Section Editor: Nancy Borkowski

OPEN

Smoothing Effect in Vital Sign Recordings: Fact or Fiction? A Retrospective Cohort Analysis of Manual and Continuous Vital Sign Measurements to Assess Data Smoothing in Postoperative Care

Hamish R. Tomlinson, BE (Hons),* Marco A. F. Pimentel, PhD,* Stephen Gerry, MSc,† David A. Clifton, PhD,* Lionel Tarassenko, PhD,* and Peter J. Watkinson, MBChB, MD‡

BACKGROUND: Data smoothing of vital signs has been reported in the anesthesia literature, suggesting that clinical staff are biased toward measurements of normal physiology. However, these findings may be partially explained by clinicians interpolating spurious values from noisy signals and by the undersampling of physiological changes by infrequent manual observations. We explored the phenomenon of data smoothing using a method robust to these effects in a large postoperative dataset including respiratory rate, heart rate, and oxygen saturation (Spo_2). We also assessed whether the presence of the vital sign taker creates an arousal effect.

METHODS: Study data came from a UK upper gastrointestinal postoperative ward (May 2009 to December 2013). We compared manually recorded vital sign data with contemporaneous continuous data recorded from monitoring equipment. We proposed that data smoothing increases differences between vital sign sources as vital sign abnormality increases. The primary assessment method was a mixed-effects model relating continuous-manual differences to vital sign values, adjusting for repeated measurements. We tested the regression slope significance and predicted the continuous-manual difference at clinically important vital sign values. We calculated limits of agreement (LoA) between vital sign sources using the Bland–Altman method, adjusting for repeated measures. Similarly, we assessed whether the vital sign taker affected vital signs, comparing continuous data before and during manual recording.

RESULTS: From 407 study patients, 271 had contemporaneous continuous and manual recordings, allowing 3740 respiratory rate, 3844 heart rate, and 3896 Spo_2 paired measurements for analysis. For the model relating continuous-manual differences to continuous-manual average vital sign values, the regression slope (95% confidence interval) was 0.04 (−0.01 to 0.10; $P = .11$) for respiratory rate, 0.04 (−0.01 to 0.09; $P = .11$) for heart rate, and 0.10 (0.07–0.14; $P < .001$) for Spo_2. For Spo_2 measurements of 91%, the model predicted a continuous-manual difference (95% confidence interval) of −0.88% (−1.17% to −0.60%). The bias (LoA) between measurement sources was −0.74 (−7.80 to 6.32) breaths/min for respiratory rate, −1.13 (−17.4 to 15.1) beats/min for heart rate, and −0.25% (−3.35% to 2.84%) for Spo_2. The bias (LoA) between continuous data before and during manual observation was −0.57 (−5.63 to 4.48) breaths/min for respiratory rate, −0.71 (−10.2 to 8.73) beats/min for heart rate, and −0.07% (−2.67% to 2.54%) for Spo_2.

CONCLUSIONS: We found no evidence of data smoothing for heart rate and respiratory rate measurements. Although an effect exists for Spo_2 measurements, it was not clinically significant. The wide LoAs between continuous and manually recorded vital signs would commonly result in different early warning scores, impacting clinical care. There was no evidence of an arousal effect caused by the vital sign taker. (Anesth Analg 2018;127:960–6)

KEY POINTS

- **Question:** Does a "smoothing effect" exist between manually recorded and electronically recorded vital sign measurements in postoperative care?
- **Findings:** Using a mixed-effects model, we found no relationship between continuous-manual differences and continuous-manual average values for heart rate and respiratory rate, and we found a weak (but clinically insignificant) relationship for oxygen saturation.
- **Meaning:** We found that clinical staff in a postoperative ward did not "smooth" vital sign values with a bias toward recording more normal readings because the differences between manual and continuous vital sign measurements were not related to the vital sign values.

From the *Department of Engineering Science, Institute of Biomedical Engineering, University of Oxford, Oxford, United Kingdom; †Centre for Statistics in Medicine, Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, Botnar Research Centre, University of Oxford, Oxford, United Kingdom; and ‡Kadoorie Centre for Critical Care Research and Education, Nuffield Department of Clinical Neurosciences, Oxford University Hospitals National Health Service (NHS) Trust, Oxford, United Kingdom.

DOI: 10.1213/ANE.0000000000003694

Accepted for publication June 19, 2018.

Funding: H.R.T. was funded by the Rhodes Trust (Rhodes Scholarship). M.A.F.P. was funded by the Research Councils United Kingdom (RCUK) Digital Economy Programme grant number EP/G036861/1 (Oxford

Manual measurements of the main vital signs—which include respiratory rate, blood pressure, heart rate, temperature, and oxygen saturation (Spo_2)—are often inaccurate.[1–3] Manual calculations of clinical risk scores are also error prone.[4–6] While automated monitoring technology exists, it is mostly confined to high-acuity patients, and manual measurement and documentation of vital signs remain the standard of care in many wards. A potential source of inaccuracy that may exist in the manual vital sign record is data smoothing.[7] Clinical staff may be biased toward vital sign values that lie within the assumed limits of normality and record vital sign values that are incorrectly normal— "smoothing" the extremes in the vital sign record. If real, this "smoothing effect"[8] may result in lost opportunities for early recognition of physiological deterioration.

