

# Poor inter-observer agreement in the measurement of respiratory rate in children: a prospective observational study

William James Daw,<sup>1,2</sup> Ruth N Kingshott,<sup>1,2</sup> Heather E Elphick<sup>1,2</sup>

**To cite:** Daw WJ, Kingshott RN, Elphick HE. Poor inter-observer agreement in the measurement of respiratory rate in children: a prospective observational study. *BMJ Paediatrics Open* 2017;1:e000173. doi:10.1136/bmjpo-2017-000173

Received 29 June 2017

Revised 6 October 2017

Accepted 10 October 2017

## ABSTRACT

**Objective** To determine the inter-observer agreement of a respiratory rate (RR) count on a child when assessed by three independent observers.

**Design** The RR of 169 children (age range: 3 days to 15 years) was measured by three independent observers over a 3-month period. The first RR was taken by different healthcare professionals (HCPs) from within the hospital using their own preferred method of measurement. A further count of RR was then taken by two observers from the research team simultaneously within 30 min of the first measurement, using the WHO-recommended method of measurement.

**Results** 507 RR measurements were taken on 169 children. Median RR showed a 4 beats per minute (bpm) difference between the HCP (median RR 32 bpm) and the researchers (median RR 28 bpm). The 95% limits of agreement between the first measurement and second and third measurements were -10.2 to 17.7 bpm and -11.4 to 18.7 bpm, respectively. For simultaneous measurements, the 95% limits of agreement were -7.1 to 7.0 bpm. 81 children had a RR  $\geq$ 95th centile for their age and an even poorer level of agreement was seen in these children than in those whose RR was within normal range. In only 27 of these 81 children (33%) did all three observers agree on the presence of a raised RR.

**Conclusions** Inter-observer agreement for the measurement of RR in children is poor. The effect that this variation has on the clinical assessment and subsequent management of a child may be significant. These findings highlight the need for a robust review of our current measurement methods and interpretation of an important vital sign.

## INTRODUCTION

The measurement of a child's vital signs including heart rate, temperature, capillary refill and respiratory rate (RR) is routine practice to all those who attend emergency departments and paediatric assessment units. The National Institute for Health and Care Excellence also recommends that these signs are recorded for all children presenting with a fever.<sup>1</sup> RR is an important vital sign used in the initial and ongoing assessment of unwell

## What is already known on this topic?

- ▶ Respiratory rate is an important vital sign used in the assessment and management of unwell children.
- ▶ Tachypnoea is a key criterion used in assessing the unwell child and guidelines for conditions such as pneumonia rely on tachypnoea in its diagnostic criteria.

## What this study hopes to add?

- ▶ Inter-observer agreement in the measurement of respiratory rate in children is poor.
- ▶ The variability seen between healthcare professional's measurements in clinical practice and observers under research conditions highlights the inaccurate methods that are being employed.
- ▶ Increased variability in measurements exist when respiratory rates are raised potentially impacting on the recognition and identification of tachypnoea.

children.<sup>2</sup> It can be used to assess a child's clinical status and as a predictor of serious deterioration.<sup>3</sup> It is also incorporated into early warning scoring systems, which are now used widely in paediatric clinical care and have been shown to accurately identify children who may be deteriorating.<sup>4-7</sup>

RR may be measured by observing abdominal or chest movements or by auscultation. Both methods have been shown to give similar results.<sup>8</sup> The current WHO standard for RR measurement is a count over a full minute by observing abdominal and chest movements.<sup>9</sup> However, in practice it is usual for a direct observation of respirations to take place for 15 seconds and then be multiplied by four, to save time. This method has been shown to lead to inaccuracies, and when compared with pneumogram measurements quadrupling a 15 second count showed up to 50% inaccuracy.<sup>10</sup>



CrossMark

<sup>1</sup>Department of Respiratory Medicine, Sheffield Children's Hospital, Sheffield, UK

<sup>2</sup>Academic Unit of Child Health, Sheffield Children's Hospital, Sheffield, UK

## Correspondence to

Dr William James Daw;  
willj.daw@gmail.com

**Table 1** 95th centile respiratory rates (RRs) by age group<sup>16</sup>

Age range	<3 months	3–6 months	6–18 months	1.5–2 years	2–8 years	8–12 years	12 years +
RR (bpm)	50	45	40	35	30	25	24

