Journal of Clinical Monitoring and Computing
https://doi.org/10.1007/s10877-021-00785-y

ORIGINAL RESEARCH

Continuous vital sign monitoring using a wearable patch sensor in obese patients: a validation study in a clinical setting

Niels Kant[1] · Guido M. Peters[2,3] · Brenda J. Voorthuis[3] · Catharina G. M. Groothuis-Oudshoorn[3] · Mark V. Koning[1] · Bart P. L. Witteman[4] · Myra Rinia-Feenstra[1] · Carine J. M. Doggen[2,3]

Received: 1 July 2021 / Accepted: 27 November 2021

Abstract

Our aim was to determine the agreement of heart rate (HR) and respiratory rate (RR) measurements by the Philips Biosensor with a reference monitor (General Electric Carescape B650) in severely obese patients during and after bariatric surgery. Additionally, sensor reliability was assessed. Ninety-four severely obese patients were monitored with both the Biosensor and reference monitor during and after bariatric surgery. Agreement was defined as the mean absolute difference between both monitoring devices. Bland Altman plots and Clarke Error Grid analysis (CEG) were used to visualise differences. Sensor reliability was reflected by the amount, duration and causes of data loss. The mean absolute difference for HR was 1.26 beats per minute (bpm) (SD 0.84) during surgery and 1.84 bpm (SD 1.22) during recovery, and never exceeded the 8 bpm limit of agreement. The mean absolute difference for RR was 1.78 breaths per minute (brpm) (SD 1.90) during surgery and 4.24 brpm (SD 2.75) during recovery. The Biosensor's RR measurements exceeded the 2 brpm limit of agreement in 58% of the compared measurements. Averaging 15 min of measurements for both devices improved agreement. CEG showed that 99% of averaged RR measurements resulted in adequate treatment. Data loss was limited to 4.5% of the total duration of measurements for RR. No clear causes for data loss were found. The Biosensor is suitable for remote monitoring of HR, but not RR in morbidly obese patients. Future research should focus on improving RR measurements, the interpretation of continuous data, and development of smart alarm systems.

Keywords Continuous monitoring · Wireless technology · Wearable electronic devices · Monitoring · Physiologic/instrumentation · Vital signs

1 Introduction

Despite medical advances, the risk of developing postoperative complications for patients undergoing surgery is ever present. A recent review found the incidence of 30-day postoperative complications to range from 5.8 to 43.5% in general surgery patients [1]. Early warning scores (EWS) have been developed to identify early signs of patient deterioration [2, 3]. Heartrate (HR) and respiratory rate (RR) are used in these EWS because these are among the first vital signs to change when deterioration occurs [4–6]. Since EWS in general wards are commonly measured intermittently, early signs of deterioration can easily be missed, possibly leading to life threatening events and even death [7].

Continuous monitoring of vital signs using conventional monitors, which connect the patient to a limited space around the bedside, may lead to worse health outcomes by restricting patient mobility [8–10]. However, continuous

Niels Kant and Guido M. Peters have contributed equally to the work.

✉ Carine J. M. Doggen
cdoggen@rijnstate.nl

1 Department of Anesthesiology and Pain Management, Rijnstate Hospital, Arnhem, The Netherlands

2 Scientific Bureau, Rijnstate Hospital, Rijnstate Research Center, Wagnerlaan 55, PO Box 9555, 6800 TA Arnhem, The Netherlands

3 Technical Medical Centre, Department of Health Technology and Services Research, University of Twente, Enschede, The Netherlands

4 Rijnstate Hospital, Vitalys Obesity Centre, Arnhem, The Netherlands

Published online: 08 December 2021

monitoring using wireless wearable devices enables patients to ambulate without restriction, and could therefore be a preferable solution for early detection of deterioration, provided that measurements are valid [10–14].

Previous studies of various wearable devices showed an accurate ECG-based heart rate measurement. The accuracy of RR was lower, depending on the method of measurement and patient characteristics [15–17]. RR may be derived from ECG, accelerometry, impedance, capnography, or a combination. The sensor that is the subject of this study, the Philips Biosensor, is an updated version of the VitalConnect HealthPatch, with a new algorithm for computing RR. It measures HR with ECG, and determines RR through a combination of ECG derivatives and accelerometry. Limitations of previous studies investigating wireless wearable sensors for continuous monitoring include non-clinical settings, non-device-based monitoring as reference, comparison with intermittent measurements, and small and often healthy study populations [15–23]. As a result, generalization of their findings to an obese population as well as to use in clinical practice is limited. As populations are becoming increasingly overweight [24], it is especially important that wearable devices are validated in obese patients, as a larger amount of subcutaneous fat might interfere with measurements. Additionally results achieved in lab settings often differ from results seen in clinical settings. Thus, validating devices in a clinical setting is imperative before implementation at a wider scale.

The objective of this study was to determine the agreement of HR and RR measured by the Philips wearable Biosensor in morbidly obese patients undergoing bariatric surgery.

Agreement was determined during surgery and in the recovery room where standard monitoring was available. A secondary aim was to assess the reliability of this sensor in terms of the duration and frequency of data loss. This was determined over the total duration of Biosensor measurements, i.e. during surgery, recovery and on the ward.

2 Methods

Ethical approval for this study was asked for and waived by the Medical Research Ethics Committee Arnhem-Nijmegen, (registration 2019–5489). The study fell outside the remit of the law for Medical Research Involving Human Subjects Act and was approved by the local ethical committee.

2.1 Study design

We conducted a prospective observational study in a large topclinical hospital in the Netherlands. All patients scheduled for bariatric surgery between June and September 2019 were screened for inclusion. Patients were eligible if they were ≥ 18 years of age, had a body mass index (BMI) of ≥ 35 kg/m^2 and were scheduled for laparoscopic Roux-en-Y gastric bypass surgery or laparoscopic sleeve gastrectomy. Patients with a pacemaker/implantable cardioverter defibrillator, allergy to adhesives, or skin deformities on the left chest were excluded. All patients received a written invitation and a phone call to participate in the study. All participants provided written informed consent. Ethical approval for this study was asked for and waived by the Medical Research Ethics Committee of Commissie Mensgebonden Onderzoek Arnhem-Nijmegen, (registration 2019–5489). The study fell outside the remit of the law for Medical Research Involving Human Subjects Act and was approved by the local ethical committee.

