Author's response to reviews

Title: Symphysis-fundus height measurement to predict small-for-gestational-age status at birth: a systematic review

Authors:

Aase Serine D Pay (aase.pay@fhi.no)
Johanna Wiik (johanna.wiik@gmail.com)
Bjørn Backe (bibac@online.no)
Bo Jacobsson (bo.jacobsson@obgyn.gu.se)
Annika Strandell (annika.strandell@vgregion.se)
Atle Klovning (atle.klovning@me.com)

Version: 2
Date: 20 January 2015

Author's response to reviews: see over
January 20, 2015

Professor Valerie Smith  
BMC Pregnancy and Childbirth

Manuscript id: 1468433674149586  
Manuscript title: Symphysis-fundus height measurement to predict small-for-gestational-age status at birth: a systematic review  
Authors: Aase Serine D Pay, Johanna Wiik, Bjørn Backe, Bo Jacobsson, Annika Strandell and Atle Klovning

Dear Professor Smith,

Thank you for the opportunity to resubmit our manuscript, “Symphysis-fundus height measurement to predict small-for-gestational-age status at birth: a systematic review” to BMC Pregnancy and Childbirth. We appreciate the time and effort that the reviewers and editors have dedicated to critiquing and commenting on our work.

We have reviewed the issues highlighted by the two reviewers and responded to each in the following pages. Page and paragraph numbers are included to indicate the location of revisions in response to each comment.

Please feel free to contact us by telephone or e-mail should you have any additional questions or concerns.

We look forward to learning of your decision.

Sincerely,

Aase Devold Pay

aase.pay@fhi.no

Norwegian Institute of Public Health  
PO Box 4404 Nydalen  
N-0403 Oslo  
Tel: +4797580912
Replies to Reviewer 1

1. Background, line 76. You write that a major goal of antenatal care is to detect and prevent FGR. It is probably more correct write …. To detect FGR and prevent adverse outcomes associated with FGR. (Discretionary revision)

Response: We have rephrased the sentence (page 4, paragraph 1).

2. The background section is very short. In my opinion, the article will be better if you explain that routine SF measurement in antenatal care is a screening test, a couple of paragraphs to explain what a screening test is meant to be, provide more information about FGR, adverse outcomes associated with the condition and if measuring SF height is a routine part of antenatal care in Western countries. You could provide information about the proportion neonates born as SGAs to indicate if this is an important public health issue. You could also give a more detailed description of how the SF measure is performed (by a tape measurer), and what “further investigation of fetal growth and well-being” is (like CTG, ultrasound to assess fetal size, umbilical cord flow, etc). (Minor revision)

Response: Thank you for pointing this out. We have included this information in the Background section of the revised manuscript (page 4, paragraph 1-3).

3. Background or Methods: Please provide a short description on the most used SF curves. (Minor revision)

Response: This information has been added to the Background section of the revised manuscript (Page 4, paragraph 2).

4. Methods/Criteria for considering studies for this review/Index test, line 115: ….., compared with a reference or gold standard in the population…….. You should explain what a reference standard is, compared to a gold standard. The reference standard is birth weight – in my opinion that a gold standard rather than a reference standard. (Discretionary revision)

Response: In studies of diagnostic accuracy, reference (or gold) standards are used to distinguish patients with the target condition from those without it. In our study, the target condition was SGA status at birth and the reference (or gold) standards were the diagnosis of SGA or FGR, defined as birth weight < 10th, 5th, or 3rd percentile, or ≥ one or two standard deviations below the mean (performed postnatally). The index test (intervention) was SF measurement. We have revised the “criteria for considering studies” section to clarify this information (page 6, paragraph 1).

5. Data collection and analyses/Assessment of methodological quality: You should present the tool used for assessment in an additional file. (Minor revision)

Response: In the revised manuscript, we have included the QUADAS-2 tool in an additional file.
6. Data collection and analyses/Statistical analyses and data synthesis: If you provide a short information about what LRs are (the increase/decrease in odds for having the disease/reference standard after a positive negative test result) and what they tell us (e.g. that a test is regarded as useful if PLR>5 and NLR<0.2 or other values you find more correct, and provide a reference for your choice of threshold), it will be helpful for clinicians and other readers who do not have extensive knowledge in how to assess screening/diagnostic tests. You should also explain DOR. (Minor revision)

Response: An LR describes how many times more likely it is that a person with the target condition will receive a particular test result than will a person without it. Categorization of LRs was adopted from Deeks et al. where PLRs > 10 or NLRs < 0.1 are considered to provide convincing diagnostic evidence [1]. The DOR is commonly used as an overall indicator of diagnostic performance and calculated as the odds of a positive test result among those with the target condition, divided by the odds of a positive test result among those without the condition. As a general rule, a DORs > 100 indicate high accuracy, values of 25-100 indicate moderate accuracy, and those < 25 indicate that the test is not useful [1]. This information has been added to the Method section of the revised manuscript (page 8, paragraph 1).

