulations or windows; in the final set of results, we investigate the performance of RR estimation at the level of the individual subject. Noting that each subject has 8-minutes of continuous data, which were processed into windows of 32s duration, an evenly-spaced time-series of up to 15 RR estimates is available for each subject if each window yields an RR estimate. However, RR estimates may not be available for every window for every method. This section investigates the

completeness of RR estimation across whole subject records to determine how the inability of a method to estimate RR for some windows might affect the care of an individual patient.

Fig. 4 shows time-series of RR estimates for two subjects from $\mathcal{D}_{c,2}$.⁶ In Fig. 4, we show results for RQIFusion with the application of a 3-point median filter as well as the three existing methods, averaging, SmartFusion, and modified SmartFusion. The case studies shown in Fig. 4 demonstrate two key improvements of RQIFusion over existing methods: (i) improved RR estimation occurs during intervals in which there is highly variable RR, as shown in Fig. 4A; and (ii) highly improved completeness of RR estimation, even when existing methods were unable to estimate RR for most of the windows for a given subject, as shown in Fig. 4B.

The latter finding is further demonstrated in Fig. 5, which shows histograms of the number of windows for which no RR estimate is available for each subject in $\mathcal{D}_{c,2}$ and \mathcal{D}_m . The figure shows that existing methods have a large number of subjects for whom there is a substantial proportion of their records without RR estimates. Fig. 5A shows that many subjects have more than 5 of 15 windows without RR estimates; this is exacerbated for data set \mathcal{D}_m , as shown in Fig. 5B, where many subjects have the majority of their records without RR estimates. This substantial data-loss could adversely affect patient safety in some applications.

IV. DISCUSSION

The current work demonstrates that RQIFusion exhibits a number of advantages over previous modulation fusion methods and provides substantially more complete RR monitoring; however, further exploration of RQIFusion is still necessary.

A. Summary of Main Points

Previous efforts to provide signal quality measures for RR-estimation applications have focused on determining whether the PPG or ECG waveforms are artifactual. We have described the use of the F1 score [25] and approaches that have focused on an individual feature of the sensor waveform, such as sinusoidality [17]. Such methods look only at the sensor waveforms themselves, and not at the respiratory waveforms derived from them—and it is the latter that are used for RR estimation. We have shown that existing SQI-based methods are a poor proxy for determining the presence of respiratory information in the modulations extracted from sensor waveforms (Fig. 2 and Table II). This is because the PPG and ECG can be non-artifactual but not contain respiratory information in their respiratory modulations [16], [17].

We have shown that the individual RQIs, $Q_{S,1} \dots Q_{S,3}$, (based on the Fourier transform, autocorrelation, and autoregression) all contain critical respiratory quality information as no single RQI consistently out-performed the others. Each of these methods was implemented to capture different anticipated features pertaining to information in the modulations of the sensor waveforms. Both $Q_{S,1}$ and $Q_{S,3}$ are based on power

⁵In this analysis, when only ECG is considered, SmartFusion is also modified so that it is run in an identical fashion except that the three modulations it considers are also the ECG modulations.

⁶The time-series of RR estimates for the other subjects in $\mathcal{D}_{c,2}$ is available in supplementary Fig. S.7

TABLE IV
MEDIAN AND MEAN ABSOLUTE ERROR VALUES AND THE PERCENTAGE OF MODULATIONS DISCARDED FOR EACH RR FUSION METHOD FOR $\mathcal{D}_{c,2}$ AND \mathcal{D}_m .

	Median (IQR)	$\mathcal{D}_{c,2}$ MAE (\pm SD)	% Dis.	Median (IQR)	\mathcal{D}_m MAE (\pm SD)	% Dis.
Avg.	0.81 (0.39 - 1.39)	1.06 \pm 0.93	0.0%	3.54 (1.72 - 5.88)	4.04 \pm 2.90	0.0%
SmartFusion	0.46 (0.23 - 0.92)	0.68 \pm 0.61	18.0%	2.83 (0.66 - 8.51)	4.61 \pm 4.50	58.8%
Mod. SmartFusion	0.65 (0.31 - 1.08)	0.76 \pm 0.56	26.7%	2.20 (1.21 - 4.51)	3.33 \pm 3.00	73.0%
RQIFusion	0.43 (0.25 - 0.77)	0.71 \pm 0.89	1.3%	0.72 (0.27 - 4.91)	3.12 \pm 4.39	23.2%