The smoothing effect has been reported to occur during anesthesia[8–13] and in acute ward monitoring.[14,15] Most studies use methods based on the comparison of vital sign values from manual and automated measurements sources.[9–15] The comparison method makes the cause of "data smoothing" unclear. Vital sign values from monitoring equipment are noisy and may be corrupted with signal artefact, so the smoothing effect may partly result from clinicians correcting spurious values.[14] Undersampling may also affect data smoothing based on the magnitude and frequency of extremal values in longitudinal records because sparsely sampled manual observations may not coincide with times of large fluctuations in vital sign values. Analysis compensating for these confounding factors is essential to discover whether the smoothing effect is of clinical relevance.

We present a secondary analysis of a large database of postoperative vital sign records to investigate data smoothing of respiratory rate, heart rate, and Spo_2. We propose that data smoothing increases the differences between continuous and manual vital sign measurements as the (absolute) value of the vital sign becomes more extreme. We tested whether differences between continuous and manual vital sign recordings are related to the average value of the 2 vital sign recordings. We also assessed agreement between continuous and manual data. Finally, we investigated whether there is an arousal effect caused by the vital sign taker.[15]

Centre for Doctoral Training in Healthcare Innovation) and Fundação para a Ciência e a Tecnologia under the grant SFRH/BD/79799/2011. S.G. was funded by a National Institute for Health Research (NIHR) Doctoral Fellowship (DRF-2016-09-073). D.A.C. was funded by the Wellcome Trust and Engineering and Physical Sciences Research Council (EPSRC) under grant number WT 088877/Z/09/Z. L.T. and P.J.W. were supported by the NIHR Biomedical Research Centre Programme, Oxford, which funded the Computer Alerting Monitoring System 2 study.

The authors declare no conflicts of interest.

Supplemental digital content is available for this article. Direct URL citations appear in the printed text and are provided in the HTML and PDF versions of this article on the journal's website (www.anesthesia-analgesia.org).

Clinical Trial Registration: International Standard Randomised Controlled Trial Number (ISRCTN) database—principal investigator: Dr Peter Watkinson; ISRCTN No: ISRCTN58660550, August 11, 2017 (Computer Alerting Monitoring System 2 study).

Reprints will not be available from the authors.

Address correspondence to Hamish R. Tomlinson, BE (Hons), Department of Engineering Science, Institute of Biomedical Engineering, University of Oxford, Old Rd Campus Research Bldg, Oxford OX3 7DQ, United Kingdom. Address e-mail to Hamish.tomlinson@eng.ox.ac.uk.

METHODS

This article adheres to the Reporting of Studies Conducted Using Observational Routinely-Collected Health Data statement, an extension of the Strengthening the Reporting of Observational Studies in Epidemiology guidelines.[16,17]

Dataset

The database for this retrospective analysis was created during the Computer Alerting Monitoring System 2 (CALMS-2) study, which was granted ethical approval (Mid and South Buckinghamshire ethics committee Research Ethics Committee: 08/H0604/79, December 9, 2008, and Leeds [West] ethics committee Research Ethics Committee: 11/YH/0056, May 20, 2011) and registered in the International Standard Randomised Controlled Trial Number (ISRCTN) database (principal investigator: P.J.W., ISRCTN No: ISRCTN58660550, August 11, 2017). This study assessed whether ambulatory physiological monitoring combined with an alerting system improved recognition and outcomes in patients after major surgery.

Vital sign data used in the CALMS-2 study were collected in a step-down postoperative ward of the Oxford University Hospitals National Health Service (NHS) Trust, Oxford, between May 2009 and December 2013. Potential participants were screened during preoperative assessment and deemed eligible if they were planned to undergo major upper gastrointestinal surgery. This category was defined as follows: oesophagectomy, oesophagogastrectomy, gastrectomy, Whipple's operation, liver resection, pancreatectomy, gastric bypass, biliary reconstruction, and splenectomy. Participants were excluded based on the following criteria: participants <16 years of age, pregnant women, participants unable to wear the required monitoring, participants without the capacity to consent, and participants who could not understand written English and for whom no translator could be found. For this secondary analysis, patients who did not receive contemporaneous bedside electronic vital sign monitoring and manual vital sign observations were excluded a priori. Written informed consent was obtained for all subjects.