Convenient electronic devices exist for the measurement of the other vital signs including pulse, blood pressure, oxygen saturation and temperature. These provide accurate and prompt measures, which are easier to achieve and also take away some of the subjectivity. RR, however, remains a subjective assessment and even if full compliance with recommendations was achieved variability can still be expected. In children this variability may be higher than in adults as they may not be as cooperative during the measurement and their RR may also vary quickly between breaths. Devices for monitoring RR exist and have entered the commercial market, but there is currently no clinically validated device available that gives an accurate and rapid measurement in acute clinical practice.<sup>11</sup>

If high levels of inconsistencies in RR measurements exist, then this will call into question the reliability of such an important vital sign. It may also impact greatly on the child, their clinical assessment and accurate identification of possible deterioration. The aim of this study was to determine the degree of inter-observer agreement in RR measurements of children presenting to a tertiary children's hospital.

## METHODS

### Study design and setting

This study was a prospective observational study conducted at Sheffield Children's Hospital.

### Participants and eligibility criteria

Children between the ages of 0–16 years with any clinical condition who had had their RR measured within the previous 30 minutes were recruited. All children were clinically stable on one of the hospital wards and had already had at least one RR measurement taken during their admission. Children were excluded if they were acutely unwell or had had any clinical intervention in the period between the initial RR measurement and the planned subsequent measurement. Participants were recruited between the months of August 2016 and October 2016.

### Sample size

The sample size was calculated based on a previous pilot study<sup>12</sup> using the statistical programme STATA v.14. To detect a bidirectional mean difference of  $\pm 2.0$  breaths/min with 90% power and a significance level of 5%, a sample size of 169 children was required. This was based on previously reported inter-observer limits of agreement in adults.<sup>13</sup> In total 169 children were recruited to the study.

### Recruitment

Participants were approached by members of the research team, and information was given to both parents and their child. There were no incentives offered to take part in the study.

### Data collection and procedure

Each participant was assigned a unique identifying number based on the order that they were recruited. Data on the participants age, sex, presenting complaint and activity status (asleep/active/awake) at the time of both the measurements were collected. The first RR taken by the healthcare professional (HCP) (RR1) was noted and the HCP was then asked as to the method and timing period used for their measurement as well as their subjective assessment of the child's activity status at the time of their measurement. A further count of RR was then taken by two different observers simultaneously within 30 minutes of the first measurement, the activity status at the time of this second measurement was again noted. These observers were members of the research team and consisted of a Paediatric Doctor (RR2) and Paediatric Respiratory Physiologist (RR3). They measured the RR using the WHO-recommended method of measurement.<sup>9</sup> The third observer was added to the study in order to assess the agreement between simultaneous measurements. All observers were blinded to each of the others' measurements.

### Statistical analysis

The inter-observer agreement was assessed by Bland-Altman analysis by calculating the mean difference between RR measurements with 95% limits of agreement (mean $\pm$ SD of the difference). Intraclass correlation coefficients were not reported as they only estimate the degree of association and do not reveal information about the individual differences between measures.<sup>14 15</sup> To assess any significant difference between different groups a Fisher r-to-z transformation was performed and differences expressed as P values. All results were analysed using SPSS V.22.0 for Mac.

The inter-observer agreement was also assessed for those children with a normal RR and for those who had a RR  $\geq$ 95th centile for their age, as defined by the resuscitation councils Advanced Paediatric Life Support guidelines (table 1).<sup>16</sup> These centile values have recently been updated and reflect those suggested by recent research findings.<sup>17</sup> A child was classified as having a raised RR when one or more of the observers measured a RR at or above the 95th centile for their age.

## RESULTS

### Participants

A total of 507 RR measurements were taken on 169 children. Fifty-three per cent of the participants were men

**Table 2** Patient characteristics (n=169)

Age in months, median, range	29 (0.1–192)
Male gender, n (%)	90 (53)
Primary presenting complaint, n (%)	
Increased work of breathing	39 (23.1)
Fever	22 (13.0)
Cough	16 (9.4)
Vomiting	20 (11.8)
Diarrhoea and vomiting	9 (5.3)
Skin complaint	8 (4.7)
Feeding difficulty	4 (2.4)
Headache	3 (1.8)
Burns	3 (1.8)
Surgical problem	9 (5.3)
Head injury	2 (1.2)
Seizure	5 (3.0)
Pain	5 (3.0)
Constipation	2 (1.2)
Planned admission/procedure	16 (9.4)
Other*	6 (3.5)

\*Included—anaphylaxis, accidental ingestion, animal bite, eye complaint and rheumatological complaint.

and the median age was 29 months. The youngest participant was 3 days and the oldest was 15 years and 11 months. The median time between the RR1 and RR2/RR3 measurements was 16 min (range: 1–30 min). [Table 2](#) shows the patient characteristics and presenting complaint and ([table 3](#)) the age range of children studied.