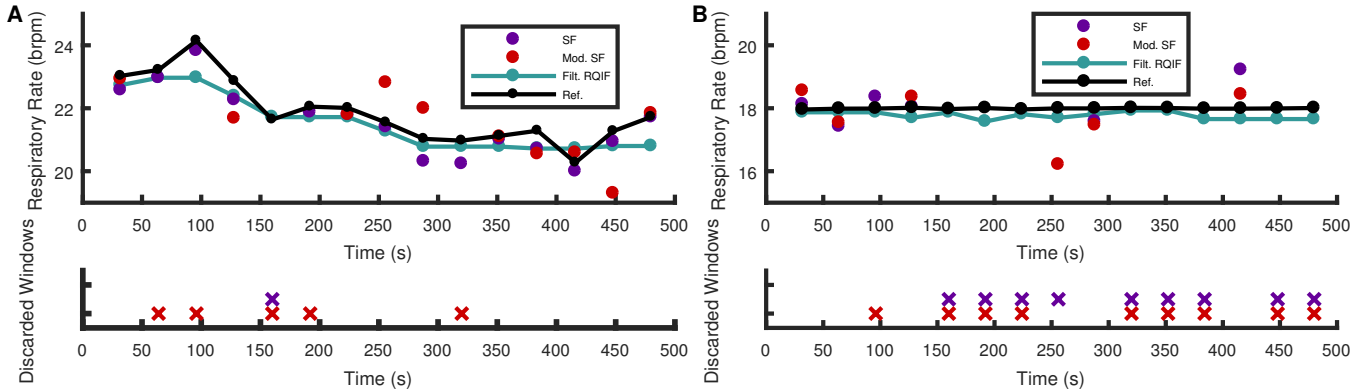


Fig. 4. (Upper row) Time-series of RR estimates for modified SmartFusion, SmartFusion, RQIFusion, and the R_{ref} values for two subjects in data set $\mathcal{D}_{c,2}$: (A) subject 23 and (B) subject 103. (Lower row) Corresponding plots showing those windows for which each method did not produce an RR estimate.

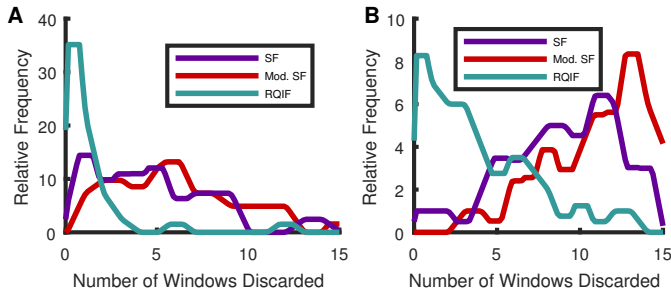


Fig. 5. The relative frequency of windows being discarded for individual subjects in \mathcal{D}_c and \mathcal{D}_m (shown in A and B, respectively), using modified SmartFusion (red), SmartFusion (purple), and RQIFusion (cyan).

spectral measures, and have been used for RR estimation previously [9], although not for the function performed by the RQIs. Comparative studies of both methods in heart-rate variability (HRV) applications have suggested that, despite methodological similarities, each provides unique information concerning the strength of the respiratory modulation [30].

Unlike the other two RQIs, $Q_{S,2}$, does not assume the sinusoidality of the respiratory signal, but assumes periodicity. While respiratory signals are generally sinusoidal, this is not always the case as breathing patterns such as long or short inspiration and positive-negative respiration, can alter the waveform [31]. $Q_{S,2}$ is proposed to cope with this.

We concluded that RQI fusion using linear regression provided the best-performing fusion model, $Q_{F,1}$. While LR is a straightforward supervised fusion method, we found that it outperformed SVR when applied to held-out data. This suggests that the more complex, non-linear SVR does not

improve results for the data sets that we considered (Fig. 2).

We defined a novel RR estimation method, RQIFusion, that depended on $Q_{F,1}$ to serve as a measure of confidence in the quality of a respiratory modulation. Our algorithm resulted in a trade-off between reducing RR estimation error and increasing data loss (Fig. 3). RQIFusion had improved performance compared to existing methods. Data-discard rates were especially high for some individual subjects using existing methods, which was substantially improved by RQIFusion (Fig. 5).

B. Advantages of the Proposed Method

In addition to improved RR estimation, RQIFusion is configurable for different applications via the selection of threshold, T . In settings where higher RR estimation accuracy is required, the value of T can be increased whereas in settings where a continuous stream of estimates is preferred, the value of T can be decreased. A potentially useful consequence of having RQIs to track the presence or absence of respiratory information in modulations of the ECG and PPG is for applications which it is helpful to understand activity of the parasympathetic nervous system on a window-by-window basis. For example, changes in the presence of respiratory sinus arrhythmia (FM of the PPG and ECG) may be associated with changes in health status of certain patient groups, such as those underlying hemodialysis [32].

