High Reliability in Respiratory Rate Assessment in Children with Respiratory Symptomatology in a Rural Area in Mozambique

by Miguel Lanaspa,[1,2] Clarissa Valim,[3,4] Sozinho Acacio,[2] Katherine Almendinger,[3] Rushdy Ahmad,[3] Roger Wiegand,[3] *and* Quique Bassat[1]

[1]*Barcelona Centre for International Health Research, Hospital Clinic, University of Barcelona, Barcelona, Spain*

[2]*Centro de Investigação em Saúde de Manhiça, Mozambique*

[3]*Division of Infectious Diseases, Broad Institute, Cambridge, Massachusetts, USA*

[4]*Department of Immunology and Infectious Diseases, Harvard School of Public Health, Boston, Massachusetts, USA*

Correspondence: Miguel Lanaspa. Rosello 132, 08036 Barcelona, Spain. E-mail: <miguel.lanaspa@cresib.cat>.

Summary

Early recognition of severe medical conditions is often based on clinical scores and vital sign measurements such as the respiratory rate (RR) count. We designed this study to determine the reliability of RR assessment counted three times during a full minute by independent observers in children in a developing country setting. A total of 55 participants were enrolled in the study. Participant ages ranged from 10 days to 7 years (median 22 months). Agreement for RR count was high (intraclass correlation coefficient of 0.95; 95% confidence interval: 0.93–0.97). Agreement for presence of tachypnea was also high (Kappa coefficient of 0.83, $p < 0.001$). However, a single reading would have misclassified 5–11% of the participants as non-tachypneic. Repeated RR counts offer reliable results if done during a full minute. Patients not fulfilling tachypnea criterion but with a high RR count should have the measurement repeated.

Key words: respiratory rate, reliability of results, tachypnea, pneumonia, child.

Background

Clinical scores and vital sign measurements are important in clinical practice to correctly assess and classify patients, especially for early recognition of severe medical conditions. Adequate clinical assessment is also part of clinical algorithms, which require the identification of simple and objective criteria, justifying the initiation of particular interventions. The measurement of specific variables and signs requires reliability (or reproducibility) and validity. Several childhood disease management guidelines, including those proposed by the World Health Organization (WHO), such as the Integrated Management of Childhood Illness manual, incorporate the respiratory rate (RR) as an important criterion for the assessment of sick children. Indeed, the assessment of RR is part of the clinical pneumonia definition, and it is also important as part of the initial assessment of severity at triage, the early recognition of potentially life-threatening diseases, as an alarm sign during transfusions and other intravenous infusions and, finally, as part of monitoring response to treatment [1–4].

Lack of access to diagnostic technology often makes clinical evaluation the only diagnostic tool available in developing settings. Tachypnea, or age-specific elevation in RR, is associated with hypoxemia. Its accurate assessment could improve the outcome of tachypneic children by prompt administration of oxygen in situations where standard technology for hypoxemia detection is unavailable [5, 6]. However, tachypnea assessment displays better predictive capacity in combination with other signs and symptoms than its isolated detection [7]. For pneumonia and other severe disease in young children, tachypnea has been shown to be an independent predictor [8–10].

Among the classic vital signs, RR is one of the few that relies only on clinical observation and does not

Acknowledgements

The authors would like to thank the participants and their legal guardians for their participation in this study.

Funding

This study was conducted as part of a larger study funded by the Bill and Melinda Gates Foundation (OPP500092) to discover biomarkers to differentiate malaria from bacterial and viral pneumonia.

doi:10.1093/tropej/fmt081 Advance Access published on 25 September 2013

commonly use an electronic confirmation. The WHO standard for RR measurement includes a count over a full minute and for this purpose, electronic chronometers have been distributed in many developing world settings [11]. Proper training is essential to ensure reliability of the measurement [12]. This count may be done by observing abdominal or chest movements or by auscultation with similar results [13]. Electronic devices to measure RR exist and have shown a variable accuracy compared with observer RR measurement [14, 15], but such devices have seldom been introduced in the developing world. Furthermore, studies assessing the interobserver agreement report a wide range of reliability for RR both in children and adults [16–19]. In those studies, the interobserver agreement expressed by the weighted kappa statistic ranged from 0.15 (slight agreement) to 0.79 (substantial agreement) if the assessments were done by two trained observers no more than 10 min apart.

