

PEER REVIEW HISTORY

BMJ Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form (<http://bmjopen.bmj.com/site/about/resources/checklist.pdf>) and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below. Some articles will have been accepted based in part or entirely on reviews undertaken for other BMJ Group journals. These will be reproduced where possible.

ARTICLE DETAILS

TITLE (PROVISIONAL)	Use of wind-up fetal Doppler versus Pinard for fetal heart rate intermittent monitoring in labour: a randomised clinical trial
AUTHORS	Montgomery, Ann; Byaruhanga, Romano; Bassani, D; Jagau, Anneke; Muwanguzi, Paul; Lawn, Joy

VERSION 1 - REVIEW

REVIEWER	Assistant Professor Yigit Cakiroglu Kocaeli University Department of Obstetrics and Gynecology Kocaeli-Turkey Infertility, Reproductive Endocrinology
REVIEW RETURNED	14-Nov-2014

GENERAL COMMENTS	<p>General comments: The article is interesting related with its topic. The manuscript needs to be improved in terms of minor revisions.</p> <p>Abstract: -The Abstract section is current and complete.</p> <p>Introduction: Introduction section is current and complete</p> <p>Methods: - Would the authors define the indications for Caesarean section? - Would the authors define the causes of high risk pregnancies exactly that resulted in exclusion from the study?</p> <p>Results: Results section is current and complete.</p> <p>Discussion: Discussion section is current and complete and gives the final message.</p> <p>References: -References are up-to-date.</p>
-------------------------	---

REVIEWER	Babović Ivana MD PhD Assistant School of Medicine, University of Belgrade, Belgrade, Serbia Department for Gynecology and Obstetrics, Clinical Center of Serbia, Belgrade, Belgrade, Serbia
REVIEW RETURNED	22-Nov-2014

GENERAL COMMENTS	<p>First I want to congratulate to Authors for investigate. Second I think that the Authors lack to more clearly described patients in both groups.I supposed that the aim of Your study is to determine the safest and cheapest monitoring labor and diagnosed acute intrapartal hypoxia.I think you must emphasize that the results you described are related to physiological pregnancies and labors.I have no information did You involve high risk patients for example diabetic mothers, mothers wih cardiac diseases etc.You excluded only fetal anomalies from Your study. Second I would like to know how You difference fetal heart rate and pulsation maternal abdominal aorta and how do You count fetal heart betas in fetal or maternal arrhythmia, used by Doppler or auscultation? Did You standardized Your results in this cases?</p>
-------------------------	---

REVIEWER	Andrew Hinde University of Southampton United Kingdom
REVIEW RETURNED	28-Nov-2014

GENERAL COMMENTS	<p>This paper is a study comparing the use of an existing method of monitoring foetal heart rates during labour using the Pinard Foetal Stethoscope (Pinard) with an alternative method using a handheld Doppler ultrasound foetal heart rate monitor (Doppler). The results show that the Doppler detects more foetal heart rate abnormalities but that this does not translate into lower stillbirth rates or neonatal mortality rates than experienced using the Pinard. I have the following comments on the paper.</p> <ol style="list-style-type: none"> 1. You mention a couple of times that care providers and the women both expressed a preference for the Doppler. This is a non-trivial factor, in my view. Given that from your results the Doppler performs no worse than the Pinard with respect to either its immediate function (detecting heart rate abnormalities) or the ultimate outcome (the survival of a healthy infant), then the preferences of the women and their carers become a significant factor. I realise you were not explicitly examining user preferences, but they seem to me to be rather more important than your paper suggests. 2. I think you need to do more to explain why you used generalised Poisson modelling. Your outcomes are - if I understand the paper correctly - binary rather than counts, and so a logistic regression model would seem more appropriate. The Poisson is bound to be an approximation for binary data as it involves a non zero probability of more than one event happening to the same person (i.e. a baby being still-born twice). I appreciate that if the rate is small then the probability of 2 or more events can be regarded as negligible, but even so - why use a model which is an approximation when there are models available that are not? 3. Table 2 can be deleted. It deals with outcomes which are not of primary interest and, apart from the Caesarean section, you do not mention any of them in the text of the paper. So, just quote the figures for the Caesarean section in the text and omit the table. 4. The results show that the Doppler detects more abnormalities than the Pinard, but these extra abnormalities detected by the Doppler do not affect mortality either immediately before or after birth. Is this because the additional abnormalities detected by the
-------------------------	---