2.2 Study protocol

All patients were equipped with a Biosensor on the upper left side of the chest by a trained nurse on the morning of surgery. The Biosensor was calibrated and connected with a relay station (IntelliVue, Philips, The Netherlands). Having received a Biosensor, patients were seated in a movable operating chair and brought to the operating room, where they were attached to routine monitoring, consisting of a 3-lead ECG, automated non-invasive blood pressure measurement and pulseoximetry (General Electric Carescape B650, GE Healthcare, United States). In the operating room, patients were anaesthetised and their trachea intubated for mechanical ventilation. Surgery was performed with patients in reversed-Trendeleburg position. After surgery, anaesthesia was discontinued and residual neuromuscular blockade was measured and antagonised, if needed. Patients' trachea were subsequently extubated and taken to the recovery room. Discharge from the recovery room to the ward was allowed when the patient had an Aldrete score > 8 and pain, nausea and other side-effects were well managed. On the ward, patients were stimulated to start ambulating as soon as possible. The Biosensor was removed from the patient's chest upon hospital discharge or when the battery was depleted.

3 Materials

The wearable Biosensor (Philips, Amsterdam, the Netherlands) is a disposable self-adhesive patch, which is designed for wirelessly measuring HR, RR, skin temperature, posture and detecting falls in patients. It is an updated version of the VitalConnect HealthPatch, which has been the subject of several prior validation studies [15, 16, 17, 19, 20, 23], with a new algorithm for computing respiratory rate. The new algorithm uses ECG derivatives and tri-axial accelerometry to compute respiratory rate, instead of ECG derivatives and

bio-impedance. The Biosensor is powered by a zinc-air battery, designed to last for 4 days of continuous monitoring. This study focused on the measurements of HR and RR. The Biosensor employs two ECG electrodes to derive QRS complexes from a single-lead ECG, based on which an algorithm calculates HR using R-R intervals [17].

RR is calculated using an algorithm based on three measurements: (1) breathing-induced change in cardiac axis, measured by ECG derived R wave amplitude, (2) chest movement, detected by 3-axis MEMS accelerometer, and (3) respiratory modulations, determined by ECG derived respiratory sinus arrhythmia [25].

The sensor sends these data to a 'relay' using a secured Bluetooth connection. The 'relay' is a device that sends these data to the 'Intellivue Guardian Solution' (IGS) server using a secured Wi-Fi connection. The relays used in this study were phones that were adapted to perform this task only. When the connection between the relay and the IGS is lost, the Biosensor is able to buffer up to ten hours of data. These data are later retrieved automatically when the connection is restored. The IGS is a software analytics program used to collect data from different Philips devices. The reference monitor used in this study is the General Electric Carescape B650, which is used in both high-dependency intensive care units and operating rooms. This monitor uses ECG to determine HR and thoracic impedance pneumography to determine RR. During surgery, the RR is determined by the settings of the ventilator machine (pre-set RR).

3.1 Data collection

The Biosensor took measurements until hospital discharge or battery depletion. The data were retrieved from the IGS and transferred to a research database, which is a secured storage within the hospital's servers that is only accessible for research collaborators. For each measurement, it was registered at what time the measurement was performed by the Biosensor, and at what time it was received by the relay device. Reference monitor measurements and patient characteristics were extracted from the electronic health record and also transferred to the research database. Patients for whom no RR measurement was generated within the first 45 min of the first HR measurement were defined as 'failed calibration', and were excluded from the analyses.

3.2 Outcomes

The primary outcome of this study was the mean absolute difference in HR and RR measurements by the wearable Biosensor in comparison to the reference monitor in the operating room and the recovery room. The mean absolute difference was also calculated for the average HR and RR over 15 min to adjust for outliers (see "Statistical analyses" section).

Secondary outcomes were the percentage of data that fell outside the predefined limits of agreement, as well as the duration and causes of data loss, and delay in transmission between the Biosensor and the relay device until hospital discharge.

3.3 Statistical analysis

The HR and RR measurements from the reference monitor were documented every 30 s. Since the Biosensor measures HR and RR every 4 s, we only used those measurements closest in time to the reference monitors' measurements. We refer to this as the 'original dataset'. In this dataset, two comparisons were possible for each minute of measurement. We used these data to calculate the mean absolute difference for each patient, for both HR and RR.

To minimise the impact of incidental measurement artefacts on the mean absolute difference, we also compared the mean of both measurements over a period of time. We deemed 15 min to be the maximum acceptable time without measurements for monitoring a postoperative patient. Therefore, we calculated the 15-min mean for each reference monitor measurement, and each Biosensor measurement closest in time. As this was repeated every 30 s, there was 14 min and 30 s of overlap between each mean. We will refer to this as the 'averaged' dataset. We excluded measurements generated during the time when patients were moved from the operating room to the recovery room. During this time, the reference monitor was disconnected from the patient, thus no comparison was possible. Note that the data from this period were included for the analyses of data loss and delays in transmission.

Bland Altman plots were created for both individual patients and the overall study population to assess the agreement at different values of HR and RR. The limits of agreement were set to 8 beats per minute (bpm) for HR and 2 breaths per minute (brpm) before the start of the study based on a 10% margin of error. Furthermore, we conducted a Clarke Error Grid analysis to determine the clinical accuracy of the Biosensor compared to the reference monitor, and the implications for treatment decisions. The zone boundaries were based on cut-off values for the Modified Early Warning Score [26], as done in another study [23].

The amount of clinically relevant data loss was calculated in minutes of total measurement duration. We considered missing data to be clinically relevant if more than 15 min of data were missing. Different potential causes of data loss, like movement, sensor detachment and failed internet connection were noted. Delays in data transmission between the Biosensor and the relay device was calculated in minutes for the entire duration of measurement,

including the operating room, the recovery room, and the ward. A power calculation was not feasible in this study due to the lack of preliminary data with the wearable Biosensor. Therefore, this study aimed to include at least 100 patients. All analyses were performed using R (version 3.6.1, www.r-project.org). Figures were created using the R-package 'ggplot2' [27], except for the Clarke Error Grid, which was produced using the R package 'ega' [28].

The statistical analysis plan was written (in Dutch) and filed with the institutional review board before data were accessed.

