The completeness, accuracy and impact on alerts, of wearable vital signs monitoring in hospitalised patients

Anthony Joseph Wilson
anthony.wilson@mft.nhs.uk
The University of Manchester
Alexander J Parker
Manchester University NHS Foundation Trust
Gareth B Kitchen
The University of Manchester
Andrew Martin
Manchester University NHS Foundation Trust
Lukas Hughes-Noehrer
The University of Manchester
Mahesh Nirmalan
Manchester University NHS Foundation Trust
Niels Peek
University of Cambridge
Glen Martin
The University of Manchester
Fiona Thistlethwaite
The Christie NHS Foundation Trust

Research Article

Keywords: Wearable sensors, Vital signs monitoring, Early warning scoring systems, Covid-19, SARS-CoV-2

Posted Date: October 17th, 2024

DOI: https://doi.org/10.21203/rs.3.rs-4976766/v1

Additional Declarations: Competing interest reported. Prof Thistlethwaite declares that she is the director of the Innovate Manchester Advanced Therapy Centre Hub, Innovate UK, which provided funding for this work. All other authors have no competing interests to declare.

Abstract

Background

Use of wearable vital signs sensors to monitor hospitalised patients is growing but uncertainty exists about completeness of data capture and accuracy of measurements. Implications for track and trigger systems are unclear.

Methods

In this observational study, adult inpatients with Covid-19 wore four wearable sensors recording heart rate/respiratory rate (HR/RR), oxygen saturation (SpO_2), axillary temperature and blood pressure (BP). Wearable vitals were paired with traditional vitals recorded concurrently. The accuracy of the wearable vitals was assessed using traditional vitals as the reference. National early warning (NEWS2) scores were calculated using wearable and traditional vitals.

Results

48 patients were monitored for 204 days with the sensors. Median sensor wear was 3.9(IQR:1.7-5.9), 3.9(IQR:1.6-5.9) and 3.8(IQR:0.9-5.9) days for HR/RR, temperature and SpO_2 respectively. The BP cuff was worn for median 1.9(IQR:0.9-3.8) days in 33 patients. Length of hospital stay was 8(IQR:6-13) days. Completeness of data capture was 84% for HR/RR, 98% for temperature, 72% for SpO_2 and 36% for BP.

There were 1632 HR, 1613 RR, 1411 temperature, 1294 SpO_2 and 51 BP wearable-traditional measurement pairs. 59.7% of HR pairs were within ±5bpm, 38.5% of RR pairs within ±3breaths/min, 24.4% of temperature pairs within ±0.3^oC, 32.9% of SpO_2 pairs within ±2% and 39.0% of BP pairs within ±10mmHg. Agreement between wearable and traditional RRs was poor at high RRs.

613 NEWS2 scores were calculated using wearable-traditional HR, RR, temperature and SpO_2 pairs. The median $NEWS2_{traditional}$ was 1(IQR:1-2) and the median $NEWS2_{wearable}$ was 4(IQR:3-6). Using traditional NEWS2 alerts as a reference, 86% (225/262) of wearable NEWS2 5+ alerts and 89% (82/92) of wearable NEWS2 7+ alerts were false positives.

Conclusions

Agreement between vital signs recorded by wearable sensors and concurrent traditional vitals is poor. Data from wearable sensors should not be used in existing track and trigger systems.

Background

Wearable vital signs sensors (WVSSs) are wireless, non-invasive devices worn by patients which permit near-continuous recording of heart rate (HR), respiratory rate (RR), oxygen saturations (SpO_2) and blood pressure (BP). In low resource settings and healthcare economies struggling with staff shortages, such sensors could automate repetitive manual vital signs measurements thereby freeing time to care. The near-continuous recording theoretically allows no deterioration to go unnoticed and permits advanced analytics to predict and detect important patient centred outcomes and potentially to personalise care. Patients and their relatives may feel a sense of security in knowing that they are always monitored. With these benefits in mind, there has been a growing interest[1] in the use of WVSSs to monitor patients in hospital but there remains uncertainty about the accuracy of the vital signs they record[2, 3]. The evidence that they have a positive impact on outcomes for hospitalised patients is not established[4].

Numerous studies have assessed the accuracy of vitals recorded by WVSSs but many have focussed on short timescales or have compared WVSSs to critical care standards of monitoring, sometimes in optimised environments[3]. A true picture of their performance requires assessment in hospital wards which are subject to the challenges and limitations of such environments.

Most hospitals use some form of track and trigger system to detect and respond to the deteriorating patient[5]. Yet clinical staff do not always follow track and trigger protocols[6] and WVSSs which could automate early warning score (EWS) completion could be valuable. However, studies which assess the impact of incorporating vital signs data from WVSSs into such systems are uncommon[7]. Instead, many authors have suggested developing new approaches to the deteriorating patient which can accommodate data from WVSSs. These have included incorporating an additional alerting step to prompt manual calculation of an EWS[8], adjusting EWS systems to reflect wearable data[9, 10] or the development of novel deterioration indices utilising wearable data[11]. A benefit of the EWS system is its simplicity and widespread adoption. New approaches add complexity. It is therefore prudent to assess what would happen if vital signs from WVSSs were simply used in existing EWS systems.

In this context, we conducted the COSMIC-19 (continuous signs monitoring in Covid-19) study which evaluated a suite of WVSSs in hospitalised patients with Covid-19 in a real-world, ward environment. Our aim was to determine the completeness and accuracy of the data recorded by the WVSSs in comparison to traditional ward vital signs. We also sought to assess how National Early Warning Score 2 (NEWS2) scores[12] would differ if calculated using vitals obtained by WVSSs instead of concurrent traditional vitals measurements.

Methods

The COSMIC-19 study was registered with clinicaltrials.gov (registration: NCT04581031). It was approved by Yorkshire and the Humber – Bradford Leeds Research Ethics Committee, UK (reference 20/YH/0156). It was funded by the Innovate Manchester Advanced Therapy Centre Hub, Innovate UK, (project ID: 6239) with additional support from the Christie Hospital Charitable Fund.

Symptomatic, adult inpatients (16 years or older) with suspected or confirmed Covid-19, who were suitable for escalation to critical care and receiving supplemental oxygen when screened were eligible for enrolment. Patients were recruited within 72 hours of hospital admission. We excluded patients unable to give informed consent, who were anticipated to die within 24 hours or those with a contraindication to wearing the WVSSs. Patients were recruited at Manchester Royal Infirmary, an academic, tertiary, metropolitan hospital in the UK with over 150,000 emergency department presentations annually.

Wearable vital signs sensors

We investigated four WVSSs which measured heart rate/respiratory rate (HR/RR), oxygen saturation (SpO_2), axillary temperature and systolic blood pressure (SBP). The four sensors were purchased from Isansys Ltd[13] and were used as part of the Isansys Patient Status Engine (PSE), a digital platform for vital signs capture and reporting which is a regulated medical device. Isansys Ltd had no input in the design or conduct of the research. The wearables are summarised in Table 1.

Participants wore the four WVSSs for up to 20 days or until they lost capacity to consent to continue or were discharged, whichever was earlier. They were free to discontinue any sensor at any time. Participants received usual care during monitoring with the WVSSs, including traditional vital signs measurements by ward staff. This included temperature

measurements using a tympanic thermometer. The clinical team and participants were blinded to the WVSS measurements.