	<p>Doppler are 'less abnormal' and so do not result in deaths? What matters for mortality are gross abnormalities and the Pinard can detect these.</p> <p>I am not an expert in the aetiology of stillbirths or neonatal deaths, but I am suggesting this as a possible explanation for the fact that neither stillbirths nor neonatal mortality were significantly lower using the Doppler than the Pinard, despite the Doppler's greater sensitivity in detecting foetal heart rate abnormalities.</p>
--	--

VERSION 1 – AUTHOR RESPONSE

Reviewer 1

1. Would the authors define the indications for Caesarean section?

(R) Indications for Caesarean delivery were failure to progress (as indicated by crossing of the action line on the partograph), abnormal FHR unresponsive to uterine resuscitation, and identification of malpresentation in labour (e.g. conversion from vertex to brow or mentum posterior). We have added this clarification to the Methods section on Participants.

2. Would the authors define the causes of high risk pregnancies exactly that resulted in exclusion from the study?

(R) The research protocol permitted clinical discretion with respect to participation. Women who presented with any condition that contra-indicated labouring were excluded from participation. We have added this point to provide clarification in the Methods..."presented with a condition that, according to the doctor on duty, contra-indicated labouring (e.g. antepartum hemorrhage); "

Reviewer 2

1. the results you described are related to physiological pregnancies and labors.I

have no information did You involve high risk patients for example diabetic mothers, mothers with cardiac diseases etc.

(R) See response above. Pregnancies with diabetes or cardiac disease were included provided there was no contra-indication to labouring.

2. Second I would like to know how You difference fetal heart rate and pulsation maternal abdominal aorta and how do You count fetal heart betas in fetal or maternal arrhythmia, used by Doppler or auscultation?

(R) We added that the maternal radial pulse was palpated, in order to clarify the clinical step of checking the maternal pulse simultaneously while listening to the fetal heart rate, to ensure that the midwife is not recording the maternal pulse, but rather picking up the fetal heart rate. We added to the Methods section..."The maternal radial pulse was simultaneously palpated to differentiate it with the FHR." FHR is counted the same for regular and arrhythmia - in beats per minute. "The FHR rhythm (regular or irregular)..." is documented in the clinical chart of the patient.

3. Did You standardized Your results in this cases?

(R) We are not sure what the reviewer means by this question.

Reviewer 3

1. You mention a couple of times that care providers and the women both expressed a preference for the Doppler. This is a nontrivial factor, in my view. Given that from your results the Doppler performs no worse than the Pinard with respect to either its immediate function (detecting heart rate abnormalities) or the ultimate outcome (the survival of a healthy infant), then the preferences of the women and their carers become a significant factor. I realise you were not explicitly examining user preferences, but they seem to me to be rather more important than your paper suggests.

(R) Agreed. It was not the primary objective of the study - however, we did find an overwhelming preference for the Doppler over the Pinard, and we anticipate that this would translate into more effectively monitoring due to diligent use by midwives, and demand by mothers. We have added to the final paragraph of the Discussion..."While assessing user satisfaction was not the objective of this study, the care providers and the women expressed preference for the Doppler, and given that the Doppler performed no worse than the Pinard in detecting abnormal FHR or in newborn survival, this should be an area of further research."

2. I think you need to do more to explain why you used generalised Poisson modelling. Your outcomes are - if I understand the paper correctly - binary rather than counts, and so a logistic regression model would seem more appropriate. The Poisson is bound to be an approximation for binary data as it involves a non zero probability of more than one event happening to the same person (i.e. a baby being still-born twice). I appreciate that if the rate is small then the probability of 2 or more events can be regarded as negligible, but even so - why use a model which is an approximation when there are models available that are not?