Table 1 Baseline characteristics of the 94 bariatric patients included for analysis

Baseline characteristics	N = 94
Gender	
Female n (%)	68 (72)
Age years (SD)	44.8 (12.1)
Body mass index kg/m^2 (SD)	42.6 (5.9)
Type of surgery n(%)	
Roux-en-Y Gastric bypass	78 (83)
Sleeve gastrectomy	16 (17)
Comorbidity	
Obstructive sleep apnea n (%)	22 (23)
Diabetes mellitus n (%)	13 (14)
Cardiovascular disease n (%)	9 (10)
Duration of surgery (minutes) mean (SD)	77 (20)
Duration of recovery (minutes) mean (SD)	86 (19)

4 Results

4.1 Participants

Participants were recruited between June 2019 and September 2019. Out of 157 patients undergoing bariatric surgery eligible for participation, who received an information letter at home, and were contacted by phone, 9 patients were excluded for medical reasons, and 26 patients refused to participate. Due to logistical problems and a high workload on the ward, 112 of these remaining 122 patients were provided with a Biosensor. Data from 18 Biosensors could not be used due to failed calibration. Calibration was successful in 94 patients. Mean age was 44.8 with a standard deviation (SD) of 12.1 years, 72% were women, 83% underwent a Roux-en-Y gastric bypass, and 17% a sleeve gastrectomy. Baseline characteristics are shown in Table 1.

4.2 Example of the measurements of a single patient

The measurements of HR and RR of one patient from the start of surgery until the patient leaves the recovery room are shown in Fig. 1. Overall, the HR and RR values of the Biosensor and the reference monitor for this patient seem comparable. The variability of the measurements is higher during the recovery period compared to during surgery. The measurements of RR during surgery show a horizontal line because the RR value is set by the ventilator machine. Bland Altman plots for both the original and the averaged dataset are shown in Fig. 2. The averaged dataset shows a lower mean difference in both HR and RR.

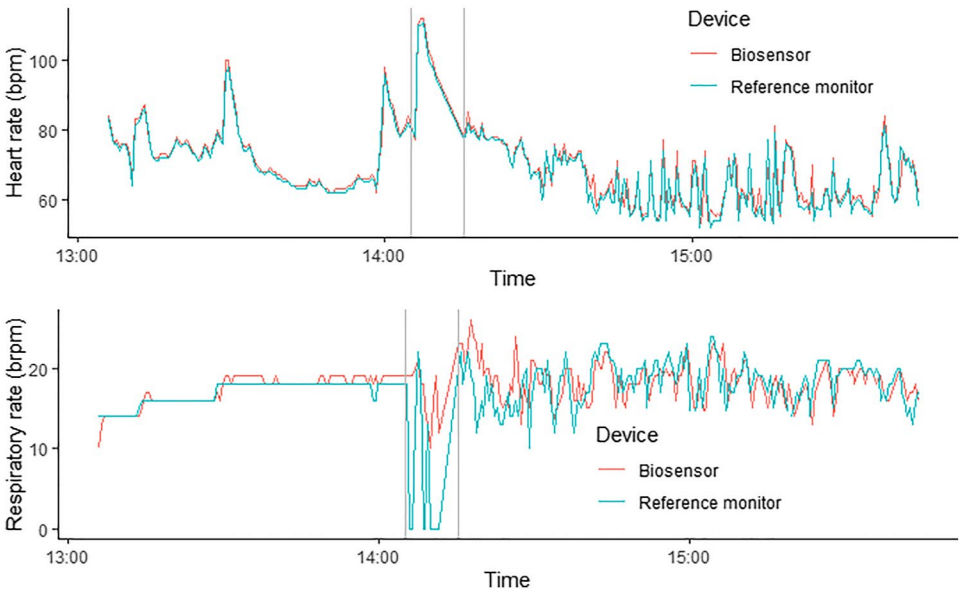

Fig. 1 Measurements of HR (*top*) and RR (*bottom*) from both the Biosensor and the reference monitor during surgery and recovery. The vertical lines indicate the period between surgery and recovery, which was excluded from analysis

