Author’s response to reviews

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

Authors:

Elizabeth L Nabiwemba (enabiwem@musph.ac.ug)
Christopher G Orach (cgorach@musph.ac.ug)
Patrick Kolsteren (pkolsteren@itg.be)

Version: 2 Date: 19 February 2013

Author’s response to reviews: see over
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?

*The prevalence of 12% used for sample size calculation was obtained from data at the hospital where the study was carried out. This was data for the period of one year preceding the study.*

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?

   *Two repeat measurements were done for each test, and these were done by the clinical supervisor of the study who is a medical doctor.*

   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.

   *Two measurements were done and the average was taken. One investigator did both measurements because it was difficult to have more than one midwife on duty especially at night.*

   iii) How was “reliability” confirmed – was a kappa statistic obtained? If so, this result should be presented in the results section.

   *Inter-observer percent agreement was calculated. Results have been added to revised manuscript.*

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

   *Student’s t-test was used, and the level of significance was p<0.05*

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.

   *The value of Pearson co-efficient between 0.71 and 0.89 was accepted as strong correlation. Spearman’s rank correlation co-efficient would have been used if the data was not normally distributed. The correlation co-efficient of determination has been added to table2 and has been discussed.*
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.

The box has been amended and a note made in the methods on the 5 babies with missing data.

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.

We have decided to report only the mean since the data was normally distributed.

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.

Both table and text amended.

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.

We have adjusted the terms used and given reference to the categorization used for the strength of r. Statistical significance of the difference of r and AUC between parameters was not tested.

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

The interpretation of likelihood ratios has been included in the discussion section.

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.
We have added another table (table 4) to incorporate the suggested additions. The current table 3 shows the number of babies above and below each cut off value and the number of babies in each group who were LBW. The sensitivity and specificity at each cut off value is also shown in this table. Table 4 shows the positive and negative likelihood ratios, and a column for the diagnostic odds ratio has been added.

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.

We have reviewed the terms used and the interpretation of the coefficient of determination ($r^2$).

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.

We have made amendments regarding sensitivity and specificity. In this particular study the low value of foot length represents the abnormal state, while the high value represents the normal state. This is unlike in most measures during diagnosis e.g. blood pressure, blood sugar heart rate where the low value represents the normal state and vice versa. During analysis the anthropometric measurements should be are coded differently. This was not done in the earlier analysis therefore the values for sensitivity and specificity were interchanged. This has been addressed and the corrected values are shown in table 3. With the correct sensitivity and specificity, we still use the cut off associated with the highest correctly classified proportion, and the sensitivity is higher than the specificity, therefore reducing the number of false negatives.

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.

We have mentioned the biases in the revised manuscript.
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.

This has been added.

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.

This has been re-phrased to reflect that the study did not evaluate ease of measurements, but it was just reported by the health workers.

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

Done.

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

Corrected.

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.

Corrections for points 21-24 have been done.

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.
Corrections done

Discussion

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

References added.

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.

Data on gestational age was not available for most babies as mothers were unsure of the dates for their LNMP.

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’
Discretionary Revisions
1. The lowest birth weight in their sample was 1370g. They excluded babies from their study if their condition was unstable at birth and figure 1 indicates that 6 babies were excluded as their birth weight was <1000g and presumably needed immediate care. The authors should clarify how many babies weighed 1000-1370g at birth and whether they were excluded due to “difficult breathing”, cyanosis etc.

There were only 6 babies less than 1000g and they were all excluded as they were visibly very weak. No other signs and symptoms were recorded. It was specified in the protocol that all babies less than 1000g will be excluded. There were no babies between 1000-1370g.

2. Many babies with birth weight 2000-2500g are growth restricted and are not necessarily in need of special care, but those <2000g at birth are at higher risk. As with the Tanzanian study they quote (ref 17), it would be useful to extend their analysis to indicate what foot length would predict a weight of <2000g. The Tanzanian study also looked at predictors of birth weight <1500g but I would guess that their number in this category would be too small to analyse.

The number of babies with birth weight <1500g were only 3 (0.42%) and <2000g were 28 (3.97%). We did not analyse these groups because of the small numbers.

Minor Essential Revisions

The references need to be revised according to standard guidelines – many of the journal names are not listed by their standard abbreviations and several references are lacking page numbers etc.

We have revised the references as per guidelines.

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

Quality of written English: Acceptable