Received: 24 December 2019 | Revised: 29 July 2020 | Accepted: 7 September 2020

DOI: 10.1111/jan.14584

REVIEW

JAN Leading Global Nursing Research WILEY

Quantitative systematic review: Sources of inaccuracy in manually measured adult respiratory rate data

Noa Kallioinen[1,2] | Andrew Hill[1,3,4] | Melany J. Christofidis[1,5] | Mark S. Horswill[1] | Marcus O. Watson[1]

[1]School of Psychology, The University of Queensland, St Lucia, QLD, Australia

[2]Institute of Cognitive Science, Osnabrück University, Osnabrück, Germany

[3]Clinical Skills Development Service, Metro North Hospital and Health Service, Herston, QLD, Australia

[4]Minerals Industry Safety and Health Centre, Sustainable Minerals Institute, The University of Queensland, St Lucia, Queensland, Australia

[5]Queensland Children's Hospital, Children's Health Queensland, South Brisbane, QLD, Australia

Correspondence
Andrew Hill, Clinical Skills Development Service, PO Box 470, Herston, QLD 4029, Australia.
Email: a.hill@psy.uq.edu.au

Funding information
The University of Queensland received research consultancy fees from the Clinical Skills Development Service, Metro North Hospital and Health Service, for some of N.K. and M.S.H.'s work.

Abstract

Aims: To identify the potential sources of inaccuracy in manually measured adult respiratory rate (RR) data and quantify their effects.

Design: Quantitative systematic review with meta-analyses where appropriate.

Data Sources: Medline, CINAHL, and Cochrane Library (from database inception to 31 July 2019).

Review Methods: Studies presenting data on individual sources of inaccuracy in the manual measurement of adult RR were analysed, assessed for quality, and grouped according to the source of inaccuracy investigated. Quantitative data were extracted and synthesized and meta-analyses performed where appropriate.

Results: Included studies ($N = 49$) identified five sources of inaccuracy. The *awareness effect* creates an artefactual reduction in actual RR, and *observation methods* involving shorter counts cause systematic underscoring. Individual RR measurements can differ substantially in either direction between observations due to *inter-* or *intra-observer variability*. *Value bias*, where particular RRs are over-represented (suggesting estimation), is a widespread problem. *Recording omission* is also widespread, with higher average rates in inpatient versus triage/admission contexts.

Conclusion: This review demonstrates that manually measured RR data are subject to several potential sources of inaccuracy.

Impact: RR is an important indicator of clinical deterioration and commonly included in track-and-trigger systems. However, the usefulness of RR data depends on the accuracy of the observations and documentation, which are subject to five potential sources of inaccuracy identified in this review. A single measurement may be affected by several factors. Hence, clinicians should interpret recorded RR data cautiously unless systems are in place to ensure its accuracy. For nurses, this includes counting rather than estimating RRs, employing 60-s counts whenever possible, ensuring patients are unaware that their RR is being measured, and documenting the resulting value. For any given site, interventions to improve measurement should take into account the local organizational and cultural context, available resources, and the specific measurement issues that need to be addressed.

KEYWORDS
documentation, nursing, quality of care, systematic reviews and meta-analyses, vital signs

1 | INTRODUCTION

Vital sign observations are a fundamental part of patient monitoring, and respiratory rate (RR), blood pressure, heart rate, temperature, and oxygen saturation are the most common metrics underlying monitoring protocols such as physiological track-and-trigger systems (Brekke et al., 2019). These protocols can support nurses and doctors in the early detection of patient deterioration and help facilitate rapid responses to adverse clinical situations (Mohammed et al., 2009; Prytherch et al., 2006; Subbe et al., 2007). There are several types of track-and-trigger system, which vary in complexity, but all essentially apply algorithms to routinely measured vital sign values to prompt appropriate actions, such as clinical escalation or increased monitoring frequency (Christofidis et al., 2015, 2016). An effective track-and-trigger system can also empower nurses to escalate by providing objective evidence of clinical deterioration to support and corroborate their more subjective or intuitive clinical assessments (Andrews & Waterman, 2005). The accuracy of vital sign data, therefore, can directly impact the usefulness of these systems, as well as potentially impacting clinical judgement and decision-making more broadly. Nevertheless, there have been few systematic reviews or meta-analyses on the topic of vital sign measurement accuracy. Exceptions are Kallioinen et al.'s (2017) investigation of the sources of inaccuracy in blood pressure measurement, and Tysinger's (2015) more general examination of the accuracy of vital sign data. Tysinger's (2015) review presented an overview of inaccuracies arising at the measurement and documentation stages. However, it did not focus specifically on RR and, unlike Kallioinen et al.'s methodology, the review protocol was not designed to comprehensively collate the research evidence for all unique sources of inaccuracy. Consequently, the results were confined to six key papers, which contrasts sharply with the 328 publications included in Kallioinen et al.'s review, despite its narrower focus on a single vital sign.

Results from systematic reviews on the topic of vital sign accuracy can help inform the interpretation and use of vital sign data, the design of track-and-trigger systems, and the formulation of clinical guidelines and educational interventions, potentially leading to improvements in the quality of patient care. However, to date, no published systematic review has presented a comprehensive evaluation of the sources of inaccuracy in RR data, although it is known to be both an important clinical indicator and subject to significant measurement issues (Lovett et al., 2005).

2 | BACKGROUND

Respiratory rate has been identified as a strong indicator of patient condition. In addition to associations with mortality (Barthel et al., 2013; Bleyer et al., 2011; Ma et al., 2011; Sinnecker et al., 2014; Strauß et al., 2014; Subbe et al., 2003), abnormal RRs are associated with cardiac arrest (Fieselmann et al., 1993; Schein et al., 1990; Subbe et al., 2003), sequential organ failure (Kenzaka et al., 2012), and escalation of care (Cardoso et al., 2014; Considine et al., 2009). There is also some evidence that, when patients' vital signs are taken in hospital wards, RR is the strongest predictor of subsequent clinical deterioration (Churpek et al., 2016).

Since RR is a useful clinical indicator, it is usually incorporated into physiological track-and-trigger systems (Gao et al., 2007; Smith et al., 2008; The ANZICS-CORE MET dose Investigators, 2012). However, studies investigating its measurement accuracy have brought into question the trustworthiness of routinely recorded RR values (e.g. Lovett et al., 2005; Philip et al., 2015). In the present paper, we therefore set out to fill the research gap identified in the Introduction by systematically reviewing the potential sources of inaccuracy in RR data evidenced in the literature.

Despite advances in automated vital sign measurement, RR remains the vital sign most commonly measured without the use of an automated device (Ansell et al., 2014; Churpek et al., 2018). For this reason, the present review focuses exclusively on the accuracy of manually measured RR data. Broadly speaking, the manual measurement process can be considered to be made up of two discrete stages: observation and documentation. Observation involves counting a patient's breaths, through visual inspection or auscultation, to determine their RR in breaths per minute. Best practice is considered to be a 60-s count using a watch or timing device (World Health Organisation, 1992, 1993). Documentation refers to the recording of the patient's RR in the appropriate region of their observation chart or electronic record. Factors that may impact the trustworthiness of RR data recorded in a clinical chart or electronic system can be associated with either of these stages and both were considered in this review.

3 | THE REVIEW

3.1 | Aims

The aims of the present study were to identify potential sources of inaccuracy in manually measured adult RR data and to quantify their effects.

3.2 | Design

A systematic literature review was conducted using *Medline* and *CINAHL* databases (via *EBSCOHost)* and the *Cochrane Library*. Due to the broadness of the research question and the diversity of study methodologies found, meta-analyses were performed only for selected sources of inaccuracy on a case-by-case basis. Details of the studies included in the review are synthesized in tabulated form and discussed in the Results section.

3.3 | Search methods

The databases were searched from inception to 31 July 2019. A broad search strategy aiming to find all English-language publications related to inaccuracies in the measurement of adult RR data was employed. The inclusion criteria were adapted from those of Kallioinen et al. (2017), and studies were eligible for inclusion if they contained all of the following:

1. Results from an empirical study relevant to the manual measurement of adult patients' RR in clinical settings;
2. Identification of at least one specific potential source of inaccuracy in the observation or documentation of RR; and
3. In relation to each source of inaccuracy identified, quantification of its prevalence or of its independent effect on documented RR values.

The database searches and their results are summarized in Table 1.

3.4 | Search outcome

The final *CINAHL*, *Medline*, and *Cochrane Library* database searches yielded a total of 7,514 results (i.e. S11 in Table 1). Figure 1 presents the subsequent study selection procedure. After 695 duplicate records were excluded, 6,819 unique records remained. Of these, a further 6,746 were excluded based on review of their titles or abstracts. Full texts were retrieved for the remaining 73 papers, as well as 11 further publications derived from the reference lists. After all 84 of these full texts were reviewed, 35 were excluded due to failure to meet the inclusion criteria for reasons such as an inappropriate sample (e.g. child patients), lack of quantitative data, or confounding of multiple sources of inaccuracy, such that independent effects could not be assessed. The remaining 49 studies were included.

3.5 | Quality appraisal

The standard quality assessment criteria for evaluating primary research papers from a variety of fields (Kmet et al., 2004) were used to assess the quality of the empirical studies included in the final review. Total scores derived from these criteria represent overall study quality taking into consideration the methodology, adequacy of reporting, and risk of biased results. However, due to variations

TABLE 1 Searches, search terms, limiters, and number of results yielded (MEDLINE, CINAHL, and Cochrane Library databases)

Search	Search terms	Limiters	Results
S11	S6 OR S7 OR S10	ALL: English Language CINAHL: exclude MEDLINE records	7,514
S10	S1 OR S9	ALL: English Language CINAHL: exclude MEDLINE records	4,415
S9	S2 AND S8	ALL: English Language CINAHL: exclude MEDLINE records	1,177
S8	S3 OR S4 OR S5	ALL: English Language CINAHL: exclude MEDLINE records	32,485
S7	TI ("vital sign*" OR "vital parameter*" OR "respiratory rate*" OR "breathing rate*" OR "respiration rate*" OR "patient assess*" OR "observation chart*" OR "early warning") AND ("respiratory rate*" OR "breathing rate*" OR "respiration rate*") AND ("measure*" OR "error*" OR "document*" OR "record*" OR "aware*" OR "bias" OR "observ*" OR "assess*" OR "neglect*" OR "missing*" OR "inaccur*" OR "accura*")	ALL: English Language CINAHL: exclude MEDLINE records	1,126
S6	("respiratory rate" OR "respiration rate" OR "breathing rate") AND ("measurement" OR "error" OR "documentation" OR "bias")	ALL: English Language CINAHL: exclude MEDLINE records	3,058
S5	(MH "Respiratory Function Tests/IS/MT/NU/ST")	ALL: English Language CINAHL: exclude MEDLINE records	4,254
S4	(MH "Triage/MT/ST")	ALL: English Language CINAHL: exclude MEDLINE records	3,968
S3	(MH "Monitoring, Physiologic/IS/MT/NU/ST")	ALL: English Language CINAHL: exclude MEDLINE records	24,459
S2	(MH "Respiration")	ALL: English Language CINAHL: exclude MEDLINE records	64,061
S1	(MH "Respiratory Rate")	ALL: English Language CINAHL: exclude MEDLINE records	3,283

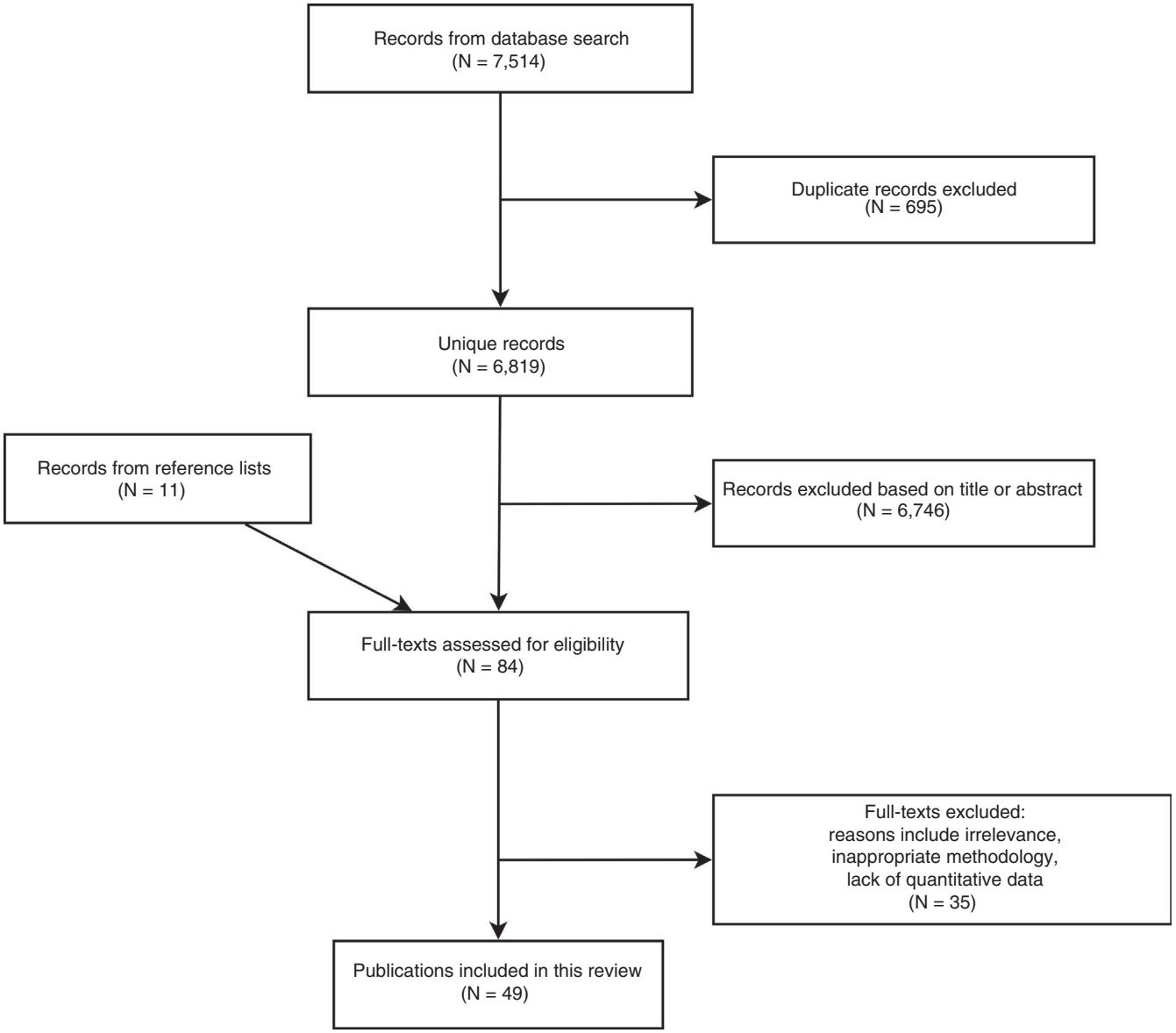


FIGURE 1 Study selection flow chart

in study design and methodology, the quality indicators for studies addressing different sources of inaccuracy are not directly comparable.