The research team monitored the participants for WVSS disconnection via a remote dashboard. Each participant's data was reviewed at least once every 24 hours. In the event of a disconnection, the researchers visited the participant to aid with wearable re-application or removal if they chose to discontinue a sensor. The research team maintained a log of all occasions when participants removed the sensors and their reasons for doing so.

Data capture

We recruited a convenience sample of 30–60 participants, informed by the number of WVSSs available for the research. No formal sample size calculations were performed.

An electronic case report form (eCRF, DataTrial Ltd, Newcastle, UK) was used to capture demographics, diagnostic information, sensor application/removal and outcomes. Traditional vital signs were downloaded from the hospital electronic track and trigger system (Patientrack, Alcidion Ltd, Aus). Traditional vitals were filtered to exclude non-physiological values. Wearable vital signs were captured via the Isansys PSE. Error codes were removed from the wearable data but no vital signs measurements were excluded.

Data analysis

The completeness of vital sign capture was reported as the percentage of time for which a vital sign was recorded by a WVSS as a proportion of the total time for which the sensor was worn. The number of participants who removed each sensor prematurely and their reasons for doing was summarised. A Kaplan Meier analysis was conducted where the event of interest was a temporary gap in wearable sensor data of varying durations. Participants were censored if they permanently removed the WVSS for any reason.

For each WVSS, the accuracy of the vital signs was assessed using traditional vital signs as a reference standard. For each traditional measurement, a corresponding wearable vital signs measurement was determined by taking the median of all wearable measurements within the preceding five minutes (for HR, RR, temperature and SpO_2) or 15 minutes (for SBP). These timescales and the approach were chosen to align with existing published work[7, 14–19]. A Bland Altman analysis[20] was performed to determine the bias and 95% limits of agreement (LoA) for each vital. The repeated nature of measurements was accounted for using the methods described by Zou et al.[21] In keeping with previous literature[3], we defined ± 5bpm, ±3breaths/min, ± 2%, ± 0.3^{o}C and ± 10mmHg a priori as clinically acceptable agreement between HR, RR, SpO_2, temperature and SBP respectively.

To assess how monitoring patients with WVSS would impact clinical alerts, a partial National Early Warning Score[12] (NEWS2) was calculated using wearable sensor measurements and compared to the NEWS2 calculated from traditional vitals. Both calculations did not include level of consciousness and air/oxygen scores. We assumed all participants were on SpO_2 scale 1. A modified Clarke Error Grid analysis[22] was also performed for each individual NEWS2 component. This quantifies how differences between corresponding wearable and traditional measurements would impact the NEWS2 component. Finally, to compare trends between successive wearable and traditional measurements, 4Q plots[23] were created for each vital sign.

All analysis was conducted using R (version 3.6)[24]. Continuous variables were assessed for normality and are presented as mean (standard deviation) or median [interquartile range]. Categorical variables are presented as the population size and the percentage (of available data) for each class.

Table 1
a summary of the wearable health monitor devices deployed in this study. BLE - Bluetooth low energy, BP = blood pressure, ECG = electrocardiogram, HR = heart rate, HRV = heart rate variability, pleth = plethysmography, PSE = patient status engine, RR = respiratory rate, SpO_2 = haemoglobin oxygen saturation.

Feature	Wearable Vital Signs Sensor			
	HR/RR Isansys Lifetouch Sensor[13]	Temperature Isansys Lifetemp Sensor[13]	SpO_2 Nonin3150 WristOx[25]	BP A&D TM2441 BP Monitor[26]
Form Factor	A small patch attached via standard ECG electrodes	A small adhesive patch	Wrist-worn module, pulse oximetry finger probe	Cuff with attached servo unit.
CE marked medical device	Yes	Yes	Yes	Yes
Wear Location	Precordium	Axilla	Wrist and fingertip	Arm
Frequency of vital sign measurement	Every minute	Every minute	Every minute	Hourly (6am-10pm), every two hours (10pm-6am)
Battery Life	72 hours.	5 days.	48 hours. 2x AAA batteries.	2x AA batteries.
Single Use?	Yes	Yes	No	No
Data Synchronisation	BLE to Samsung Galaxy Tablet. Subsequent data upload via WiFi/cellular from tablet to remote server (Isansys LifeGuard Server)			
Software Used	Isansys PSE software.			
Metrics recorded	Accelerometer (activity and posture), ECG (single lead), HR, HRV, RR.	Temperature (skin - axilla)	HR (pleth) SpO_2 (finger)	Non-invasive BP

Results

Study participants

Figure 1 summarises screening and recruitment. 179 eligible patients were identified and 48 took part. Most (43/48) were recruited between July 2020 and March 2021, during the first and second UK waves of the coronavirus pandemic. Five were recruited between July 2021 and February 2022, during the delta and omicron waves in the UK[27]. In 47 participants, SARS-CoV-2 infection was confirmed on nasopharyngeal swab (lateral flow or polymerase chain reaction), one participant had symptoms consistent with Covid-19 and high clinical suspicion of infection but without a positive screening result.

Table 2 summarises the demographics and clinical characteristics of study participants, stratified according to whether they were admitted to critical care. 32/48 (66.7%) were male and 32/48 (66.7%) were from non-Caucasian ethnicities.

	Overall	Not admitted to ICU	Admitted to ICU
	N=48	n=37	n=11
Male sex	32 (66.7)	25 (67.6)	7 (63.6)
Age	51.0 [40, 58]	51 [36, 60]	48 [42, 56]
Ethnicity			
Arabic	3 (6.2)	1 (2.7)	2 (18.2)
Indian/Pakistani	20 (41.7)	12 (32.4)	8 (72.7)
Black	8 (16.7)	8 (21.6)	0 (0.0)
Mixed	1 (2.1)	1 (2.7)	0 (0.0)
Caucasian	16 (33.3)	15 (40.5)	1 (9.1)
BMI kg/m^2 N=43	29.7 [27.0, 34.7]	30.2 [26.7, 34.0]	29.1 [28.2, 37.3]
Type 2 diabetes mellitus	12 (25.0)	8 (21.6)	4 (36.4)
Hypertension	12 (25.0)	8 (21.6)	4 (36.4)
Ischaemic heart disease	2 (4.2)	2 (5.4)	0 (0.0)
CKD (stage 2+)	4 (8.3)	2 (5.4)	2 (18.2)
Heart failure (NYHA 3+)	1 (2.1)	1 (2.7)	0 (0.0)
COPD	3 (6.2)	2 (5.4)	1 (9.1)
Asthma	13 (27.1)	9 (24.3)	4 (36.4)
Smoking status			
Current smoker	4 (8.3)	2 (5.4)	2 (18.2)
Ex-smoker	9 (18.8)	7 (18.9)	2 (18.2)
Never smoked	32 (66.7)	25 (67.6)	7 (63.6)
Unknown	3 (6.2)	3 (8.1)	0 (0.0)
Rockwood frailty score			
1	15 (31.2)	11 (29.7)	4 (36.4)
2	17 (35.4)	15 (40.5)	2 (18.2)
3	14 (29.2)	10 (27.0)	4 (36.4)
4	2 (4.2)	1 (2.7)	1 (9.1)

Table 2: demographics and clinical characteristics of patients recruited into the study. BMI = body mass index, CKD = chronic kidney disease classification, COPD = chronic obstructive pulmonary disease, ICU = intensive care unit, NYHA = New York Heart Association classification.