V. CONCLUSION

The present work describes the development and application of RQIs, a novel pre-processing step in RR estimation that can quantify the level of confidence in the quality of extracted

respiratory modulations. The algorithm extracts the AM, FM, and BW from both PPG and ECG signals and provides a LR-based fused RQI from individual RQIs derived from the FFT, autocorrelation, and AR of the modulations. The model was trained with a subset of the CapnoBase data set and verified using another independent subset of the CapnoBase data set and a subset of the MIMIC II data set. RR estimation for each time window was conducted using a tunable threshold to set the data discard percentage where there was a trade-off between RR estimation accuracy and data retention. The results showed a marked improvement in both estimation accuracy and retention compared to previous RR estimation fusion methods, including SmartFusion. This work marks an important development as RQIs consider the respiratory-specific quality of a modulation and can be incorporated into pre-existing RR estimation algorithms to improve performance. However, further validation of the proposed method is still required (see limitations of present study in section S.12). Therefore, further studies of this method should focus on the application to larger sets of ambulatory populations and the incorporation of this method into more clinically relevant applications such prediction of in-hospital catastrophic deterioration.

DATA ACCESS STATEMENT

This manuscript is in compliance with the UK Research Councils Common Principles on Research Data Policy, and the data used in this research are openly available from public sources as described in the text.