(R) We originally conducted the analysis using both logistic and Poisson regression with closely comparable results for the three primary outcomes. However, odds ratios are often interpreted incorrectly by non-specialist (e.g. OR=2 is often interpreted as 2x the risk), and relies on the rare disease assumption to approximate risk ratio. While the rare disease assumption likely holds true for

neonatal death and stillbirth, it would not for detection of FHR abnormalities. Therefore, in the interest of consistent reporting across the three primary outcome measures, and greater interpretability, we chose the Poisson regression with robust variance to estimate incidence risk ratio. We have added a reference in the Methods section, which illustrates worked examples of logistic versus Poisson, and directed the reader to this article to explain our choice of statistical methods. (Barros, BMC Med Res Methodol 2003)

3. Table 2 can be deleted. It deals with outcomes which are not of primary interest and, apart from the Caesarean section, you do not mention any of them in the text of the paper. So, just quote the figures for the Caesarean section in the text and omit the table.

(R) Agreed. Deleted table 2, and text edited accordingly. If BMJ Open editors wish, the original Table 2 can be a supplementary web appendix to demonstrate that randomization did in fact result in a balanced sample with respect to potential confounders.

4. The results show that the Doppler detects more abnormalities than the Pinard, but these extra abnormalities detected by the Doppler do not affect mortality either immediately before or after birth. Is this because the additional abnormalities detected by the Doppler are 'less abnormal' and so do not result in deaths? What matters for mortality are gross abnormalities and the Pinard can detect these.

I am not an expert in the aetiology of stillbirths or neonatal deaths, but I am suggesting this as a possible explanation for the fact that neither stillbirths nor neonatal mortality were significantly lower using the Doppler than the Pinard, despite the Doppler's greater sensitivity in detecting foetal heart rate abnormalities.

(R) This is a good point but it is unlikely to the case. Midwives are trained in, and this was reviewed with them as part of the study protocol, the identification of abnormal fetal heart rate pattern versus atypical or query abnormal fetal heart rate pattern. Query abnormal are those initial signs of concern that then resolve with minimal intervention (but require on-going monitoring and vigilance). Therefore 'less abnormal' FHR patterns identified should resolve after a short period of on-going monitoring (and not be categorized at abnormal for the purpose of the study) or be defined as abnormal if it remains unresolved. The list of abnormal FHR patterns in the Methods section is fairly objective and obvious - so we assumed that documentation of abnormal FHR for the purpose of the study was consistent in both the Pinard and Doppler group.

VERSION 2 – REVIEW

REVIEWER	Andrew Hinde Department of Social Statistics and Demography University of Southampton United Kingdom
REVIEW RETURNED	18-Dec-2014

GENERAL COMMENTS	I am happy with the way the authors have addressed points 1 and 3 in my original report. Point 4 was merely a comment and the authors have chosen to ignore it - which is their prerogative! I am still not convinced by their response to point 2. They have cited an article by
-------------------------	---

	<p>Barros and Hirakata (2003) in support of their decision to use a generalized Poisson model rather than a logistic model. Barros and Hirakata do not like odds ratios. They prefer prevalence ratios and they worry that the two may be confused by readers. Clearly, when the risk of a phenomenon is not small, odds ratios and prevalence ratios are different. But this is because they are measuring different things not because one is wrong and the other is correct. Odds ratios can always be converted to prevalence ratios - and they should be where the risk of the event is not small. Moreover Barros and Hirakata do not deal with the logical issue that the Poisson model will predict a non-zero probability of the same person dying more than once.</p> <p>My interpretation of the question is that where the risk of an event is not small, prevalence ratios and odds ratios diverge. Prevalence ratios are easier to understand intuitively than odds ratios, so it would be better to quote them in results. In this case, my preference would be to use a logistic model and to convert odds ratios to prevalence ratios (which can be done using predicted probabilities). Where the risk of an event is very small, then it does not really matter which model is used, as the chance of it happening twice to the same person is negligible and odds ratios and prevalence ratios are almost the same.</p> <p>In the case of the analysis reported in the paper, the risk of detection of an abnormality is not all that small, but the risk of stillbirth or death is very small.</p> <p>I should prefer not to use a Poisson model for the detection of abnormalities. However, since I doubt that the results would be any different with another model, and since the interest is really only in whether the difference in detection rates between the Pinard and the Doppler is statistically significant and not in quantifying the difference, I am happy to let this go.</p>
--	---