Completeness of data capture from wearable sensors

The median time from admission to wearable sensor application was 27 (IQR:22–46) hours. The median length of hospital stay for each participant was 8 (IQR:6–13) days. The total number of patient-days of sensor wear was 202, 200, 204 and 82 days for the HR/RR, temperature, SpO_2 and BP WVSSs respectively (Table 3). This represented a median duration of wear of approximately 4 days per participant for the HR/RR, temperature and SpO_2sensors. The duration of wear for the BP cuff was median 1.9 days per participant. This was due to connectivity difficulties and because many participants either declined the sensor or requested early removal. The duration of wearable application was similar when calculated from the data recorded by each wearable or from the sensor log maintained by the research team (appendix 1).

Table 3
duration of wear (first to last valid wearable vital signs measurement), completeness of data capture and reasons for wearable removal. The blood pressure cuff was not applied to 15 participants at their request. BP = non-invasive blood pressure, SpO_2 = oxygen saturations.

	Wearable Sensor (N = 48)			
	HR/RR (LifeTouch)	Temperature (LifeTemp)	SpO_2 (Nonin PulseOx)	BP (A&D TM2441)
Duration of sensor wear (days/participant)	3.9 [1.7, 5.9]	3.9 [1.6, 5.9]	3.8 [0.9, 5.9]	1.9 [0.9–3.8]
Overall completeness (%) of wearable sensor data	81.2	92.1	68.6	38.4
Completeness (%) per participant	83.8 [64.1, 95.7]	97.7 [79.7, 99.8]	72.3 [61.7, 87.2]	35.8 [16.3, 47.6]
Reason for device removal				
Discharge from hospital	25 (52.1)	23 (47.9)	24 (50.0)	9 (18.8)
Critical care, loss of capacity	4 (8.3)	4 (8.3)	3 (6.3)	-
Participant request	19 (39.6)	21 (43.8)	21 (43.8)	22 (45.8)
Never applied	-	-	-	15 (31.3)
Other	-	-	-	2 (4.2)

Figure 2 summarises the results of the Kaplan Meier analysis of gaps of varying duration in wearable sensor data. The analysis was limited for SBP due to the intermittent nature of measurements (maximum frequency of wearable recordings: 1–2 hourly). The median survival time without data loss and the percentage of patients without data loss in 24 hours are summarised in appendix 2.

Accuracy of wearable sensor measurements

After creating pairs of wearable and traditional vital signs measurements (within the 5/15min epochs, see methods), there were 1633, 1614, 1412, 1294 and 59 pairs of HR, RR, temperature, SpO_2 and SBP measurements respectively. 59.7% HR pairs were within ± 5bpm, 38.5% of RR pairs were within ± 3breaths/min, 24.4% of temperature pairs were within ± 0.3^oC, 32.9% of SpO_2 pairs were within ± 2% and 39.0% of SBP pairs were within ± 10mmHg. The correlation coefficients for each vital sign are displayed in Table 4 (see appendix 3 for scatterplots).

Figure 3 displays the Bland Altman plots for HR, RR, temperature and SpO_2 stratified according to whether the participant was on a ward or in critical care. The corresponding plot for SBP is available in appendix 4. There was constant variation between wearable and traditional vital signs measurements across the measurement range for HR and SpO_2. RR measurements showed a systematic difference at high mean RRs in critical care, with traditional RR measurements being higher than wearable measurements. A sensitivity analysis (appendix 5) identified that this difference was due to data from five participants who had higher traditional RR measurements than other participants. Temperature measurements also showed a systematic difference at low mean temperatures with traditional measurements being higher than wearable measurements. There were insufficient SBP measurement pairs to comment on the variation in measurements. Table 4 displays the bias and limits of agreement for each vital sign. Not all participants wore all sensors and only participants with at least two measurement pairs are included in this analysis.

Table 4

Bland Altman metrics for each wearable vital sign measurement compared to traditional vital signs measurements as the reference standard. Bias = wearable - traditional (95% confidence interval), LoA = 95% limits of agreement (95% confidence interval). HR = heart rate, RR = respiratory rate, SBP = systolic blood pressure, SpO_2 = oxygen saturations, Temp = temperature. *any participants with only a single measurement pair for the vital sign concerned are excluded to enable calculation of the 95% confidence intervals.

	Vital sign				
	HR (bpm)	RR (/min)	Temp (^oC)	SpO_2 (%)	SBP (mmHg)
Participants*	46	46	47	43	10
Measurement pairs	1632	1613	1411	1294	51
Pairs/participant	24 [9,39]	24 [9,39]	26 [10,42]	24 [9,32]	5 [3,5]
Correlation coefficient (Pearson r)	0.77 (0.75 to 0.79)	0.19 (0.15 to 0.24)	0.33 (0.28 to 0.38)	0.51 (0.47 to 0.55)	0.24 (-0.02 to 0.47)
Bias	-0.2 (0.3 to -0.7)	-1.8 (-1.4 to -2.2)	-1.0 (-1.0 to -1.1)	-2.8 (-2.6 to -2.9)	-0.7 (6.9 to -8.3)
Upper LoA	21.3 (23.3 to 19.6)	14.2 (17.0 to 12.0)	1.7 (2.1 to 1.4)	3.5 (4.1 to 2.9)	37.6 (55.5 to 27.9)
Lower LoA	-21.6 (-19.9 to -23.7)	-17.8 (-15.6 to -20.6)	-3.8 (-3.5 to -4.2)	-9.0 (-8.5 to -9.6)	-39.0 (-29.2 to -56.8)

Impact of wearable vital signs measurements on alerts

Amongst the pairs of traditional and wearable vital signs measurements, there were only 31 instances when HR, RR, temperature, SpO_2 and SBP were simultaneously available to calculate a partial NEWS2 (see methods). Appendix 6 summarises the differences in partial NEWS2 scores calculated from these five vital signs and the impact on the rate of NEWS2 5 + or 7 + alerts by each method. This small number of NEWS2 scores reflects that many participants did not wear the BP cuff or chose to remove it early.

In contrast, there were 613 instances when HR, RR, temperature and SpO_2 were simultaneously available to calculate a partial NEWS2. The median NEWS2 by traditional methods was 1 [IQR: 1–2] and by wearable methods was 4 [IQR: 3–6]. Table 5summarises the number of times when a NEWS2 5 + or 7 + alert would have been generated by each method. At

our institution, NEWS2 5 + would typically alert the ward based medical team, whilst 7 + would typically generate a critical care response.

*Table 5: confusion matrix for NEWS2 scores calculated from 4 vital signs (heart rate, respiratory rate, temperature and SpO_2) using paired traditional and wearable vital signs. A positive NEWS2 score is considered a score of 5+ or 7+ respectively. *On 11 occasions, vitals from wearable sensors identified a 5+ NEWS2 event at the same time as traditional measurements but also identified a 5+ NEWS2 event in the preceding 12 hours which was not detected by traditional measurements. We considered this to represent early detection of deterioration and therefore a true positive. $^{\$}$Similarly, there were 5 instances of early detection of a 7+ NEWS2 event. NEWS2 = national early warning score 2.*

		NEWS2 5+ (4 vitals, traditional)		NEWS2 7+ (4 vitals, traditional)	
		Positive	Negative	Positive	Negative
NEWS2 5+ (4 vitals, wearable)	Positive	26 + 11*	225		
	Negative	6	345		
NEWS2 7+ (4 vitals, wearable)	Positive			5 + 5$^{\$}$	82
	Negative			2	519

Appendix 7 displays the differences in NEWS2 component scores for each pair of vital signs measurements. The corresponding modified Clarke Error Grids for each vital sign are displayed in appendix 8. The greatest agreement in NEWS2 component scores was observed for HR (85.1%). For all other NEWS2 component scores agreement was less than 50%.