Accuracy and reliability of vital sign measurements are also important in clinical research. Vital signs are often used as inclusion or exclusion criteria, case definitions or outcome/endpoint criteria. Thus, vital signs used as criteria for clinical scores or for research purposes should be easily measured, consistent throughout time and show little or no interobserver variability. Furthermore, when clinical signs are used in the eligibility criteria or are part of the outcome, an assessment of interobserver agreement should be undertaken including all study observers. The purpose of this study was to determine the reliability of RR count when assessed by different observers in the context of a multicentric study where tachypnea was an inclusion criterion.

Methods

Study design

This observational study was designed to determine the reliability of RR measurements recorded by three independent observers. Patients assessed were part of a larger study investigating the etiology of respiratory distress. These measurements were done in the context of pre-recruitment screening of an Institutional Review Board-approved multicenter trial (RAPID project, Health Bioethics National Committee, Mozambique, ref. 262/CNBS/10).

Study site

The study was conducted at the Manhiça District Hospital (MDH) by Centro de Investigação em Saúde de Manhiça (CISM), located in Manhiça, Maputo Province, in Southern Mozambique. CISM runs two surveillance system platforms: (i) a demographic surveillance system in its study area; and (ii) a morbidity surveillance system based on five health centers for outpatient visits and one referral hospital, the MDH. Information on all pediatric outpatient visits and admissions to hospital, including the RR count, is collected through standardized questionnaires. MDH has >70 000 pediatric outpatient visits and >3500 pediatric admissions per year. Pediatric medical assistance is based on a two-stage model—on arrival to hospital, outpatients are initially assessed by a clinician (normally a medical agent, i.e. a clinician trained within the Mozambican health system for a minimum of 1 year) and discharged with a diagnosis and a treatment plan. Should any of those patients show signs of severe disease or require specialized assessment (i.e. by a medical 'technician' or a medical doctor), they are transferred to the short-term–hospitalization facility where more specialized pediatric care can be provided before discharge or hospital admission.

Selection of participants

All consecutive children aged <10 years, who attended the emergency room between 7 AM and 5 PM during the first week of January 2011 with referred respiratory symptoms (based on guardian's report), were included in this study.

Methods of measurement

RR was measured independently three times within 30 min following the WHO standard of 60 s observation by three different health personnel: one medical agent and two study health assistants. Time was controlled using UNICEF-donated chronometers or a mobile phone timer set to count backward from 60 to 0 s. The medical agent was not aware that further RR counts would be done. All RR measurements were made in a blinded manner so that none of the observers would know previous measurements. Whenever possible, the RR was measured while the child was calm. Information about the behavior of the child (calm, feeding, crying) was recorded for the second and third measurements.

Data collection

Participants were given an identification number based on the recruitment order. Data were collected in a de-identified manner directly on a study log and then tabulated into an Excel spreadsheet.

Definitions

Tachypnea was defined according to the age-related WHO-proposed definitions as RR >60 breaths per minute (bpm) if the child is younger than 2 months of age; RR >50 bpm if the child age is between 2 months (included) and 12 months (excluded); RR >40 bpm if the child age is between 1 year (included) and 5 years (excluded); and RR >30 bpm if the child age is between 5 years (included) and 10 years (excluded). These definitions are consistent with a recently published systematic review of normal, age-specific, ranges for respiratory and heart rates in pediatric populations [20]. Most studies were carried

out in developed countries; however, the reference ranges of heart and RRs estimated from a study in the UK may also be applied to children in developing world settings [21].

Statistical analysis

Demographic characteristics and RR counts were summarized by medians, interquartile ranges and proportions. We compared the distribution of RRs between subjects crying and not crying during the observation through Wilcoxon rank sum tests. Pairwise agreement between RR counts was assessed through Lin's concordance correlation and intraclass correlation coefficients with 95% confidence intervals (95% CIs). We also estimated association between readings through Pearson correlation coefficients and linear regression of $\log_e$ transformed RRs. All analyses were conducted in R 2.15.3 packages icc and epiR [22, 23].

Results

Characteristics of study subjects

A total of 55 participants were enrolled in the study. Participant ages ranged from 10 days to 7 years, with a median age of 22 months. Among them, 40 (72.7%) had fever or history of fever, 38 (69.1%) had referred cough and 15 (27.3%) had referred breathing difficulty. Thirty-two (58.2%) participants were male, and most presented with malaria or lower and upper respiratory tract infections (data not shown).