### Initial RR measurement

The initial RR was most often measured and recorded by a nurse (88%), who had varying levels of experience. [Table 4](#) shows the breakdown of HCPs taking the first RR and [table 5](#) shows the method of measurement that they used.

### Respiratory rates

RR measurements ranged from 11 to 65 breaths/min. [Figure 1](#) shows the variability between measurements for the three observers. RR1 had a median of 32 bpm (IQR: 24–40 bpm) RR2 a median of 28 bpm (IQR: 21–37 bpm) and RR3 a median of 28 bpm (IQR: 21–36 bpm). The

**Table 3** Age range of participants (n=169)

	n (%)
0–1 years	47 (28%)
1–2 years	29 (17%)
2–5 years	46 (27%)
5–12 years	30 (18%)
12 years+	17 (10%)

**Table 4** HCP taking RR1 (n=169)

	n (%)
Paediatric nurse band 5	82 (49%)
Paediatric nurse band 6	57 (34%)
Paediatric nurse band 7	9 (5%)
Paediatric healthcare worker	7 (4%)
Student nurse	14 (8%)

RR for some individual subjects was highly variable. The largest difference in a subject's RR from a measurement taken simultaneously (RR 2 and RR3) was 14 bpm.

### Agreement between different observers

When the RR measured by the HCP (RR1) was compared with the RR measured by the first observer (RR2, Paediatric Doctor) Bland-Altman analysis showed a mean difference of 3.8 with 95% limits of agreement of –10.2 to 17.7. When the RR measured by the HCP (RR1) was compared with the RR measured by the second observer (RR3, Paediatric Respiratory Physiologist) Bland-Altman analysis showed a mean difference of 3.7 with 95% limits of agreement of –11.4 to 18.7. When the RR measured by the simultaneous observers (RR2 and RR3) was compared, Bland-Altman analysis showed a mean difference of –0.1 with 95% limits of agreement of –7.1 to 7.0. [Figure 2](#) shows the Bland-Altman plots for each of these.

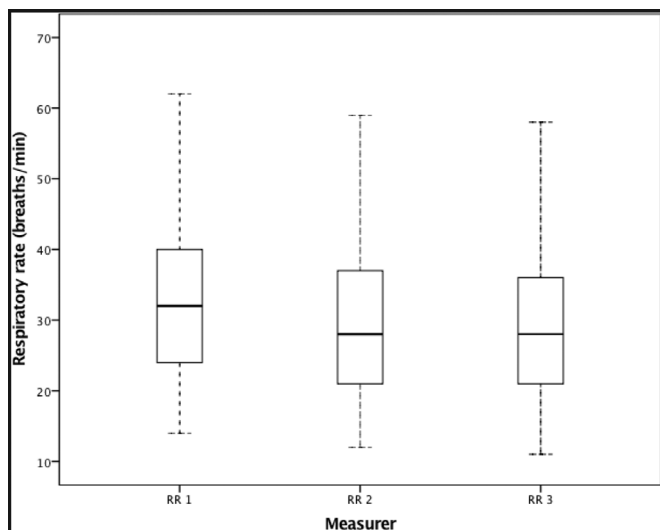
There was no significant difference observed in the pairwise agreements between measurements taken closer in time, within 0–10 min (49 children) and those taken further apart, within 20–30 min (69 children). With a mean difference of 3.7 and 95% limits of agreement of –9.9 to 17.4 for measurements taken closer in time and a mean difference of 3.6 and 95% limits of agreement –9.8 to 17.1 for those taken further apart (RR1–RR2 P=0.516, RR1–RR3 P=0.905). There was also no difference in agreement between measurements when stratifying for seniority of the HCP taking RR1.

For 26 participants (15%), the subjective assessment of the child's activity status during the measurement was different between the first and second/third RR measurements. Children whose activity status remained the same (143 children) showed a mean difference of 3.8 with 95% limits of agreement of –11.4 to 19.0 and children

**Table 5** Method of measurement (n=169)

	n (%)
Observation 10 s	11 (7%)
Observation 15 s	125 (74%)
Observation 30 s	16 (9%)
Observation 60 s	12 (7%)
Palpation 30 s	4 (2%)
Palpation 60 s	1 (<1%)

\*Observation/palpation of chest and abdominal movements.