In the step-down ward used in the CALMS-2 study, high-risk patients are admitted after a period of elective intensive care unit stay, while patients with a lower risk of complication are admitted to the ward immediately after surgery. High-risk patients typically receive an increased level of care for the first 2–48 hours of their ward stay, during which time they undergo conventional bedside monitoring, consisting of continuously measured respiratory rate, heart rate, and Spo_2 (Philips M3046A/Intellivue MP50 clinical monitor; Philips Healthcare, Best, the Netherlands). Respiratory rate was measured by impedance pneumography, heart rate was derived from the electrocardiogram, and Spo_2 was measured by pulse oximetry. Clinical staff also made manual measurements of blood pressure and temperature, typically at hourly intervals. After this initial period, these patients then join the other patients on the ward in receiving general-ward-level care, which consists of manual vital sign recording typically at 4-hour intervals. The standard of care for measuring respiratory rate on the ward is by counting chest wall movements, and heart rate and Spo_2 measurements were likely to be transposed from the monitor screen. In the CALMS-2 study, ambulatory monitoring of heart rate and Spo_2 was undertaken. However, for consistency, we restricted this analysis to manually recorded values that could be compared with contemporaneous values from bedside monitoring.

Manual vital sign measurements documented on paper-based bedside charts were double entered into an electronic database. A third researcher reconciled differences with access to the original charts.[5] Continuous vital sign data from the bedside monitors were saved directly from the monitor every second. The final dataset obtained for this study consisted of vital sign records of respiratory rate, heart rate, and Spo_2 from continuous bedside monitoring equipment and manual vital sign observations.

Manual and continuous vital sign values were compared based on the timestamps taken from paper records and the computer-generated timestamps from the patient monitor. Clinical staff manually recorded blood pressure measurements at the time of observation, and these were also logged in the automatically generated data record. We set the manual observation time for all vital signs to the computer-generated timestamps for blood pressure (while checking that data were correctly matched). This calibration method ensured that the data considered contemporaneous from manual and continuous measurement sources were synchronized, and thus could appropriately be used for comparison.

Statistical Analysis

We summarized the number of vital sign measurements included per patient using the sample median and interquartile range.

We sampled the continuous data at the time of each manual observation to create paired measurements of continuous and manual vital signs. We sampled the continuous data by extracting the median value of a 5-minute window centered at the time of manual observation. We used this methodology (also known as "median filtering") to summarize the continuous data without the effects of measurement noise or short-term variance while retaining long-term vital sign trends. We included all manual observations with contemporaneous periods of continuous data, allowing patients to provide multiple observations in our analysis. We selected a 5-minute window to reflect clinical practice, in line with previous work.[13,15] We undertook sensitivity analyses by recomputing the primary assessment method for windows of 1–10 minutes.

To obtain the differences between the 2 measurement sources, we subtracted the manual vital sign from the continuous vital sign in each measurement pair. We also calculated the average of the continuous and manual vital signs for each measurement pair. We modeled the relationship between the differences and the averages of the measurement recordings using a linear mixed-effects model, with averages as a fixed effect and subject as a random effect to adjust for repeated measurements. We note that there may be factors other than a smoothing effect that could affect the relationship between the continuous-manual difference and the continuous-manual average, notably the accuracy of the continuous monitoring equipment. However, given our use of gold standard monitoring equipment and median filtering to remove transient outliers, the continuous data were likely to be equally reliable over the range of measurement values. Thus, a linear relationship between the continuous-manual difference and the continuous-manual average would be best explained by a clinical bias in the manual data, allowing the smoothing effect to be tested.

The regression slope of the fixed effect in the mixed-effects model represents the increase in the continuous-manual difference for a 1-unit increase in the continuous-manual average. Statistical significance of the regression slope was calculated with an F test (type III with Kenward–Roger degrees of freedom approximation),[18] using a significance of 0.05 to reject the null hypothesis that the regression slope was 0 (no relationship). To assess the clinical significance of the linear relationship, we simulated the mean differences and 95% confidence intervals (CIs) predicted by our model at vital sign values corresponding to important clinical thresholds. We used respiratory rate values of 8 and 24 breaths/min, heart rate values of 40 and 130 beats/min, and an Spo_2 value of 91%, which are the maximum National Early Warning Score limits for each vital sign.[19]

We generated Bland–Altman plots for the continuous and manual vital sign data by plotting the differences between the 2 measures (y-axis) against the averages of the 2 measures (x-axis). The bias was calculated as the mean difference between continuous and manual measurement pairs. Limits of agreement (LoAs) were calculated using a mixed-effects model, including a subject random effect to adjust for repeated measurements.[20] CIs (95%) were calculated for the bias and the LoAs using the method recommended by Bland and Altman.[21] Horizontal lines were included on the Bland–Altman plots to show the bias and LoAs. The regression line from the linear mixed-effect model was plotted to visualize the relationship between the differences and the averages.

We assessed whether there is an arousal effect caused by the vital sign taker by comparing continuous data before and during the time of manual observation. For the continuous data before the observation, we used the median value of a 15-minute window ending 5 minutes before the time of observation. For the continuous data during the time of observation, we used the median of a 5-minute window, as described in previous paragraphs. These methods replicate those of Taenzer et al,[15] to allow comparison. The selection of the continuous data is shown schematically for heart rate in Figure 1. We prepared Bland–Altman plots for the "before" and "during" measurements using the same methodology described in the previous paragraph.