Appendix 9 displays the 4Q plots which assess the correlation in differences between successive vital sign measurements by traditional and wearable techniques. There was a moderate correlation between HR measurements recorded on the ward (r = 0.39) and in critical care (r = 0.40), between SpO_2 measurements recorded in critical care (r = 0.35) and between SBP measurements recorded on the ward (r = 0.45). In all other cases there was little or no correlation.

Discussion

In this observational study of WVSSs in hospitalised patients with Covid-19 we monitored HR, RR, temperature and SpO_2 in 48 individuals for a median duration of 3.8–3.9 days. Non-invasive, intermittent, automated BP measurement was poorly tolerated and suffered technical challenges such that the median duration of monitoring was 1.9 days in 33 out of 48 participants.

Our findings align with previous research suggesting that wearable patch/wrist-based vital signs sensors can achieve comprehensive data capture in an inpatient setting with modest (once daily) intervention to maintain the devices. Studies in which WVSSs have been validated against traditional vitals, in ward-based settings and for prolonged periods (> 48 hours) are uncommon but offer real world evidence about WVSS performance. Completeness of data capture by wearable sensors in such studies ranges from 76–96% for patch-based, chest HR/RR sensors[28, 29] and from 50–68% for wrist worn pulse oximeters[30, 31]. Our results are similar, and the variation between studies may be attributed to different patient populations and different levels of experience amongst patients, researchers and clinical teams in using and maintaining the sensors. Isolation measures due to Covid-19, may have also limited opportunities for sensor

maintenance. Even the lowest rate of data capture in our study (SpO_2, 68.6%) equates to continuous vitals for 16 out of every 24 hours, far exceeding what could be captured by traditional methods.

Our survival analysis found that there would be no gaps in the data of over 4 hours duration for 87.9%, 81.5% and 97.5% of patients using the HR/RR, SpO_2 and temperature sensors in the first 24 hours respectively. As four hours is often the interval between nurse measured vital signs[32] it suggests that the time nursing staff would need to spend troubleshooting/reapplying such devices would be acceptable. In high-risk surgical populations in the Netherlands, Breteler and colleagues[18, 19] found even greater durability of data capture with wearable sensors. Our results extend this finding to a UK setting with medical inpatients who are subject to isolation restrictions.

In contrast to the chest and axillary patch sensors we acknowledge that the data capture of BP measurements in our study was poor. For cuff-based, BP sensors, data capture rates of 44–63% have been recorded[30, 33] and our rates were lower than this due to technical difficulties and patient discomfort. This limits the conclusions which we can draw about the A&D TM2441 device as a useful wearable monitor but perhaps stresses the importance of considering if such devices are truly acceptable to patients as an automated, wearable technique. Inflation of a BP cuff can be uncomfortable, and it may be more tolerable when recorded by a nurse.

We observed wide limits of agreement between traditional and wearable vital signs such that differences in only 59.7% of HR, 38.5% of RR, 32.9% of SpO_2, 24.4% of temperature and 39.0% of SBP pairs were within pre-defined clinically acceptable limits which was lower than previous studies[34–36]. However, our findings agree with existing bias/LoA estimates where HR[15, 16], RR[16] and temperature[16, 37] measurements have been validated in similar circumstances and where the repeated nature of measurements in the same patient is accounted for. It is difficult to comment on the validity of SBP measurements given the paucity of measurements in our study. In keeping with previous work[3], we found that the patch-based, chest sensor tended to underestimate high RRs in critically ill patients who were transferred to the ICU. The patch-based, axillary temperature sensor also frequently recorded lower temperatures than a traditional tympanic thermometer.

The reasons for imperfect agreement between wearable vitals and nurse recorded vitals are myriad. It is well recognised[38] that vital signs recorded by healthcare professionals are impacted by poor measurement technique, value bias in recording and a Hawthorne effect during measurement[39]. Arguably therefore, WVSSs may offer a truer reflection of a patient's ongoing, non-observed physiological state. However, systematic errors in wearable vital signs measurements may also play an important role. In our study, the temperature differences we observed may be because estimates of core temperature recorded by an axillary skin sensor are different to estimates recorded by tympanic thermometers used at our institution. Similarly, RRs determined by a wearable sensor based on an algorithm utilising chest impedance and R-R variation may have been subject to systematic error in Covid-19 patients in whom a relative bradyarrhythmia has been observed[40] (potential for error due to Nyqist sampling limit) and potentially lower than expected chest wall excursion (potential for error due to smaller variation in chest impedance). Acknowledging these inherent differences, some authors have suggested that WVSS should not be compared to nurse recorded "spot" vitals[19]. We disagree, because the comparison serves to illustrate the impact that using WVSSs could have on existing patient deterioration alerting mechanisms in hospitals.

In our study, NEWS2 scores derived from HR, RR, temperature and SpO_2 measurements were typically higher when calculated from WVSSs than traditional vitals. Differences in RR, SpO_2 and temperature NEWS2 component scores were responsible for most of this difference (appendix 7). Compared to NEWS2 scores generated by traditional vitals, 85.9% (225/262) of NEWS2 5 + and 89.1% (82/92) of NEWS2 7 + from the WVSSs were false positives. False negatives were rare, 1.7% (6/351) and 0.4% (2/521) for NEWS2 5 + and 7 + respectively. In keeping with our results, Weenk and

colleagues found that two WVSS systems studied on surgical and medical wards[7] both returned higher modified early warning scores (MEWS) than traditional nurse recorded vital signs.

Our findings suggest that vital signs obtained from WVSSs should not be used to directly replace traditional vital signs in track and trigger systems that utilise NEWS2. Compared to NEWS2 scores derived from concurrently recorded traditional vitals the majority of wearable NEWS2 alerts would be false alarms. We also highlight that more work may be needed to confirm the validity of wearable vital signs measurements in settings of illness (very high RR), where a patient's physiology may be quite different from healthy volunteers.

We propose that a different approach is needed to adopt WVSSs into care pathways for the deteriorating patient. This could include scheduled reviews of wearable sensor trends by healthcare providers[41] as opposed to automated alerts, re-development of EWS thresholds for specific use with wearable sensor data[9, 10] or development of novel, broadly applicable deterioration indices using the granular data which wearable sensors provide or some of the additional metrics which they record. Examples include the presence or absence of micro-events[42] within the continuous data, trend information[43] and heart rate variability[44].

This work was a real-world clinical evaluation of WVSSs in a challenging clinical setting. The duration of monitoring is similar to the longest periods of monitoring in previous wearable studies. As monitoring was continued into the ICU for some patients, we were also able to assess sensor performance in critically ill patients with more extreme physiology. Furthermore, we studied the clinical implications of adopting wearable sensors both in terms of maintenance requirements to prevent large gaps in data and the implications of using the wearable data in NEWS2 calculations.

We acknowledge several limitations. Firstly, we studied a population with Covid-19 in which there was a preponderance of Black and Asian participants and a high rate of admission to critical care.[45] This may limit the generalisability of our findings to wider populations of inpatients. Our ward staff were blinded to wearable sensor data and not involved in sensor maintenance. Taken together with the fact that we studied a group at high risk of deterioration, with significant isolation and infection control demands, this may mean that future, deployment of wearable sensors in broader inpatient groups could achieve better data capture.