3.6 | Data abstraction

One reviewer (NK) conducted the initial analysis of the 84 potentially relevant full-text articles. For each study that met the inclusion criteria, details of the study design, observers, patients, and results were collected and tabulated, as was further information specific to the relevant source of inaccuracy. In cases where a single study investigated both manual and device-assisted measurement of RR, only data relevant to manual measurement were extracted. The quality appraisals were also conducted at this stage and the results were tabulated. The entire process of data abstraction (including quality appraisal) was then independently audited by another reviewer (MJC) and disagreements were resolved through discussion until complete agreement was reached.

3.7 | Synthesis

Findings related to each independent source of inaccuracy were tabulated separately and trends were investigated. Discussion of individual studies and trends is presented in a narrative form, while quantitative data, overall results from the quality appraisals, and further details of each study are presented in Tables 2–6. Full data from the quality appraisals can also be found in the Tables S1–S5. Due to the scarcity of comparable studies for some potential sources of inaccuracy, meta-analytic techniques were performed on a case-by-case basis for specific sources only. All such statistical analyses were

conducted in *R* (R Core Team, 2018) using the *metafor* package for meta-analysis techniques (Viechtbauer, 2010).

4 | RESULTS

4.1 | Study characteristics

The 49 studies included in this review were from 16 different countries. Most were from Australia, the UK, or the USA and investigated the RRs of medical or surgical patients in health-care centres. A minority involved either healthy volunteers or mock patients.

Studies were categorized as investigating either (a) the observation of RR or (b) its documentation, with some publications addressing one aspect of each topic or multiple aspects of either one. Within these broad categories, the studies yielded evidence for a total of five distinct sources of inaccuracy in RR data. These included three sources of inaccuracy introduced at the observation stage, namely (a) the observation method employed, (b) inter- and intra-observer variability in RR measurements, and (c) the patient's awareness of being observed (similar to the white-coat effect in blood pressure measurement; see Kallioinen et al., 2017). The remaining sources of inaccuracy related to the documentation stage were (d) value bias and (e) recording omission.

4.2 | Observation of respiratory rate

Thirteen empirical studies reported the quantitative effects of one or more sources of inaccuracy potentially introduced into manually measured RR data at the observation stage. Substantial effects of *observation method* and the *awareness effect* were apparent, along with a smaller effect of *inter-observer variability*.

4.2.1 | Observation method

Nine studies investigated the impact of one or more 'usual care' methods for taking RR observations (e.g. 15 s or 30 s counts) by comparing the accuracy of data obtained using those methods versus a criterion standard (Bianchi et al., 2013; Hill et al., 2018; Hooker et al., 1989; Lovett et al., 2005; Nielsen et al., 2015; Philip et al., 2015; Rimbi et al., 2019; Takayama et al., 2019; Worster et al., 2003) (Table 2 & Table S1).

In the studies by Philip et al. (2015) and Nielsen et al. (2015), observers were presented with a series of videos showing mock patients breathing at a range of different RRs (6, 30, and 72 breaths/min; and 5, 10, 15, 30, and 60 breaths/min respectively). Usual care measurements were then compared with these pre-determined criterion values (Table 2). However, the pattern of mean differences across criterion values was inconsistent between the two studies and no significance testing was conducted. In all other studies, criterion values were derived from observations by experienced or trained observers using standardized methods (typically 60 s counts) conducted in real time, or by video analysis.

Five studies reported significance testing on the mean difference between usual care and criterion standard measurement. Rimbi et al. (2019), Takayama et al. (2019), Hill et al. (2018), and Hooker et al. (1989) all compared RRs derived from 15 s counts (multiplied by four) with criterion standard 60 s counts. Three of these studies (Hill et al., 2018; Hooker et al., 1989; Takayama et al., 2019) reported that 15 s counts significantly underestimated RR by around 2 breaths/min on average. Rimbi et al. (2019) reported a smaller mean difference for nurses whose manual observations were supported by a mobile application that they tapped each time they observed an inspiration, and which totalled the breaths automatically. Although this mean difference was statistically non-significant, the researchers nevertheless found that, for patients with severely abnormal RRs in particular, the 15 s counts led to a substantial incidence of underscoring within their track-and-trigger system, resulting in potential failures to identify at-risk patients. In two of these studies (Hill et al., 2018; Rimbi et al., 2019), 30 s counts were also compared against the criterion-measured counts, yielding the same general pattern of results but with RR underestimated to a lesser degree. In the fifth study, Worster et al. (2003) found no significant difference between an unspecified 'standard practice' assessment and a less rigorous criterion measure based on 30 s auscultation.

However, mean differences only indicate potential systematic biases, not the extent to which an individual usual care observation may vary from a criterion standard. For this, the *limits of agreement* can be informative. Wider limits of agreement indicate greater variation in the association between usual care and criterion standard observations. Among the nine studies, the widest reported range of 95% limits of agreement for usual care observations was −30.9–+20.0 (Philip et al., 2015). Since the relevant criterion value was known to be 72 breaths/min, this indicates that 95% of all usual care observations were somewhere in the broad range between 41.1 and 92 breaths/min. On the other hand, the narrowest range was −0.49–+1.83 for a criterion value of 30 breaths/min (Nielsen et al., 2015), which bucked the general trend for the limits of agreement to be wider for higher known values in a given study (Nielsen et al., 2015; Philip et al., 2015). Philip et al. (2015) also evaluated the accuracy of "spot check" estimates (opportunities for 12 s counts without a timer) and found that the limits of agreement were much wider than when observers had the opportunity to conduct longer usual care counts of up to 60 s. This, combined with the demonstrated tendency for short observation periods to bias RRs towards underestimation (Hill et al., 2018; Takayama et al., 2019), provides clear evidence that a full 60 s count is necessary to determine an accurate RR.

Only three of the studies reported information on the experience level of the observers; however, patient characteristics were consistently well reported. The results indicate that manual RR measurements are highly variable between and within observation methods; however, without knowing the characteristics of the observers, it is

TABLE 2 Studies reporting effects of observation method on respiratory rate data

Authors (Year) Country	Observers (*N*, mean age, M/F), experience	Patients (*N*, mean age, M/F)	Observation method	Criterion standard (mean value)	Mean deviation from criterion (95% limits of agreement)	Sig.	Study Quality
Bianchi et al. (2013) USA	Triage providers (NR, NR, NR), no experience information	Emergency patients with acuity level 2 to 5 (191, median = 43 y, 107/84)	15 s count; if not sufficient, then 60 s count	WHO standard measurement by 1 trained researcher through 60 s observation or 60 s auscultation (NR)	+0.3 (−8.0 to +8.3) estimated from Bland-Altman plot	NR	73%
Hill et al. (2018) Australia	Raters measuring from video recording (2, NR, NR), trained in rating procedure, unaware of study goal	Healthy population (41, 20.07 y, 8/33); Healthy population (41, 19.51 y, 8/33)	15 s count 30 s count	60 s count by the same observers (15.39 breaths/min)	−2.19 −0.95	<0.0001 <0.0001	95%
Hooker et al. (1989) USA	Triage nurses (NR, NR, NR)	Triage patients (110, 38 ± 17 y, 57/53)	15 s count	60 s count by medical students (20.1 breaths/min)	−1.7 (NR)	<0.0001	90%
Lovett et al. (2005) USA	Triage nurses (NR, NR, NR), no experience information, aware of study goal	Triage patients (135, range = 18–89 y, 74/81 + 4 unspecified)	Standard triage assessment	WHO standard measurement by 7 trained research assistants by 60 s auscultation or 60 s observation if auscultation not possible (18.9 breaths/min)	+0.45 (−8.6 to +9.5)	NR	85%
Nielsen et al. (2015) Denmark	15 Nurses and 3 nursing assistants (NR, median = 42 y, NR), median 18 y of experience	Mock patient (1, 30 y, 1/0)	Opportunity for 60 s count	known value of mock patient RR (5 breaths/min)	0.33 (−1.01 to +1.68)	NR	95%
				known value of mock patient RR (10 breaths/min)	−1.61 (−2.98 to − 0.24)	NR	
				known value of mock patient RR (15 breaths/min)	−4.89 (−6.65 to − 3.13)	NR	
				known value of mock patient RR (30 breaths/min)	+0.67 (−0.49 to +1.83)	NR	
Philip et al. (2015) UK	Doctors (54, NR, NR) 18 with <1 y of experience, 20 with 2–10 y of experience, 12 with >10 y of experience	Mock patient (1, NR, NR)	30 s count or 60 s count, as per usual practice	known value of mock patient RR (6 breaths/min)	+2.46 (−3.2 to +8.1)	NR	91%
				known value of mock patient RR (30 breaths/min)	−0.02 (−11.7 to +11.6)	NR	
				known value of mock patient RR (72 breaths/min)	−5.43 (−30.9 to +20.0)	NR	
			'Spot' check (12 s estimation with no timer)	known value of mock patient RR (6 breaths/min)	+4.42 (−3.9 to +12.8)	NR	
				known value of mock patient RR (30 breaths/min)	−0.28 (−24.8 to +24.3)	NR	
				known value of mock patient RR (72 breaths/min)	−19.18 (−67.2 to +28.9)	NR	
				known value of mock patient RR (60 breaths/min)	+4.39 (−6.50 to +15.27)	NR	

(Continues)

TABLE 2 (Continued)

Authors (Year) Country	Observers (*N*, mean age, M/F), experience	Patients (*N*, mean age, M/F)	Observation method	Criterion standard (mean value)	Mean deviation from criterion (95% limits of agreement)	Sig.	Study Quality
Rimbi et al. (2019) Uganda	Nurses [aided by mobile app] (2, NR, NR)	Acutely ill medical patients at hospital admission (321, 49.6 y, 126/195) [770 total recordings]	15 s count	60 s count by the same observers (18.1 breaths/min)	−1.22 (−7.16 to +4.72)	ns	100%
			30 s count		−0.46 (−3.89 to +2.97)	ns	
Takayama et al. (2019) Japan	Nurses (57, 40.3 y, 5/52), mean 19.1 y of experience	Mock patient (1, 48 y, 1/0)	15 s count	60 s count by the same observers (24.5 breaths/min)	−2.1 (−7.7 to 3.5)	< 0.001	91%
			Measure duration of one breath and divide into 60		+0.5 (−4.5 to +5.5)	< 0.001	
Worster et al. (2003) Canada	Full-time emergency nurses (6, NR, NR), trained in triage, unaware of study goal	Triage patients (72, 39.46 y, 34/38)	'Standard practice' assessment	Measurement by 1 physician investigator by 30 s auscultation (NR)	+0.36 (NR)	ns	75%

Abbreviation: NR, not reported.

difficult to fully establish the underlying causes, given the somewhat inconsistent results across studies (Table 2). Although meta-analysis on limits of agreement is possible (see Tipton & Shuster, 2017), it was not considered appropriate in this instance due to the varying definitions of both the criterion standards and comparison measurements in the studies.

4.2.2 | Inter- and intra-observer variability

Two studies investigated *inter*-observer variability (Dinh et al., 2013; Lim et al., 2002) and/or *intra*-observer variability (Lim et al., 2002) in RR measurements (Table 3 & Table S2). Given the small number of studies and their divergent methodologies, meta-analysis was not considered appropriate.

Dinh et al. (2013) reported a significant difference between RRs measured first by Emergency Medical Services clinicians and subsequently by Emergency Department staff. However, the mean magnitude of this difference was smaller than 1 breath/min. Lim et al. (2002) made three different comparisons: two simultaneous observers (nurses vs. study investigator), two observations by different observers (both nurses) taken 15 min apart, and two observations by the same nurses separated by 15 min. The mean differences were all small (≤0.1 breath/min), indicating no systematic directional difference and their statistical significance was not reported. However, once again, the limits of agreement provide further information. Across both studies, the widest 95% limits of agreement were −12–+10 (between Emergency Medical Services and the Emergency Department; Dinh et al., 2013), while the narrowest were −4.2–+4.4 (for two simultaneous observers; Lim et al., 2002).

It should be noted that the differences reported in these two studies cannot be attributed entirely to the observers themselves. Dinh et al.'s (2013) study and two of Lim et al.'s (2002) three conditions investigated observations that were taken at different times. Only in one condition of Lim et al.'s (2002) study was variability between simultaneous observers examined, yielding the most consistent measurements of all. Thus, inter- and intra-observer variability may represent a combination of factors, including variability in the performance of individual observers, context-related differences, and genuine changes or variability in RR over time. Regardless of the precise underlying factors, it is clear that RR measurements can vary substantially between subsequent or simultaneous observations.

4.2.3 | Awareness effect

Three studies reported the effect of awareness of measurement on individuals' RRs (Han et al., 1997; Hill et al., 2018; Western & Patrick, 1988; Table 4 and Table S3). Hill et al. (2018) investigated the RRs measured manually from video recordings of healthy volunteers. The other two studies employed automated devices to measure RR. Nevertheless, these studies were still regarded as meeting the criterion of being *relevant* to the manual measurement of RR. This is

TABLE 3 Studies reporting inter- and/or intra-observer variability in respiratory rate measurements

60.375 pt	Observers (*N*, age, M/F)	Patients (*N*, age, M/F)	Comparison and mean values	Comparison type	Mean deviation (95% limits of agreement)	Sig.	Study Quality
Dinh et al. (2013) Australia	Ambulance service personnel/ emergency department staff (NR, NR, NR)	Adult emergency patients arriving via emergency services (1,181, 43 y ± 20, 825/356)	Emergency Medical Services (19 breaths/min) versus Emergency Department (18 breaths/min)	Inter-observer	−0.55 (−12 to +10)	<0.001	95%
Lim et al. (2002) UK	Nurses/ study investigator (NR, NR, NR)	Adult medical ward or lung function department patients (245, 68.2 y ± 18, 139/106)	Nurse (22.1 breaths/min) versus study investigator simultaneously (22.0 breaths/min) [*n* = 49]	Inter-observer	+0.1 (−4.2 to +4.4)	NR	95%
			Nurse (20.9 breaths/min) versus different nurse 15 min later (20.9 breaths/min) [*n* = 58]	Inter-observer	0 (−5.7 to +5.7)	NR	
			Nurse (24.1 breaths/min) versus same nurse 15 min later (24.1 breaths/min) [*n* = 136]	Intra-observer	+0.04 (−4.9 to +4.9)	NR	

Abbreviation: NR, not reported.

because individuals' *actual* RRs were expected to be affected by this source of inaccuracy and hence it would have an impact on RR data regardless of whether the measurements were taken automatically or manually.

In the study by Western and Patrick (1988), participants' awareness of their own respiration was heightened by asking them to count their breaths in threes, while in the studies by Hill et al. (2018) and Han et al. (1997), participants were explicitly told that their RRs were being observed or recorded. RRs recorded during these periods were compared with those recorded without intervention.