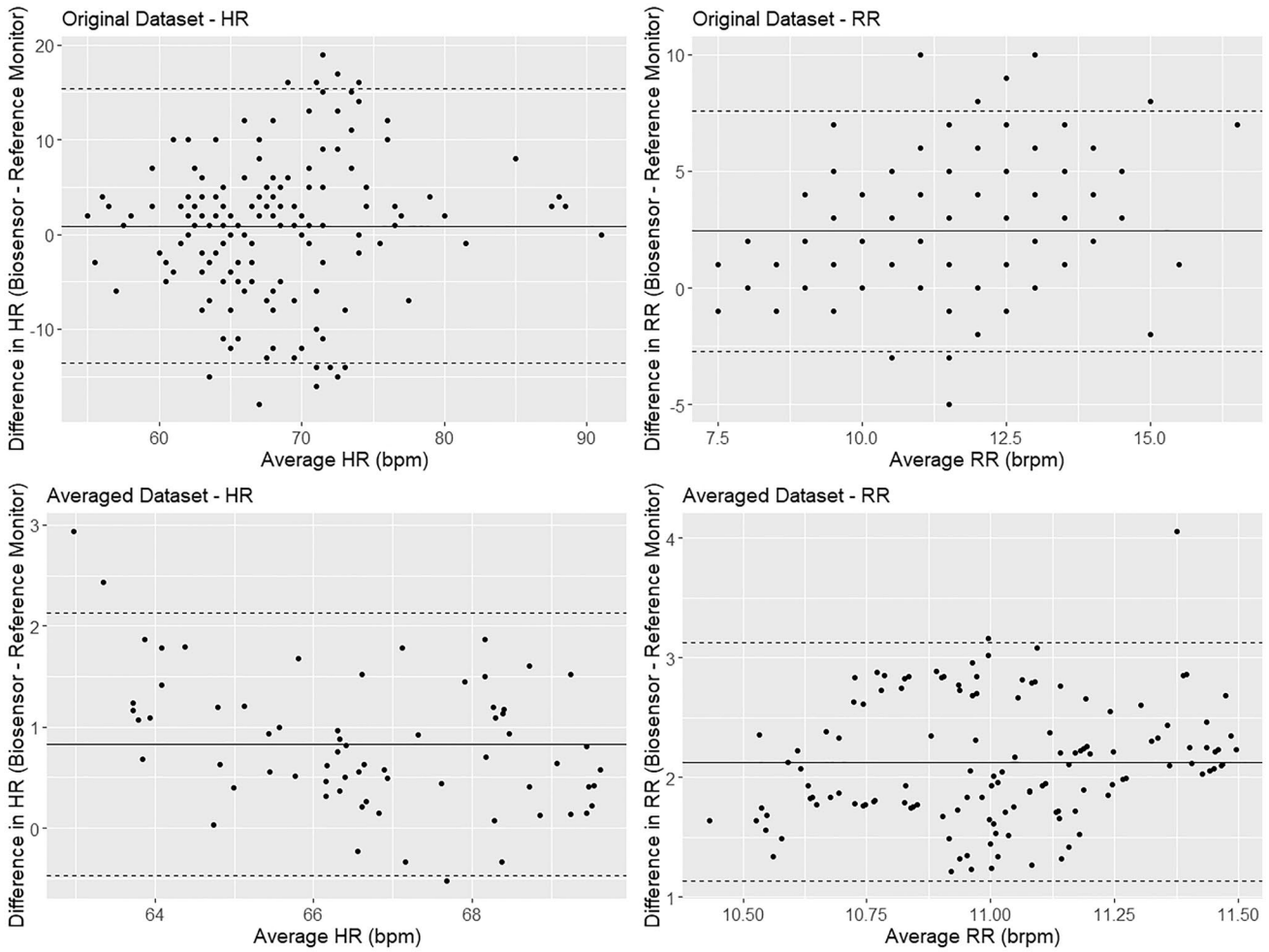

Fig. 2 Bland Altman plots for both the original and averaged dataset in one patient during recovery. The solid line shows the mean difference, the dotted lines show the 95% confidence intervals

4.3 Heart rate

A total of 4954 min of HR measurements during surgery and 7642 min during recovery in 94 patients were compared to reference monitor measurements. The mean absolute difference of HR between the Biosensor and the reference monitor was 1.26 bpm (SD 0.84) during surgery and 1.84 bpm (SD 1.22) during recovery. In the averaged dataset a smaller mean absolute difference of 1.04 bpm (SD 0.46) during surgery and 1.45 bpm (SD 0.54) during recovery (Table 2) were found. All measurements of both the original and the averaged dataset were within the limits of agreement of 8 bpm. No clear association between the average value of the measurement and the difference between the measurements can be seen, as shown in Fig. 3. The Clarke Error Grid analysis (Fig. 4) shows that adequate treatment decisions would have been taken in 99.98% of cases if they had been based on the biosensor (Table 3), for both the original and the averaged datasets. Three measurements (out of 14,143) were located within Region D (Table 3) for the original dataset, indicating that patient safety might have been compromised if those data points had been used in clinical decision making. For the averaged dataset two measurements were located in Region D (Table 3).

4.4 Respiratory rate

A total of 4811 min of respiratory rate measurements during surgery and 7312 min during recovery in 94 patients were compared to reference monitor measurements. The mean absolute difference of respiratory rate between the Biosensor and the reference monitor was 1.78 brpm (SD 1.90) during surgery and 4.24 brpm (SD 2.75) during recovery. Using the averaged dataset resulted in a smaller mean difference for RR of 1.62 brpm (SD 1.73) during surgery and 3.34 brpm (SD 2.52) during recovery (Table 2). During surgery 72% of the mean differences of the original dataset and 76% of the mean differences of the averaged dataset were within the 2

Table 2 Mean differences from heart rate and respiratory rate for both the original and averaged dataset

		Mean absolute difference in beats/minute (SD)	Median absolute difference in beats/minute (minimum–maximum)
Heart rate	*During surgery*		
	Original dataset	1.26 (0.84)	1.00 (0.72–6.40)
	Averaged dataset	1.04 (0.46)	0.91 (0.65–3.46)
	During recovery		
	Original dataset	1.84 (1.22)	1.31 (0.72–6.44)
	Averaged dataset	1.45 (0.54)	1.33 (0.61–4.15)
Respiratory rate	*During surgery*		
	Original dataset	1.78 (1.90)	1.19 (0.07–10.72)
	Averaged dataset	1.62 (1.73)	1.06 (0.07–8.04)
	During recovery		
	Original dataset	4.24 (2.75)	3.52 (0.93–16.00)
	Averaged dataset	3.34 (2.52)	2.48 (0.44–12.16)