Secondly, we note that some of our participants did not wear the devices for a long time, especially the blood pressure cuff. Further qualitative work is needed to understand the acceptability of wearable devices to differing patient groups as this may aid their optimal deployment. Finally, the precise time that traditional vital signs measurements were taken was not recorded, reflecting that there may be a delay between measurements and recording in the electronic patient record by our nursing staff. Whilst this could impact our validation results, the guidance at our institution is that recording of vital signs should be performed promptly, making it reasonable to assume that measurements were made in the preceding 5 minutes.

Conclusion

Wearable vital signs sensors have potential to improve the care of patients in hospital but the best way to deploy them in existing healthcare systems remains unclear. In this study, 48 inpatients with Covid-19 wore four wearable sensors for a median of almost four days. The BP cuff was poorly tolerated and suffered technical difficulties. Completeness of data capture from the other three sensors was in keeping with previous work and a survival analysis found no gaps in data collection of over 4 hours in over 80% of patients in the first 24 hours. Compared to nurse recorded vital signs, the validity of data capture by the wearable sensors was poor. Less than 60% of HR measurements, 40% of RR and SpO_2 measurements and 25% of temperature measurements fell within pre-defined clinically acceptable limits. There were systematic differences in measurements at high RRs and low temperatures. As a result, NEWS2 alerts from wearable

sensor data were frequently false positives when compared to NEWS2 alerts from traditional vitals. As it stands, vital signs from wearable sensors cannot directly replace traditional vital signs measurements in existing track and trigger systems.

Declarations

Ethics approval and consent to participate

The COSMIC-19 study was registered with clinicaltrials.gov (registration: NCT04581031, registered 4th May 2020). It was approved by Yorkshire and the Humber – Bradford Leeds Research Ethics Committee, UK (reference 20/YH/0156). All participants gave written consent to participate.

Consent for publication

Not applicable.

Availability of data and materials

The COSMIC-19 dataset is available on request subject to appropriate information governance and data sharing agreements.

Competing interests

Prof Thistlethwaite declares that she is the director of the Innovate Manchester Advanced Therapy Centre Hub, Innovate UK, which provided funding for this work. All other authors have no competing interests to declare.

Funding

Funded by the Innovate Manchester Advanced Therapy Centre Hub, Innovate UK, (project ID: 6239) with additional support from the Christie Hospital Charitable Fund.

Authors contributions

- Anthony Wilson - conceptualisation, data curation, formal analysis, investigation, methodology, project administration, resources, software, validation, visualisation, writing – original draft, writing – review & editing
- Alexander Parker - data curation, writing - review & editing
- Gareth Kitchen - conceptualisation, methodology, project administration, writing – review & editing
- Andrew Martin – conceptualisation
- Lukas Hughes-Noehrer - writing – review & editing
- Mahesh Nirmalan - supervision, writing – review & editing
- Niels Peek - methodology, supervision, writing – review & editing
- Glen Martin - supervision, writing – review & editing
- Fiona Thistlethwaite - conceptualisation, funding acquisition, project administration, resources, supervision, writing – review & editing

Acknowledgements

The authors acknowledge Manchester University NHS Foundation Trust's Clinical Data Science Unit who supported extraction of traditional vital signs measurements from the electronic patient record.

Authors' information

Not applicable

References

1. Michard F, Sessler DI. Ward monitoring 3.0. *Br J Anaesth* [Internet] Elsevier; 2018 [cited 2018 Nov 4]; **121**: 999–1001 Available from: http://www.ncbi.nlm.nih.gov/pubmed/30336872
2. Patel V, Orchanian-Cheff A, Wu R. Evaluating the Validity and Utility of Wearable Technology for Continuously Monitoring Patients in a Hospital Setting: Systematic Review. *JMIR Mhealth Uhealth* [Internet] JMIR Mhealth Uhealth; 2021 [cited 2024 Apr 23]; **9** Available from: https://pubmed.ncbi.nlm.nih.gov/34406121/
3. Chan PY, Ryan NP, Chen D, McNeil J, Hopper I. Novel wearable and contactless heart rate, respiratory rate, and oxygen saturation monitoring devices: a systematic review and meta-analysis. *Anaesthesia* [Internet] Anaesthesia; 2022 [cited 2024 Apr 23]; **77**: 1268–80 Available from: https://pubmed.ncbi.nlm.nih.gov/35947876/
4. Areia C, Biggs C, Santos M, et al. The impact of wearable continuous vital sign monitoring on deterioration detection and clinical outcomes in hospitalised patients: a systematic review and meta-analysis. *Crit Care* [Internet] BioMed Central; 2021 [cited 2021 Oct 24]; **25**: 1–17 Available from: https://ccforum.biomedcentral.com/articles/10.1186/s13054-021-03766-4
5. Gao H, McDonnell A, Harrison DA, et al. Systematic review and evaluation of physiological track and trigger warning systems for identifying at-risk patients on the ward. *Intensive Care Med* [Internet] Springer; 2007 [cited 2022 Feb 21]; **33**: 667–79 Available from: https://link.springer.com/article/10.1007/s00134-007-0532-3
6. Eddahchouri Y, Koeneman M, Plokker M, et al. Low compliance to a vital sign safety protocol on general hospital wards: A retrospective cohort study. *Int J Nurs Stud* [Internet] Int J Nurs Stud; 2021 [cited 2024 Jul 3]; **115** Available from: https://pubmed.ncbi.nlm.nih.gov/33517078/
7. 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* [Internet] European Resuscitation Council, American Heart Association, Inc., and International Liaison Committee on Resuscitation.~Published by Elsevier Ireland Ltd; 2019 [cited 2021 Nov 17]; **136**: 47–53 Available from: https://doi.org/10.1016/j.resuscitation.2019.01.017
8. Stellpflug C, Pierson L, Roloff D, et al. Continuous Physiological Monitoring Improves Patient Outcomes. *Am J Nurs* [Internet] Am J Nurs; 2021 [cited 2021 Nov 28]; **121**: 40–6 Available from: https://pubmed.ncbi.nlm.nih.gov/33755624/
9. van Ede ES, Scheerhoorn J, Bonomi AG, Buise MP, Bouwman RA, Nienhuijs SW. Continuous remote monitoring in post-bariatric surgery patients: development of an early warning protocol. *Surg Obes Relat Dis* [Internet] Surg Obes Relat Dis; 2022 [cited 2024 Jul 3]; **18**: 1298–303 Available from: https://pubmed.ncbi.nlm.nih.gov/35850957/
10. van der Stam JA, Mestrom EHJ, Nienhuijs SW, et al. A wearable patch based remote early warning score (REWS) in major abdominal cancer surgery patients. *European Journal of Surgical Oncology* W.B. Saunders; 2023; **49**: 278–84
11. Watkinson PJ, Pimentel MA, Clifton L, et al. Early detection of physiological deterioration in post-surgical patients using wearable technology combined with an integrated monitoring system: a pre- and post-interventional study. *medRxiv* [Internet] Cold Spring Harbor Laboratory Press; 2020 [cited 2021 Nov 17]; 2020.12.01.20240770 Available from: https://www.medrxiv.org/content/10.1101/2020.12.01.20240770v1
12. National Early Warning Score (NEWS) 2 | RCP London [Internet]. [cited 2021 Jun 6]. Available from: https://www.rcplondon.ac.uk/projects/outputs/national-early-warning-score-news-2
13. Welcome to Isansys [Internet]. [cited 2023 Jun 24]. Available from: https://www.isansys.com/

