LB-5488

Diagnostic challenges in childhood pneumonia: lessons learned in selecting a reference standard for validating new respiratory rate counting aids

Kevin Baker[1,2], Charlotte Ward [1], Sarah Marks[1], Dawit Getachew[3], Tedila Habte[3], Cindy McWhorter[4], Paul LaBarre[4], Max Petzold[5], Karin Källander[2,6]

[1]Malaria Consortium, London, UK; [2] Karolinska Institutet, Stockholm, Sweden; [3]Malaria Consortium, Ethiopia; [4]UNICEF Supply Division, Copenhagen, Denmark; [5]University of Gothenburg, Sweden; [6]UNICEF Programme Division, New York, U.S.A.

Key messages

- There is evidenced variability in humans counting respiratory rate (RR) in children under five, even with training and documented standardisation, in real time or with a video.
- Reference standard and test device measurements should be simultaneous and use the same measurement methodology.
- Agreement measures should reflect the true performance of the test device, including under or over diagnosing compared to the reference standard.

Introduction

In the absence of aetiology-based tools for diagnosing pneumonia, it is essential that community health workers (CHWs) correctly ascertain a child's RR and classify fast breathing according to WHO guidelines.

Different tools for counting RR have recently been developed. One important step toward introduction of new RR counting devices is to understand their accuracy. In the absence of gold standard technology for counting RR, it has been proposed that the agreement between the test device and a reference standard be evaluated. However, little data currently exist to guide selecting the most appropriate reference standard and measures of agreement.

Methods

The Pneumonia Diagnostics Project (PDP) tested four manual RR counters with CHWs across four countries in sub-Saharan Africa and Southeast Asia from February-June 2015. Reference standard was a continuous respiratory patient monitor with Phasein ISA CO2 capnography counting RR.

Another project, the Acute Respiratory Infection Diagnostic Aid (ARIDA) project, tested ChARM, an automated RR counter against a reference standard of two to four pediatricians (video expert panel (VEP)) who counted the child's breaths in 60 seconds from a video recording. As a secondary outcome, RR from an expert counter (EC) in real time was assessed. The study took place at St Paul's hospital in Addis Ababa, Ethiopia between April and May 2017.

Results

Table 1 shows the agreement between all test devices and the references, both automated and manual, from the PDP.

- Root mean squared difference (RMSD) ranges from 8.7 to 15.8 bpm (breaths per minute)
- Mean difference (bias) ranged from -0.5 to 5.5 bpm
- Kappa values range from 0.41 to 0.49 -(weak)

Table 2 shows the interrater agreement between two clinicians counting RR with and without video assistance (VEP and EC respectively) in the ARIDA study.

- RMSD was lower for two VEPs by 2.4 bpm (6.6 vs. 4.2 bpm)
- RR classification was similar for both groups (VEP vs VEP and EC vs EC, strong) but only moderate for EC vs VEP (0.69)

Table 1: Agreement results for PDP study by device and agreement measures

Name of agreement measure	Number of observations	Mean difference or 'bias' (bpm), 95% CI	Root mean square difference	Kappa value (Standard error) (Interpretation)
Agreement result: MK2 ARI vs continuous monitor	322	-0.6; 95% CI 3.8 to -0.2	12.2	0.49 (0.05) (weak)
Agreement result: RR vs continuous monitor	304	5.5; 95% CI 3.2 to 7.8	15.8	0.44 (0.06) (weak)
Agreement result: Respirometer vs continuous monitor	626	-0.5; 95% CI -2.1 to 1.2	14.7	0.41 (0.04) (weak)
Agreement result: Beads vs continuous monitor	172	-1.9; 95% CI -3.8 to -0.2	8.7	0.41 (0.07) (weak)

Table 2: Interrater agreement between human counters in ChARM study

	Root mean square difference	Positive percent agreement (%) (95%CI)	Negative percent agreement (%) (95% CI)	Kappa (interpretation)
VEP 1 vs VEP 2 (n=105)	4.2	92.9 (82.7, 98)	91.8 (80.4, 97.7)	0.85 (strong)
EC vs. EC (n=37)	6.6	82.4 (56.6, 96.2)	100 (83.2, 100)	0.83 (strong)
EC vs. VEP (n=98)	5.3	92.6 (82.1, 97.9)	75 (59.7, 86.8)	0.69 (moderate)

ARIDA agreement study pre-test – a child is being assessed by the ChARM device and by an expert clinician using the MK2 ARI timer. The assessment is being recorded on video for VEP review.

Conclusions

1. Given the lack of agreement between VEP members and ECs conducting manual counting in the ChARM study, neither of the two measurements can be considered gold standard for RR counting, and will therefore not be suitable when compared against automated respiratory rate counters.
2. While the findings from the ChARM study indicate that agreement between humans is better if they have videos to look at, the findings from this study cannot support the measurement of ChARM agreement, given that ChARM measures a different breath sequence than the manual human counters, and because of the large difference observed in the assessment of human expert counters.
3. The PDP study shows low levels of agreement between the test devices and the reference standards in terms of RR counting and classification.
   - The manual RR counters tested provide a low level of support to CHWs
   - The continuous monitor was not validated in U5 children

Further studies are required to continue the development of appropriate reference methods for new respiratory rate counting aids.

For more information
Kevin Baker, k.baker@malariaconsortium.org
Malaria Consortium
The Green House, Cambridge Heath Road, London, E9 2DA
www.malariaconsortium.org

Acknowledgements
The authors would like to thank "la Caixa" Foundation for their financial support for the ARIDA field trials and to express gratitude to the research teams in Ethiopia. Further, we would like to thank the SNNPR Regional Health bureau and the Federal Ministry of Health in Ethiopia for supporting the study.