Hill et al. (2018) reported a significant decrease in mean RR (−2.13 breaths/min) when patients were aware of the measurement. Han et al. (1997) and Western and Patrick (1988) both reported significant increases in inspiratory time and Han et al. (1997) reported a significant increase in expiratory time, when awareness of breathing was heightened. In combination, the mean changes in inspiratory and expiratory time in these two studies corresponded to mean decreases in RR of −2.4 and −1.1 breaths/min respectively. Hence, all of the available evidence, regardless of whether RR was measured manually or automatically, suggests that awareness of observation causes a decrease in measured RR as an artefact of the observation process, such that it becomes less reflective of the patient's clinical condition. However, no meta-analysis was conducted on these studies due to their small numbers and diverse methodologies.

4.3 | Documentation of respiratory rate

Thirty-seven publications that identified sources of inaccuracy related to the documentation of RR were included. Two specific sources were apparent: *value bias* and *recording omission*.

4.3.1 | Value bias

It has been suggested that particular values are often over-represented in manually measured RR data because they represent estimates rather than actual measurements (Badawy et al., 2017; Cooper et al., 2013; Keene et al., 2017; Semler et al., 2013). Eight studies reported on values that appeared to be over-represented among recorded RRs (Badawy et al., 2017; Bianchi et al., 2013; Cooper et al., 2013; Granholm et al., 2016; Keene et al., 2017; Mukkamala et al., 2008; Pedersen et al., 2018; Semler et al., 2013) (Table 5 and Table S4). Only three of these studies provided direct evidence of value bias by comparing the recorded prevalence of over-represented values with criterion-measured comparison data (Bianchi et al., 2013; Mukkamala et al., 2008; Semler et al., 2013). For example, Semler et al. found that, in combination, values of 18 or 20 breaths/min accounted for 71.8% of recorded RRs—significantly more than the 13.0% indicated by the corresponding criterion-measured data. This means that *unsubstantiated* instances of these two values accounted for 58.8% of all recorded RRs. Similar patterns of over-representation were reported by Mukkamala et al. (for 20 or 18 breaths/min) and Bianchi et al. (for 16 or 18 breaths/min). Although significance tests were not conducted, *unsubstantiated* instances of two values again represented over 50% of RRs in each study.

In studies that lacked criterion-measured comparison data, weaker indirect evidence of value bias was presented in the form of clusters of particular values that were often seen as inherently unlikely (e.g. Cooper et al., 2013) or else compared with expected distributions and found to deviate (i.e. Badawy et al., 2017; Pedersen et al., 2018).

We conducted a meta-analysis on frequency data from all eight studies (employing an unweighted random effects model

TABLE 4 Studies reporting the effect of awareness on respiratory rate

Authors (Year) Country	Measurement method	Patients (N, age, M/F)	Awareness comparison	Inspiratory time difference (seconds)	Sig.	Expiratory time difference, (seconds)	Sig.	Approximate mean breath/min difference	Sig.	Study Quality
Han et al. (1997) Belgium	Automated device: Respitrace (inductance plethysmography)	Healthy population (74, 33 y, range = 21–63 y, 34/40)	Told machine is recording versus not told machine is recording	+0.2 (2.1 s vs. 1.9 s)	<0.001	+0.2 (2.8 s vs. 2.6 s)	<0.001	−1.1 (12.2 vs. 13.3)	N/A	89%
Hill et al. (2018) Australia	Manual measurement from video recordings	Healthy population (41, 20.07 y, 8/33); healthy population (41, 19.51 y, 8/33)	Aware of respiratory rate monitoring by researcher versus unaware	NR	N/A	NR	N/A	−2.13 (13.28 vs. 15.40)	<0.0001	100%
Western and Patrick (1988) UK	Automated device: *Respitrace* (inductance plethysmography)	Healthy population (18, range = 17–59 y, 18/0)	Asked to count breaths in threes versus not asked	+0.36 (1.90 s vs. 1.54 s)	<0.01	+0.36 (2.68 s vs. 2.32 s)	ns	−2.4 (13.1 vs. 15.5)	N/A	82%

Abbreviation: NR, not reported.

using the double arcsine transformation; Freeman & Tukey, 1950; Miller, 1978). As each study presented data from a different hospital with a different population of clinicians and patients, an unweighted model, which does not weight studies by sample size, was considered the most appropriate. This type of model most adequately allows for generalization to future studies with different samples (Hall & Rosenthal, 2018).

Since several studies (such as Semler *et al.*) only reported combined frequencies for two RR values, we analysed the combined percentage frequency of the two most commonly recorded values in each study. The meta-analysis indicated an overall predicted proportion of 71.5% [95% CI: 63.2%, 79.1%; 95% Prediction Interval: 46.2%, 91.1%; $\tau^2 = 0.016$] (see Figure 2). Although it is impossible to know how many of these values were the result of value bias, it is worth noting that the frequencies reported in the three studies that did include criterion-measured comparison data (i.e. Bianchi et al., 2013; Mukkamala et al., 2008; Semler et al., 2013) all fell neatly within the 95% confidence interval (range = 68.7–75.4%), suggesting that they were not atypical of the studies as a whole.

The two specific RR values that were the most common varied from study-to-study, suggesting local cultural or systemic influences. Across the eight studies, the values frequently included 20 or 18 breaths/min, with six instances each, followed by 16 breaths/min with three instances. Appearing only once was 15 breaths/min, a value that is notable in two ways. First, unlike the other common values, it cannot be derived from a 30 s count (or a 15 s count, unlike 20 or 16 breaths/min), making its high prevalence all the more implausible. Second, the two most common values in this study (15 and 20 breaths/min) were the lowest and highest values that receive a score of 1 on the relevant hospital's track-and-trigger system, which may have contributed to their local over-representation.

Taken together, these studies (especially those that employed criterion-measured comparisons) provide clear support for the suggestion that there is a tendency to bias RR data by recording values that are thought to be common or 'normal' (e.g. Badawy et al., 2017; Semler et al., 2013). However, no study reported detailed characteristics of the observers, such as experience. This impedes attempts to pinpoint the reasons for such a tendency and potential targets for interventions.

4.3.2 | Recording omission

Thirty-two studies were identified that reported frequencies for the omission of RRs from patient charts or electronic records. Two of the studies were large-scale audits (Table 6 & Table S5). Ramgopal et al. (2018) assessed vital sign data recorded during patient transit by Emergency Medical Services staff from 20 different agencies over a 9-month period and reported an omission rate of only 1.85% for RR among 346,863 records. In an even larger study, Pedersen et al. (2018) audited the electronic records of ward-based inpatients at 11 hospitals and reported an omission rate of 4.39% among 2.84 million records made over a 12-month period.

TABLE 5 Studies reporting potential value bias in the documentation of respiratory rate

Authors (Year) Country	Observers (*N*, age, M/F)	Criterion measurement	Patients (*N*, age, M/F)	Time of measurements	Total number of recordings	Most frequently reported values (breaths/min)	Frequency of values (vs. criterion-measured data, where available)	Proportion of values (vs. criterion-measured data, where available)	Sig.	Study Quality
Badawy et al. (2017) USA	Hospital staff (NR, NR, NR)	No criterion measurement	Non-ICU hospitalisations (36,966, 61.7 y, 16902/20064)	During admission	220,665	18 or 20	165,499	75.0%	N/A	89%
Bianchi et al. (2013) USA	Triage providers (NR, NR, NR)	60-s count by investigator	Emergency patients with acuity level 2 to 5 (191, median = 43 y, 107/84)	During triage	191	16 or 18	144 (vs. 39)	75.4% (vs. 20.4%)	NR	73%
Cooper et al. (2013) Australia	Hospital staff (NR, NR, NR)	No criterion measurement	Medical/surgical patients (474, 66.5 y, NR)	During admission	464	16 18	186 198	40.1% 42.7%	N/A N/A	89%
Granholm et al. (2016) Denmark	Hospital staff (NR, NR, NR)	No criterion measurement	Ward patients (50, 71.5 y, 23/27)	During ward rounds	289	16 18 20	44 62 117	15.2% 21.5% 40.5%	N/A N/A N/A	89%
Keene et al. (2017) South Africa	Hospital staff (NR, NR, NR)	No criterion measurement	Acute trauma ward patients (181, NR, NR)	6:00 a.m. vitals on morning after ward admission	181	15 20	48 106	26.5% 58.6%	N/A	89%
Mukkamala et al. (2008) USA	Nurses (NR, NR, NR)	60-s count by medical students	General medicine patients (NR, NR, NR)	Entire day shifts in a 5 day period	467	20 18	234 (vs. 13) ~87[a] (vs. ~64[a])	50.1% (vs. 2.8%) ~18.6%[a] (vs. ~13.7%[a])	NR NR	89%
Pedersen et al. (2018) Denmark	Hospital staff (NR, NR, NR)	No criterion measurement	Ward patients (168,496, median = 60 y, 77508/90988)	During ward rounds	2,710,946	16 18 20	~797,000[a] ~461,000[a] ~464,000[a]	~29.4%[a] ~17.0%[a] ~17.1%[a]	N/A	83%
Semler et al. (2013) USA	Hospital staff (NR, NR, NR)	60-s count by resident physicians	Internal medicine patients (361, NR, NR)	Immediately prior and subsequent to audit measurement	361	18 or 20	259 (vs. 47)	71.8% (vs. 13.0%)	<0.001	89%

Abbreviation: NR, not reported.

[a]Value extracted from graph.

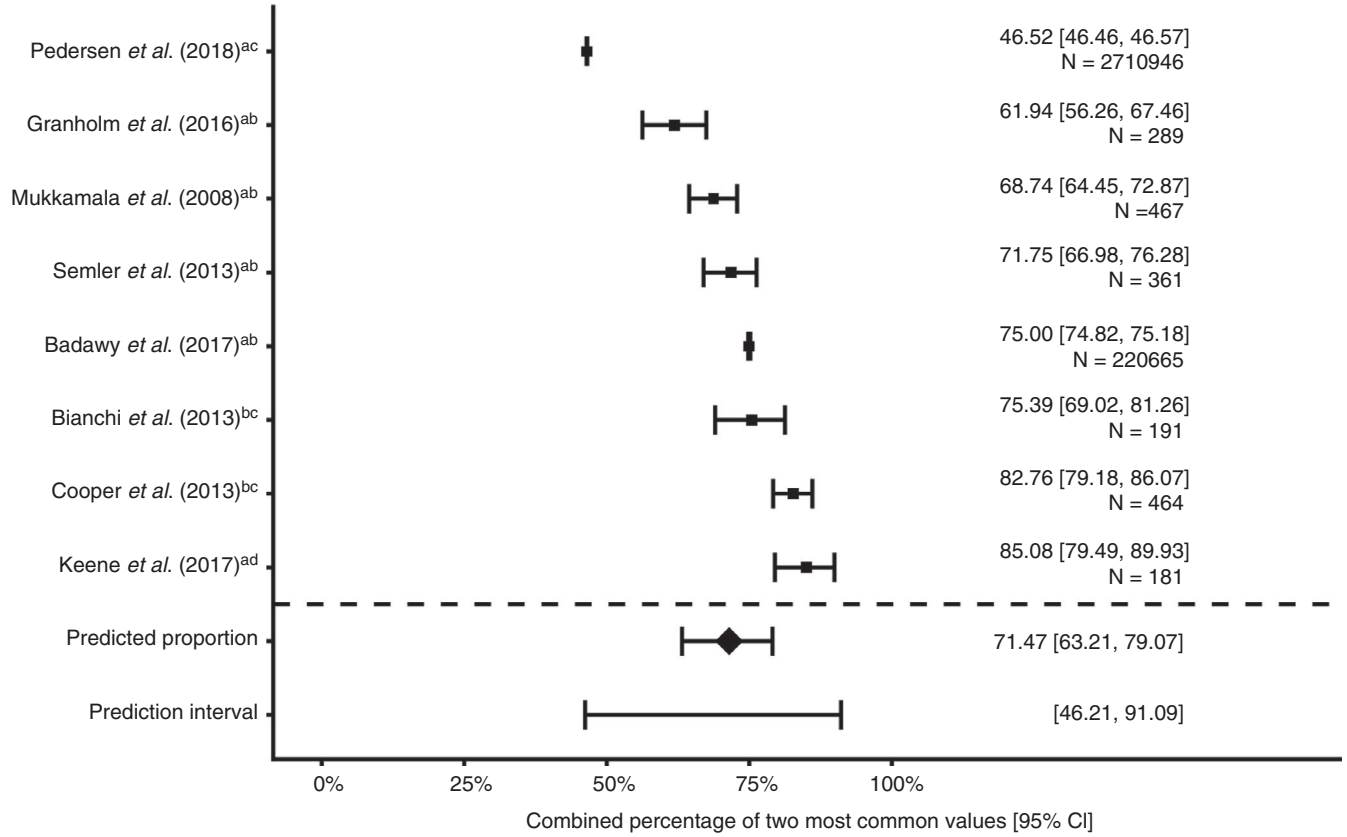


FIGURE 2 Over-represented respiratory rate values: Forest plot of combined frequencies for the two most commonly-recorded values in each study (For each study, superscripts represent the two relevant values: $a = 20$, $b = 18$, $c = 16$, $d = 15$ breaths/min.)

However, the 13 studies that reported smaller-scale general audits of hospital inpatients' vital sign documentation yielded omission rates ranging from 9.4–100% (excluding rates obtained after interventions; Table 6 and Table S5). Among these, five studies employed 12, 24, or 48 hr audit periods (Cretikos et al., 2007; Hall et al., 2003; McBride et al., 2005; Odell et al., 2007; Van Leuvan & Mitchell, 2008), four used longer audit periods of at least 1 week (Cahill et al., 2011; Edwards & Murdin, 2001; Helliwell et al., 2002; McGain et al., 2008), and three examined the entirety of patients' stays in the ward (Rosen et al., 2015; Smith & Oakey, 2006) or the emergency department (Parkes, 2011). In addition, a single point prevalence study examined RR omissions during night-time shifts compared with daytime shifts across 41 intensive care units but found no significant diurnal variation (Sundararajan et al., 2016). Of the 14 studies that reported hospital-based general audits, the 12 that reported both the total number of possible recordings and the number of omissions (i.e. all except McGain et al., 2008 and McBride et al., 2005) were included in a meta-analysis of proportions, using the same method as previously. The predicted proportion of omissions based on this meta-analysis was 58.1% [95% CI: 41.0%, 74.2%; 95% Prediction Interval: 1.9%, 100%, $\tau^2 = 0.129$] (Figure 3a).

Nine studies examined recordings made in the time leading up to particular patient events, yielding omission rates ranging from 17.3–99.1% (again excluding those obtained after interventions; Table 7 & Table S6). These included adverse events in general (Chen et al., 2009; Cretikos et al., 2007; MERIT study investigators, 2005), cardiac and/or respiratory arrest (Hodgetts et al., 2002; Kenward et al., 2001; Nurmi et al., 2005), admission of inpatients to the intensive care unit (Goldhill et al., 1999; Jonsson et al., 2011), or clinical concern about potential deterioration requiring overnight medical review (Gordon & Beckett, 2011). With the exception of one study that only reported omissions made in the 15 min immediately preceding an event (MERIT study investigators, 2005), most investigated the shift or 24-hr period preceding an event (sometimes excluding the final 15 min). A meta-analysis was also conducted on the proportion of omissions reported in these studies. The meta-analysis yielded a predicted proportion of 47.8% [95% CI: 26.8%, 69.2%; Prediction Interval: 0%, 99.5%, $\tau^2 = 0.122$] (Figure 3b). Notably, a further study that focussed on the emergency department treatment of asthma attacks found a substantial omission rate of 72.9% at the severity evaluation stage (Linares et al., 2006).