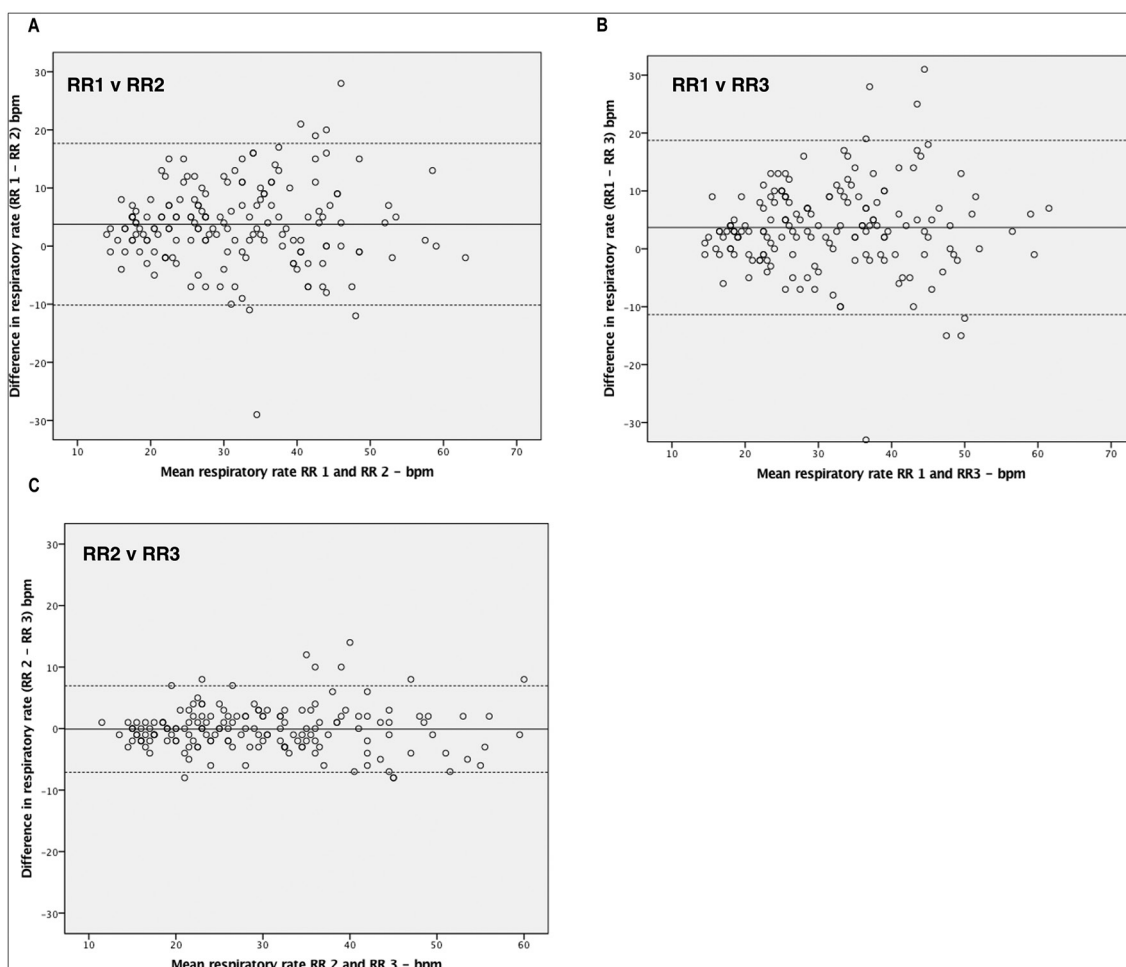


**Figure 1** Box plot showing the variability of RR measurements for each observer (RR1, RR2, RR3). The solid line in the middle of the box represents the median. The boxes span the interquartile range and the whiskers extend to +1.5 the interquartile range.

whose activity status differed (26 children) showed a mean difference of 3.0 with 95% limits of agreement of  $-11.3$  to  $17.3$ . This was not a statistically significant difference (RR1–RR2  $P=0.269$ , RR1–RR3  $P=0.210$ ).