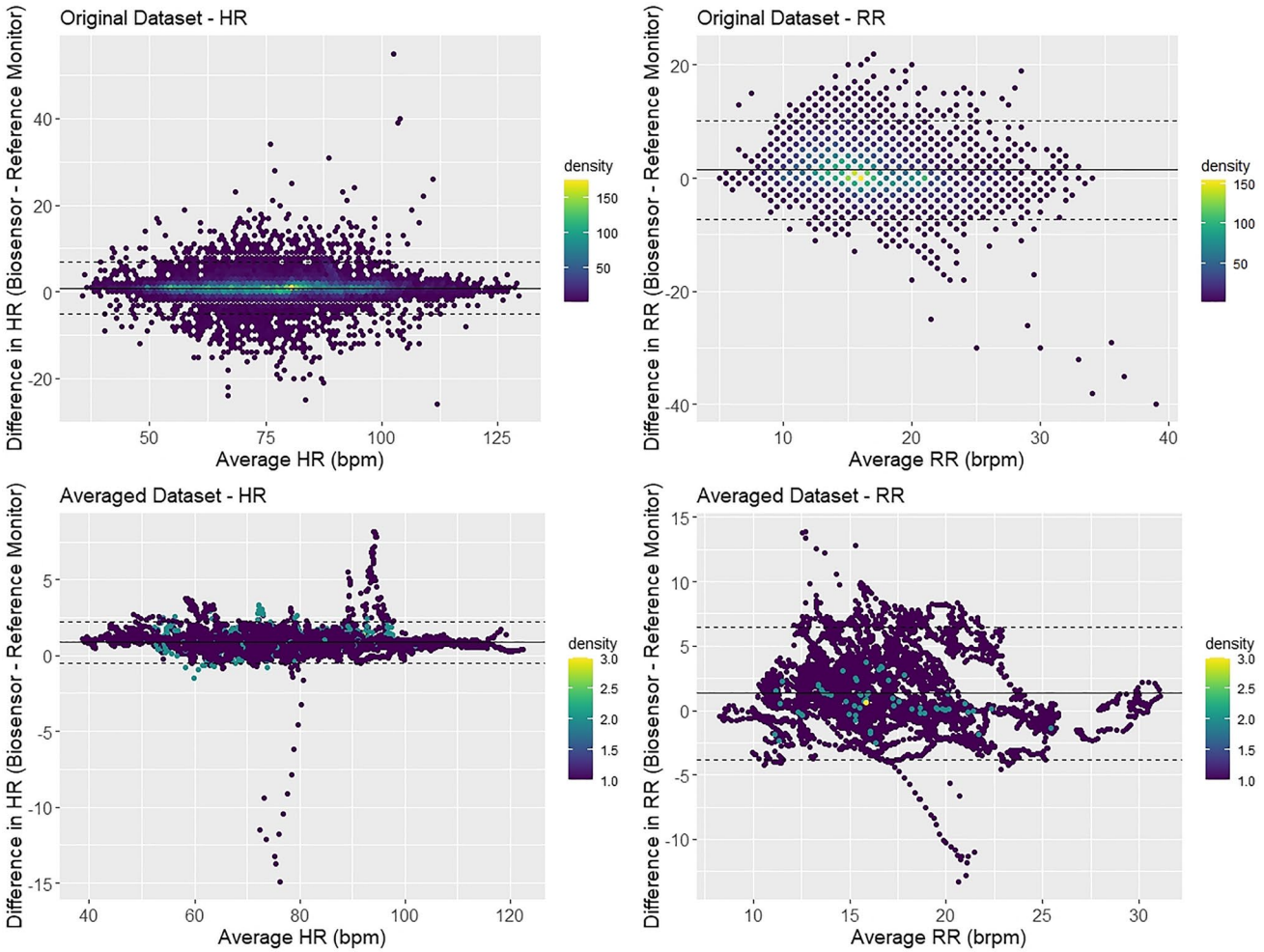

Fig. 3 Bland Altman plots for both the original and averaged dataset in all patients during recovery. The solid line shows the mean difference, the dotted lines show the 95% confidence intervals. Increasing density of points is illustrated by the change of colour from purple to yellow

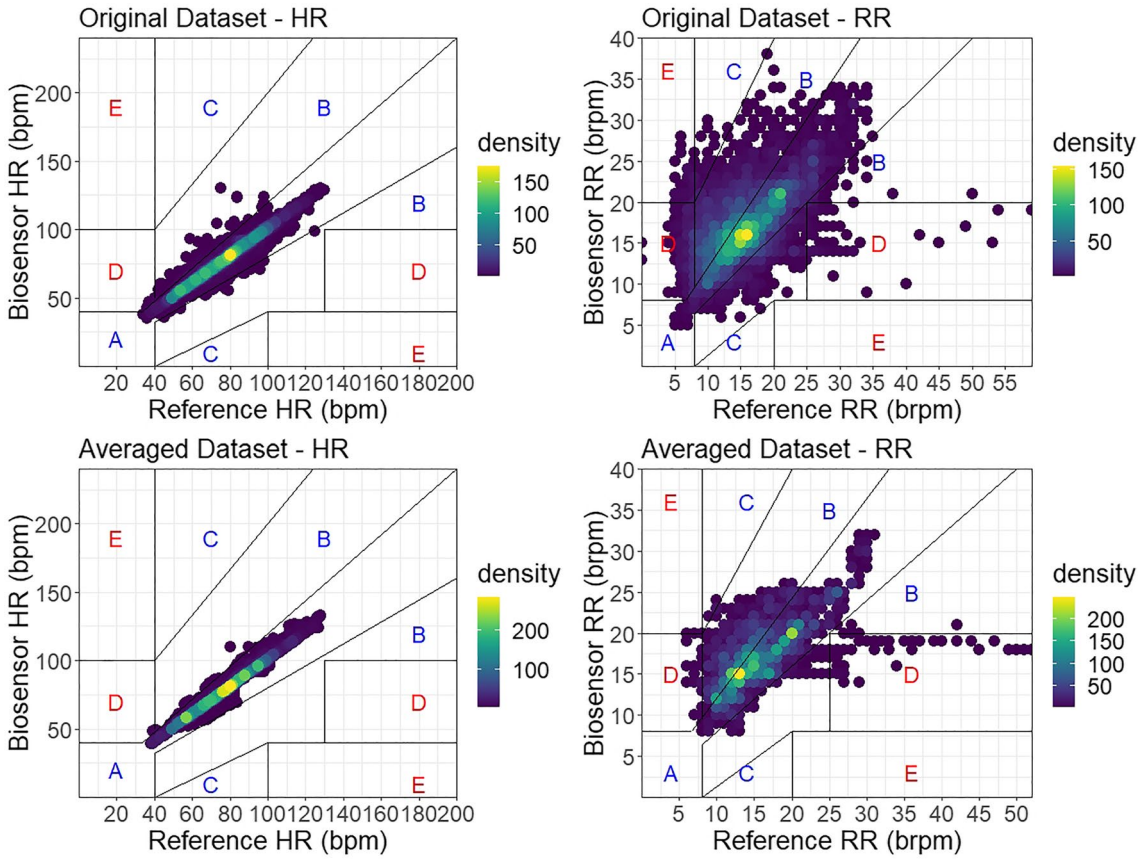

Fig. 4 Clarke Error Grid analysis showing the clinical accuracy of the HR and RR measurements of the Biosensor compared with the reference monitor. Each colored dot represents a measurement pair. The color intensity is proportional to the number of observations. Region A contains points within 20% of the reference monitor; region B encloses points outside 20% of the reference, but not leading to unnecessary treatment; region C is composed of points leading to unnecessary treatment; region D indicates a potentially dangerous failure to detect bradypnoea or tachypnoea, and region E represents points where events are confused (e.g., bradypnoea with tachypnoea)

Table 3 Results of the Clarke Error Grid analysis, showing the number of measurements in Regions A through E for HR and RR, for both the original and the averaged datasets