Eight additional studies investigated the frequency of missing RRs from observations taken during triage or hospital admission (Bergrath et al., 2011; Considine et al., 2006; Cooper et al., 2013; Crandon et al., 2008; Gerdtz et al., 2013; O'Reilly et al., 2012), or immediately after admission (Armstrong et al., 2008; Keene et al., 2017). Omission rates ranged from 0.8%-81.5% (excluding any postintervention rates; Table 8 and Table S7). Data from these studies were included in a third meta-analysis of proportions, which yielded a predicted proportion of omissions of 21.6% [95% CI: 6.2%, 42.7%; Prediction Interval: 0.0%, 84.7%; $\tau^2 = 0.108$] (see Figure 3c).

TABLE 6 Studies reporting the percentage of respiratory rate measurements omitted from records in general audits of vital sign documentation

Authors (Year) Country	Observers (*N*, age, M/F)	Patients (*N*, age, M/F)	Track-and-trigger system/ chart type/ intervention timepoint[a]	Period of recording	Total possible recordings	Number of omissions	Omission rate	Study Quality
Cahill et al. (2011) Australia	Hospital staff in three wards [medical/surgical, surgical, medical] (NR, NR, NR)	Ward patients, timepoint 1 (104, NR, NR)	(1) Before intervention	All recordings from discharges in 14-day period	2,557	1,335	52.2%	89%
		Ward patients, timepoint 2 (147, NR, NR)	(2) 2 weeks after intervention (new track and trigger chart + education)		2,435	54	2.2%	
		Ward patients, timepoint 3 (119, NR, NR)	(3) 3 months after intervention		2,250	34	1.5%	
Cretikos et al. (2007) Australia	Hospital staff (NR, NR, NR)	Patients without adverse event (520, 69 y, 299/221)	(1) No track-and-trigger system	Within 24-hr before audit	520	123	23.7%	89%
Edwards and Murdin (2001) UK	Hospital staff (NR, NR, NR)	Patients on wards (NR, NR, NR)	NR	Within two-week period	274	115	42.0%	78%
Hall et al. (2003) UK	Hospital staff on eight medical wards (NR, NR, NR)	Patients on eight medical wards (~200, NR, NR)	(1) Prior to Critical Care Outreach education	Within 12-hr before audit	~200	~164 (based on graph)	82%	78%
		Patients on eight medical wards (~175, NR, NR)	(2) Post-CCO education	Within 12-hr before audit	~175	~24 (based on graph)	14%	
		Patients on eight medical wards (~200, NR, NR)	(3) Post-CCO education, 6 month follow-up	Within 12-hr before audit	~200	~36 (based on graph)	18%	
Helliwell et al. (2002) UK	Hospital staff (NR, NR, NR)	Medical/surgical patients (344, NR, NR)	NR	Within each day in 19-day period	1,545	912	59.0%	83%
McBride et al. (2005)[b] UK	Hospital staff (NR, NR, NR)	Patients on various wards, timepoint 1 (1,251, NR, NR).	(1) Wards in group 1: No specific system (pre-intervention)	Within 24-hr before audit	NR	NR	81.6%	81%
			(2) Wards in group 2: No specific system (pre-intervention)		NR	NR	59.3%	
		Patients on various wards, timepoint 2 (1,234, NR, NR).	(3) Wards in group 1: Newly designed chart + education		NR	NR	18.6%	
			(4) Wards in group 2: MEWS + newly designed chart + education		NR	NR	43.6%	
		Patients on various wards, timepoint 3 (600, NR, NR)	(5) Wards in group 1: MEWS + newly designed chart + education		NR	NR	8.6%	
			(6) Wards in group 2: MEWS + newly designed chart + education		NR	NR	9.1%	

(Continues)

TABLE 6 (Continued)

Authors (Year) Country	Observers (*N*, age, M/F)	Patients (*N*, age, M/F)	Track-and-trigger system/ chart type/ intervention timepoint[a]	Period of recording	Total possible recordings	Number of omissions	Omission rate	Study Quality
McGain et al. (2008)[b] Australia	Hospital staff (NR, NR, NR)	Major surgery patients, hospital A (42, median = 69.5 y, 30/12)	NR	During first 7 days post-operation	NR	NR	15.4% (overall)	83%
		Major surgery patients, hospital B (42, median = 66.5, 31/11)						
		Major surgery patients, hospital C (42, median = 64.0 y, 26/16)						
		Major surgery patients, hospital D (43, median = 63.0 y, 19/24)						
		Major surgery patients, hospital E (42, median = 62.5 y, 29/13)						
Odell et al. (2007) UK	Hospital staff (NR, NR, NR)	Ward patients hospital A, time point 1 (296, NR, NR)	(1) No specific system	Within 24-hr before audit	296	267	90.2%	89%
		Ward patients hospital B, timepoint 1 (191, NR, NR)	(2) No specific system	Within 24-hr before audit	191	191	100%	
		Ward patients hospital A, timepoint 2 (315, NR, NR)	(3) R-MEWS + education piloted on surgical wards	Within 24-hr before audit	315	224	71.1%	
		Ward patients hospital B, timepoint 2 (267, NR, NR)	(4) No specific system	Within 24-hr before audit	267	260	97.4%	
		Ward patients hospital A, timepoint 3 (222, NR, NR)	(5) R-MEWS + education in all wards	Within 24-hr before audit	316	94	29.7%	
		Ward patients hospital B, timepoint 3 (117, NR, NR)	(6) R-MEWS + education in all wards	Within 24-hr before audit	277	160	57.8%	
		Ward patients hospital A, timepoint 4 (328, NR, NR)	(7) R-MEWS + education in all wards	Within 24-hr before audit	328	53	16.2%	
		Ward patients hospital B, timepoint 4 (168, NR, NR)	(8) R-MEWS + education in all wards	Within 24-hr before audit	168	96	57.1%	
		Ward patients hospital A, timepoint 5 (317, NR. NR)	(9) R-MEWS + education in all wards	Within 24-hr before audit	317	39	12.3%	
		Ward patients hospital B, timepoint 5 (163, NR, NR)	(10) R-MEWS + education in all wards	Within 24-hr before audit	163	67	41.1%	
Parkes (2011) UK	Hospital staff (NR, NR, NR)	Emergency patients (594, NR, NR)	Unspecified	Entire stay in emergency department	594	422	71.0%	89%
Pedersen et al. (2018) Denmark	Ward staff (NR, NR, NR)	Ward patients (168,496, median = 60 y, 77508/90988)	NEWS	Chart audit of all data recorded over a 12 month period	2,835,331	124,385	4.39%	83%

(Continues)

TABLE 6 (Continued)

Authors (Year) Country	Observers (*N*, age, M/F)	Patients (*N*, age, M/F)	Track-and-trigger system/ chart type/ intervention timepoint[a]	Period of recording	Total possible recordings	Number of omissions	Omission rate	Study Quality
Ramgopal et al. (2018)[b] USA	Emergency medical service staff from 20 agencies (NR, NR, NR)	Adult emergency patients (349,863, NR, 152375/196042)	Electronic patient care record	Chart audit of all data recorded during transit to hospital over a 9 month period	346,863	6,378	1.84%	89%
Rosen et al. (2015) Sierra Leone	Site A: ward nurses (NR, NR, NR)	Ward patients, timepoint 1 (65, NR, NR)	(1) Pre- Failure Mode Analysis and Effects intervention	Chart audit of entire stay	651	483	74.2%	89%
	Site B: ward nurses (NR, NR, NR)	Ward patients, timepoint 1 (30, NR, NR)	(2) No intervention	Chart audit of entire stay	203	19	9.4%	
	Site A: ward nurses (NR, NR, NR)	Ward patients, timepoint 2 (82, NR, NR)	(3) Post-FMEA intervention	Chart audit of entire stay	891	507	56.9%	
	Site B: ward nurses (NR, NR, NR)	Ward patients, timepoint 2 (51, NR, NR)	(4) No intervention	Chart audit of entire stay	355	72	20.3%	
Smith and Oakey (2006) UK	Hospital staff (NR, NR, NR)	Patients with Legionnaire's disease (89, median = 64.7 y, NR) and patients without (100, 61.0 y, NR)	EWS	During hospital stay	3,739	982	26.3%	89%
Sundararajan et al. (2016) New Zealand	Daytime hospital staff from 41 hospitals (NR, NR, NR)	ICU patients (48, median = 62.5 y, 32/16)	Unspecified RRT system	All hourly daytime obs on chart at time of census	480	312	65.0%	89%
	Night-time hospital staff from 41 hospitals (NR, NR, NR)			All hourly night-time obs on chart at time of census	672	442	65.8%	
Van Leuvan and Mitchell (2008) Australia	Hospital staff (NR, NR, NR)	Patients in various wards (62, median = 67 y, 33/29)	General observation chart	Within 48-hr audit period	422	321	76.1%	89%

Abbreviation: NR, not reported.

[a]For studies that provide separate data on multiple systems/charts, subgroups or timepoints, the bracketed numbers in this column can be used to cross-reference Figure 3a.

[b]This study was not included in the meta-analysis (see Results section).