#### Agreement between observers in children with RR $\geq 95$ th centile

A total of 48% (81 children) of all the measurements would have been classified as being at or above the 95th centile for the child's age,<sup>16</sup> by one or more of the three observers. Of these children in only 33% (27 children) did all three observers agree that the RR would have been at or above the 95th centile. In 28% (15 children) the HCP (RR1) would not have classified the child as having a raised RR, but one or both of the other observers would have done. Notably, when comparing measurements between the researchers and the HCP, in these children the agreement was statistically significantly different from the children whose RR was classified as being within the normal range by all of the observers. This indicated that at higher RRs less agreement between measurements was seen. However, for simultaneous measurements, there



**Figure 2** Bland-Altman plots assessing pairwise agreement for respiratory measurements by A. RR 1 and RR 2 B. RR 1 and RR 3 C. RR 2 and RR 3. The x-axis represents the mean values of the two measurements and the y-axis the difference between the two. The solid line shows the mean bias and the dashed lines the 95% CI based on the standard deviation of the distribution.

was no significant difference observed. Table 6 shows the 95% limits of agreement for the different groups along with the P values indicating the significance in the difference in agreement. Figure 3 shows the associated Bland-Altman plots.

## DISCUSSION

This study has examined the inter-observer agreement of RR measurements in children as encountered in day-to-day clinical practice in the UK. We have shown from 507 RR recordings that there is poor agreement between measurements when taken by a HCP in usual clinical practice, compared with researchers using the recommended WHO method within 30 min. Median RR showed a 4 bpm difference with median measurement from the HCP being 32 bpm and median for the researchers being 28 bpm. This could be explained by measurements often being taken over a duration of 15 s in clinical practice and being multiplied by four, resulting in an overestimate of 4 bpm due to observers invariably rounding values up rather than down. There was, however, a wide variability in agreement with 95% limits of agreement indicating that measurements in clinical practice may have varied from 11 beats below to 18 beats above the standardised WHO method. There was better agreement between the two researchers taking simultaneous measurements, but even then there was a difference of up to 14 bpm. In children with a RR  $\geq$ 95th centile for their age, there was an even poorer level of agreement seen than in children whose RR was within normal range, and in only 33% of children did all three observers agree on the presence of a raised RR.

The available studies to date assessing the inter-observer agreement of RR report a wide range of inter-observer variability in both children and adults.<sup>13 18–23</sup> This may reflect the heterogeneity of the studies, with many assessing the variability in RR measurements as part of a wider clinical score. Some studies only looked at small convenience samples and some looked at very narrow age ranges or specific clinical conditions only. Variation in assessments may also exist due to changes in the clinical status of the patient between measurements, which many of the studies do not account for, comparing measurements taken up to 6 or even 8 hours later.<sup>19 22</sup> Most studies in children report good agreement on the presence of

tachypnoea.<sup>19 22 23</sup> We have attempted to produce a study that could address these issues and bring a more conclusive answer.