	HR		RR	
Region	Original	Averaged	Original	Averaged
A	14,062 (99.4%)	13,873 (99.9%)	4999 (62.4%)	5353 (68.3%)
B	78 (0.6%)	8 (0.1%)	2673 (33.4%)	2414 (30.8%)
C	–	–	25 (0.3%)	6 (0.1%)
D	3 (0.0%)	2 (0.0%)	236 (2.9%)	62 (0.8%)
E	–	–	74 (0.9%)	5 (0.1%)
Total	14,143	13,883	8007	7840

brpm limit of agreement. During recovery the numbers for the original and the averaged dataset were 24% and 40% respectively. Figure 3 shows that there is no clear correlation between the absolute value of the measurement and the difference between the measurements. Clarke Error Grid analysis (Fig. 4) showed that using the RR measurements from the biosensor would have resulted in adequate treatment decisions in 95.8% of cases for the original dataset, and in 99.1% of cases for the averaged dataset. Patient safety might have been compromised in 4.2% of cases if the original dataset had been used to take clinical decisions (Table 3). Using the averaged dataset, this would have been true for 0.9% of measurements (Table 3).

4.5 Missing data

From hospital admission to hospital discharge, the Biosensors recorded a total of 176,063 min of HR data. The Biosensors showed a total of 2200 min of clinically relevant data loss for HR (1.3% of the total measurement duration). For RR, 172,613 min of data were recorded, of which 7730 min were lost (4.5% of the total). The proportion of data loss was greater on the ward than during surgery and recovery for both HR and RR. The data loss was caused by 10 different Biosensors. Table 4 provides a comprehensive

Table 4 Duration and frequency of clinically relevant data loss in minutes of the 10 Biosensors in which it occurred

	HR	RR
Total measurement duration	176,063	172,613
Total data loss	2200 (1.3%)	7730 (4.5%)
Measurement duration surgery/recovery	14,644	14,136
Data loss surgery/recovery	47 (0.32%)	222 (1.6%)
Measurement duration ward	161,356	158,477
Data loss ward	2153 (1.3%)	7508 (4.7%)
Data loss frequency (Median [Min–Max])	2 [1–10]	1 [1–10]
Data loss duration (Median [Min–Max])	26 [16–1436]	49 [16–1592]

overview of the clinically relevant data loss that occurred. Among the patients who received a biosensor, there were eight instances of Biosensors detaching from the patient during the measurement period. No clear association between movement, defined as the change in posture, and data loss could be found. Loss of internet connection, interference of other monitoring devices, cardiac arrhythmias and supine position of the patient were investigated but also provided no clear explanation for data loss.

Delays in the transmission from the Biosensor to the relay of HR data occurred 1322 times over 88 patients, for a total duration of 74,660 min of data (42.4% of all measurements). In 7.4% the delay was longer than 15 min, which occurred in 39 patients and accounted for 3758 min of data (2.1% of measurements). RR data was delayed 1549 times over 87 patients, accounting for 143,609 min of data (83.2% of measurements). In 71.5% of delays, the duration exceeded 15 min, which occurred in 87 patients and accounted for 141,221 min of data (81.8% of all measurements).

5 Discussion

In this study we compared the HR and RR measurements of a wearable biosensor to a reference monitor in 94 morbidly obese patients in a perioperative setting. The agreement of HR was within the predetermined limits for clinical use. However, RR showed significantly less agreement and exceeded these limits. The amount of data loss was low, but the cause of the data loss remains unclear.

All HR measurements by the wearable biosensor fell within the limits of agreement of ± 8 bpm compared to the reference monitor, although it did overestimate HR somewhat overall. This is consistent with findings of previous studies [16, 22, 23]. Agreement of RR measurements were within predefined limits of agreement of ± 2 brpm 72% and 76% of the time during surgery for the original and averaged datasets, respectively. Previous studies, using limits of agreement of 3 brpm, showed similar agreement [21, 23]. However, in the recovery room, agreement was reduced to 24% and 40%. Like HR, RR was also consistently overestimated compared to reference monitor measurements. This deviation can partly be explained by the limitations of impedance pneumography, used by the reference monitor. However, a previous study showed that the mean difference of the Biosensor compared to capnography was 3.5 breaths/minute (± 5.2 breaths/minute) on average [21]. This suggests that the deviation of the RR measurements of the Biosensor cannot be explained by the limitations of impedance pneumography alone. The results from the pre-set RR during surgery, which showed a smaller mean absolute difference, supports these findings. The remaining difference in agreement of RR measurements in comparison to the reference monitor, is most likely caused by the Biosensor's measurement method for RR. The Biosensor uses a combination of three derivatives to measure RR: R wave amplitude that measures change of cardiac axis during breathing, 3-axis MEMS accelerometer detecting chest movement and ECG derived respiratory sinus arrhythmia that measures respiratory modulations. It is possible that the calculation algorithm does not sufficiently correct for imprecision of measurements.

Despite the fact that a large percentage of the Biosensor's RR measurements fell outside of the limits of agreement, Clarke Error Grid analysis showed that the effect on treatment decisions would be greatly reduced compared with its predecessor, the VitalConnect HealthPatch. RR measurements taken by the Biosensor might have compromised patient safety for 4.2% of measurements using the original data, and for 1% of measurements using the averaged dataset, while this was 22.6% for the HealthPatch [23]. This finding the difficulty in validating devices where software updates can lead to major changes in device performance.

Data loss was 1.3% for HR and 4.5% for RR over the total measurement duration. This amount of data loss is lower than that reported in previous studies, which reported 6% to 13.1% data loss [16, 23]. We did not find an explanation for the data loss that occurred. No association between patient movement, internet connection, interference with other monitoring devices, cardiac arrhythmias and supine position of the patient and data loss was apparent. It is possible that data loss of both HR and RR was caused by minor disconnections of the ECG electrodes from the patients' skin. To generate a RR measurement, the Biosensor uses derivatives from the ECG signal and acquires calibration after an adequate HR signal is generated. However, the HR measurement only requires an adequate connection of the ECG electrodes to the patient's skin. In other words, an RR measurement can only be generated with an HR measurement. This is demonstrated by the fact that data loss only occurred for RR alone or HR and RR at the same time. The reason for the data

loss of RR alone remains unclear but could be explained by failed recalibration of the Biosensor, which is automatically initiated by the Biosensor when the patient shows a certain amount of movement.