We did not perform sample size calculations for this study because it is a retrospective cohort study of an existing dataset—we used all available data, noting that our dataset was larger than those used in most previous analyses of the smoothing effect.[9–13,15,22] We also could not identify the statistical power of the study since previous studies of "the smoothing effect" have not used Bland–Altman analyses (preventing estimation of approximate population regression slopes). We instead provided CIs for all reported outcomes to demonstrate the precision of our results, as recommended by Goodman and Berlin.[23]

RESULTS

Patient inclusion is shown in Figure 2. Concurrent manual and continuous vital sign measurements were available for respiratory rate from 263 patients (3740 paired

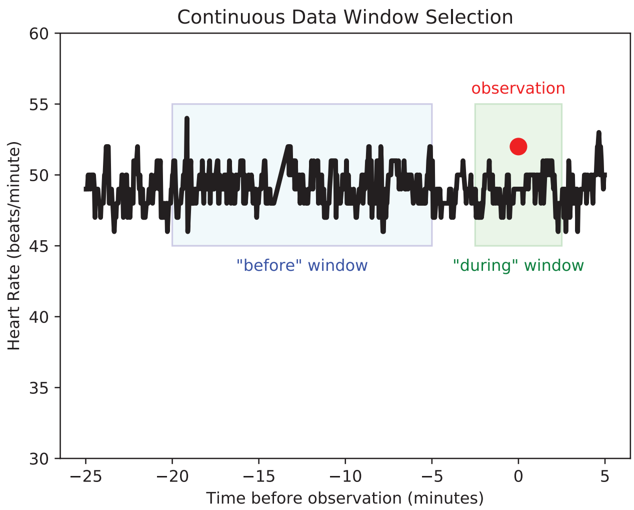

Figure 1. An example heart rate observation, for which the "before" window is shown in blue, the "during" window is shown in green, and the manual observation (synchronized to the time of the blood pressure measurement) is shown in red.

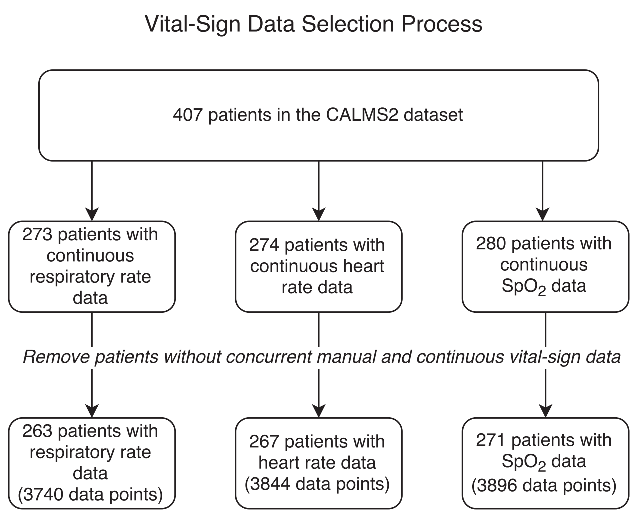

Figure 2. The vital sign data selection process. CALMS-2 indicates Computer Alerting Monitoring System 2; Spo_2, oxygen saturation.

measurements), heart rate from 267 patients (3844 paired measurements), and Spo_2 from 271 patients (3896 paired measurements). The median (interquartile range) number of observations for each patient was 11 (7–18) respiratory rate, 11 (7–18) heart rate, and 11 (7–19) Spo_2 measurements.

The mixed-effect model regression slope (95% CI) between the continuous-manual difference and the continuous-manual average was 0.04 (−0.01 to 0.10; P = .11) for respiratory rate, 0.04 (−0.01 to 0.09; P = .11) for heart rate, and 0.10 (0.07–0.14; P < .001) for Spo_2 (Figure 3 and Table). The mean differences (95% CI) predicted by the model were −1.14 (−1.57 to −0.71) and −0.44 (−1.57 to 0.04) breaths/min for respiratory rates of 8 and 24 breaths/min, respectively. Likewise, the differences were −2.93 (−4.75 to −1.11) and 0.63 (−4.75 to 3.26) beats/min for heart rates of 40 and 130 beats/min, respectively, and −0.88% (−1.17% to −0.60%) for Spo_2 values of 91%.

The bias (LoA) between the continuous and manual data was −0.74 (−7.80 to 6.32) breaths/min for respiratory rate, −1.13 (−17.4 to 15.1) beats/min for heart rate, and −0.25% (−3.35% to 2.84%) for Spo_2 (Figure 3 and Table).

The bias (LoA) between continuous data before and during manual recording was −0.57 (−5.63 to 4.48) breaths/min for respiratory rate, −0.71 (−10.2 to 8.73) beats/min for heart rate, and −0.07% (−2.67% to 2.54%) for Spo_2 (Figure 4 and Table). CIs (95%) for the bias and LoAs are reported in the Table.