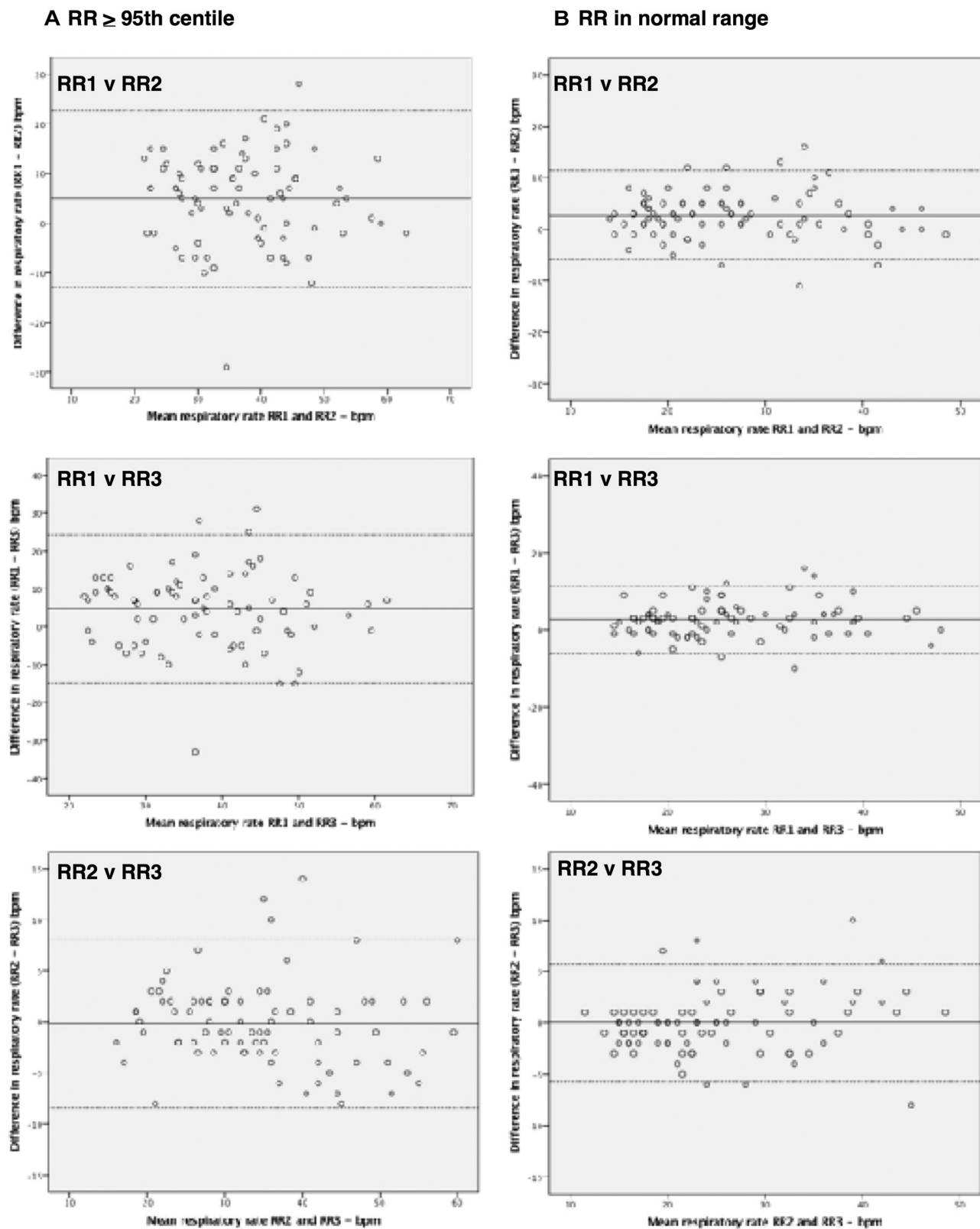
Many previous studies analyse and present their data by assessing the correlation between different measurements. However, there are no such studies in children reporting the agreement in RR measurements. One study in adults reported the limits of agreement in RR measurements for the same observer as being  $-4.86$  to  $4.94$  breaths/min and  $-5.7$  to  $5.7$  breaths/min for different observers.<sup>13</sup> We report much wider limits of agreement in children. This may be due to the nature of measuring a RR in a child, where the measurement often involves the observation of complex respiratory patterns in uncooperative subjects.

Overall the first measurement appeared to overestimate the RR, reflected by the mean measurements from each observer. This was likely to be due to the method of measurement used. In only 7% of measurements by the HCP was a 60 s RR count used. It is widely known this leads to inaccurate measurements.<sup>10 24</sup> RR1 was often a nurse and, to save time, nursing staff will often observe a RR for 15 s and multiply the result by four to get a value of breaths per minute. This would inevitably lead to an error of up to four breaths per minute as the observer would naturally round up rather than down. The agreement between the first and second, and first and third measurements was poorer than that of the simultaneousness measurements. This may be explained partly by the fact that RR1 was measured by multiple different HCPs whereas the second RR (RR2 and RR3) was consistently measured by the same observers. Although this may lead to a degree of variability between measurements it in fact enabled the RR1 measurement to represent current clinical practice where multiple HCPs will take a child's RR and as such gives a true indication of the variability that exists. The difference in measurements between the count by the HCP and a WHO standard count could have been anything from 11 breaths less to 18 breaths more per minute. This is potentially a significant level of variation in the context of clinical practice and it may have had clear implications on the sickness score given to the child and also on their subsequent clinical management.