14. Van Der Stam JA, Mestrom EHJ, Scheerhoorn J, et al. Accuracy of vital parameters measured by a wearable patch following major abdominal cancer surgery. *European Journal of Surgical Oncology* [Internet] Elsevier; 2022 [cited 2022 Apr 14]; **48**: 917–23 Available from: http://www.ejso.com/article/S0748798321007848/fulltext
15. Cheng SM, Chan JJI, Tan CW, Lu E, Sultana R, Sng BL. Use of wireless respiratory rate sensor monitoring during opioid patient-controlled analgesia after gynaecological surgery: A prospective cohort study. *Indian J Anaesth* [Internet] Wolters Kluwer Medknow Publications; 2021 [cited 2022 Apr 18]; **65**: 146–52 Available from: https://journals.lww.com/ijaweb/Fulltext/2021/02000/Use_of_wireless_respiratory_rate_sensor_monitoring.10.aspx
16. Downey C, Ng S, Jayne D, Wong D. Reliability of a wearable wireless patch for continuous remote monitoring of vital signs in patients recovering from major surgery: A clinical validation study from the TRaCINg trial. *BMJ Open* [Internet] British Medical Journal Publishing Group; 2019 [cited 2021 Nov 17]; **9**: e031150 Available from: https://bmjopen.bmj.com/content/9/8/e031150
17. Kant N, Peters GM, Voorthuis BJ, et al. Continuous vital sign monitoring using a wearable patch sensor in obese patients: a validation study in a clinical setting. *J Clin Monit Comput* [Internet] [cited 2022 Apr 14]; Available from: https://doi.org/10.1007/s10877-021-00785-y
18. Breteler MJM, Huizinga E, Van Loon K, et al. 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* [Internet] BMJ Open; 2018 [cited 2022 Feb 21]; **8**: e020162 Available from: https://pubmed.ncbi.nlm.nih.gov/29487076/
19. Breteler MJMM, KleinJan EJ, Dohmen DAJJ, et al. Vital Signs Monitoring with Wearable Sensors in High-risk Surgical PatientsA Clinical Validation Study. *Anesthesiology* [Internet] American Society of Anesthesiologists; 2020 [cited 2020 Jan 21]; **132**: 424–39 Available from: http://journals.lww.com/10.1097/ALN.0000000000003029
20. Bland JM, Altman DG. Measuring agreement in method comparison studies. *Stat Methods Med Res* [Internet] Sage PublicationsSage CA: Thousand Oaks, CA; 1999 [cited 2022 Jan 3]; **8**: 135–60 Available from: https://journals.sagepub.com/doi/abs/10.1177/096228029900800204
21. Zou GY. Confidence interval estimation for the Bland-Altman limits of agreement with multiple observations per individual. *Stat Methods Med Res* [Internet] Stat Methods Med Res; 2013 [cited 2022 Jun 7]; **22**: 630–42 Available from: https://pubmed.ncbi.nlm.nih.gov/21705434/
22. Clarke WL, Cox D, Gonder-Frederick LA, Carter W, Pohl SL. Evaluating Clinical Accuracy of Systems for Self-Monitoring of Blood Glucose. *Diabetes Care* [Internet] American Diabetes Association; 1987 [cited 2021 Dec 28]; **10**: 622–8 Available from: http://diabetesjournals.org/care/article-pdf/10/5/622/515265/10-5-622.pdf
23. Montenij LJ, Buhre WF, Jansen JR, Kruitwagen CL, De Waal EE. Methodology of method comparison studies evaluating the validity of cardiac output monitors: a stepwise approach and checklist. *BJA: British Journal of Anaesthesia* [Internet] Oxford Academic; 2016 [cited 2021 Nov 27]; **116**: 750–8 Available from: https://academic.oup.com/bja/article/116/6/750/2566462
24. Team RC. R: A language and environment for statistical computing. [Internet]. Vienna, Austria: R Foundation for Statistical Computing; 2019. Available from: http://www.r-project.org/
25. WristOx2® Model 3150 with USB - Nonin [Internet]. [cited 2024 Apr 23]. Available from: https://www.nonin.com/products/3150-usb/
26. TM-2441 | Professional | Medical | Products | A&D [Internet]. [cited 2024 Apr 23]. Available from: https://www.aandd.jp/products/medical/professional/2441.html
27. Cases in the UK | Coronavirus in the UK [Internet]. [cited 2023 Jun 24]. Available from: https://coronavirus.data.gov.uk/details/cases
28. Haahr-Raunkjaer C, Skovbye M, Rasmussen SM, et al. Agreement between standard and continuous wireless vital sign measurements after major abdominal surgery: a clinical comparison study. *Physiol Meas* [Internet] Physiol Meas; 2022 [cited 2023 Apr 11]; **43** Available from: https://pubmed.ncbi.nlm.nih.gov/36322987/

29. Elvekjaer M, Aasvang EK, Olsen RM, et al. Physiological abnormalities in patients admitted with acute exacerbation of COPD: an observational study with continuous monitoring. *J Clin Monit Comput* [Internet] Springer; 2020 [cited 2021 Nov 18]; **34**: 1051–60 Available from: https://link.springer.com/article/10.1007/s10877-019-00415-8
30. Mølgaard J, Rasmussen SS, Eiberg J, Sørensen HBD, Meyhoff CS, Aasvang EK. Continuous wireless pre- and postoperative vital sign monitoring reveal new, severe desaturations after vascular surgery. *Acta Anaesthesiol Scand* [Internet] John Wiley & Sons, Ltd; 2023 [cited 2023 Apr 12]; **67**: 19–28 Available from: https://onlinelibrary.wiley.com/doi/full/10.1111/aas.14158
31. Rasmussen PS, Aasvang EK, Olsen RM, et al. Continuous peripheral perfusion index in patients admitted to hospital wards – An observational study. *Acta Anaesthesiol Scand* [Internet] John Wiley & Sons, Ltd; 2021 [cited 2024 Apr 23]; **65**: 257–65 Available from: https://onlinelibrary.wiley.com/doi/full/10.1111/aas.13711
32. Dall'Ora C, Griffiths P, Hope J, et al. How long do nursing staff take to measure and record patients' vital signs observations in hospital? A time-and-motion study. *Int J Nurs Stud* [Internet] Elsevier; 2021 [cited 2024 Apr 23]; **118** Available from: /pmc/articles/PMC8249906/
33. Haahr-Raunkjaer C, Mølgaard J, Elvekjaer M, et al. Continuous monitoring of vital sign abnormalities; association to clinical complications in 500 postoperative patients. *Acta Anaesthesiol Scand* John Wiley and Sons Inc; 2022;
34. 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* [Internet] JMIR Mhealth Uhealth; 2017 [cited 2021 Nov 23]; **5** Available from: https://pubmed.ncbi.nlm.nih.gov/28679490/
35. Lee PJ. Clinical evaluation of a novel respiratory rate monitor. *J Clin Monit Comput* [Internet] Springer Netherlands; 2016 [cited 2023 Feb 20]; **30**: 175–83 Available from: https://link.springer.com/article/10.1007/s10877-015-9697-4
36. Liu Y, Liu C, Gao M, et al. Evaluation of a wearable wireless device with artificial intelligence, iThermonitor WT705, for continuous temperature monitoring for patients in surgical wards: a prospective comparative study. *BMJ Open* [Internet] 2020 [cited 2022 Apr 15]; **10**: 39474 Available from: http://bmjopen.bmj.com/
37. Zsuzsanna Balla H, Theodorsson E, Ström JO. Evaluation of commercial, wireless dermal thermometers for surrogate measurements of core temperature. *Scand J Clin Lab Invest* [Internet] Scand J Clin Lab Invest; 2019 [cited 2023 Mar 28]; **79**: 1–6 Available from: https://pubmed.ncbi.nlm.nih.gov/30882250/
38. Kallioinen N, Hill A, Christofidis MJ, Horswill MS, Watson MO. Quantitative systematic review: Sources of inaccuracy in manually measured adult respiratory rate data. *J Adv Nurs* [Internet] Wiley-Blackwell; 2021 [cited 2024 Apr 23]; **77**: 98 Available from: /pmc/articles/PMC7756810/
39. Kallioinen N, Hill A, Christofidis MJ, Horswill MS, Watson MO. Quantitative systematic review: Sources of inaccuracy in manually measured adult respiratory rate data. *J Adv Nurs* [Internet] Wiley-Blackwell; 2021 [cited 2024 May 14]; **77**: 98 Available from: /pmc/articles/PMC7756810/
40. Jung L-Y, Kim J-M, Ryu S, Lee C-S. Relative bradycardia in patients with COVID-19. *International Journal of Arrhythmia 2022 23:1* [Internet] BioMed Central; 2022 [cited 2024 Apr 23]; **23**: 1–5 Available from: https://arrhythmia.biomedcentral.com/articles/10.1186/s42444-022-00073-z
41. Leenen JPL, Rasing HJM, Van Dijk JD, Kalkman CJ, Schoonhoven L, Patijn GA. Feasibility of wireless continuous monitoring of vital signs without using alarms on a general surgical ward: A mixed methods study. *PLoS One* Public Library of Science; 2022; **17**
42. Duus CL, Aasvang EK, Olsen RM, et al. Continuous vital sign monitoring after major abdominal surgery—Quantification of micro events. *Acta Anaesthesiol Scand* [Internet] 2018 [cited 2021 Nov 20]; **62**: 1200–8 Available from: https://onlinelibrary.wiley.com/doi/full/10.1111/aas.13173
43. van Rossum MC, Vlaskamp LB, Posthuma LM, et al. Adaptive threshold-based alarm strategies for continuous vital signs monitoring. *J Clin Monit Comput* [Internet] Springer Science and Business Media B.V.; 2021 [cited 2022 Apr 14]; 1–11 Available from: https://link.springer.com/article/10.1007/s10877-021-00666-4