Results for the sensitivity analysis of the primary assessment method for different window sizes (during the manual observation) are given in Supplemental Digital Content 1, Table 1, http://links.lww.com/AA/C522.

DISCUSSION

We found no clinically relevant smoothing effect in postoperative care. If vital sign data were smoothed, then manual measurements would be recorded above abnormally low and below abnormally high continuous measurements, causing continuous-manual differences to be related to vital sign values. However, differences between manual and continuous heart rate and respiratory rate measurements were not related to the average values because the regression slope was not significantly different to 0. Differences between manual and continuous measurements of Spo_2 were related to the measurement value. However, this relationship was clinically insignificant (Figure 3), with minor differences for low values of Spo_2. For example, for Spo_2 measurements of 91%, the model predicted a continuous-manual difference of −0.88% (−1.17% to −0.60%), which is within the measurement error of most pulse oximeters.[24] The LoAs (95% CI) between continuously and manually recorded vital signs were large: 14.1 (13.8–14.4) breaths/min, 32.5 (31.9–33.3) heartbeats/min, and 6.2% (6.1%–6.3%) Spo_2 between LoAs, suggesting that these recordings cannot be used interchangeably. We found no evidence of an arousal effect from the vital sign taker. The bias between continuous vital sign values recorded before and during manual observation was less than a single breath, heartbeat, or percentage Spo_2.

Our sensitivity analysis showed that window size affected the relationship between continuous-manual differences and averages but not to a clinically meaningful extent (Supplemental Digital Content 1, Table 1, http://links.lww.com/AA/C522). As window size was reduced, lessening the effect of median filtering, the differences predicted by the model increased, although only by 1–2 heartbeats or breaths. Increasing the window size does not affect our findings, suggesting that the choice of a 5-minute window (also chosen by Reich et al[13] and Taenzer et al[15]) appropriately removes artefact without increasing the sample size beyond what is plausible for bedside measurement. Notably, 4 previous studies found a smoothing effect using automated monitoring signals without temporal averaging, relying on manufacturer settings for artefact removal.[8,9,11,12] Sapo et al[14] demonstrated that Spo_2 values <90% are associated with poor signal quality, suggesting that the effects in these studies may be due to clinicians correctly removing spurious values.

Further methodological differences may explain why our results contrast with the previous literature suggesting a smoothing effect.[8–10,12,13,22] Four studies compared the

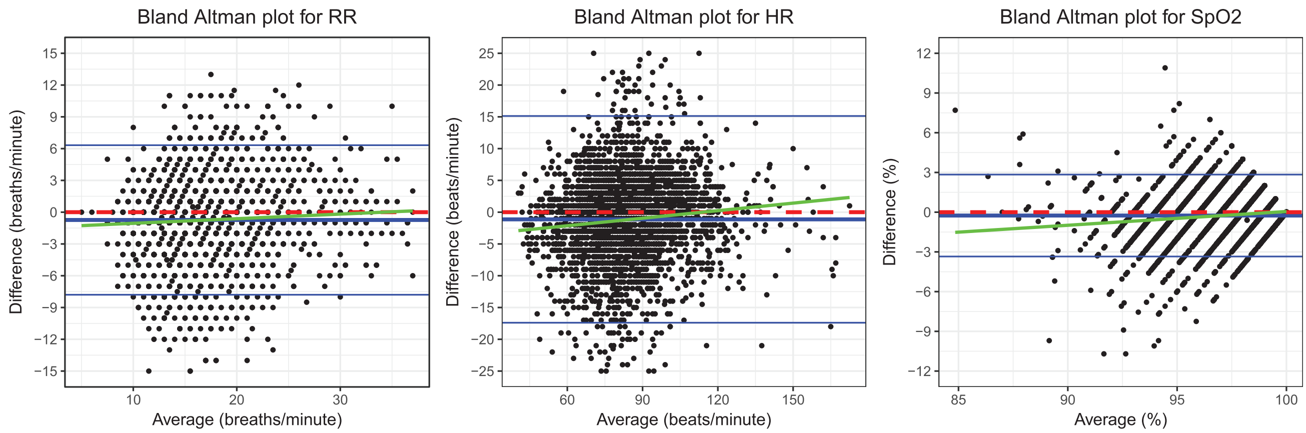

Figure 3. Bland–Altman plots for RR, HR, and Spo_2 showing limits of agreement between the continuous data and the manual observation (continuous data–manual observation). Bias and limits of agreement are shown with blue lines, the regression line is shown in green, and a dashed red line shows the zero y-intercept. HR indicates heart rate; RR, respiratory rate; Spo_2, oxygen saturation.