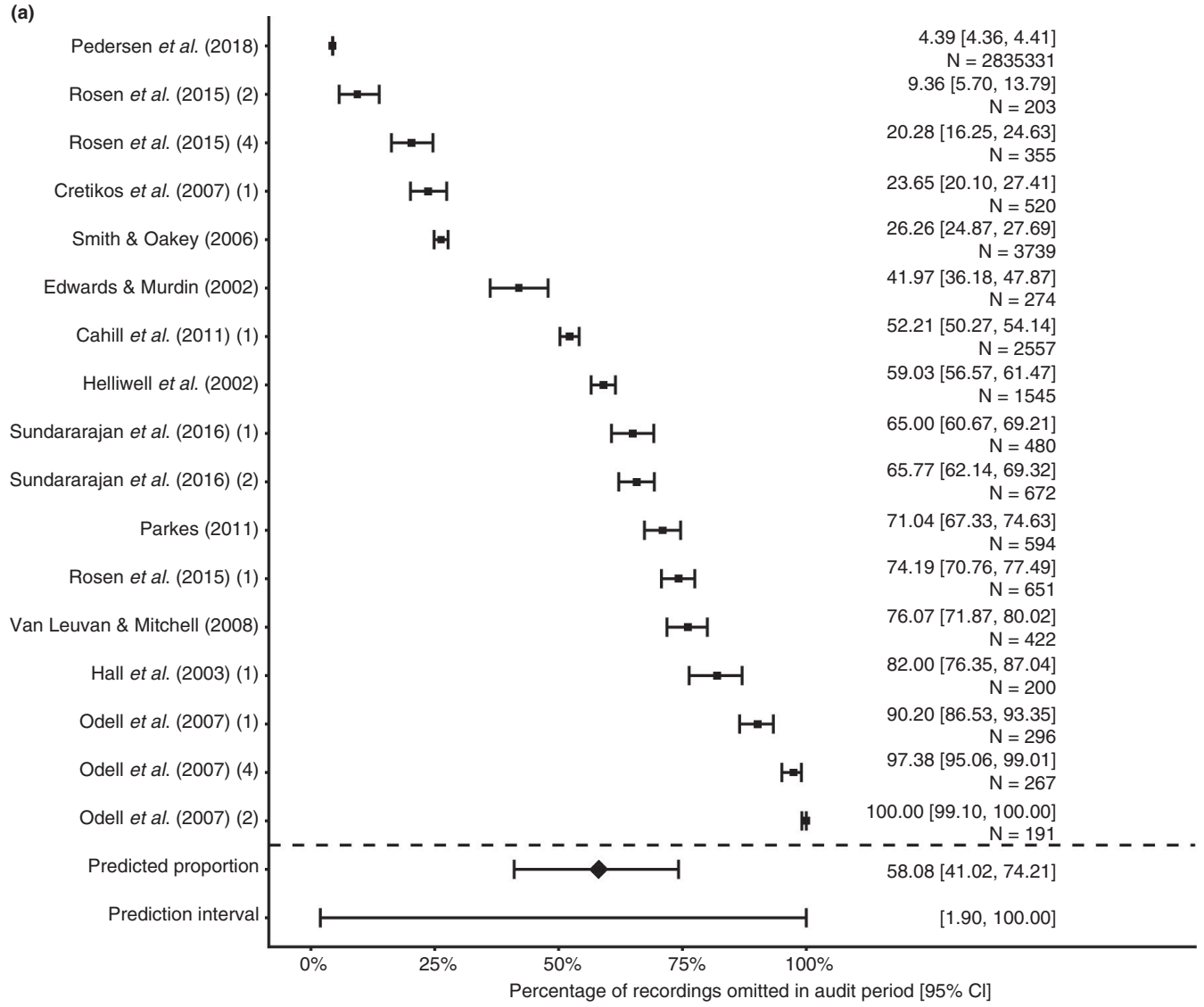


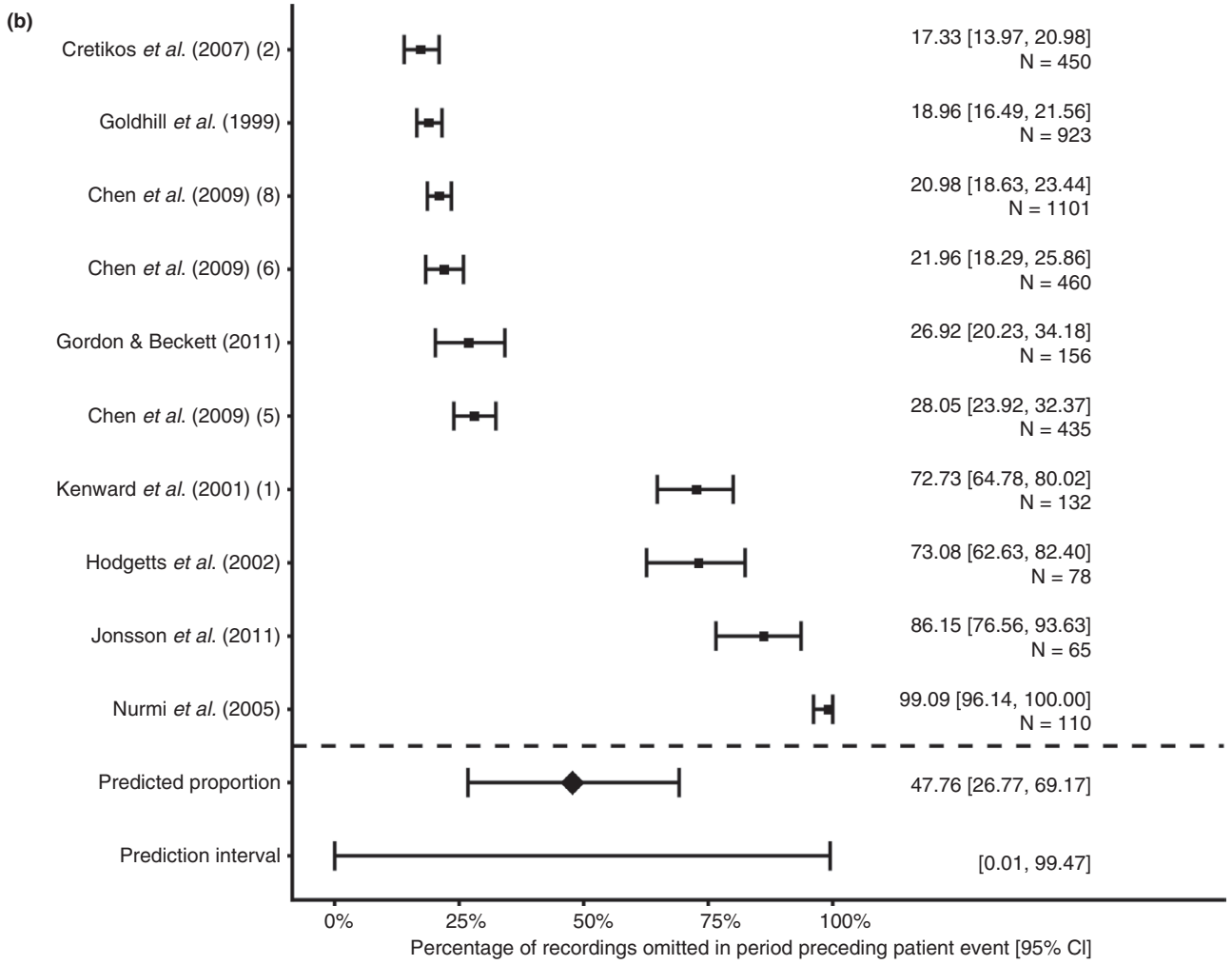
FIGURE 3 Recording omission: Forest plots of the percentage of respiratory rate measurements omitted from recorded vital sign observations: (a) in hospital-based general audits of inpatient vital sign documentation; (b) prior to inpatient adverse events or clinical deterioration; and (c) during triage or admission to hospital, or immediately after admission (all plots exclude rates obtained after interventions; bracketed numbers correspond to numbering in Tables 6–8 for studies that reported multiple datasets)

5 | DISCUSSION

This systematic review identified five distinct potential sources of inaccuracy in adult patients' manually measured RR data. Three of these—*observation method*, *inter-/intra-observer variability*, and the *awareness effect*—relate to the way RR observations are conducted. The two remaining sources—*value bias* and *recording omission*—relate to the documentation of RR data.

Of all five sources of inaccuracy, the *awareness effect* yielded the clearest and most consistent results. All relevant studies found mean reductions in RR when participants' attention was drawn to their respiration or they were explicitly made aware that it was being observed (Table 4). Similarly, regarding *observation method*, there was clear evidence that shorter counts (e.g. 15 or 30 s vs. 60 s) led to systematic mean underscoring of RR and that this may be particularly problematic for patients with severely abnormal RRs (Table 2). In addition, the limits of agreement between usual care observations and criterion standard measurements were often wide, indicating that individual usual care observations could be subject to substantial measurement error in either direction, which generally tended to be more pronounced for higher known criterion values. Likewise, the limits of agreement for *inter- and intra-observer variability* indicated that individual RR measurements can vary substantially in either direction between observations, although the mean differences were small or negligible (Table 3).

The results from studies of over-represented RR values conducted in a range of countries suggested that *value bias* is a widespread problem. The actual values that appeared to be biased varied between study locations (with 20 and 18 breaths/min being the most common). However, evidence from studies with



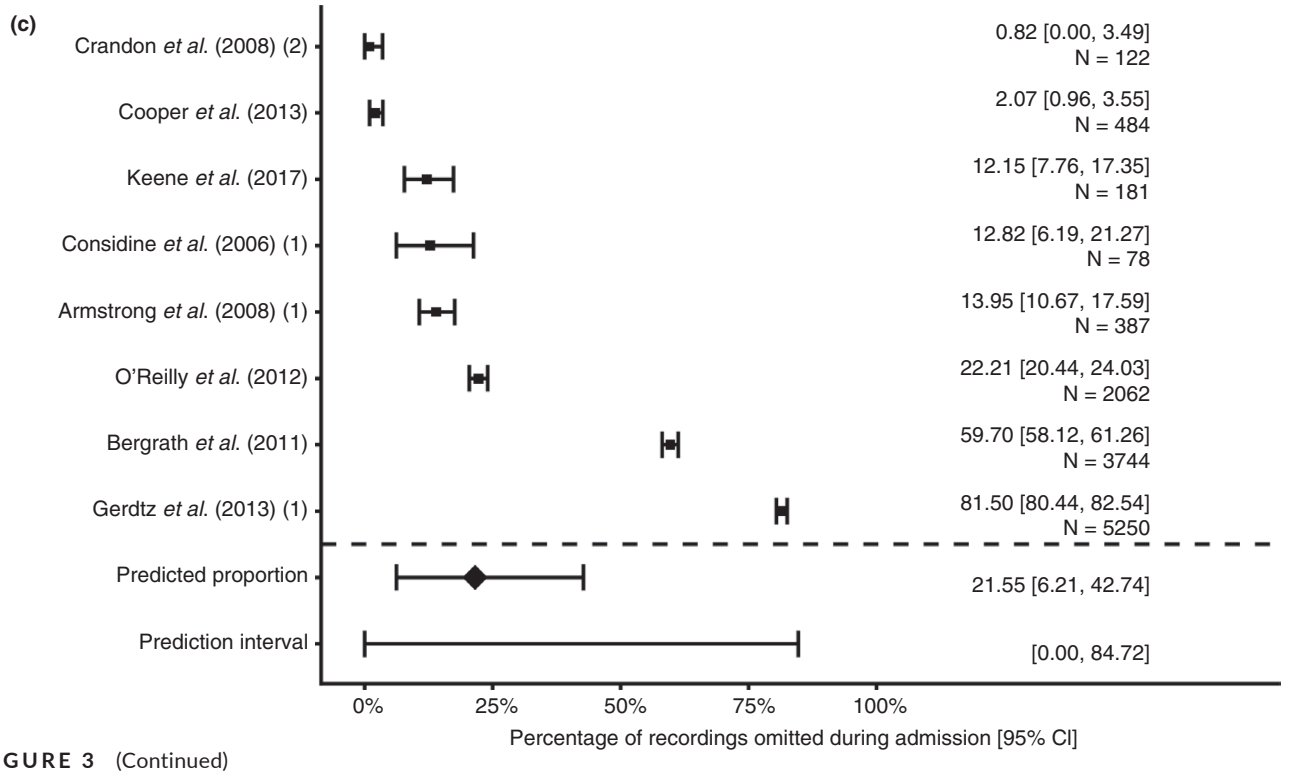


FIGURE 3 (Continued)

TABLE 7 Studies reporting the percentage of respiratory rate measurements omitted from records prior to patient events or clinical deterioration

Authors (Year) Country	Observers (*N*, age, M/F)	Patients (*N*, age, M/F)	Track-and-trigger system/ chart type/ intervention timepoint[a]	Period of recording	Total possible recordings	Number of omissions	Omission rate	Study Quality
Chen et al. (2009) Australia	Hospital staff [MET hospitals] (NR, NR, NR)	Ward patients, timepoint 1 (NR, NR, NR)	(1) Pre-intervention	Within 15 min before adverse event	435	309	71%	89%
	Hospital staff [control hospitals] (NR, NR, NR)	Ward patients, timepoint 1 (NR, NR, NR)	(2) No intervention		460	377	82%	
	Hospital staff [MET hospitals] (NR, NR, NR)	Ward patients, timepoint 2 (NR, NR, NR)	(3) Post introduction of MET + education		2,291	1,604	70%	
	Hospital staff [control hospitals] (NR, NR, NR)	Ward patients, timepoint 2 (NR, NR, NR)	(4) No intervention		1,101	980	89%	
	Hospital staff [MET hospitals] (NR, NR, NR)	Ward patients, timepoint 1 (NR, NR, NR)	(5) Pre-intervention	Between 15 min & 24-hr before adverse event	435	122	28%	
	Hospital staff [control hospitals] (NR, NR, NR)	Ward patients, timepoint 1 (NR, NR, NR)	(6) No intervention		460	101	22%	
	Hospital staff [MET hospitals] (NR, NR, NR)	Ward patients, timepoint 2 (NR, NR, NR)	(7) Post introduction of MET + education		2,291	573	25%	
	Hospital staff [MET hospitals] (NR, NR, NR)	Ward patients, timepoint 2 (NR, NR, NR)	(8) No intervention		1,101	231	21%	
Cretikos et al. (2007) Australia	Hospital staff (NR, NR, NR)	Patients with adverse event (450, 69 y, 264/186)	(2) No track-and-trigger system	Between 15 min and 24-hr before adverse event	450	78	17.3%	89%
Goldhill et al. (1999) UK	Hospital staff (NR, NR, NR)	Patients admitted to ICU (923, NR, NR)	(1) Unspecified observation charts	Within 24 hr before ICU admission	923	~175 (based on graph)	19.0%	83%
Gordon and Beckett (2011) UK	Hospital staff (NR, NR, NR)	Ward patients requiring overnight medical review due to SEWS score trigger or nursing staff concern (121, NR, NR)	SEWS	Overnight observations from night when medical review was required	156	42	26.9%	89%
Hodgetts et al. (2002) USA	Hospital staff (NR, NR, NR)	Patients with potentially avoidable cardiac arrest (78, NR, NR)	NR	Within 24-hr before cardiac arrest	78	57	73.0%	83%
Jonsson et al. (2011) Iceland	Hospital staff (NR, NR, NR)	Acute medical or surgical patients (65, 65 y, 37/28)	MEWS	During shift prior to unplanned admission to ICU	65	56	86.2%	89%

(Continues)

TABLE 7 (Continued)

Authors (Year) Country	Observers (*N*, age, M/F)	Patients (*N*, age, M/F)	Track-and-trigger system/ chart type/ intervention timepoint[a]	Period of recording	Total possible recordings	Number of omissions	Omission rate	Study Quality
Kenward et al. (2001) UK	Hospital staff (NR, NR, NR)	Patients with respiratory or cardiac arrest (132, NR, NR)	(1) No specific system	Within 24-hr before cardiac or respiratory arrest	132	96	72.7%	78%
	Hospital staff (NR, NR, NR)	Patients without respiratory or cardiac arrest (132, NR, NR)	(2) Post-MET team introduction + education	NR	132	15	11.4%	
Linares et al. (2006)[b] Spain	Hospital staff (NR, NR, NR)	Asthma patients presenting at the emergency department (46, NR, NR)	Acute asthma clinical guides	During asthma attack episode	48	35	72.9%	89%
MERIT study investigators (2005)[b] Australia	Hospital staff in 23 different hospitals (NR, NR, NR)	Patients with adverse event without an NFR order (NR, NR, NR)	11 hospitals without MET, 12 hospitals with MET	Within 15-min before adverse event	5,899	3,657 (missing RR, HR and BP)	62%	83%
Nurmi et al. (2005) Finland	Hospital staff (NR, NR, NR)	Cardiac arrest patients in four hospitals (110, 68 y, 64/46)	Not specified	Within 24-hr before cardiac arrest	110	109	99.1%	83%

Abbreviation: NR, not reported.

[a]For studies that provide separate data on multiple systems/charts, subgroups or timepoints, the bracketed numbers in this column can be used to cross-reference Figure 3b.

[b]This study was not included in the meta-analysis (see Results section).

TABLE 8 Studies reporting the percentage of respiratory rate measurements omitted from records during triage or admission to hospital, or immediately after admission

Authors (Year) Country	Observers (*N*, age, M/F)	Patients (*N*, age, M/F)	Track-and-trigger system/ chart type/ intervention timepoint[a]	Period of recording	Total possible recordings	Number of omissions	Omission rate	Study Quality
Armstrong et al. (2008) UK	Triage nurses (NR, NR, NR)	Emergency admissions (387, 56 y, NR)	(1) Custom emergency department notes	Initial assessment within 15 min of arrival	387	54	13.9%	89%
			(2) Custom emergency department notes	Repeat assessment within 60 min of arrival	387	193	49.9%	
Bergrath et al. (2011) Germany	EMS doctors (43, NR, NR)	Emergency patients (3,744, NR, NR)	Mainz Emergency Evaluation Score	During admission to emergency department	3,744	2,235	59.7%	94%
Considine et al. (2006) Australia	Hospital staff (NR, NR, NR)	Emergency department patients, timepoint 1 (78, >18 y, NR)	(1) Emergency Department Observation Chart; before intervention	During emergency admission	78	10	12.8%	89%
		Emergency department patients, timepoint 2 (74, >18 y, NR)	(2) Emergency Department Observation Chart; after intervention	During emergency admission	74	23	31.1%	
Cooper et al. (2013) Australia	Hospital staff (NR, NR, NR)	Medical/surgical patients (484, 66.5 y, NR)	Unspecified MET system	During admission	484	10	2.1%	89%
Crandon et al. (2008) Jamaica	Referring hospital staff (NR, NR, NR)	Patients being transferred between hospitals (122, 27.8 y, 97/25)	(1) NR	Before departure from referring hospital	122	110	90.2%	94%
	Receiving hospital staff (NR, NR, NR)		(2) NR	On arrival at receiving hospital	122	1	0.8%	
Gerdtz et al. (2013) Australia	Triage nurses (122, NR, NR)	Triage patients, timepoint 1 (5,250, documented: median = 45 y, 876/814; undocumented: median = 41 y, 1,978/1,577)	(1) No specific system (baseline)	During triage	5,250	4,279	81.5%	94%
		Triage patients, timepoint 2 (4,975, NR, NR)	(2) Electronic triage interface restructured		4,975	2,960	59.5%	
		Triage patients, timepoint 3 (4,801, NR, NR)	(3) + audit feedback		4,801	2,837	59.1%	
		Triage patients, timepoint 4 (4,828, NR, NR)	(4) + triage education sessions		4,828	2,617	54.2%	
		Triage patients, timepoint 5 (5,008, documented: median = 44 y, 2,160/1,964; undocumented: median = 42 y, 514/356)	(5) 12 month follow up		5,008	1,552	31.0%	
Keene et al. (2017) South Africa	Hospital staff (NR, NR, NR)	Acute trauma ward patients (181, NR, NR)	MEWS	6:00 a.m. vitals on morning after ward admission	181	22	12.15%	89%

(Continues)

TABLE 8 (Continued)

Authors (Year) Country	Observers (*N*, age, M/F)	Patients (*N*, age, M/F)	Track-and-trigger system/ chart type/ intervention timepoint[a]	Period of recording	Total possible recordings	Number of omissions	Omission rate	Study Quality
O'Reilly et al. (2012) Australia	Hospital staff (NR, NR, NR)	Major trauma patients (2,520, NR, NR)	Unspecified systems	Admission	2,062	458	18.1%	83%

Abbreviation: NR, not reported.