Respiratory rate

RR was recorded three times (RR1, RR2 and RR3) for each patient and ranged from 20 to 89 (medians of ~40–42). The spread of RR at each time was similar (Fig. 1). Agreement across the three readings was high (Intraclass Correlation Coefficient of 0.95; 95% CI: 0.93–0.97). A total of 29 (53%) subjects fulfilled tachypnea criterion in RR1, 26 (47%) in RR2 and 22 (40%) in RR3. Agreement for presence of tachypnea recorded in the three readings was also high with Kappa coefficient of 0.83 ($p < 0.001$). However, having only the RR3 reading would have misclassified six (11%) subjects as non-tachypneic, and having only RR2 would have misclassified three (5%) subjects as non-tachypneic. If only the RR2 and RR3 measurements were taken, three subjects (5%) would have been classified as non-tachypneic.

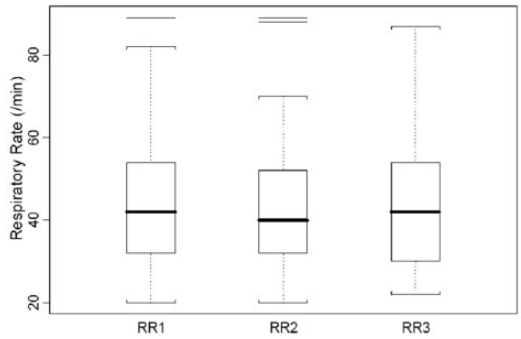

FIG. 1. Variability of the first (RR1), second (RR2), and third (RR3) RR readings in 55 children. The solid line in the middle of the box represents the median. The boxes span the interquartile range and whiskers extend to ±1.5 the interquartile range.

Pairwise agreement between measurements taken closer in time (first vs. second and second vs. third) was slightly larger than more distant measurements (first vs. third), but differences were not statistically significant. The concordance correlation between first and second was 0.95 (95% CI = 0.92, 0.97), between second and third was 0.98 (95% CI = 0.97, 0.99) and between first and third was 0.93 (95% CI = 0.89, 0.96). Inspection of Bland Altman plots (Fig. 2) suggests absence of systematic bias. Pairwise associations are presented in Fig. 3.

Despite the high agreement and correlations across readings, RR was highly variable for some subjects across readings. The 'between' readings standard deviation had a median of ~2 (Interquartile range (IQR) = 1.2–3.2). The RR measurements within a subject differed from 1 to 18 units (minimum = 1; IQR = 2–6; median = 4). This did not seem to be associated with either age (Spearman correlation coefficient of 0.13 and *p*-value of slope of regression of 0.72) or the magnitude of the first RR reading (Spearman correlation coefficient of 0.19 and *p*-value of slope of regression of 0.24). Differences of RR measurements within a subject were slightly higher when the child was crying during one of the measurements (median largest difference of 5.6; IQR = 3–8) than when not crying (median largest difference 5, IQR 3–8), but these differences were not statistically significant (*p*-value Wilcoxon test = 0.25). The largest differences between two measurements within a subject were not recorded for children that were crying during one of the measurements.

Discussion

Clinical assessments, often prone to the observer's subjectivity, require a minimum reproducibility to be useful. Variability can be expected from subjective assessments, which involve the assessment of complex respiratory patterns or signs, but objective data such as vital signs are also subject to variability. As classic examples, the interrater agreement in assessing systolic blood pressure or tachypnea is low [17]. When assessing RR, variability is often high owing to both interobserver variability and the lack of reliable electronic devices for its measurement