Table. Results for the Bland–Altman Analysis Comparing Continuous Vital Sign Data to Manual Observations for Respiratory Rate, Heart Rate, and Oxygen Saturation

	Vital Sign		
Bland–Altman Analysis	**Respiratory Rate (breaths/min)**	**Heart Rate (beats/min)**	**Spo_2 (%)**
Sample size			
Patients (observations)	263 (3740)	267 (3844)	271 (3896)
Continuous-manual			
Bias (CI)	−0.74 (−0.82 to −0.66)	−1.13 (−1.32 to −0.94)	−0.25 (−0.29 to −0.22)
Lower LoA (CI)	−7.80 (−7.94 to −7.66)	−17.4 (−17.7 to −17.1)	−3.35 (−3.41 to −3.29)
Upper LoA (CI)	6.32 (6.18–6.46)	15.1 (14.8–15.5)	2.84 (2.78–2.90)
Regression slope[a] (CI), *P* value[b]	0.04 (−0.01 to 0.10), .11	0.04 (−0.01 to 0.09), .11	0.10 (0.07–0.14), <.001
"Before" – "During"			
Bias (CI)	−0.57 (−0.63 to −0.51)	−0.71 (−0.82 to −0.60)	−0.07 (−0.10 to −0.04)
Lower LoA (CI)	−5.63 (−5.72 to −5.53)	−10.2 (−10.3 to −9.97)	−2.67 (−2.72 to −2.62)
Upper LoA (CI)	4.48 (4.38–4.58)	8.73 (8.55–8.92)	2.54 (2.49–2.59)

Abbreviations: CI, confidence interval; LoA, limits of agreement; Spo_2, oxygen saturation.

[a]Regression slope from the linear mixed-effect model, which represents the increase in the continuous-manual difference for a 1-unit increase in the continuous-manual average.

[b]Statistical significance of the regression slope between the differences and the average of continuous and manual vital sign data was calculated with an *F* test (type III with Kenward–Roger degrees of freedom approximation),[18] using a significance of .05 to reject the null hypothesis that the regression slope was 0.

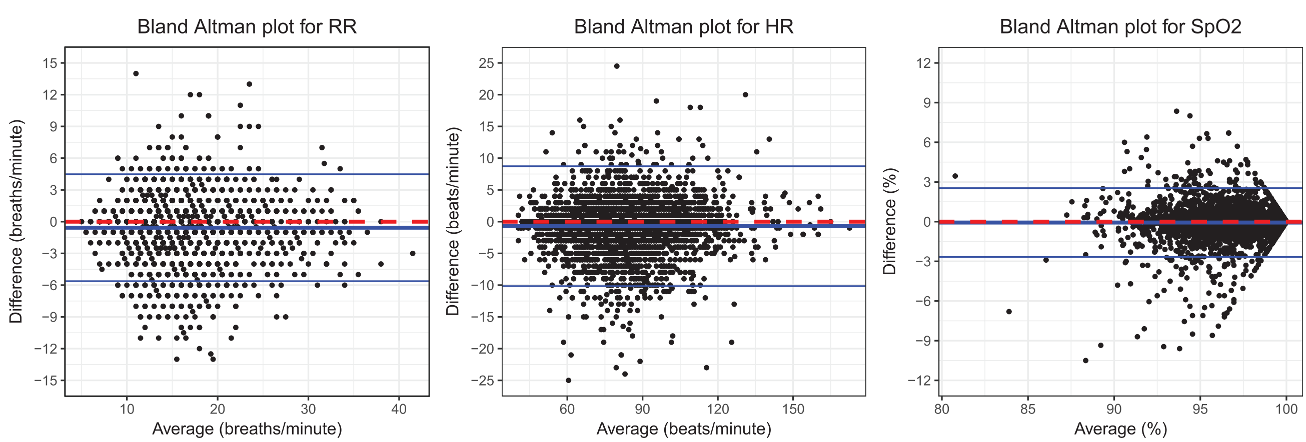

Figure 4. Bland–Altman plots for RR, HR, and Spo_2 showing limits of agreement between continuous data sampled before the observation and continuous data sampled during the observation ("before" data – "during" data). Bias and limits of agreement are shown with blue lines, and a dashed red line shows the zero y-intercept. HR indicates heart rate; RR, respiratory rate; Spo_2, oxygen saturation.

magnitude and frequency of extreme values in manual and automated vital sign measurements where the automated measurement had higher measurement rates.[9,10,12,22] Frequently sampled signals are more likely to capture transient extreme measurements than sparsely sampled signals, partly explaining the discrepancies found in these articles. Furthermore, 1 study compared manual and automated measures from different patient cohorts,[13] while others only presented data from automated measurements[8] or used simulated measurements from mannequins.[22]

In contrast to our Spo_2 findings, Taenzer et al[15] reported a difference of 6.5% between continuous and manual measurements of Spo_2 <90% in general medical and postoperative wards. The method used continuous data to group Spo_2 measurements >90% or <90%, thus comparing the differences between measurement sources to the values of the continuous source. Bland and Altman[25] have shown that this process introduces a false correlation to the data. Hence, we compared the differences to the average values as recommended.[25,26] For readers interested in this effect, we have replicated our analyses, comparing the difference against the continuous measure (and finding false correlations) in Supplemental Digital Content 2, Figure 1, http://links.lww.com/AA/C523.[25]

Manually recorded vital sign measurements varied widely from the continuous measurements. Respiratory rate is difficult to measure clinically,[2,3] so the high variance in the differences between continuous and manual measurements is perhaps not unexpected. However, as early warning scores commonly include respiratory rate ranges between 2 scores of 4 breaths/min or less,[19,27] these differences would commonly impact clinical care. The LoAs for heart rate and Spo_2 were also wide and again would commonly result in different early warning scores. These results are important for those seeking to automate early warning scores,[28,29] which have not been designed for use with continuous data, so patients would clearly alert differently.