[a]For studies that provide separate data on multiple systems/charts, subgroups or timepoints, the bracketed numbers in this column can be used to cross-reference Figure 3c.

criterion-measured comparison data consistently found that over half of all recorded RR values were potentially the product of bias and all studies showed that the two most common values appeared to be suspiciously prevalent (Table 5 and Figure 2). Finally, widespread evidence of *recording omission* was also revealed (Table 6). Meta-analyses showed that omission rates tended to be substantially higher among hospital-based general audits of inpatient observations (58.1%) and patient records completed prior to inpatient adverse events or clinical deterioration (47.8%), compared with data collected during triage, admission to hospital, or immediately after admission (21.6%; Figure 3). In comparison with *value bias*, however, there was also substantially more variation between study sites.

In interpreting the findings of this review, it is worth noting that some of the identified sources of inaccuracy could potentially have cumulative effects on individual RR observations. For example, if a 15-s count is conducted on a patient who is aware that their respiration is being observed, then the combination of an artefactual reduction in their actual RR due to the *awareness effect* and underscoring due to the *observation method* could potentially create a substantially misleading impression of the patient's clinical condition (Hill et al., 2018). This issue could also be exacerbated by *inter- or intra-observer* variability. Furthermore, the prevalence of *value bias* may call into question the trustworthiness of all recorded RR data at affected sites, while *recording omission* has a further impact on the utility of RR as a clinical indicator. Consequently, hospitals that rely on the manual measurement of RR may need to address all of these issues to maximize the clinical utility of RR data, including its use in the context of track-and-trigger scoring systems designed to facilitate the early recognition of deteriorating patients.

5.1 | Implications for nurses

In terms of nursing practice, the results of this review emphasize the value of conducting 60-s counts whenever it is practicable to do so and of ensuring that patients are unaware that their RR is being measured. They also highlight that *recording omission* is a widespread and important problem, and the need to count RR rather than simply estimating it. Hence, it is vital that nursing education addresses all of these points, as well as instilling an appropriate counting technique to minimize unwanted variability in RR data. Nevertheless, despite being *necessary*, training alone may not be *sufficient* to eliminate the measurement issues identified in this review if their underlying causes are not considered and addressed.

5.2 | Underlying causes

In health care, administrators and educators often assume that safety or compliance issues can be remedied through training interventions designed to change the behaviour of clinical staff; however, in reality, such efforts are unlikely to lead to sustained

improvements unless other elements of the surrounding system are optimized first (Russ et al., 2013). Hence, knowledge of the underlying mechanisms for the sources of inaccuracy identified in this review may allow for better targeted and more effective interventions.

A particular challenge in determining the underlying causes of these issues is the marked lack of reported information about the individuals whose observations were studied (e.g. nurses). Authors typically reported sufficient information on patient participants (when applicable), yet this was rarely the case for observers, even when *observation method* or *inter-/intra-observer variability* was the main focus of the study. While there may be pragmatic obstacles to its inclusion, particularly for large-scale chart-review studies, detailed observer information would be a useful addition to future primary research publications on these topics.

There is some suggestion in the literature that many instances of *value bias* may reflect estimates—rather than measurements—of RR (Badawy et al., 2017; Keene et al., 2017; Semler et al., 2013). Similarly, some instances of *recording omission* may also represent failure to measure RR in the first place. Recent studies have pointed to a range of potential underlying causes for the general neglect of RR measurement in hospital care (Ansell et al., 2014; Elliott, 2016; Hogan, 2006). These include: perceived or actual lack of time; lack of knowledge and training; lack of automated measurement; and the perceived unimportance of RR (despite actually being a strong predictor of clinical deterioration; Churpek et al., 2016). Similarly, time constraints may cause clinicians who do conduct formal RR counts to favour shorter count durations (e.g. 15 s). This may cause RR to be underestimated, as demonstrated in the present review. In addition, neglect of vital sign monitoring more generally has been found to be influenced by shift length, with longer shifts leading to less frequent monitoring (Dall'Ora et al., 2019) and therefore more *recording omissions*. Clearly, some of these causes are rooted in systemic and cultural factors and consequently cannot be remedied through training alone.

5.3 | Potential solutions to improve the accuracy of respiratory rate data

A range of options are available to hospitals seeking to improve the accuracy of their RR data. In this section, we discuss these in light of our results. However, it is important to appreciate that there is no 'one size fits all' solution. Rather, the best solution—or combination of solutions—for a particular site will depend on the organizational and cultural context, the available resources, and the extent to which particular measurement issues prevail locally. For example, it is clear from the review that, although *recording omission* is a serious and widespread problem and it is generally more prevalent in inpatient wards, some individual ward sites have very low omission rates (and vice versa for emergency departments). It is also important to consider any unintended consequences for patient care of any proposed intervention or change.

5.3.1 | Automated measurement

For sites with substantial and pervasive issues around the manual measurement or documentation of RR, one potential option that addresses many of the possible underlying causes outlined above is to transition to a fully automated measurement method. In theory, the right technology could eliminate all sources of error discussed in this review (provided it is also interoperable with an electronic record system). Even the *awareness effect* could potentially be avoided if it is not readily apparent when measurements are occurring, or if continuous monitoring allows the patient to habituate to measurement.

Unfortunately, current mainstream methods for automated RR measurement (i.e. spirometry, capnometry, and impedance pneumography) are not practical for the mass routine monitoring of inpatients due to a range of factors, such as patient discomfort (e.g. capnometry), interference to natural breathing (e.g. spirometry), and resource intensiveness (all methods; Liu et al., 2019). However, RR measurement is ripe for digital disruption and over a dozen alternative technological methods have already been proposed (for a recent review, see Liu et al., 2019). Several of these can potentially be incorporated into clothing or other body-worn devices and some do not even require physical contact with the patient (Liu et al., 2019). Although many of these technologies are in the early stages of development and only a couple can be implemented at relatively low cost (Liu et al., 2019), progress to date suggests that unobtrusive automated monitoring of all ward patients' RRs will be a viable option for some sites in the relatively near future. As well as directly addressing the RR measurement issues outlined in this review, such technologies may also have additional indirect benefits if they alleviate some of the time pressure experienced by nursing staff in busy hospital environments, freeing up more time for high-quality nurse-patient interactions. These interactions, in turn, may lead to the identification of additional care needs (Cardona-Morrell et al., 2016), improved recovery times (Castillo & Sánchez-Sosa, 2002), and greater patient satisfaction (Evans, 2016).

However, even when such technologies become widely available, they will not necessarily be a panacea for sites where monitoring is currently suboptimal. First, although these devices may be affordable for some hospitals, they are likely to be prohibitively expensive for many, particularly in the developing world. Second, there may be a risk that nurse–patient contact actually decreases if none of the patients' vital signs need to be measured manually and time pressures continue to increase. Finally, there is a risk of deskilling if clinicians are no longer well-practiced at manual measurement. This could potentially place deteriorating patients at elevated risk if and when technologies fail. For example, some clinical information systems have proven vulnerable to prolonged network failures (Berinato, 2003; Flinders, 2015) or ransomware attacks (Collier, 2017). Indeed, the catastrophic failure of most or all computers, mobile phones, and other sensitive electronic equipment—as well as sustained power outages—are very real possibilities if a solar superstorm on the scale of the 1859 *Carrington Event* were to strike the modern world (Jonas, 2015). As unlikely as this may sound to some readers, the current (at time of writing) COVID-19 pandemic

illustrates the need for health systems to be resilient to low-probability, high-impact events such as this. Hence, technological solutions do not eliminate the need for clinicians to acquire and maintain skills in the manual measurement of vital signs, including RR.

5.3.2 | Measurement aids

Various devices have also been developed to aid, rather than replace, the manual measurement of RR. Some are relatively low-technology devices, such as the World Health Organization's *ARI Timer* and *ARI Timer MK2* (Gan et al., 2015; WHO &WHO, 1992, 1993). These devices are simple timers that indicate when 60 s have elapsed. The target users are primarily community healthcare workers in developing countries, and there has not been widespread adoption in hospitals. More recently, however, more sophisticated mobile phone applications have been developed. Most of these use the 'one-press-per-breath' (OPB) concept, where an observer presses a button on the phone each time the patient takes a breath. The application, rather than the observer, keeps track of the breath count. Notably, some of these applications, such as *RRate* (Karlen et al., 2015), do not require an entire 60-s count. Instead, the RR is derived from the average amount of time between breaths (Karlen et al., 2015). There is evidence that this method can potentially allow for adequately accurate RR measurement in a fraction of the 60 s recommended for manual measurement (Gan et al., 2015). Further, in a study by Black et al. (2015), four mobile phone applications were compared: three that employed OPB (of various lengths or number of breaths) and a simple 60-s timer. A 20-breath OPB and a 60-s OPB produced the most accurate measurements and both out-performed the 60-s timer.

Because mobile applications can be used on existing Android or iOS personal devices, there is a smaller barrier to adoption than with special devices like the ARI Timer, at least in the developed world. Importantly, the benefit of accurate measurement in shorter counting periods may also overcome one of Elliot's (2016) reported reasons for the neglect of RR, namely lack of time. Hence, the use of such applications may reduce the prevalence of *recording omissions* and *value bias*, as well as reducing *inter-* and *intra-observer variability*. Nevertheless, the *awareness effect* remains a potential issue. Compared with taking a simple count unobtrusively in one's head, tapping each time the patient breathes may actually draw more attention to the fact that RR is being assessed. Hence, future research should investigate whether the use of OPB applications increases the *awareness effect*. As with fully automated systems, the potential for technological failures also means that manual measurement skills cannot be replaced entirely.

5.3.3 | Track-and-trigger systems

The introduction of paper-based or electronic track-and-trigger systems may also have an impact on the measurement of RR. For example, as well as providing baseline data on *recording omission* rates, three studies included in this review also investigated whether introducing a track-and-trigger system would reduce them (Cahill et al., 2011; McBride et al., 2005; Odell et al., 2007; Table 6). In all three studies, the frequency of missing RR values was shown to decrease, falling by up to 78% (Odell et al., 2007). Furthermore, in a simulation study by Ludikhuize et al. (2011), hospital nurses were more likely to seek out the RRs of mock patients if they had prior experience using a track-and-trigger system in their ward. This may indicate that the importance of RR as a clinical indicator is better understood when observers have been exposed to track-and-trigger systems. However, in each of these studies, it is unclear to what extent the results were attributable to the track-and-trigger system itself, or to the training that accompanied its introduction.

Although a track-and-trigger system (especially an electronic system) may provide a forcing function for the recording of vital signs (Hogan et al., 2019), including RR—thus potentially reducing *recording omissions*—there is no guarantee that any manually measured RR values documented therein are accurate. Unfortunately, none of the studies of *recording omission* discussed above assessed the accuracy of the values that *were* recorded against criterion-measured data. However, the general point is illustrated by Cooper et al.'s (2013) study, which yielded one of the lowest omission rates (2.1%) but also one of the highest combined percentages for the two most common commonly recorded values (82.76%). If nursing staff are time-poor, an unintended consequence of using a track-and-trigger system may be that there is a perverse incentive to estimate RR, resulting in an increased incidence of *value bias*. This may potentially undermine the track-and-trigger system itself by causing RR to be scored incorrectly (Badawy et al., 2017).

5.3.4 | Other clinical workplace changes

A further five of the studies that provided baseline *recording omission* data also investigated the impact of alternative interventions involving changes to clinical work systems (Chen et al., 2009; Considine et al., 2006; Gerdtz et al., 2013; Kenward et al., 2001; Rosen et al., 2015; Table 6). However, none investigated whether reduced omission rates were accompanied by a concomitant decrease in accuracy or an increase in the prevalence of *value bias*.

Two of these studies examined the impact of introducing a Medical Emergency Team (MET), yielding mixed results for *recording omissions*. The study by Chen et al. (2009) produced weak evidence of slightly lower omission rates at sites where a MET had been introduced. In contrast, Kenward et al. (2001) found a substantial reduction in *recording omissions* after the introduction of a MET; however, they attributed it primarily to the accompanying education around the early detection of clinical deterioration.

One study investigated the effect of redesigning an electronic triage interface to better align with the clinical decision-making process and to make data entry less effortful (Gerdtz et al., 2013). Documentation of RR more than doubled following the interface

change and further increased after triage education sessions were conducted that emphasized the importance of vital signs for recognizing ill health and deterioration at triage. Another study investigated the effect of an intervention comprising the introduction of written nursing practice standards for the initial assessment of emergency department patients with supporting education, but found an unexplained increase in RR *recording omissions* (Considine et al., 2006). Finally, one study employed a *Failure Mode and Effects Analysis* process, where the researchers worked with frontline staff to: (a) identify current and potential risks to the completion of vital sign monitoring and documentation in their workplace; (b) articulate the potential effects of these risks; and (c) identify potential solutions appropriate to that specific site (Rosen et al., 2015). The proposed solutions were then presented to the facility's leadership, and RR *recording omissions* decreased significantly in the following 2 months.

5.3.5 | Training interventions

Presently, sources such as nursing textbooks provide conflicting advice about the duration of manual RR counts (Hill et al., 2018). However, it is clear from the *observation method* results that 60-s counts should be preferred. Nevertheless, enshrining this advice in clinical training will not ensure that it is adhered to if issues such as time pressure and interruptions make it impractical to do so, in which case a technology-based solution may be preferable. Similarly, in the context of manual measurement, *inter-* and *intra-observer variability* can be addressed via training in the appropriate protocols, but it may not always be possible to enact these protocols consistently given competing pressures. Of all five sources of inaccuracy, the *awareness effect* is probably the most amenable to a training-based solution. Indeed, nursing texts are relatively consistent in advising that patients should not be made aware that their RR is being measured (Hill et al., 2018).

It is also clear from some of the studies discussed above that education may provide a useful adjunct to interventions that involve improvements to clinical work systems, such as those aimed at reducing *recording omissions*. In particular, there was some evidence that there is value in training staff not just in protocols for the accurate measurement of vital signs, but also their central importance to the detection of clinical deterioration (e.g. Gerdtz et al., 2013; Kenward et al., 2001). Similarly, another of the included studies showed that education provided by critical care outreach teams reduced the incidence of *recording omissions* on wards already using a track-and-trigger scoring system (Hall et al., 2003; Table 6). Finally, a further potentially fruitful use of training is for familiarizing staff with new technologies (Russ et al., 2013), such as automated measurement devices or measurement aids.

5.4 | Strengths and Limitations

This systematic review is the first to focus on identifying and quantifying potential sources of inaccuracy in manually measured RR data. The integrity of the review was ensured by adhering to the PRISMA guidelines and the use of two reviewers. In addition, all included papers were subjected to a standardized quality appraisal, the full results of which have been included in the supplementary materials for transparency. The identification of sources of inaccuracy is an important first step towards improving measurement quality, with the potential to inform interventions and further investigations.

