Reviewer's report

Title: Determining an anthropometric surrogate measure for identifying low birth weight babies in Uganda: a hospital based cross sectional study

Version: 1 Date: 28 December 2012

Reviewer: Alan Horn

Reviewer's report:

Method
Section title: Study design, data collection and analysis

a) Sample size calculations: The sample size was calculated using a LBW prevalence of 12% - this was the exact prevalence in the actual study. A LBW prevalence of 16% in a previous Ugandan study was cited in the introduction – why was this value not used in sample size and post-test probability calculations?

b) Repeat measurements: Repeat measurements were done for every 1 in 30 study participants to check reliability:
   i) How many repeat measurements were done for each test and by whom?
   ii) Was only one measurement by one investigator taken on those infants who were not one of those 1 in 30 having repeat measurements? Please clarify.
   iii) How was “reliability” confirmed – was a kappa statistic obtained? If so, this result should be presented in the results section.
   iv) How was the appropriateness of repeating 1 in 30 tests derived?

c) Differences between means: How were differences between means calculated and what significance levels were used? (for normal and non-normally distributed data ?)

d) Pearson co-efficients: What value of Pearson co-efficients were accepted as strong correlation? What test would have been used if the data were not normally distributed? The correlation coefficient of determination r2 is also relevant here.

e) ROC and AUC analysis: What statistical tests and significance levels were used to determine whether the differences between the AUC were significant - was the study appropriately powered to detect these differences?

Results

6. Fig 1: The box showing the 42 excluded babies should be amended to include the 5 with missing data and state in the methods that infants with missing (presumably descriptive demographic) data were also excluded.

7. Descriptive characteristics: By convention, mean (SD) conveys descriptive data
if normally distributed and median (range / IQR) conveys non-normally distributed data – the authors have given both values and have included 95%CI instead of range/IQR. This should be amended.

8. Table 1: Table 1 should reflect p values using a superscript and footnote. The text should reflect the use of student’s t-test in the methods section and also a statement that this test was (presumably) used for normally distributed data.

9. Table 2: The text compares r and AUC as listed in table 2 using terms such as “strongest”, “weakest” and “best” – before using such terms, the authors need to state if and how the differences between the parameters were shown to be statistically significant – if not then these statements must be qualified to that effect.

10. Table 3: The authors draw conclusions about differences between summary measures in Table 3 without referencing the method used to establish statistical significance.

11. Table 3: The authors discuss interpretation of the likelihood ratios – this discussion should rather be in the methods and/or discussion.

12. Table 3: The authors reference Table 3 as presenting data that shows how a foot length < 7.9 cm increases the probability of LBW from 12% to 66% - but the table does not present sufficient data to easily calculate post test probability. The table should provide additional columns showing the number of infants above and below each cut off value and the numbers of infants in each group who were LBW. (ie include a 2x2 table for each test/cut off). This would then provide the reader with sufficient data to more easily calculate PPV/post test probability.

13. Table 3: The authors provide LR+, but the diagnostic odds ratio (LR+/LR-) is a more appropriate single measure of the effectiveness of a screening test – this should be included in table 3 and referred to in the discussion.

Discussion

14. Correlation: The authors describe the correlation \( r=0.76 \) as “strong” and lesser correlations are described as “fairly strong”. \( r=0.76 \) is usually described as “moderate” at best, further demonstrated by the associated \( r^2 = 0.58 \) – hence only 58% of the variation in \( y \) can be explained by the linear relationship between \( x \) and \( y \). The interpretation should be reviewed.

15. Last paragraph: The authors state that they have minimized the number of false negatives rather than the number of false positives – but the higher specificity compared to sensitivity with a cut off of 7.9cm does NOT support their claim. The authors have chosen cut-off values to maximize the correct identification of infants. However, the sensitivity associated with this cutoff is only 82.6% with a higher specificity of 94.1%. It is more important to have a high sensitivity than a high specificity in a screening test, particularly when the test is designed to identify infants who will benefit from interventions that are basic and/or low cost, but potentially life-saving. In this setting, a cut off that allows a
higher sensitivity than specificity is appropriate. Suggestion: Use the cut off associated with the highest “correctly classified” proportion where the sensitivity is also greater than the specificity.

Limitations

16. The potential for bias by having a single investigator performing both index and gold standard tests should be stated, the number of measurements taken for each test (see earlier comment) and relative agreement between each assessor should also be discussed. Further potential for bias exists when a reference measurement is derived rather than tested prospectively, this should also be mentioned.

Conclusion

17. The conclusion should specifically state that further research should prospectively evaluate the specified cut-off of “x” cm with trained community health workers.

18. The authors claim that the study showed that foot length was easier to measure – but they did not scientifically evaluate ease of measurement or inter-individual variation with the different measurement types. Hence their finding can only be that the health workers reported that that foot length was the easiest measurement.

Minor Essential Revisions

Abstract

19. The abstract should state that a simpler measurement than weight is required because of lack of scales.

20. The statement “with a mean of…..and 85(12%) babies weighing less than..”, does not read well.

Introduction

21. P1, Line 4: “resulted into” – review the grammar
22. P6: Line 7: “context-specific” should be hyphenated
23. P7: Line 1: “hospital-based” should be hyphenated (and future occurrences)
24. The reason for requiring a measurement other than a weight is not made clear: The fourth paragraph in the discussion on p 10, beginning, “Uganda has launched..”, would address this if included in the introduction.

Method

25. P7 line 17: “using a hard transparent…” review the grammar
26. P7: The description of the 3-day training process occurs twice on this page – a single mention is sufficient.
Discussion

27. P9 line 19: “Previous studies…….”. This sentence as well as the one that follows should be appropriately referenced.

Discretionary revisions

28. The known association between foot length and gestational age could be discussed with reference to the fact that foot length may miss “wasted” babies. If available, data on gestational age and ponderal index may be useful to determine the reasons for false positives and false negatives. If this data is not available, discussion around this aspect could be included in the limitations section.

Level of interest: An article whose findings are important to those with closely related research interests

Quality of written English: Needs some language corrections before being published

Statistical review: No, the manuscript does not need to be seen by a statistician.

Declaration of competing interests:

'I declare that I have no competing interests'