In addition to the data loss that occurred, some data was retrieved later in time. This delay was mostly caused by the loss of connection between the Biosensor and the relay device. The Biosensor is able to buffer measurements up to 10 h when connection is lost, and the missing data was received later. While delays longer than 15 min accounted for only 2.1% of data for HR, 81.8% of RR data was delayed for more than 15 min. This makes the biosensor unsuitable for situations wherein the maximum permissible interval between measurements is 15 min. It is unclear why there is such a discrepancy between the delays in HR and RR measurements.

Investigating an obese population provides new information on the agreement of the measurements of wearable devices in patients with a large quantity of subcutaneous fat. Another strength of this study is the large number of patients which provides information about the clinical applicability of wearable sensors. By validating the RR for both a fixed RR (during surgery) and a variable RR (after surgery), the sensor is compared with both a pre-set RR and impedance pneumography. Because impedance pneumography itself has its limitations, the pre-set RR provides additional information about the agreement of the RR measurements.

In contrast to the conventional bedside monitors, the Biosensor allows patients to increase their physical activity beyond the limited space of the bedside, which itself might positively influence the patient's recovery. However, since the measurements used for validation took place during and right after surgery, patients were under sedation or still recovering from sedation and therefore patient movement was limited. Although this provided a controlled setting to measure vital signs, it might not be representative for situations in which patients have increased physical activity. An earlier study found that increased physical activity might decrease the agreement of HR and RR measurements from the Biosensor [21].

The results of the current study should be interpreted in the context of some limitations. For one, we only included morbidly obese patients, meaning that generalizability to non-obese populations may be limited. Furthermore, we used impedance pneumography as a reference standard, rather than the gold standard of capnography. We chose to do so because although capnography is the gold standard, snoring–which is highly prevalent among (morbidly) obese populations–is known to cause measurement artefacts, and because impedance pneumography is often used in practice, despite being less accurate than capnography. Moreover, we were not able to validate the Biosensor's measurements in the general ward, while ambulating, or at home. It should be noted that validating any sensor in these settings in a meaningful way is difficult, since patients are typically not monitored in these settings. As pointed out in a recent editorial, guidelines or at least a consensus statement on how to best conduct validation studies in these settings may be helpful [13]. Finally, we did not compute the sensitivity and specificity of the biosensor in terms of its ability to detect tachycardia, bradycardia, tachypnea, and bradypnea. Our study was not designed for that purpose, and as a consequence it was not powered to estimate sensitivity and specificity. As such we did not have enough data in the low and high ranges to provide a reliable estimate.

While the Biosensor has the benefits of wireless connectivity, limited data loss, and reliable HR measurement, it does not reliably measure RR in obese patients with limited movement. Yet, it may be possible to use the data generated by the Biosensor to predict patient deterioration, in addition to detecting when HR and RR measurements exceed pre-set limits of agreement. Recent studies have shown promising results in terms of predicting adverse events using alarms that were personalised, based on time trends, or based on risk spikes [29, 30]. Such approaches may be feasible even if a sensor overestimates the value of a vital sign, as long as it does so consistently, as the true value is less important using these approaches. Moreover, it may be that HR alone is enough to indicate deterioration, though predictions based on more parameters may be more accurate. Future studies should focus on improving RR measurements by wireless wearable sensors, the interpretation of continuous data, and development of early alarm systems which can be applied in clinical practice. Further research is needed to determine whether the Biosensor can be used to detect deterioration based on HR and RR, despite the limited agreement of RR with the reference monitors used in this study.

Acknowledgements NK and GMP contributed equally to this work. The authors would like to thank the following persons from Rijnstate Hospital: José W.J.M. Geurts PhD and Pascal S.H. Smulders MSc, Department of Anesthesiology and Pain Management, for logistical support, Sjoerd J.A. Boogaard and Heleen M. Schoorl, Department of Information Technology, for technical support, Marieke J. Bosch, and Laura N. Deden, Vitalys Obesity Centre for the recruitment of patients. Finally, we would like to thank all the nurses from the Department of Bariatric Surgery, and patients for participating in this study.

Funding Biosensors were provided free of charge by Philips. Other than that, support was provided solely from hospital and university sources.

Availability of data and material Data is available upon reasonable request

Code availability We used the statistical software R (version 3.6.1, www.r-project.org). Figures were created using the R-package 'ggplot2'.

Declarations

Conflict of interest Biosensors were provided free of charge by Philips. Other than that, support was provided solely from hospital and university sources. The authors declare no competing interests.

Ethical approval
Ethical approval for this study was asked for and waived by the Medical Research Ethics Committee Arnhem-Nijmegen, (registration 2019–5489). The study fell outside the remit of the law for Medical Research Involving Human Subjects Act and was approved by the local ethical committee.

Consent to participate Written informed consent was obtained from all participants.

Consent for publication Written informed consent for publication of their data to be used for scientific publication was obtained from all participants.