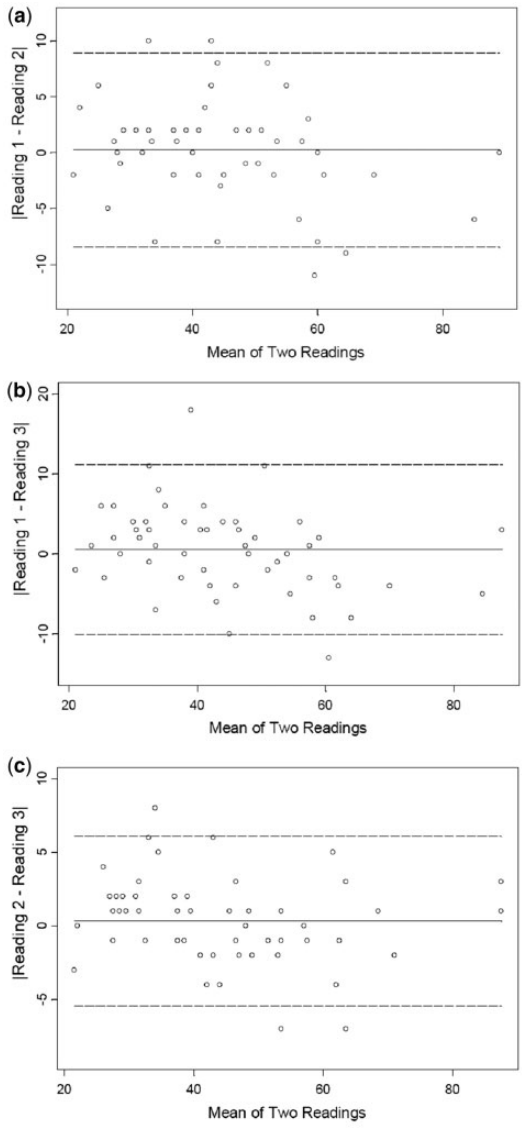

FIG. 2. Bland-Altman plots assessing pairwise agreement between readings—(a) reading 1 and 2; (b) reading 1 and 3; (c) reading 2 and 3. The y axis represents the bias (absolute value) and the x axis the mean values of the two measurements of RR. The solid line represents the mean bias across all subjects and the dashed lines represent the 95% CI, as based on the standard deviation of the distribution of biases.

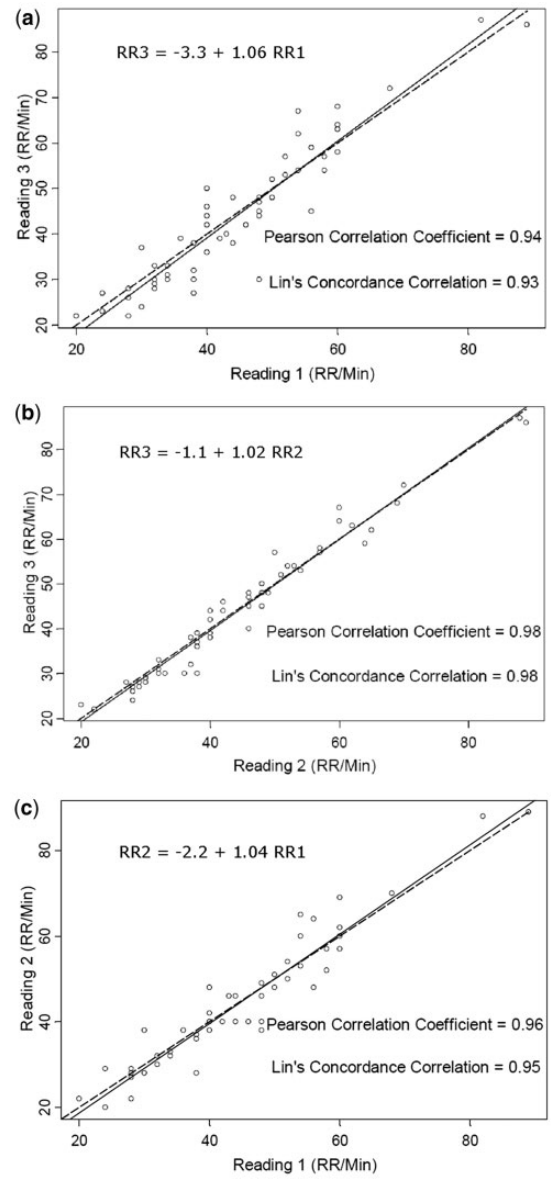

FIG. 3. Scatterplot of the pairwise correlations between two RR readings—(a) first and third readings (RR1 and RR3); (b) second and third readings (RR2 and RR3); and (c) first and second readings (RR1 and RR2). Dashed line represents the line of perfect concordance (intercept of 0 and slope of 1). The solid line represents the observed association between the two measurements obtained in a linear regression (coefficients as shown in equation). Pearson correlation coefficient and Lin's concordance coefficient are shown.

in routine clinical practice. Such uncertainty becomes important because RR is a critical sign for early recognition of life-threatening conditions. Electronic devices usually extrapolate RR over a minute from the timing of the last few breathing cycles, which

may not be accurate enough in cases of irregular breathing patterns or among young children. Errors in the assessment of RR could mislead clinicians in recognizing severe disease and result in over- or undertreatment of patients.