Our study is limited by its single-center design because practices may vary between hospital wards and institutions. There may have been transcription errors in the value or timing of manual vital sign observations. This effect was minimized by double data entry and synchronizing observation times with computer-generated timestamps of blood pressure. One bedside monitoring provider was used in this study, so we could not assess differences between monitors. If we had found a relationship between the continuous-manual differences and averages, then this would have prevented exploration of whether the relationship could be explained by monitor inaccuracy, rather than clinician smoothing. Because there is no clinically significant relationship, this is not a significant issue for our findings. Our results are not influenced by undersampling because the measurement pairs of manual and continuous data sample the same physiology. The strengths of our article are the large dataset used in comparison to previous work and the analysis of 3 different vital signs.

We have provided evidence against the existence of a smoothing effect in postoperative care. However, this phenomenon may still exist in the context of anesthesia. It is possible that errors in vital sign documentation increase when clinicians record vital signs from memory after the measurement has been taken.[9,10] This may be more pervasive in acute episodes during surgery and may not apply to postoperative wards, where nursing staff are available to measure and document vital signs simultaneously. Further research should investigate the smoothing effect during anesthesia, using the Bland–Altman method of assessing agreement between measurement sources, and comparing manually recorded measures to the median of a continuous window. If confirmed, the effects of the large differences between manually recorded and continuous vital sign measurements on early warning scores require investigation before the measurement and recording of these variables can be safely automated for early warning score computation.

CONCLUSIONS

We found no evidence that a clinically significant smoothing effect exists for respiratory rate, heart rate, or Spo_2 in postoperative care. We found no evidence of an arousal effect caused by the vital sign taker. Differences between manually recorded and continuous measures of respiratory rate, heart rate, and Spo_2 were frequently large, suggesting that the methods cannot be used interchangeably.

ACKNOWLEDGMENTS

The authors thank several individuals for their work on the Computer Alerting Monitoring System 2 study. Breda Lynch, Theresa Saunders, Sarah Vollam, and Deborah Evans conducted data collection; Jacqueline Birks undertook the statistical analysis and assisted in protocol development; and Julie Darbyshire assisted in proof reading. The authors would also like to thank Dr Lei Clifton for her work in preparing the dataset. Furthermore, the study could not have been undertaken without the support of the clinical staff on the upper gastrointestinal ward. The authors would like to thank the participants of the Computer Alerting Monitoring System 2 study for their commitment to research.

DETAILS OF ACKNOWLEDGED INDIVIDUALS

Breda Lynch, Research Nurse, Oxford University Hospitals NHS Trust, Oxford, United Kingdom; Theresa Saunders, Research Nurse, Oxford University Hospitals NHS Trust, Oxford, United Kingdom; Deborah Evans, Research Nurse, Oxford University Hospitals NHS Trust, Oxford, United Kingdom; Sarah Vollam, Research Nurse, Oxford University Hospitals NHS Trust, Oxford, United Kingdom; Jacqueline Birks, MA, MSc, National Institute for Health Research–Oxford Biomedical Research Centre (NIHR–OXBRC) Senior Medical Statistician, Centre for Statistics in Medicine, Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford, Oxford, United Kingdom; Julie Darbyshire, BA, MA, MSc, Critical Care Research Programme Manager, Oxford University Hospitals NHS Trust, Oxford, United Kingdom; and Dr Lei Clifton, BSc, MSc, PhD, Medical Statistician, Centre for Statistics in Medicine, Botnar Research Centre, Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford, Oxford, United Kingdom.

DISCLOSURES

Name: Hamish R. Tomlinson, BE (Hons).
Contribution: This author helped conduct the analysis and prepare the manuscript.
Name: Marco A. F. Pimentel, PhD.
Contribution: This author helped with conception and design of the analysis and drafting the manuscript.
Name: Stephen Gerry, MSc.
Contribution: This author helped with conception and design of the statistical methodology, and drafting the manuscript.
Name: David A. Clifton, PhD.
Contribution: This author helped with data acquisition and drafting the manuscript.
Name: Lionel Tarassenko, PhD.
Contribution: This author helped with conception and design of the analysis, data acquisition, and drafting the manuscript.
Name: Peter J. Watkinson, MBChB, MD.

Contribution: This author helped with conception and design of the analysis, data acquisition, and drafting the manuscript.
This manuscript was handled by: Nancy Borkowski, DBA, CPA, FACHE, FHFMA.