Nevertheless, existing research on the topic of RR measurement accuracy is notably sparse. While there were a total of 49 studies identified, 32 of these related to a single source of inaccuracy, namely *recording omission*. The remaining sources of inaccuracy were only investigated in two to nine studies each. Thus, the sources of inaccuracy presented in this review should be taken primarily as a representation of the available literature and not an exhaustive list of potentially influential factors. In addition, it should be noted that all studies of the *awareness effect* were conducted with healthy volunteers. Hence, it would be particularly valuable for future research to replicate these findings with clinical populations to clarify the extent to which they will generalize, given that inpatients are likely to have more abnormal RR values and more irregular breathing patterns than healthy participants.

6 | CONCLUSION

RR is an important clinical indicator, but this review has identified five sources of inaccuracy related to its observation or documentation. When not subject to *recording omission*, manually measured RR data may still be inaccurate due to *value bias*, the *observation method* used, *inter/intra-observer variability*, or the *awareness effect*. In some cases, a single RR measurement may even be affected by several of these factors. Hence, clinicians should interpret recorded RR data cautiously unless systems are in place to ensure its accuracy. For nurses, this includes ensuring that patients are unaware that their RR is being measured, counting rather than estimating RRs, employing 60-s counts whenever possible, and documenting the resulting value. Future research should investigate the underlying causes of the sources of inaccuracy identified in this review, other potential sources that have not previously been investigated, and methods to improve observation and documentation quality. For any given site, the solutions employed should be tailored to take into account the local organizational and cultural context, the available resources, and the specific measurement issues that need to be addressed.

CONFLICT OF INTEREST

The fifth author has one relevant patent issued (*methods and means of physiological monitoring*), but none of the articles cited in this review relate directly to it. All other authors declare no conflict of interest.

AUTHOR CONTRIBUTIONS

All authors have agreed on the final version and meet at least one of the following criteria (recommended by the ICMJE (http://www.icmje.

org/recommendations/)): (1) substantial contributions to conception and design, acquisition of data, or analysis and interpretation of data; (2) drafting the article or revising it critically for important intellectual content.

ORCID

Andrew Hill https://orcid.org/0000-0002-9111-3240

REFERENCES

Andrews, T., & Waterman, H. (2005). Packaging: A grounded theory of how to report physiological deterioration effectively. *Journal of Advanced Nursing*, *52*, 473–481. https://doi.org/10.1111/j.1365-2648.2005.03615.x

Ansell, H., Meyer, A., & Thompson, S. (2014). Why don't nurses consistently take patient respiratory rates? *British Journal of Nursing*, *23*(8), 414–418. https://doi.org/10.12968/bjon.2014.23.8.414

Armstrong, B., Walthall, H., Clancy, M., Mullee, M., & Simpson, H. (2008). Recording of vital signs in a district general hospital emergency department. *Emergency Medicine Journal*, *25*(12), 799–802. https://doi.org/10.1136/emj.2007.052951

Badawy, J., Nguyen, O. K., Clark, C., Halm, E. A., & Makam, A. N. (2017). Is everyone really breathing 20 times a minute? Assessing epidemiology and variation in recorded respiratory rate in hospitalised adults. *BMJ Quality & Safety*, *26*(10), 832–836. https://doi.org/10.1136/bmjqs-2017-006671

Barthel, P., Wensel, R., Bauer, A., Muller, A., Wolf, P., Ulm, K., Huster, K. M., Francis, D. P., Malik, M., & Schmidt, G. (2013). Respiratory rate predicts outcome after acute myocardial infarction: A prospective cohort study. *European Heart Journal*, *34*(22), 1644–1650. https://doi.org/10.1093/eurheartj/ehs420

Bergrath, S., Skorning, M., Rörtgen, D., Beckers, S. K., Brokmann, J. C., Mutscher, C., & Rossaint, R. (2011). Is paper-based documentation in an emergency medical service adequate for retrospective scientific analysis? An evaluation of a physician-run service. *Emergency Medicine Journal*, *28*(4), 320–324. https://doi.org/10.1136/emj.2009.086538

Berinato, S. (2003). *All systems down*. Retrieved from: http://www.computerworld.com/article/2581420/disaster-recovery/all-systems-down.html.

Bianchi, W., Dugas, A. F., Hsieh, Y.-H., Saheed, M., Hill, P., Lindauer, C., Terzis, A., & Rothman, R. E. (2013). Revitalizing a vital sign: Improving detection of tachypnea at primary triage. *Annals of Emergency Medicine*, *61*(1), 37–43. https://doi.org/10.1016/j.annemergmed.2012.05.030

Black, J., Gerdtz, M., Nicholson, P., Crellin, D., Browning, L., Simpson, J., Bell, L., & Santamaria, N. (2015). Can simple mobile phone applications provide reliable counts of respiratory rates in sick infants and children? An initial evaluation of three new applications. *International Journal of Nursing Studies*, *52*(5), 963–969. https://doi.org/10.1016/j.ijnurstu.2015.01.016

Bleyer, A. J., Vidya, S., Russell, G. B., Jones, C. M., Sujata, L., Daeihagh, P., & Hire, D. (2011). Longitudinal analysis of one million vital signs in patients in an academic medical center. *Resuscitation*, *82*(11), 1387–1392. https://doi.org/10.1016/j.resuscitation.2011.06.033

Brekke, I. J., Puntervoll, L. H., Pedersen, P. B., Kellett, J., & Brabrand, M. (2019). The value of vital sign trends in predicting and monitoring clinical deterioration: A systematic review. *PLoS One*, *14*(1), e0210875–https://doi.org/10.1371/journal.pone.0210875

Cahill, H., Jones, A., Herkes, R., Cook, K., Stirling, A., Halbert, T., Yates, A., Lal, S., Gardo, A., Donnelly, R., & Gattas, D. J. (2011). Introduction of a new observation chart and education programme is associated with higher rates of vital-sign ascertainment in hospital wards. *BMJ Quality & Safety*, *20*(9), 791–796. https://doi.org/10.1136/bmjqs.2010.045096

Cardona-Morrell, M., Prgomet, M., Lake, R., Nicholson, M., Harrison, R., Long, J., Westbrook, J., Braithwaite, J., & Hillman, K. (2016). Vital signs monitoring and nurse–patient interaction: A qualitative observational study of hospital practice. *International Journal of Nursing Studies*, *56*, 9–16. https://doi.org/10.1016/j.ijnurstu.2015.12.007

Cardoso, F. S., Karvellas, C. J., Kneteman, N. M., Meeberg, G., Fidalgo, P., & Bagshaw, S. M. (2014). Respiratory rate at intensive care unit discharge after liver transplant is an independent risk factor for intensive care unit readmission within the same hospital stay: A nested case-control study. *Journal of Critical Care*, *29*(5), 791–796. https://doi.org/10.1016/j.jcrc.2014.03.038

Castillo, J., & Sánchez-Sosa, J. (2002). Well-being and medical recovery in the Critical Care Unit: The role of the nurse-patient interaction. *Salud Mental*, *25*(2), 21–31.

Chen, J., Hillman, K., Bellomo, R., Flabouris, A., Finfer, S., & Cretikos, M. (2009). The impact of introducing medical emergency team system on the documentations of vital signs. *Resuscitation*, *80*(1), 35–43. https://doi.org/10.1016/j.resuscitation.2008.10.009

Christofidis, M. J., Hill, A., Horswill, M. S., & Watson, M. O. (2015). Less is more: The design of early-warning scoring systems affects the speed and accuracy of scoring. *Journal of Advanced Nursing*, *71*(7), 1573–1586. https://doi.org/10.1111/jan.12618

Christofidis, M. J., Hill, A., Horswill, M. S., & Watson, M. O. (2016). Observation chart design features affect the detection of patient deterioration: A systematic experimental evaluation. *Journal of Advanced Nursing*, *72*(1), 158–172. https://doi.org/10.1111/jan.12824

Churpek, M. M., Adhikari, R., & Edelson, D. P. (2016). The value of vital sign trends for detecting clinical deterioration on the wards. *Resuscitation*, *102*, 1–5. https://doi.org/10.1016/j.resuscitation

Churpek, M. M., Snyder, A., Twu, N. M., & Edelson, D. P. (2018). Accuracy comparisons between manual and automated respiratory rate for detecting clinical deterioration in ward patients. *Journal of Hospital Medicine*, *13*(7), 486–489. https://doi.org/10.12788/jhm.2914

Collier, R. (2017). NHS ransomware attack spreads worldwide. *Canadian Medical Association Journal*, *189*(22), E786–E787. https://doi.org/10.1503/cmaj.1095434

Considine, J., Potter, R., & Jenkins, J. (2006). Can written nursing practice standards improve documentation of initial assessment of ED patients? *Australasian Emergency Nursing Journal*, *9*(1), 11–18. https://doi.org/10.1016/j.aenj.2006.03.004

Considine, J., Thomas, S., & Potter, R. (2009). Predictors of critical care admission in emergency department patients triaged as low to moderate urgency. *Journal of Advanced Nursing*, *65*(4), 818–827. https://doi.org/10.1111/j.1365-2648.2008.04938.x

Cooper, S., Cant, R., & Sparkes, L. (2013). Respiratory rate records: The repeated rate? *Journal of Clinical Nursing*, *23*(9–10), 1236–1238. https://doi.org/10.1111/jocn.12234

Crandon, I. W., Harding, H. E., Williams, E. W., & Cawich, S. O. (2008). Inter-hospital transfer of trauma patients in a developing country: A prospective descriptive study. *International Journal of Surgery*, *6*(5), 387–391. https://doi.org/10.1016/j.ijsu.2008.07.004

Cretikos, M., Chen, J., Hillman, K., Bellomo, R., Finfer, S., & Flabouris, A. (2007). The objective medical emergency team activation criteria: A case–control study. *Resuscitation*, *73*(1), 62–72. https://doi.org/10.1016/j.resuscitation.2006.08.020

Dall'Ora, C., Griffiths, P., Redfern, O., Recio-Saucedo, A., Meredith, P., & Ball, J. (2019). Nurses' 12-hour shifts and missed or delayed vital signs observations on hospital wards: Retrospective observational study. *British Medical Journal Open*, *9*(1), e024778. https://doi.org/10.1136/bmjopen-2018-024778

Dinh, M. M., Oliver, M., Bein, K., Muecke, S., Carroll, T., Veillard, A.-S., Gabbe, B. J., & Ivers, R. (2013). Level of agreement between prehospital and emergency department vital signs in trauma patients. *Emergency Medicine Australasia*, *25*(5), 457–463. https://doi.org/10.1111/1742-6723.12126

Edwards, S. M., & Murdin, L. (2001). Respiratory rate–an under-documented clinical assessment. *Clinical Medicine*, *1*(1), 85. https://doi.org/10.7861/clinmedicine.1-1-85

Elliott, M. (2016). Why is respiratory rate the neglected vital sign? A narrative review. *International Archives of Nursing and Health Care*, *2*(3), 1–4. https://doi.org/10.23937/2469-5823/1510050

Evans, E. C. (2016). Exploring the nuances of nurse-patient interaction through concept analysis: Impact on patient satisfaction. *Nursing Science Quarterly*, *29*(1), 62–70. https://doi.org/10.1177/0894318415614904

Fieselmann, J. F., Hendryx, M. S., Helms, C. M., & Wakefield, D. S. (1993). Respiratory rate predicts cardiopulmonary arrest for internal medicine inpatients. *Journal of General Internal Medicine*, *8*(7), 354–360. https://doi.org/10.1007/BF02600071

Flinders, K. (2015). *Network failure crashed frontline services at London hospital*. Retrieved from: http://www.computerweekly.com/news/4500247512/Network-failure-crashed-frontline-services-at-London-hospital

Freeman, M. F., & Tukey, J. W. (1950). Transformations related to the angular and the square root. *The Annals of Mathematical Statistics*, *21*(4), 607–611. https://doi.org/10.2307/2236611

Gan, H., Karlen, W., Dunsmuir, D., Zhou, G., Chiu, M., Dumont, G. A., & Ansermino, J. M. (2015). The performance of a mobile phone respiratory rate counter compared to the WHO ARI timer. *Journal of Healthcare Engineering*, *6*(4), 691–704. https://doi.org/10.1260/2040-2295.6.4.691

Gao, H., McDonnell, A., Harrison, D. A., Moore, T., Adam, S., Daly, K., Esmonde, L., Goldhill, D. R., Parry, G. J., Rashidian, A., Subbe, C. P., & Harvey, S. (2007). Systematic review and evaluation of physiological track and trigger warning systems for identifying at-risk patients on the ward. *Intensive Care Medicine*, *33*(4), 667–679. https://doi.org/10.1007/s00134-007-0532-3

Gerdtz, M. F., Waite, R., Vassiliou, T., Garbutt, B., Prematunga, R., & Virtue, E. (2013). Evaluation of a multifaceted intervention on documentation of vital signs at triage: A before-and-after study. *Emergency Medicine Australasia: EMA*, *25*(6), 580–587. https://doi.org/10.1111/1742-6723.12153

Goldhill, D. R., White, S. A., & Sumner, A. (1999). Physiological values and procedures in the 24 h before ICU admission from the ward. *Anaesthesia*, *54*(6), 529–534. https://doi.org/10.1046/j.1365-2044.1999.00837.x

Gordon, C. F., & Beckett, D. J. (2011). Significant deficiencies in the overnight use of a Standardised Early Warning Scoring system in a teaching hospital. *Scottish Medical Journal*, *56*(1), 15–18. https://doi.org/10.1258/smj.2010.010009

Granholm, A., Pedersen, N. E., Lippert, A., Petersen, L. F., & Rasmussen, L. S. (2016). Respiratory rates measured by a standardised clinical approach, ward staff and a wireless device. *Acta Anaesthesiologica Scandinavica*, *60*(10), 1444–1452. https://doi.org/10.1111/aas.12784

Hall, J. A., & Rosenthal, R. (2018). Choosing between random effects models in meta-analysis: Units of analysis and the generalizability of obtained results. *Social and Personality Psychology Compass*, *12*(10), e12414. https://doi.org/10.1111/spc3.12414.

Hall, S., Williams, E., Richards, S., Subbe, C., & Gemmell, L. (2003). Waiting to exhale: Critical Care Outreach and recording of ventilatory frequency. *British Journal of Anaesthesia*, *90*, 570P–571P.

Han, J. N., Stegen, K., Cauberghs, M., & Van de Woestijne, K. P. (1997). Influence of awareness of the recording of breathing on respiratory pattern in healthy humans. *The European Respiratory Journal*, *10*(1), 161–166. https://doi.org/10.1183/09031936.97.10010161

Helliwell, V., Hadfield, J., & Gould, T. (2002). Documentation of respiratory rate for acutely sick hospital inpatients – An observational study. *Intensive Care Medicine*, *28*, S21.

Hill, A., Kelly, E., Horswill, M. S., & Watson, M. O. (2018). The effects of awareness and count duration on adult respiratory rate measurements: An experimental study. *Journal of Clinical Nursing*, *27*(3–4), 546–554. https://doi.org/10.1111/jocn.13861

Hodgetts, T. J., Kenward, G., Vlackonikolis, I., Payne, S., Castle, N., Crouch, R., Ineson, N., & Shaikh, L. (2002). Incidence, location and reasons for avoidable in-hospital cardiac arrest in a district general hospital. *Resuscitation*, *54*(2), 115–123. https://doi.org/10.1016/s0300-9572(02)00098-9

Hogan, H., Hutchings, A., Wulff, J., Carver, C., Holdsworth, E., Welch, J., & Black, N. (2019). A qualitative study of staff perspectives on measures to prevent patient deterioration. *Health Services & Delivery Research*, *7*(2), 29–41.