Studies assessing the reliability of RR count present highly variable results [15, 16, 19]. Similar to other studies, the RR range in the present study is high. However, an important distinction is that while most other studies display low to moderate interrater agreement, the present study expresses high agreement, as indicated by the high Pearson correlation coefficient between the first and the second reading, the second and the third and even between the first and the third. Despite such high interrater agreement, between 5 and 11% of the study subjects still would have been misclassified as non-tachypneic if only 1 or even 2 measurements had been done. Because tachypnea is a key criterion for severity assessment, a single RR measurement appears to lack the required sensitivity for such an important indicator. The decrease in agreement over time has been described in a population of older children and adults where the k value for tachypnea assessment was 0.79 if the time interval between assessments was <10 min, and 0.48 if it was >10 min [17]. In the present study, there was no significant decrease in agreement when comparing readings closer in time (RR1 vs. RR2) with readings done over a longer interval (RR1 vs. RR3).

The child's behavior did not seem to affect the RR count. Indeed, the mean RR was 43.8 for calm children, 44 if crying and 41.1 if feeding. The differences were not statistically significant. This should reassure clinicians worried about the validity of the RR counted in situations where it may be usual to face crying children. In our study, 83 (75.5%) of the assessments were done while the child was calm, 15 (13.6%) while the child was crying and 12 (10.9%) while feeding. These figures do not vary when specifically assessing children <24 months of age. In the younger group, however, mean RR seemed to be slightly higher if feeding but the differences were not significant. A full report of all the RR measurements is available in Table 1.

Compared with other studies, the interobserver agreement was high between the first assessment and the other two. It is important to mention that the first observer was not aware that the RR would be counted again and compared with the first count, which could have affected his accuracy in counting the RR, and neither the second nor third observers were aware of the initial RR count. What seems to explain this high agreement is the measurement method. In MDH, the standard practice is to count the RR for 1 min, and the three readings done for the present study were done following the same method. This result is consistent with other published studies where the agreement was slight or fair if the RR

TABLE 1
Full report of RR measurements (RR1, RR2 and RR2) and the behavior of the child during measurement (RR2 and RR3)

Age (months)	RR1	RR2	Behavior RR2	RR3	Behavior RR3
23	34	32	Calm	31	Calm
24	30	38	Calm	37	Calm
1	48	49	Calm	48	Calm
27	38	37	Calm	38	Calm
29	44	40	Feeding	38	Calm
1	54	53	Calm	54	Calm
1	48	46	Calm	47	Calm
2	56	48	Calm	45	Calm
25	28	29	Calm	28	Calm
24	40	42	Calm	44	Calm
4	43	46	Calm	40	Calm
5	32	32	Crying	33	Calm
26	46	40	Calm	42	Calm
6	58	57	Calm	57	Calm
24	44	46	Calm	48	Calm
9	52	54	Calm	53	Calm
10	48	46	Feeding	45	Feeding
22	60	69	Calm	68	Calm
28	40	48	Calm	50	Crying
10	68	70	Feeding	72	Calm
11	34	32	Calm	33	Calm
11	40	40	Calm	42	Calm
36	32	32	Calm	30	Calm
49	28	22	Feeding	22	Calm
11	58	52	Calm	54	Feeding
11	60	62	Crying	63	Calm
15	40	38	Calm	36	Crying
13	52	50	Crying	57	Calm
50	38	28	Calm	27	Calm
52	28	27	Calm	28	Calm
13	38	37	Calm	32	Calm
48	32	30	Calm	29	Calm
15	42	40	Calm	39	Calm
16	38	36	Crying	30	Calm
47	40	42	Calm	46	Calm
16	48	38	Calm	30	Calm
46	50	51	Calm	52	Calm
17	30	28	Crying	24	Feeding
17	60	57	Calm	58	Calm
19	82	88	Calm	87	Feeding
20	56	64	Crying	59	Calm
20	89	89	Calm	86	Calm
22	36	38	Crying	39	Crying
22	50	48	Calm	48	Feeding
65	24	20	Calm	23	Calm
59	24	29	Feeding	27	Feeding
0	60	60	Calm	64	Crying
64	54	60	Calm	67	Calm
67	54	65	Crying	62	Crying
73	20	22	Calm	22	Calm
0	48	40	Calm	44	Feeding
75	32	30	Calm	28	Calm
72	32	30	Calm	28	Calm
87	28	28	Crying	26	Crying
85	34	33	Calm	30	Calm

assessment was done over 15 or 30 s and then multiplied by 2 or 4, respectively; but it was moderate to almost perfect if the assessments were done over a full minute [11].