References

1. Tevis SE, Kennedy GE. Postoperative complications and implications on patient-centered outcomes. J Surg Res. 2013;181:106–13. https://doi.org/10.1016/j.jss.2013.01.032.
2. Beth Smith ME, Chiovaro JC, O'Neil M, et al. Early warning system scores for clinical deterioration in hospitalized patients: a systematic review. Ann Am Thorac Soc. 2014;11:1454–65. https://doi.org/10.1513/AnnalsATS.201403-102OC.
3. Alam N, Hobbelink EL, van Tienhoven AJ, Van de Ven PM, Jansma EP, Nanayakkara PWB. The impact of the use of the Early Warning Score (EWS) on patient outcomes: a systematic review. Resuscitation. 2014;85:587–94. https://doi.org/10.1016/j.resuscitation.2014.01.013.
4. Cretikos MA, Bellomo R, Hillman K, Chen J, Finfer S, Flabouris A. Respiratory rate: the neglected vital sign. Med J Aust. 2008;188:657–9. https://doi.org/10.5694/j.1326-5377.2008.tb01825.x.
5. Subbe CP, Davies RG, Williams E, Rutherford P, Gemmell L. Effect of introducing the Modified Early Warning score on clinical outcomes, cardio-pulmonary arrests and intensive care utilisation in acute medical admissions. Anaesthesia. 2003;58:797–802. https://doi.org/10.1046/j.1365-2044.2003.03258.x.
6. Fieselmann JF, Hendryx MS, Helms CM, Wakefield DS. Respiratory rate predicts cardiopulmonary arrest for internal medicine inpatients. J Gen Intern Med. 1993;8:354–60. https://doi.org/10.1007/BF02600071.
7. Young MP, Gooder VJ, McBride K, James B, Fisher ES. Inpatient transfers to the intensive care unit: delays are associated with increased mortality and morbidity. J Gen Intern Med. 2003;18:77–83. https://doi.org/10.1046/j.1525-1497.2003.20441.x.
8. Pashikanti L, Von Ah D. Impact of early mobilization protocol on the medical-surgical inpatient population: an integrated review of literature. Clin Nurse Spec. 2012;26:87–94. https://doi.org/10.1097/NUR.0b013e31824590e6.
9. Markey DW, Brown RJ. An interdisciplinary approach to addressing patient activity and mobility in the medical-surgical patient. J Nurs Care Qual. 2002;16:1–12. https://doi.org/10.1097/00001786-200207000-00002.
10. Joshi M, Ashrafian H, Aufegger L, Khan S, Arora S, Cooke G, Darzi A. Wearable sensors to improve detection of patient deterioration. Expert Rev Med Dev. 2019;16:145–54. https://doi.org/10.1080/17434440.2019.1563480.
11. Boer C, Touw HR, Loer SA. Postanesthesia care by remote monitoring of vital signs in surgical wards. Curr Opin Anaesthesiol. 2018;31:716–22. https://doi.org/10.1097/ACO.0000000000000650.
12. Watkins T, Whisman L, Booker P. Nursing assessment of continuous vital sign surveillance to improve patient safety on the medical/surgical unit. J Clin Nurs. 2016;25:278–81. https://doi.org/10.1111/jocn.13102.
13. Saugel B, Hoppe P, Khanna AK. Automated continuous noninvasive ward monitoring: validation of measurement systems is the real challenge. Anesthesiology. 2020;132:407–10. https://doi.org/10.1097/ALN.0000000000003100.
14. Michard F, Kalkman CJ. Rethinking patient surveillance on hospital wards. Anesthesiology. 2021;135:531–40. https://doi.org/10.1097/ALN.0000000000003843.
15. Izmailova ES, McLean IL, Bhatia G, et al. Evaluation of wearable digital devices in a phase I clinical trial. Clin Transl Sci. 2019;12:247–56. https://doi.org/10.1111/cts.12602.
16. Breteler MJMM, Huizinga E, van Loon K, Leenen LPH, Dohmen DAJ, Kalkman CJ, Blokhuis TJ. Reliability of wireless monitoring using a wearable patch sensor in high-risk surgical patients at a step-down unit in the Netherlands: a clinical validation study. BMJ Open. 2018;8:e020162. https://doi.org/10.1136/bmjopen-2017-020162.
17. Selvaraj N, Nallathambi G, Moghadam R, Aga A. Fully disposable wireless patch sensor for continuous remote patient monitoring. Conf Proc IEEE Eng Med Biol Soc. 2018. https://doi.org/10.1109/EMBC.2018.8512569.
18. Koenders N, Seeger JPH, Van Der Giessen T, et al. Validation of a wireless patch sensor to monitor mobility tested in both an experimental and a hospital setup: a cross-sectional study. PLoS ONE. 2018;13:e0206304. https://doi.org/10.1371/journal.pone.0206304.
19. Weenk M, Koeneman M, van de Belt TH, Engelen LJLPG, Van Goor H, Bredie SJH. Wireless and continuous monitoring of vital signs in patients at the general ward. Resuscitation. 2019;136:47–53. https://doi.org/10.1016/j.resuscitation.2019.01.017.
20. Weenk M, van Goor H, Frietman B, et al. Continuous Monitoring of Vital Signs Using Wearable Devices on the General Ward: Pilot Study. JMIR mHealth uHealth. 2017;5:91. https://doi.org/10.2196/mhealth.7208.
21. Li T, Divatia S, McKittrick J, Moss J, Hijnen NM, Becker LB. A pilot study of respiratory rate derived from a wearable biosensor compared with capnography in emergency department patients. Open Access Emerg Med. 2019;11:103–8. https://doi.org/10.2147/OAEM.S198842.
22. Selvaraj N, Nallathambi G, Kettle P. A novel synthetic simulation platform for validation of breathing rate measurement. Conf Proc IEEE Eng Med Biol Soc. 2018. https://doi.org/10.1109/EMBC.2018.8512352.
23. Breteler MJM, KleinJan EJ, Dohmen DAJ, et al. Vital signs monitoring with wearable sensors in high-risk surgical patients: a clinical validation study. Anesthesiology. 2019;132:424–39. https://doi.org/10.1097/ALN.0000000000003029.
24. Finucane MM, Stevens GA, Cowan MJ, et al. National, regional, and global trends in body-mass index since 1980: systematic analysis of health examination surveys and epidemiological studies with 960 country-years and 9.1 million participants. Lancet. 2011;377:557–67. https://doi.org/10.1016/S0140-6736(10)62037-5.
25. Chan AM, Ferdosi N, Narasimhan R. Ambulatory respiratory rate detection using ECG and a triaxial accelerometer. Annu Int Conf IEEE Eng Med Biol Soc. 2013. https://doi.org/10.1109/EMBC.2013.6610436.

26. Subbe CP, Kruger M, Rutherford P, Gemmel L. Validation of a modified Early Warning Score in medical admissions. QJM. 2001;94:521–6.
27. Wickham H. ggplot2: elegant graphics for data analysis. New York: Springer; 2016.
28. Schmolze D (2017). ega: Error Grid Analysis. R package version 2.0.0. https://CRAN.R-project.org/package=ega
29. Keim-Malpass J, Clark MT, Lake DE, Moorman JR. Towards development of alert thresholds for clinical deterioration using continuous predictive analytics monitoring. J Clin Monit Comput. 2020;34(4):797–804. https://doi.org/10.1007/s10877-019-00361-5.
30. Van Rossum MC, Vlaskamp LB, Posthuma LM, Visscher MJ, Breteler MJM, Hermens HJ, Kalkman CJ, Preckel B. Adaptive threshold-based alarm strategies for continuous vital signs monitoring. J Clin Monit Comput. 2021. https://doi.org/10.1007/s10877-021-00666-4.