A limitation of our study is that all three measurements were not recorded simultaneously. This would have been

**Table 6** Agreement of measurements based on respiratory rate (RR) range

Observer	RR range (no)	95% limits of agreement (mean difference)	Significance (P value)
RR1 vs RR2	$\geq$ 95th centile (81)	$-12.9$ to $22.7$ (4.9)	P=0.0002
	Normal range (88)	$-5.9$ to $11.3$ (2.7)	
RR1 vs RR3	$\geq$ 95th centile (81)	$-14.8$ to $24.4$ (4.8)	P=0.0001
	Normal range (88)	$-6.0$ to $11.4$ (2.7)	
RR2 vs RR3	$\geq$ 95th centile (81)	$-8.4$ to $8.1$ ( $-0.1$ )	P=0.184
	Normal range RR (88)	$-5.7$ to $5.7$ ( $-0.03$ )	



**Figure 3** Bland-Altman plots assessing pairwise agreement for measurements for children whose RR was  $>$ 95th centile (A) and those whose RR was within normal range (B).

possible, but we opted to delay the researchers' observations until the HCP measurement had taken place so that actual clinical practice could be recorded. If the HCP had been aware of the researchers taking the RR

simultaneously with them, this could have altered their method of measurement and would not have truly reflected their actual practice, leading to a bias in our results.

Importantly, there was no statistical change seen in agreement when comparing readings closer in time with readings over a longer time interval. The maximum time limit between the first and second/third measurements was 30 min. The child's RR may have changed in this time, depending on the child's activity and underlying illness, this could potentially produce a variation in measurements. However, this upper time limit between measurements remains less than or equal to previous studies.<sup>19 22</sup> Also changes in the activity status of the child between measurements did not affect agreement and therefore we do not believe that the time difference significantly affected our results.

We also showed that the agreement between simultaneous measurements using the WHO-recommended method of measurement could have been anything from seven breaths less to seven breaths more per minute. Previous studies have reported high correlation between measurements taken over 1 min,<sup>10 23</sup> but they have not explored the agreement. These limits of agreement are significantly better than that between the first and second, and first and third measurements. This once again reiterates the importance of using the correct method of measurement. However, RR remains a somewhat subjective measure and this level of agreement may still hold significance within clinical practice.

An important finding from this study was that in the 81 children identified as having a RR that was  $\geq 95$ th centile by one or more of the observers in only 33% of these did all three observers agree. Despite an overall higher median reading (32 bpm vs 28 bpm), in 28% (15 children) the HCP (RR1) would not have classified the child as having a raised RR, whereas one or both of the other observers would have done. In children with faster RRs there were poorer levels of agreement seen than in children whose RR was within the normal range. For simultaneous measurements, there was no significant difference observed. This difference may again have reflected the fact that all three measurements were not taken simultaneously, or may have reflected the differences in measurement methods. However, the differences are concerning for clinical practice. Tachypnoea is a key criterion used in assessing the unwell child and is important in low/middle-income countries where guidelines for conditions such as pneumonia rely on tachypnoea in its diagnostic criteria. It is, therefore, clinically important that tachypnoea is recognised and can be accurately identified with a single RR measurement.

The results from our study bring into question our reliance on the accuracy of a RR measurement, as it is currently measured in clinical practice. In the light of recent recommendations suggesting new reference ranges for RR<sup>17</sup> we must remember that these data come from measurements obtained by HCPs in clinical practice performing an observed count. Even if many of these measurements were performed using the WHO-recommended method there is still a degree of variation that may exist. A robust assessment of the impact that this

variation may have on clinical assessment and management of children along with a re-emphasis on recommendations for improvement of its measurement are needed. A review of education tools and measurement techniques, including introduction of objective technological solutions is required.

## CONCLUSION

RR measurements in children vary significantly between different observers. This is likely to have clear consequences in clinical practice and needs further evaluation. Variability in measurements is even greater in children with high RR ( $\geq 95$ th centile), potentially impacting on the recognition and identification of tachypnoea. The variability seen between HCPs in clinical practice and observers under research conditions highlights that the inaccurate methods that are being employed at the frontline of clinical care are affecting the reliability of an important vital sign that is relied on to make critical clinical decisions. For such an important vital sign there clearly needs to be a minimum degree of reproducibility. Paediatric HCPs will benefit from further education on their technique, with a particular emphasis being placed on performing a measurement over 60s. Even researchers using the recommended criteria achieved suboptimal agreement and the introduction of more objective measures including using medical devices to measure RR needs to be considered. These findings highlight the need for a robust review of the clinical impact of inconsistencies in measurements, as well as our current reliance and interpretation of such an important vital sign.

**Contributors** HEE conceived of the study. WJD and RNK initiated the study design and implemented the study. All authors contributed to refinement of the study protocol and approved the final manuscript.