Conclusion

In settings where routine examination of RR includes a 1 min evaluation, repeated respiratory counts within a time lapse of no more than 30 min offers reliable results with little interobserver variability, regardless of whether the child is calm, crying or feeding. RR counts not fulfilling tachypnea criterion but falling near the threshold should be repeated if the clinical management of the patient depends on this single observation. For clinical studies or trials in which RR is either an inclusion criterion or part of an assessed outcome, it is advisable to include a fully chronometered 1 min-long count, or preferably, calculate the average of multiple counts.

References

1. American College of Chest Physicians/Society of Critical Care Medicine Consensus Conference: definitions for sepsis and organ failure and guidelines for the use of innovative therapies in sepsis. Crit Care Med 1992;20:864–74.
2. Cooper RJ, Schriger DL, Flaherty HL, *et al*. Effect of vital signs on triage decisions. Ann Emerg Med 2002;39: 223–32.
3. 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.
4. WHO. Emergency triage assessment and treatment (ETAT). Geneva: World Health Organization, 2005.
5. Opiyo N, English M. What clinical signs best identify severe illness in young infants aged 0-59 days in developing countries? A systematic review. Arch Dis Child 2011;96:1052–9.
6. Usen S, Weber M, Mulholland K, *et al*. Clinical predictors of hypoxaemia in Gambian children with acute lower respiratory tract infection: prospective cohort study. BMJ 1999;318:86–91.
7. Usen S, Webert M. Clinical signs of hypoxaemia in children with acute lower respiratory infection: indicators of oxygen therapy. Int J Tuberc Lung Dis 2001;5: 505–10.
8. Group YICSS. Clinical signs that predict severe illness in children under age 2 months: a multicentre study. Lancet 2008;371:135–42.
9. Mulholland EK, Simoes EA, Costales MO, *et al*. Standardized diagnosis of pneumonia in developing countries. Pediatr Infect Dis J 1992;11:77–81.
10. Singhi S, Dhawan A, Kataria S, Walia BN. Validity of clinical signs for the identification of pneumonia in children. Ann Trop Paediatr 1994;14:53–8.
11. 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.
12. Gadomski AM, Khallaf N, el Ansary S, Black RE. Assessment of respiratory rate and chest indrawing in children with ARI by primary care physicians in Egypt. Bull World Health Organ 1993;71:523–7.
13. Singhi S, Bhalla AK, Bhandari A, Narang A. Counting respiratory rate in infants under 2 months: comparison between observation and auscultation. Ann Trop Paediatr 2003;23:135–8.
14. Bianchi W, Dugas AF, Hsieh YH, *et al*. Revitalizing a vital sign: improving detection of tachypnea at primary triage. Ann Emerg Med 2013;61:37–43.
15. Lovett PB, Buchwald JM, Sturmann K, Bijur P. The vexatious vital: neither clinical measurements by nurses nor an electronic monitor provides accurate measurements of respiratory rate in triage. Ann Emerg Med 2005;45:68–76.
16. Chan A, Langley J, Leblanc J. Interobserver variability of croup scoring in clinical practice. Paediatr Child Health 2001;6:347–51.
17. Edmonds ZV, Mower WR, Lovato LM, Lomeli R. The reliability of vital sign measurements. Ann Emerg Med 2002;39:233–7.
18. Liu LL, Gallaher MM, Davis RL, *et al*. Use of a respiratory clinical score among different providers. Pediatr Pulmonol 2001;37:243–8.
19. Wang EE, Milner RA, Navas L, Maj H. Observer agreement for respiratory signs and oximetry in infants hospitalized with lower respiratory infections. Am Rev Respir Dis 1992;145:106–9.
20. Fleming S, Thompson M, Stevens R, *et al*. Normal ranges of heart rate and respiratory rate in children from birth to 18 years of age: a systematic review of observational studies. Lancet 2011;377:1011–8.
21. Wallis LA, Maconochie I. Age related reference ranges of respiratory rate and heart rate for children in South Africa. Arch Dis Child 2006;91:330–3.
22. Gamer M, Lemon J, Fellows I, Singh P. irr: Various Coefficients of Interrater Reliability and Agreement 2012, R package version 0.84. http://CRAN.R-project.org/package=irr.
23. Stevenson M, Nunes T, Sanchez J, *et al*. epiR: An R package for the analysis of epidemiological data 2013, R package version 0.9-48. http://CRAN.R-project.org/package=epiR.