REFERENCES

- [1] R. Schein, N. Hazday, M. Pena, B. Ruben, and C. Sprung, "Clinical antecedents to in-hospital cardiopulmonary arrest," *Chest Journal*, vol. 98, no. 6, pp. 1388–1392, 1990.
- [2] J. F. Fieselman, M. S. Hendryx, C. M. Helms, and D. S. Wakefield, "Respiratory rate predicts cardiopulmonary arrest for internal medicine inpatients," *J Gen Intern Med*, vol. 8, no. 7, pp. 354–60, 1993.
- [3] D. R. Goldhill, S. A. White, and A. Sumner, "Physiological values and procedures in the 24 h before ICU admission from the ward," *Anaesthesia*, vol. 54, no. 6, pp. 529–34, 1999.
- [4] J. McBride, D. Knight, J. Piper, and G. B. Smith, "Long-term effect of introducing an early warning score on respiratory rate charting on general wards," *Resuscitation*, vol. 65, no. 1, pp. 41–4, 2005.
- [5] C. H. Van Leuvan and I. Mitchell, "Missed opportunities? An observational study of vital sign measurements," *Crit Care Resusc*, vol. 10, no. 2, pp. 111–15, 2008.
- [6] E. A. Hooker, D. J. O'Brien, D. F. Danzl, J. A. Barefoot, and J. E. Brown, "Respiratory rates in emergency department patients," *J Emerg Med*, vol. 7, no. 2, pp. 129–32, 1989.
- [7] J. Hogan, "Why don't nurses monitor the respiratory rates of patients?" *Br J Nurs*, vol. 15, no. 9, pp. 489–92, 2006.
- [8] M. A. Cretikos, R. Bellomo, K. Hillman, J. Chen, S. Finfer, and A. Flabouris, "Respiratory rate: the neglected vital sign," *Medical Journal of Australia*, vol. 188, no. 11, p. 657, 2008.
- [9] P. Charlton, D. Birrenkott, T. Bonnici, M. Pimentel, A. Johnson, A. J., L. Tarassenko, P. J. Watkinson, R. Beale, and D. Clifton, "Respiratory rate estimation from the electrocardiogram and photoplethysmogram: A review," *Reviews in Biomedical Engineering*, In Press. [Online]. Available: <http://ieeexplore.ieee.org/document/8081839/>
- [10] T. S. Larson and W. J. Brady, "Electrocardiographic monitoring in the hospitalized patient: a diagnostic intervention of uncertain clinical impact," *Am J Emerg Med*, vol. 26, no. 9, pp. 1047–55, 2008.
- [11] J. A. Walsh, E. J. Topol, and S. R. Steinhubl, "Novel wireless devices for cardiac monitoring," *Circulation*, vol. 130, no. 7, pp. 573–581, 2014.
- [12] P. Hubner, A. Schober, F. Sterz, P. Stratil, C. Wallmueller, C. Testori, D. Grassmann, N. Lebl, I. Ohrenberger, H. Herkner, and C. Weiser, "Surveillance of patients in the waiting area of the department of emergency medicine," *Medicine*, vol. 94, no. 51, p. e2322, 2015.
- [13] F. El-Amrawy and M. I. Nounou, "Are currently available wearable devices for activity tracking and heart rate monitoring accurate, precise, and medically beneficial?" *Health Inform Res*, vol. 21, no. 4, pp. 315–320, 2015.
- [14] D. Meredith, D. Clifton, P. Charlton, J. Brooks, C. Pugh, and L. Tarassenko, "Photoplethysmographic derivation of respiratory rate: a review of relevant physiology," *Journal of medical engineering & technology*, vol. 36, no. 1, pp. 1–7, 2012.
- [15] D. Birrenkott, "Respiratory quality indices for ECG- and PPG-derived respiratory data," Technical Report, 2016.
- [16] F. Yasuma and J. Hayano, "Respiratory sinus arrhythmia: why does the heartbeat synchronize with respiratory rhythm?" *Chest*, vol. 125, no. 2, pp. 683–90, 2004.
- [17] S. Nemati, A. Malhotra, and G. D. Clifford, "Data fusion for improved respiration rate estimation," *EURASIP Journal on Advances in Signal Processing*, vol. 2010, no. 1, p. 926305, 2010.
- [18] P. H. Charlton, T. Bonnici, L. Tarassenko, J. Alastruey, D. A. Clifton, R. Beale, and P. J. Watkinson, "Extraction of respiratory signals from the electrocardiogram and photoplethysmogram: technical and physiological determinants," *Physiol Meas*, vol. 38, no. 5, pp. 669–690, 2017.
- [19] W. Karlen, S. Raman, J. M. Ansermino, and G. A. Dumont, "Multiparameter respiratory rate estimation from the photoplethysmogram," *Biomedical Engineering, IEEE Transactions on*, vol. 60, no. 7, pp. 1946–1953, 2013.
- [20] W. Karlen, M. Turner, E. Cooke, G. Dumont, and J. Ansermino, "Capnobase: Signal database and tools to collect, share and annotate respiratory signals," in *Annual Meeting of the Society for Technology in Anesthesia, 2010.*, Conference Proceedings, p. 25.
- [21] A. L. Goldberger, L. A. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C. K. Peng, and H. E. Stanley, "Physiobank, physiotoolkit, and physionet: components of a new research resource for complex physiologic signals," *Circulation*, vol. 101, no. 23, pp. E215–20, 2000.
- [22] M. Saeed, M. Villarreal, A. T. Reisner, G. Clifford, L. W. Lehman, G. Moody, T. Heldt, T. H. Kyaw, B. Moody, and R. G. Mark, "Multiparameter intelligent monitoring in intensive care II: a public-access intensive care unit database," *Crit Care Med*, vol. 39, no. 5, pp. 952–60, 2011.
- [23] M. Pimentel, E. Johnson, P. Charlton, D. Birrenkott, P. Watkinson, L. Tarassenko, and D. Clifton, "Towards a robust estimation of respiratory rate from pulse oximeters," *IEEE Transactions on Biomedical Engineering*, vol. 64, no. 8, pp. 1914–1923, 2017.
- [24] J. Li, J. Jin, X. Chen, W. Sun, and P. Guo, "Comparison of respiratory-induced variations in photoplethysmographic signals," *Physiol Meas*, vol. 31, no. 3, pp. 415–25, 2010.
- [25] M. Pimentel, M. Santos, D. Springer, and G. Clifford, "Heart beat detection in multimodal physiological data using a hidden semi-Markov model and signal quality indices," *Physiol Meas*, vol. 36, no. 8, pp. 1717–27, 2015.
- [26] S. A. Shah, S. Fleming, M. Thompson, and L. Tarassenko, "Respiratory rate estimation during triage of children in hospitals," *Journal of Medical Engineering & Technology*, vol. 39, no. 8, pp. 514–524, 2015.
- [27] B. Hjorth, "EEG analysis based on time domain properties," *Electroencephalogr Clin Neurophysiol*, vol. 29, no. 3, pp. 306–10, 1970.
- [28] D. B. Springer, "Mobile phone-based rheumatic heart disease detection," Thesis, 2015.
- [29] R. Takalo, H. Hytti, and H. Ihalainen, "Tutorial on univariate autoregressive spectral analysis," *J Clin Monit Comput*, vol. 19, no. 6, pp. 401–10, 2005.
- [30] D. Chemla, J. Young, F. Badilini, P. Maison-Blanche, H. Affres, Y. Lecarpentier, and P. Chanson, "Comparison of fast Fourier transform and autoregressive spectral analysis for the study of heart rate variability in diabetic patients," *Int J Cardiol*, vol. 104, no. 3, pp. 307–13, 2005.
- [31] N. A. Bergman, "Effects of varying respiratory waveforms on gas exchange," *Anesthesiology*, vol. 28, no. 2, pp. 390–5, 1967.
- [32] D. Meredith, Y. Borhani, S. Sutherland, L. Hills, S. Fleming, D. Clifton, A. Thornley, L. Tarassenko, and C. Pugh, "Monitoring of vital signs during haemodialysis," in *British Renal Association Conference, 2010.*, Conference Proceedings, p. 355.