**Funding** This study was funded by The Children's Hospital Charity in April 2016.

**Competing interests** None declared.

**Ethics approval** The National Research Ethics Service Committee Yorkshire and the Humber.

**Provenance and peer review** Not commissioned; externally peer reviewed.

**Open Access** This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>

© Article author(s) (or their employer(s) unless otherwise stated in the text of the article) 2017. All rights reserved. No commercial use is permitted unless otherwise expressly granted.

## REFERENCES

1. NICE. Feverish illness in children—assessment and initial management in children younger than 5 years. 2007.
2. Gandevia SC, McKenzie DK. Respiratory rate: the neglected vital sign. *Med J Aust* 2008;189:532.
3. Subbe CP, Davies RG, Williams E, *et al*. Effect of introducing the modified early warning score on clinical outcomes, cardiovascular arrests and intensive care utilisation in acute medical admissions. *Anaesthesia* 2003;58:797–802.

4. Duncan H, Hutchison J, Parshuram CS. The pediatric early warning system score: a severity of illness score to predict urgent medical need in hospitalized children. *J Crit Care* 2006;21:271–8.
5. Bradman K, Maconochie I. Can paediatric early warning score be used as a triage tool in paediatric accident and emergency? *Eur J Emerg Med* 2008;15:359–60.
6. Parshuram CS, Hutchison J, Middaugh K. Development and initial validation of the bedside paediatric early warning system score. *Crit Care* 2009;13:R135.
7. Akre M, Finkelstein M, Erickson M, et al. Sensitivity of the pediatric early warning score to identify patient deterioration. *Pediatrics* 2010;125:e763–e769.
8. Singhi S, Bhalla AK, Bhandari A, et al. Counting respiratory rate in infants under 2 months: comparison between observation and auscultation. *Ann Trop Paediatr* 2003;23:135–8.
9. WHO. *Integrated management of childhood illness: caring for newborns and children in the community. Manual for the community health worker*. Geneva: World Health Organization, 2002.
10. Simoes EA, Roark R, Berman S, et al. Respiratory rate: measurement of variability over time and accuracy at different counting periods. *Arch Dis Child* 1991;66:1199–203.
11. Daw W, et al. Medical devices for measuring respiratory rate in children: a review. *J of Adv in Bio Eng and Tech* 2016;3:21–7.
12. Daw W, Elphick H. 2015. Pilot Study: interobserver variation in respiratory rate measurements at sheffield children's hospital emergency department.
13. Lim WS, Carty SM, Macfarlane JT, et al. Respiratory rate measurement in adults—how reliable is it? *Respir Med* 2002;96:31–3.
14. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986;1:307–10.
15. Giavarina D. Understanding bland altman analysis. *Biochem Med* 2015;25:141–51.
16. Group, A.L.S. *Advanced Paediatric Life Support: the practical approach*. 6th ed: WileyBlackwell, 2004.
17. O'Leary F, Hayen A, Lockie F, et al. Defining normal ranges and centiles for heart and respiratory rates in infants and children: a cross-sectional study of patients attending an Australian tertiary hospital paediatric emergency department. *Arch Dis Child* 2015;100:733–7.
18. Wang EE, Milner RA, Navas L, et al. Observer agreement for respiratory signs and oximetry in infants hospitalized with lower respiratory infections. *Am Rev Respir Dis* 1992;145:106–9.
19. Wang EE, Law BJ, Stephens D, et al. Study of interobserver reliability in clinical assessment of RSV lower respiratory illness: a Pediatric Investigators Collaborative Network for Infections in Canada (PICNIC) study. *Pediatr Pulmonol* 1996;22:23–7.
20. Chan A, Langley J, Leblanc J. Interobserver variability of croup scoring in clinical practice. *Paediatr Child Health* 2001;6:347–51.
21. Liu LL, Gallaher MM, Davis RL, et al. Use of a respiratory clinical score among different providers. *Pediatr Pulmonol* 2004;37:243–8.
22. Gajdos V, Beydon N, Bommenel L, et al. Inter-observer agreement between physicians, nurses, and respiratory therapists for respiratory clinical evaluation in bronchiolitis. *Pediatr Pulmonol* 2009;44:754–62.
23. Lanaspá M, Valim C, Acacio S, et al. High reliability in respiratory rate assessment in children with respiratory symptomatology in a rural area in Mozambique. *J Trop Pediatr* 2014;60:93–8.
24. Berman S, Simoes EA, Lanata C. Respiratory rate and pneumonia in infancy. *Arch Dis Child* 1991;66:81–4.