Hogan, J. (2006). Why don't nurses monitor the respiratory rates of patients? *British Journal of Nursing*, *15*(9), 489–492. https://doi.org/10.12968/bjon.2006.15.9.21087

Hooker, E. A., O'Brien, D. J., Danzl, D. F., Barefoot, J. A., & Brown, J. E. (1989). Respiratory rates in emergency department patients. *The Journal of Emergency Medicine*, *7*(2), 129–132. https://doi.org/10.1016/0736-4679(89)90257-6

Jonas, F. M. (2015). Solar flares. In J. N. Pelton, & F. Allahdad (Eds.), *Handbook of cosmic hazards and planetary defense* (pp. 243–257). Springer International Publishing.

Jonsson, T., Jonsdottir, H., Möller, A. D., & Baldursdottir, L. (2011). Nursing documentation prior to emergency admissions to the intensive care unit. *Nursing in Critical Care*, *16*(4), 164–169. https://doi.org/10.1111/j.1478-5153.2011.00427.x

Kallioinen, N., Hill, A., Horswill, M., Ward, H., & Watson, M. (2017). Sources of inaccuracy in the measurement of adult patients' resting blood pressure in clinical settings: A systematic review. *Journal of Hypertension*, *35*(3), 421–441. https://doi.org/10.1097/HJH.0000000000001197

Karlen, W., Dunsmuir, D., & Ansermino, J. M. (2015). Efficiency of respiratory rate measurements: Comment on Black et al. *International Journal of Nursing Studies*, *52*(7), 1279–1280. https://doi.org/10.1016/j.ijnurstu.2015.03.014

Keene, C. M., Kong, V. Y., Clarke, D. L., & Brysiewicz, P. (2017). The effect of the quality of vital sign recording on clinical decision making in a regional acute care trauma ward. *Chinese Journal of Traumatology*, *20*(5), 283–287. https://doi.org/10.1016/j.cjtee.2016.11.008

Kenward, G., Hodgetts, T. J., & Castle, N. (2001). Time to put the R back in TPR. *Nursing times*, *97*, 32.

Kenzaka, T., Okayama, M., Kuroki, S., Fukui, M., Yahata, S., Hayashi, H., Kitao, A., Sugiyama, D., Kajii, E., & Hashimoto, M. (2012). Importance of vital signs to the early diagnosis and severity of sepsis: Association between vital signs and sequential organ failure assessment score in patients with sepsis. *Internal Medicine*, *51*(8), 871–876. https://doi.org/10.2169/internalmedicine.51.6951

Kmet, L. M., Cook, L. S., & Lee, R. C. (2004). *Standard quality assessment criteria for evaluating primary research papers from a variety of fields*. Alberta Heritage Foundation for Medical Research. https://doi.org/10.7939/R37M04F16

Lim, W. S., Carty, S. M., Macfarlane, J. T., Anthony, R. E., Christian, J., Dakin, K. S., & Dennis, P. M. (2002). Respiratory rate measurement in adults–how reliable is it? *Respiratory Medicine*, *96*(1), 31–33. https://doi.org/10.1053/rmed.2001.1203

Linares, T., Campos, A., Torres, M., & Reyes, J. (2006). Medical audit on asthma in an emergency department. *Allergologia Et Immunopathologia*, *34*(6), 248–251. https://doi.org/10.1157/13095872

Liu, H., Allen, J., Zheng, D., & Chen, F. (2019). Recent development of respiratory rate measurement technologies. *Physiological Measurement*, *40*(7), 1–27. https://doi.org/10.1088/1361-6579/ab299e

Lovett, P. B., Buchwald, J. M., Stürmann, K., & Bijur, P. (2005). The vexatious vital: Neither clinical measurements by nurses nor an electronic monitor provides accurate measurements of respiratory rate in triage. *Annals of Emergency Medicine*, *45*(1), 68–76. https://doi.org/10.1016/j.annemergmed.2004.06.016

Ludikhuize, J., de Jonge, E., & Goossens, A. (2011). Measuring adherence among nurses one year after training in applying the Modified Early Warning Score and Situation-Background-Assessment-Recommendation instruments. *Resuscitation*, *82*(11), 1428–1433. https://doi.org/10.1016/j.resuscitation.2011.05.026

Ma, H. M., Tang, W. H., & Woo, J. (2011). Predictors of in-hospital mortality of older patients admitted for community-acquired pneumonia. *Age and Ageing*, *40*(6), 736–741. https://doi.org/10.1093/ageing/afr087

McBride, J., Knight, D., Piper, J., & Smith, G. B. (2005). Long-term effect of introducing an early warning score on respiratory rate charting on general wards. *Resuscitation*, *65*(1), 41–44. https://doi.org/10.1016/j.resuscitation.2004.10.015

McGain, F., Cretikos, M. A., Jones, D., Van Dyk, S., Buist, M. D., Opdam, H., Pellegrino, V., Robertson, M. S., & Bellomo, R. (2008). Documentation of clinical review and vital signs after major surgery. *The Medical Journal of Australia*, *189*(7), 380–383. https://doi.org/10.5694/j.1326-5377.2008.tb02083.x

MERIT study investigators (2005). Introduction of the medical emergency team (MET) system: A cluster-randomised controlled trial. *The Lancet*, *365*(9477), 2091–2097. https://doi.org/10.1016/s0140-6736(05)66733-5

Miller, J. J. (1978). The inverse of the Freeman-Tukey double arcsine transformation. *The American Statistician*, *32*(4), 138. https://doi.org/10.1080/00031305.1978.10479283

Mohammed, M. A., Hayton, R., Clements, G., Smith, G., & Prytherch, D. (2009). Improving accuracy and efficacy of early-warning scores in acute care. *British Journal of Nursing*, *18*, 18–23. https://doi.org/10.12968/bjon.2009.18.1.32072

Mukkamala, S. G., Gennings, C., & Wenzel, R. P. (2008). R = 20: Bias in the reporting of respiratory rates. *The American Journal of Emergency Medicine*, *26*(2), 237–239. https://doi.org/10.1016/j.ajem.2007.05.001

Nielsen, L. G., Folkestad, L., Brodersen, J. B., & Brabrand, M. (2015). Inter-observer agreement in measuring respiratory rate. *PLoS One*, *10*(6), e0129493. https://doi.org/10.1371/journal.pone.0129493

Nurmi, J., Harjola, V. P., Nolan, J., & Castrén, M. (2005). Observations and warning signs prior to cardiac arrest. Should a medical emergency team intervene earlier? *Acta Anaesthesiologica Scandinavica*, *49*(5), 702–706. https://doi.org/10.1111/j.1399-6576.2005.00679.x

O'Reilly, G. M., Cameron, P. A., & Jolley, D. J. (2012). Which patients have missing data? An analysis of missingness in a trauma registry. *Injury*, *43*(11), 1917–1923. https://doi.org/10.1016/j.injury.2012.07.185

Odell, M., Rechner, I. J., Kapila, A., Even, T., Oliver, D., Davies, C., Milsom, L., Forster, A., & Rudman, K. (2007). The effect of a critical care outreach service and an early warning scoring system on respiratory rate recording on the general wards. *Resuscitation*, *74*(3), 470–475. https://doi.org/10.1016/j.resuscitation.2007.01.035

Parkes, R. (2011). Rate of respiration: The forgotten vital sign. *Emergency Nurse*, *19*(2), 12–17. https://doi.org/10.7748/en2011.05.19.2.12.c8504

Pedersen, N. E., Rasmussen, L. S., Petersen, J. A., Gerds, T. A., Østergaard, D., & Lippert, A. (2018). A critical assessment of early warning score records in 168,000 patients. *Journal of Clinical Monitoring and Computing*, *32*(1), 109–116. https://doi.org/10.1007/s10877-017-0003-5

Philip, K. E. J., Pack, E., Cambiano, V., Rollmann, H., Weil, S., & O'Beirne, J. (2015). The accuracy of respiratory rate assessment by doctors in a London teaching hospital: A cross-sectional study. *Journal of Clinical Monitoring and Computing*, *29*(4), 455–460. https://doi.org/10.1007/s10877-014-9621-3.

Prytherch, D. R., Smith, G. B., Schmidt, P., Featherstone, P. I., Stewart, K., Knight, D., & Higgins, B. (2006). Calculating early-warning scores – A classroom comparison of pen and paper and hand-held computer methods. *Resuscitation*, *70*, 173–178. https://doi.org/10.1016/j.resuscitation.2005.12.002

R Core Team. (2018). *R: A language and environment for statistical computing (Version 3.5.2)*, Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org

Ramgopal, S., Elmer, J., Escajeda, J., & Martin-Gill, C. (2018). Differences in prehospital patient Assessments for pediatric versus adult patients. *The Journal of Pediatrics*, *199*, 200–205.e6. https://doi.org/10.1016/j.jpeds.2018.03.069

Rimbi, M., Dunsmuir, D., Ansermino, J. M., Nakitende, I., Namujwiga, T., & Kellett, J. (2019). Respiratory rates observed over 15 and 30 seconds compared with rates measured over 60 seconds. Practice-based evidence from an observational study of acutely ill adult medical patients during hospital admission. *QJM*, *112*(7), 513–517. https://doi.org/10.1093/qjmed/hcz065

Rosen, M. A., Chima, A. M., Sampson, J. B., Jackson, E. V., Koka, R., Marx, M. K., Kamara, T. B., Ogbuagu, O. U., & Lee, B. H. (2015). Engaging staff to improve quality and safety in an austere medical environment: A case-control study in two Sierra Leonean hospitals. *International Journal for Quality in Health Care*, *27*(4), 320–327. https://doi.org/10.1093/intqhc/mzv030

Russ, A. L., Fairbanks, R. J., Karsh, B.-T., Militello, L. G., Saleem, J. L., & Wears, R. L. (2013). The science of human factors: Separating fact from fiction. *BMJ Quality & Safety*, *22*, 802–808. https://doi.org/10.1136/bmjqs-2012-001450

Schein, R. M. H., Hazday, N., Pena, M., Ruben, B. H., & Sprung, C. L. (1990). Clinical antecedents to in-hospital cardiopulmonary arrest. *Chest*, *98*(6), 1388–1392. https://doi.org/10.1378/chest.98.6.1388

Semler, M. W., Stover, D. G., Copland, A. P., Hong, G., Johnson, M. J., Kriss, M. S., Otepka, H., Wang, L. I., Christman, B. W., & Rice, T. W. (2013). Flash mob research: A single-day, multicenter, resident-directed study of respiratory rate. *Chest*, *143*(6), 1740–1744. https://doi.org/10.1378/chest.12-1837

Sinnecker, D., Dommasch, M., Barthel, P., Müller, A., Dirschinger, R. J., Hapfelmeier, A., Huster, K. M., Laugwitz, K.-L., Malik, M., & Schmidt, G. (2014). Assessment of mean respiratory rate from ECG recordings for risk stratification after myocardial infarction. *Journal of Electrocardiology*, *47*(5), 700–704. https://doi.org/10.1016/j.jelectrocard.2014.04.021

Smith, A. F., & Oakey, R. J. (2006). Incidence and significance of errors in a patient "track and trigger" system during an epidemic of Legionnaires' disease: Retrospective casenote analysis. *Anaesthesia*, *61*(3), 222–228. https://doi.org/10.1111/j.1365-2044.2005.04513.x

Smith, G. B., Prytherch, D. R., Schmidt, P. E., & Featherstone, P. I. (2008). Review and performance evaluation of aggregate weighted 'track and trigger' systems. *Resuscitation*, *77*(2), 170–179. https://doi.org/10.1016/j.resuscitation.2007.12.004

Strauß, R., Ewig, S., Richter, K., König, T., Heller, G., & Bauer, T. T. (2014). The prognostic significance of respiratory rate in patients With pneumonia. *Deutsches Ärzteblatt International*, *111*(29–30), 503–508. https://doi.org/10.3238/arztebl.2014.0503

Subbe, C. P., Davies, R. G., Williams, E., Rutherford, P., & Gemmell, L. (2003). Effect of introducing the Modified Early Warning Score on clinical outcomes, cardio-pulmonary arrests and intensive care utilisation in acute medical admissions. *Anaesthesia*, *58*(8), 797–802. https://doi.org/10.1046/j.1365-2044.2003.03258.x

Subbe, C. P., Gao, H., & Harrison, D. A. (2007). Reproducibility of physiological track-and-trigger warning systems for identifying at-risk patients on the ward. *Intensive Care Medicine*, *33*, 619–624. https://doi.org/10.1007/s00134-006-0516-8

Sundararajan, K., Flabouris, A., Thompson, C., & Seppelt, I. (2016). Hospital overnight and evaluation of systems and timelines study: A point prevalence study of practice in Australia and New Zealand. *Resuscitation*, *100*, 1–5. https://doi.org/10.1016/j.resuscitation.2015.11.029

Takayama, A., Takeshima, T., Nakashima, Y., Yoshidomi, T., Nagamine, T., & Kotani, K. (2019). A comparison of methods to count breathing

frequency. *Respiratory Care*, *64*(5), 555–563. https://doi.org/10.4187/respcare.06451

The ANZICS-CORE MET dose Investigators. (2012). Rapid Response Team composition, resourcing and calling criteria in Australia. *Resuscitation*, *83*(5), 563–567. https://doi.org/10.1016/j.resuscitation.2011.10.023

Tipton, E., & Shuster, J. (2017). A framework for the meta-analysis of Bland-Altman studies based on a limits of agreement approach. *Statistics in Medicine*, *36*(23), 3621–3635. https://doi.org/10.1002/sim.7352

Tysinger, E. L. (2015). How vital are vital signs? A systematic review of vital sign compliance and accuracy in nursing. *Journal of Science and Medicine*, *1*, 68–75.

Van Leuvan, C. H., & Mitchell, I. (2008). Missed opportunities? An observational study of vital sign measurements. *Critical Care and Resuscitation*, *10*(2), 111–115.

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, *36*(3), 1–48. https://doi.org/10.18637/jss.v036.i03

Western, P. J., & Patrick, J. M. (1988). Effects of focusing attention on breathing with and without apparatus on the face. *Respiration Physiology*, *72*(1), 123–130. https://doi.org/10.1016/0034-5687(88)90084-9

World Health Organisation. (1992). *Programme for the control of acute respiratory infections: Fifth programme report, 1990–1991*. Switzerland.

World Health Organisation. (1993). *Programme for the control of acute respiratory infections: Interim programme report, 1992*. Switzerland.

Worster, A., Elliott, L., Bose, T. J., & Chemeris, E. (2003). Reliability of vital signs measured at triage. *European Journal of Emergency Medicine*, *10*(2), 108–110. https://doi.org/10.1097/00063110-200306000-00008

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: 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*. 2021;77:98–124. https://doi.org/10.1111/jan.14584