7. Data collection and analyses/Investigation of heterogeneity: Could you write a few paragraphs about assessment of clinical heterogeneity, and add the results in the result section? (Major revision)

Response: Both clinical and statistical heterogeneity were evaluated. Assessment of clinical heterogeneity involved comparison of SF reference curves, cut-off criteria used to identify abnormal results, and SGA definitions. Assessment of statistical heterogeneity involved visual inspection of forest plots, and calculation of the inconsistency index ($I^2$) which quantifies the proportion of variation across included studies that is due to heterogeneity, rather than chance [2]. In accordance with your suggestion, we have added this information to the Methods section (page 8, paragraph 3). The $I^2$ value was typically high (98%; 95% CI, 97-99%). Given the small number of included studies (and thus low statistical power), subgroup analyses and covariate hierarchical modeling to investigate heterogeneity were not performed (Result section - page 10, paragraph1).

8. Results, lines 182-184: 721 citations were excluded, probably during the first screening of titles and abstracts. Why were they excluded? Please provide full references in Additional file 1. A flow-chart showing the selection process of eligible studies would be useful. (Minor revision)

Response: Initial database searches retrieved 722 citations, of which 525 citations remained after duplicates were removed (Figure 1). Screening of the titles and abstracts identified 51 potentially relevant articles that were retrieved in full-text format. Forward and backward citation tracking did not result in the identification of additional relevant articles. Eight articles were included in final analyses. Additional file 3 lists the reasons for excluding 43 articles on the basis of study population, design or outcome measures. This information has been added to the revised manuscript (page 9, paragraph 1).

9. Results, Table 1: Please provide information in the table about quality assessment, and if the study was of high quality or not. (Minor revision)
Response: The quality of each included study was assessed using the QUality Assessment of Diagnostic Accuracy Studies (QUADAS-2) checklist [3, 4]. The use of an overall quality score is not recommended for review authors, as different shortcomings may generate different magnitudes of bias, even in opposing directions, making it very difficult to attach sensible weights to each quality item. One way to summarize the quality assessment is to tabulate the results of individual QUADAS items for each single study [5]. We have thus presented quality assessment information in a separate table (Table 2).

10. Results, Figure 1: The figure is not very useful, it is better if you provide more information on each study in Table 1 (Discretionary revision)

Response: We have removed this figure.

11. Results, Figure 3: The information in the Figure is overlapping with information in Tables 2-4. You may consider removing it. (Discretionary revision)

Response: We agree and have removed the figure.

12. Conclusion/Implications for practice, lines 313-315: The paragraph “Other techniques that could improve upon this limitation (e.g., cardiotocography, Doppler ultrasound, and biophysical profile score) have not been implemented in the routine prenatal care setting.” It is unclear what you mean, please explain.

Response: Thank you. We have revised the sentence. Other techniques that could improve upon this limitation (e.g., routine ultrasound in the third trimester) have not been implemented in the routine prenatal care setting [6] (page 13, paragraph 2).

Replies to Reviewer 2

Minor essential revisions:

1. Methods. Searches: The search criteria should be more specific. All databases must be mentioned accurately so that a search can be replicated by others, “general bibliographic databases such as” is not specific enough. Did the authors check reference lists or contact others to identify missing studies?

Response: Electronic databases (PubMed, Medline, Embase, CINAHL, Cochrane Library, and SweMeD) were searched to identify eligible diagnostic studies from the earliest year possible to September 2014. The search strategy was developed for PubMed and modified for use in other databases (see Additional file 1). The reference lists of all included publications and relevant systematic reviews were checked and forward citation searches were performed. This information has been added to the revised manuscript (page 6, paragraph 2).
2a. Methods - Index test: It is not clear when and how frequent SF is measured in different studies (every week, every second week etc.), does this mean that the procedure is relatively consistent across the included studies? Is the frequency of measurement related to diagnostic accuracy?

Response: Many parameters involving the performance of SF height measurement, such as technique, frequency of measurement, and performer’s experience, potentially affect test accuracy [7]. All studies included in the systematic review used the same index test; SF measurement compared with the SF distribution of the population; however, we did not have detailed information about the test conditions, preventing exploration of the effects of potential differences in methods. This information has been added to the Discussion section of the revised manuscript (page 12, paragraph 1).