44. Jansen C, Chatterjee DA, Thomsen KL, et al. Significant reduction in heart rate variability is a feature of acute decompensation of cirrhosis and predicts 90-day mortality. *Aliment Pharmacol Ther* [Internet] Blackwell Publishing Ltd; 2019 [cited 2021 Nov 28]; **50**: 568–79 Available from: https://onlinelibrary.wiley.com/doi/full/10.1111/apt.15365
45. Docherty AB, Harrison EM, Green CA, et al. Features of 20 133 UK patients in hospital with covid-19 using the ISARIC WHO Clinical Characterisation Protocol: prospective observational cohort study. *BMJ* [Internet] British Medical Journal Publishing Group; 2020 [cited 2023 Aug 14]; **369** Available from: https://www.bmj.com/content/369/bmj.m1985

Figures

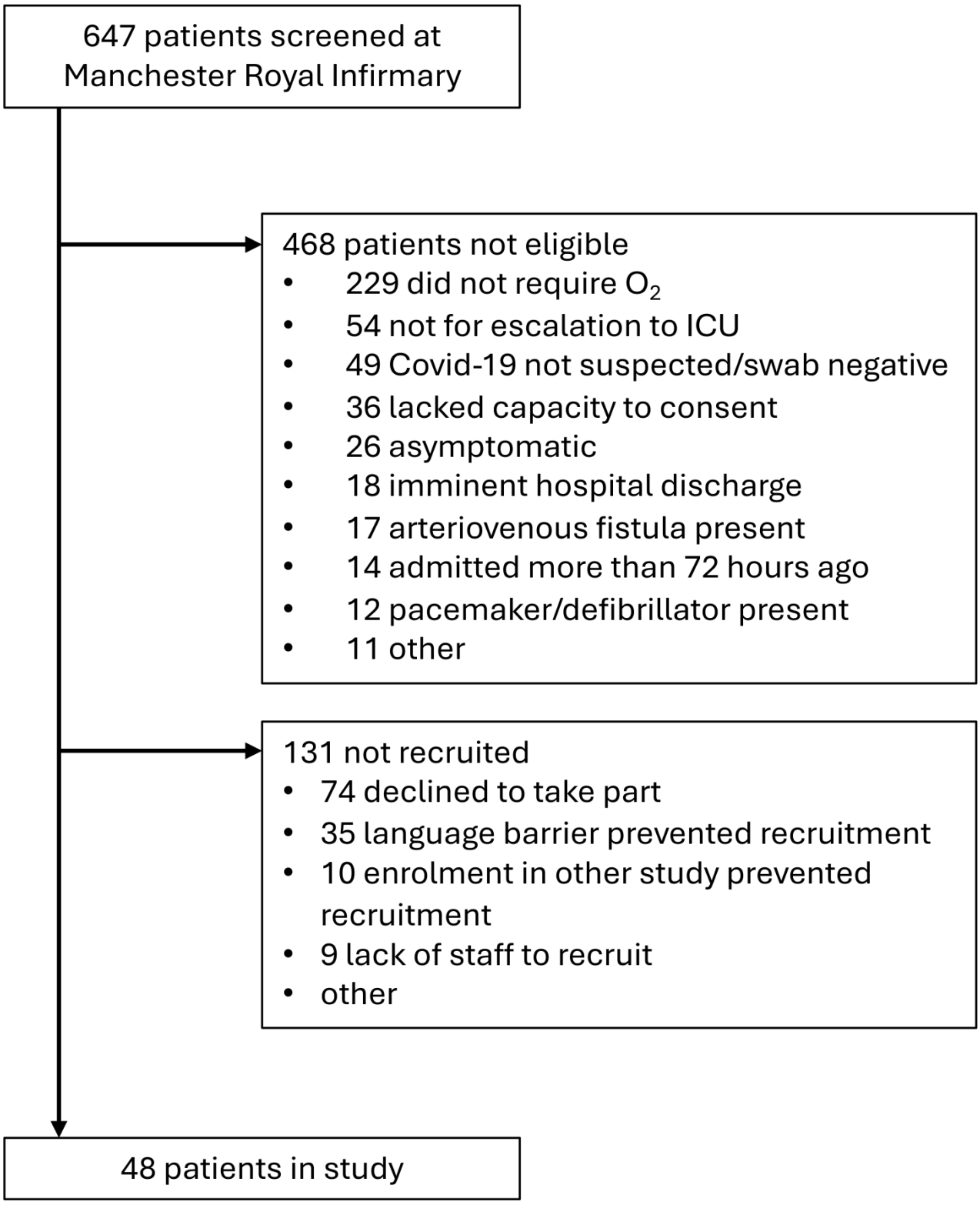

Figure 1

screening and recruitment to the study

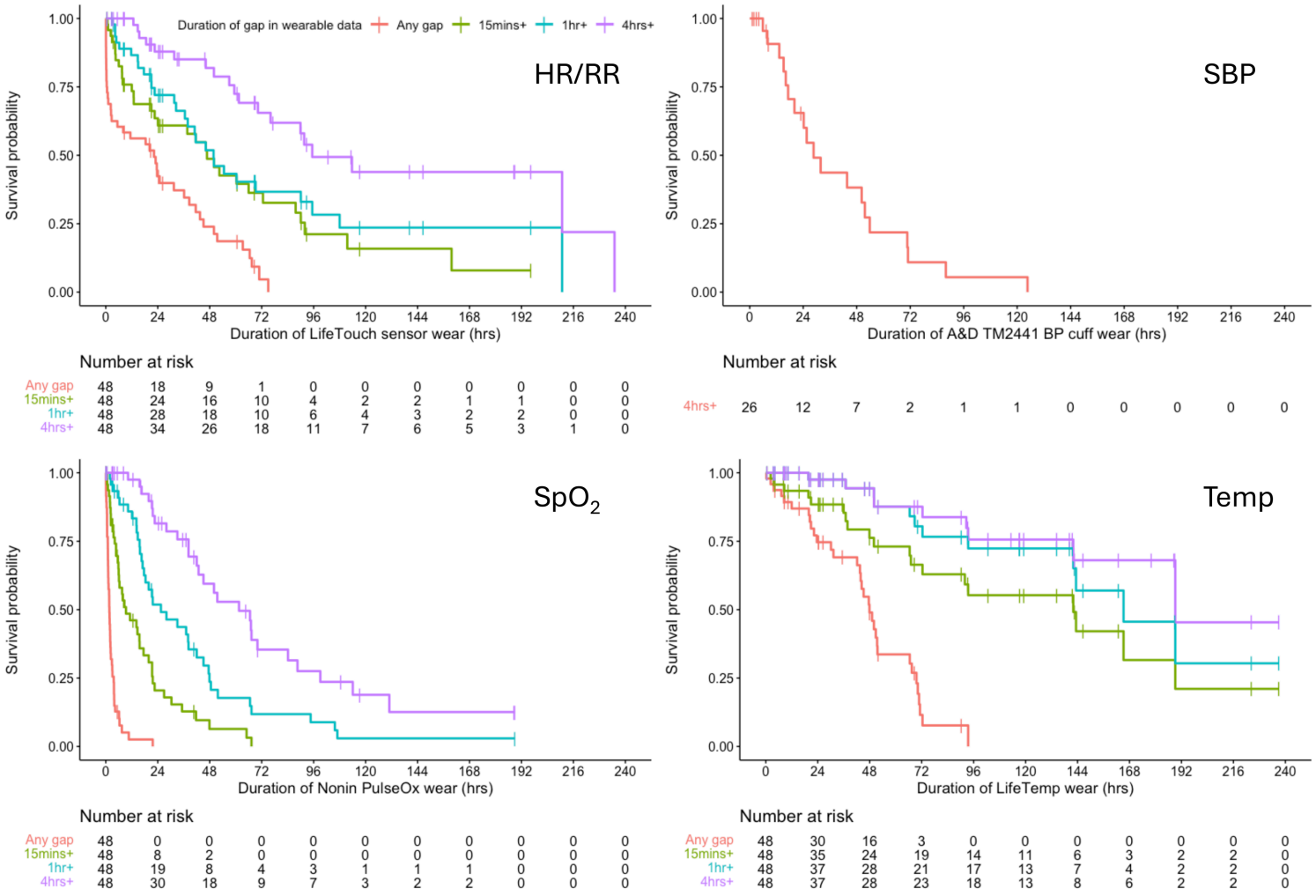

Figure 2

Kaplan Meier analysis of gaps of varying duration in wearable sensor data. HR/RR = heart rate/respiratory rate, SpO_2 = oxygen saturation, SBP = systolic blood pressure, Temp = temperature.

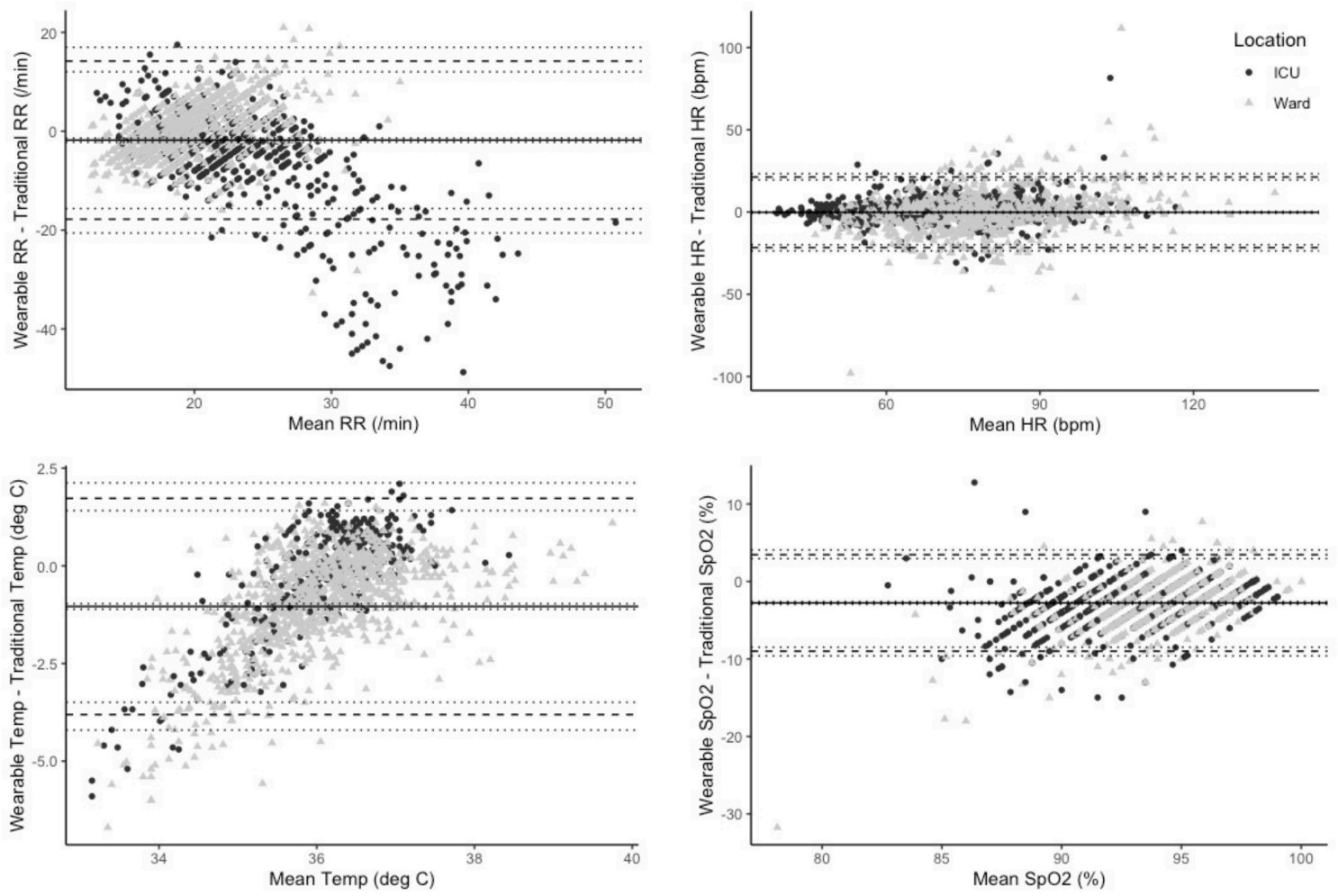

Figure 3

Bland Altman plots corrected for repeated measures. HR = heart rate, RR = respiratory rate, SpO2 = oxygen saturations, Temp = temperature

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SDCWilsonValidationofWearableSensorsBMC.pdf