REVIEWER	Babović Ivana MD PhD Assistant School of Medicine, University of Belgrade, Serbia Department for Gynecology and Obstetrics, Clinical Center of Serbia, Belgrade, Serbia
REVIEW RETURNED	20-Dec-2014

GENERAL COMMENTS	<p>Congratulations for the authors for this research.</p> <p>I think that the most problem of this study is the fetomaternal morbidity. We can see that you perform Pinard auscultation or Doppler in physiological term pregnancies and deliveries. But in your low income country, as you cited, we have a problem with high risk term pregnancies. Perinatal outcome must be some different if you included that pregnancies, especially late preterm deliveries before 34th gestational weeks.</p> <p>Second problem is comparison more accurate method such as Doppler to subjective method as Pinard auscultation FHR. Pinard auscultation was complicated in obesity, hydramnion, especially between contractions. The average fetal BFHR is conditionally accurate in Pinard auscultation group compared to Doppler group. In section Material and Methods you did not inform as about Doppler measurements (in Doppler study we usually used some indexes for example PI or RI which inform as about fetal oxygenation before delivery.)</p>
-------------------------	---

VERSION 2 – AUTHOR RESPONSE

Reviewer: 3

Reviewer Name Andrew Hinde

Institution and Country Department of Social Statistics and Demography

University of Southampton

United Kingdom

Please state any competing interests or state 'None declared': None declared.

1. In the case of the analysis reported in the paper, the risk of detection of an abnormality is not all that small, but the risk of stillbirth or death is very small.

I should prefer not to use a Poisson model for the detection of abnormalities. However, since I doubt that the results would be any different with another model, and since the interest is really only in whether the difference in detection rates between the Pinard and the Doppler is statistically significant and not in quantifying the difference, I am happy to let this go.

(R) All good points listed in the reviewers second response. We will likely reconsider how we conduct future analyses for our clinical trials based on these valid points. As the reviewer is satisfied with letting it go, we did make edits to the manuscript.

Reviewer: 2

Reviewer Name Babović Ivana MD PhD Assistant

Institution and Country School of Medicine, University of Belgrade, Serbia

Department for Gynecology and Obstetrics, Clinical Center of Serbia, Belgrade, Serbia

Please state any competing interests or state 'None declared': none declared

1. I think that the most problem of this study is the fetomaternal morbidity. We can see that you perform Pinard auscultation or Doppler in physiological term pregnancies and deliveries. But in your low income country, as you cited, we have problem high risk term pregnancies. Perinatal outcome must be some different if you included that pregnancies, especially late preterm deliveries before 34th gestational weeks.

(R) Agreed, the outcomes could be quite different if we include preterm deliveries and this would be a topic for further study. Our study, as stated in our Methods section, included only term and postterm pregnancies (≥ 37 weeks) in which labouring was not contra-indicated. As this point is explicit in the Methods section, the external validity of the study is transparent. Therefore, we have not made edits to the manuscript.

2. Second problem is comparison more accurate method such as Doppler to subjective method as Pinard auscultation FHR. Pinard auscultation was complicated in obesity, hydramnion, especially between contractions. The average fetal BFHR is conditionally accurate in Pinard auscultation group compare to Doppler group.

In section Material and Methods you did not inform as about Doppler measurements (in Doppler study we usually used some indexes for example PI or Ri which inform as about fetal oxygenation before delivery.)

(R) To clarify, the Doppler is a handheld device used to auscultate the FHR while the woman is in labour. It is not to be confused with umbilical artery ultrasound (also called Doppler flow studies) used to assess fetal well-being in the context of antenatal care for high-risk pregnancies. In our study, both the handheld Doppler and the Pinard used the same measurements (baseline measured in beats per minute, presence or absence of accelerations/decelerations etc.) In the Methods sections, we

describe how the FHR using either Doppler or Pinard is assessed in labour by the research midwife. The method of assessment is the same for both the intervention and control group. As the reviewer is referring to a different tool in obstetrics, we did not edit the manuscript.