3. Results - Methodological quality. The included studies were all published in the eighties or early nineties. All included studies were appraised according to the QUADAS-criteria, and all were judged to fulfil most criteria (i.e. low risk of bias on most domains). This is not usually the case, and it makes me wonder whether the QUADAS-criteria were applied too loosely. Of course, it is likely that the authors are perfectly right, but the process is not described, and as a reader it is difficult to understand how the authors have applied the quality criteria. It would certainly help with some information (e.g. supplementary table) that a) says something general about what kind of methodological flaws that can lead to biased conclusions when studying the question of interest b) states what concern the authors had when risk of bias was assessed to unclear or high.

Response: 3a) Two review authors (ASDP, JW) assessed the quality of each included study using the QUality Assessment of Diagnostic Accuracy Studies (QUADAS-2) checklist [3, 4]. The QUADAS-2 checklist asks signaling questions in four risk of bias domains relating to patient selection, index test, reference standard, and flow and timing. Each domain is assessed in terms of risk of bias, and the first three domains are also assessed in terms of applicability. The review authors classified each item as “yes” (adequately addressed), “no” (inadequately addressed), or “unclear” (inadequate detail presented to allow a judgment to be made) (page 7, paragraph 3). The QUADAS-2 tool is shown in Additional file 2.

Response 3b) Based on the inclusion criteria, no included study had a case-control design [8-15]. All studies avoided inappropriate exclusions. Six of the eight studies used consecutive or random recruitment of participants. The two remaining studies [13, 15] did not report such information and were considered to be at unclear risk of patient selection bias. Most studies had a low risk of bias due to patient flow and timing; seven of eight studies involved the analysis of all recruited participants and one analysis included 78% of recruited participants [15]. Studies included in this review had a low risk of bias for the conduct of the reference standard. All studies used pre-specified index test thresholds. No study reported blinding to test results, but birth weight is objective and should not result in bias. Regarding the applicability of studies to the review questions, no study raised concern about the index test, reference standard or patient selection (page 9, paragraph 3).

4. Discussion. As the authors say, the SF height measuring has a low sensitivity and many false positives. I agree with the authors that when this leads to over-referral it is probably of less concern than failing to identify pregnancies at risk. However, I miss a discussion round the issue that the
referral can lead to increased anxiety for the women. It is crucial that the women are well informed about the limitations of the test and that midwives and clinicians are sensitive to this and that the women are well informed.

Response: Primary screening should emphasize the importance of sensitivity over specificity to identify almost all at-risk participants. No test is perfect and there will always be problems with incorrect results, e.g., anxiety and unnecessary intervention due to a false-positive result or a false sense of security caused by a false-negative result. A positive SF screening result can usually be confirmed or refuted with further evaluation of fetal growth and well-being by a specialist. This information has been added to the revised manuscript (page 13, paragraph 1).

5. Discussion. Women, I think, have a higher Body mass index now in general, than they did thirty years ago. What does that imply for the test today? Is it necessary to develop a new curve, or can we anticipate that SF height measuring and BMI are independent variables? Should such a curve be the same for all be specific to different groups of women? I’d like more details on how new studies should be designed.

Response: In a previous study, we presented a new SF reference curve with a pattern that differs from previous Scandinavian reference curves, reflecting changes in the pregnant population [16]. We also reported on the influences of maternal weight, height, age, parity, and fetal sex. Booking weight was the most important maternal characteristic, with a difference of 2 cm between the lightest and heaviest weight groups. Observational studies suggest that adjusted (customized) SF charts may improve the detection of a SGA neonate. In one study, use of customized charts resulted in improved sensitivity for an SGA neonate (48% vs. 29%) compared with abdominal palpation [17]. Use of customized charts was also associated with fewer referrals for investigation and fewer admissions. An audit from Great Britain also showed that use of customized SF charts detected 36% of SGA neonates compared with only 16% when customized charts were not used [18]. However, the benefit of adjusted vs. unadjusted SF curves remains to be demonstrated. In the revised manuscript, we suggest that further studies including larger numbers of patients and better standardized reporting criteria are desirable. The accuracy of adjusted vs. unadjusted SF curves also needs to be evaluated (page 13, paragraph 2).

Discretionary revisions:

6. Background. I miss information of results from the Cochrane Review (ref number 11) – what were the conclusions from this review and why was it necessary to do a new one?

Response: This information has been added to the Background section (page 4, paragraph 3).

7a. Result section. Included studies: PRISMA suggest that the search results and the process from searching to including studies is shown in a trial flow chart. The authors should consider doing this.

Response: Thank you. A flow chart has been added to the revised manuscript (Figure 1).

7b. The authors use the term article(s) instead of study sometimes in the text. I suggest the term study is used all over when referring to the included studies.
Response: We have changed this usage in the revised manuscript.

References