REFERENCES

1. Leuvan CH, Mitchell I. Missed opportunities? An observational study of vital sign measurements. *Crit Care Resusc*. 2008;10:111–115.
2. Lovett PB, Buchwald JM, Stürmann K, Bijur P. The vexatious vital: neither clinical measurements by nurses nor an electronic monitor provides accurate measurements of respiratory rate in triage. *Ann Emerg Med*. 2005;45:68–76.
3. Hooker EA, O'Brien DJ, Danzl DF, Barefoot JA, Brown JE. Respiratory rates in emergency department patients. *J Emerg Med*. 1989;7:129–132.
4. Prytherch DR, Smith GB, Schmidt P, et al. Calculating early warning scores: a classroom comparison of pen and paper and hand-held computer methods. *Resuscitation*. 2006;70:173–178.
5. Clifton DA, Clifton L, Sandu DM, et al. "Errors" and omissions in paper-based early warning scores: the association with changes in vital signs: a database analysis. *BMJ Open*. 2015;5:e007376.
6. Subbe CP, Gao H, Harrison DA. Reproducibility of physiological track-and-trigger warning systems for identifying at-risk patients on the ward. *Intensive Care Med*. 2007;33:619–624.
7. Reich DL, Wax DB. Mind the gap. *Anesth Analg*. 2011;113:214–215.
8. Block FE Jr. Normal fluctuation of physiologic cardiovascular variables during anesthesia and the phenomenon of "smoothing." *J Clin Monit*. 1991;7:141–145.
9. Cook RI, McDonald JS, Nunziata E. Differences between handwritten and automatic blood pressure records. *Anesthesiology*. 1989;71:385–390.
10. Lerou JG, Dirksen R, van Daele M, Nijhuis GM, Crul JF. Automated charting of physiological variables in anesthesia: a quantitative comparison of automated versus handwritten anesthesia records. *J Clin Monit*. 1988;4:37–47.
11. Zollinger RM Jr, Kreul JF, Schneider AJ. Man-made versus computer-generated anesthesia records. *J Surg Res*. 1977;22:419–424.
12. Thrush DN. Are automated anesthesia records better? *J Clin Anesth*. 1992;4:386–389.
13. Reich DL, Wood RK Jr, Mattar R, et al. Arterial blood pressure and heart rate discrepancies between handwritten and computerized anesthesia records. *Anesth Analg*. 2000;91:612–616.
14. Sapo M, Wu S, Asgari S, et al. A comparison of vital signs charted by nurses with automated acquired values using waveform quality indices. *J Clin Monit Comput*. 2009;23:263–271.
15. Taenzer AH, Pyke J, Herrick MD, Dodds TM, McGrath SP. A comparison of oxygen saturation data in inpatients with low oxygen saturation using automated continuous monitoring and intermittent manual data charting. *Anesth Analg*. 2014;118:326–331.
16. Benchimol EI, Smeeth L, Guttmann A, et al; RECORD Working Committee. The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) statement. *PLoS Med*. 2015;12:e1001885.
17. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP; STROBE Initiative. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *Lancet*. 2007;370:1453–1457.
18. Kenward MG, Roger JH. Small sample inference for fixed effects from restricted maximum likelihood. *Biometrics*. 1997;53:983.
19. McGinley A, Pearse RM. A National Early Warning Score for acutely ill patients. *BMJ*. 2012;345:e5310.
20. Carstensen B, Simpson J, Gurrin LC. Statistical models for assessing agreement in method comparison studies with replicate measurements. *Int J Biostat*. 2008;4:Article 16.
21. Bland JM, Altman DG. Measuring agreement in method comparison studies. *Stat Methods Med Res*. 1999;8:135–160.
22. Devitt JH, Rapanos T, Kurrek M, Cohen MM, Shaw M. The anesthetic record: accuracy and completeness. *Can J Anaesth*. 1999;46:122–128.
23. Goodman SN, Berlin JA. The use of predicted confidence intervals when planning experiments and the misuse of power when interpreting results. *Ann Intern Med*. 1994;121:200–206.
24. Milner QJ, Mathews GR. An assessment of the accuracy of pulse oximeters. *Anaesthesia*. 2012;67:396–401.
25. Bland JM, Altman DG. Comparing methods of measurement: why plotting difference against standard method is misleading. *Lancet*. 1995;346:1085–1087.
26. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*. 1986;327:307–310.
27. Tarassenko L, Clifton DA, Pinsky MR, Hravnak MT, Woods JR, Watkinson PJ. Centile-based early warning scores derived from statistical distributions of vital signs. *Resuscitation*. 2011;82:1013–1018.
28. Subbe CP, Duller B, Bellomo R. Effect of an automated notification system for deteriorating ward patients on clinical outcomes. *Crit Care*. 2017;21:52.
29. Medtronic.com. Vital Sync™ Virtual Patient Monitoring Platform 2.6. Available at: http://www.medtronic.com/covidien/en-us/products/health-informatics-and-monitoring/vital-sync-virtual-patient-monitoring-platform-2–6.html. Published 2017. Accessed October 24, 2017.