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.

ARTICLE DETAILS

TITLE (PROVISIONAL)	Impact of preterm birth on maternal wellbeing and women's perceptions of their baby: a population based survey
AUTHORS	Henderson, Jane; Carson, Claire; Redshaw, Maggie

VERSION 1 - REVIEW

REVIEWER	Rafał Bobiński University of Bielsko-Biała, Faculty of Health Sciences Poland
REVIEW RETURNED	13-Jun-2016

GENERAL COMMENTS	I read the manuscript of the publication entitled: Impact of preterm birth on maternal wellbeing and women's perceptions of their baby: a population based survey. The topic is interesting and the utilised methodology is sound. However, there is a need for a better definition of the studied population as well as a more thorough and comprehensive discussion of the results. 1. Page 5, line 14: "Women reported gestational age at delivery…" The question is whether the women were absolutely sure about the gestational age. Supposing they were, how was the age calculated? AGA (appropriate for gestational age) fetuses can only be identified after gestational age has been confirmed by ultrasound performed in the first half of pregnancy, therefore such information should have been provided by the Authors together with a reference to the ultrasound methodology. What we are certain about is the fact that women were classified into appropriate groups. A better description of the population would make the results and the conclusions more useful. 2. Table 1.: What does „37+" mean? Are overdue pregnancies also included? Were babies classified between 10th-90th percentile? The Authors should have also specified whether they had included LGA babies (>90 percentile); in this case LGA babies should have either been excluded from the discussion or added as a separate group. In order to better characterise their population, the Authors could have at least provided data about the APGAR scale of the newborns. 3. Page 5, line 15: Did they receive corticosteroids during pregnancy? The chapter 'Discussion' is the weakest part of the manuscript. It includes numerous repetitions from the chapter 'Results'. The discussion is a minor contribution to decent research work. It only includes references to other results. No bold attempt at interpretation has been made, which leaves the reader unsatisfied and disappointed. The title of the journal, British Medical Journal, suggests that it focuses, among other issues, on the problems of a widely

	understood medical mechanism. Therefore, a reference should have been made to such challenging issues and the interesting problems presented in the manuscript ought to have been discussed in such a context. Then the article would make sense, would be more interesting and recognisable, and it would gain the opportunity to be cited. If it were up to me, I would accept the publication on condition that the content of the chapter 'Discussion' was edited and the amendments suggested above were made.

REVIEWER	Antje Horsch University of Lausanne Switzerland
REVIEW RETURNED	23-Jun-2016

GENERAL COMMENTS	Thank you for inviting me to review this interesting study reporting data from the UK Millennium Cohort Study, which makes an important contribution to the existing literature. Important strengths of this study are the large sample size (which allows to control for several confounding factors) of a nationally representative cohort. Important weaknesses of the study are the use of a non-validated questionnaire to measure symptoms of anxiety, depression, fatigue, three selective PTSD symptoms, four questions relating to early contact with the baby and development of the mother-infant relationship and asking mothers to retrospectively respond whether they had experienced symptoms at 10 days, 1 month, and 3 months. I have some major concerns that need to be addressed and some clarifications are needed: (1) Given the design of the data reported here is cross-sectional, any causal language should be avoided, e.g. "impact of x on y" or "x affects y". (2) Outcomes were measured at 3 months rather than 3 months corrected age, as is common practice in preterm populations. This means that for the mothers, the time since the birth of their baby will vary greatly (with a range of approximately 4 months between the extremely premature babies at the one end and the post-term babies at the other end of the spectrum). Evidence shows that the time parents had to adjust to the birth of their (premature) baby plays an important factor with regard to their psychological adjustment. This needs to be discussed. (3) If I understand correctly, then mothers received one questionnaire at 3 months but they had to retrospectively respond whether they had experienced symptoms at 10 days, 1 month, and 3 months. The validity of responses is questionable given the possibility of a recall bias. (4) The description of the non-validated symptom checklist lacks detail and no rationale for the selection of items is given. How can anxiety and depression be assessed with one item, when the only validated screening questions for depression (Whooley questions) consist of 2 items? How can one question when mothers "first felt that their baby really belonged to them" measure mother-infant relationship? (5) How was the "number of positive/negative adjectives used by

	mother about baby" measured? Was it a pre-defined list or were mothers ask to generate their own list of adjectives? In case of the latter, was the number of possible adjectives restricted? Again, sufficient detail and a rationale are needed. (6) An a-priori power analysis is not reported. (7) For the results in Table 2, it is not clear to me which groups were compared here (please clarify in the data analysis section). I would also suggest changing the way the results in Table 2 are presented in order to make clear which group differences were significantly different. (8) The authors had no access to objective medical record data and therefore have to rely on the accuracy of self-reported health problems related to the mother herself and her baby. This needs to be discussed as a limitation. Minor issues: (1) I would suggest quoting the number of preterm babies in 2014 (the year in which the data was collected) rather than presenting a figure from 2010 (abstract). (2) In the abstract, "statistically significant differences" are reported without explaining which groups were compared. (3) The word "closely" is repeated twice in the second sentence of "Strengths and limitations". (4) Please give test statistic and p-value for the comparison of responders and non-responders. (5) In the discussion it is stated that mothers of preterm and very preterm babies had more depressive symptoms, although this was not significant. I suggest taking this out (as the difference was not significant). (6) In the discussion (p. 15) the authors mention the term "attachment" for the first time without defining it.

VERSION 1 – AUTHOR RESPONSE

Reviewer 1 Rafał Bobiński, University of Bielsko-Biała, Faculty of Health Sciences, Poland

I read the manuscript of the publication entitled: Impact of preterm birth on maternal wellbeing and women's perceptions of their baby: a population based survey. The topic is interesting and the utilised methodology is sound.

Thank you for your positive comment.

However, there is a need for a better definition of the studied population as well as a more thorough and comprehensive discussion of the results.

Table 1 already gives considerable detail of the studied population. We have added detail to the text on the very high proportions of women in the study sample having first trimester US scans (95%) and later anomaly scans (99%).

1. Page 5, line 14: ”Women reported gestational age at delivery…” The question is whether the women were absolutely sure about the gestational age. Supposing they were, how was the age calculated? AGA (appropriate for gestational age) fetuses can only be identified after gestational age has been confirmed by ultrasound performed in the first half of pregnancy, therefore such information should have been provided by the Authors together with a reference to the ultrasound methodology.

No clinical or medical record data were available in this survey based study. Thus we do not have information on the ultrasound methodology. The data sources are clearly described in the methods section.

We have added a note to the discussion (p16 1st para), emphasising that the evidence is that gestational age is well reported by women and referenced the evidence on this point. As indicated above we have added detail about the proportions of women having US scans in the results section.

What we are certain about is the fact that women were classified into appropriate groups. A better description of the population would make the results and the conclusions more useful.

The reviewer acknowledges that the classification of groups is appropriate.

2.Table 1.: What does „37+” mean? Are overdue pregnancies also included? Were babies classified between 10th-90th percentile?

The Tables have been clarified, ‘37+’ has been replaced with ‘37 or more’. This group did include overdue pregnancies and this has been clarified (p5 2nd para).

The authors should have also specified whether they had included LGA babies (>90 percentile); in this case LGA babies should have either been excluded from the discussion or added as a separate group.

The focus in the paper was preterm birth and outcomes for mothers and thus SGA and LGA babies were included in the analyses on prematurity as is made clear in the methods.

In order to better characterise their population, the Authors could have at least provided data about the APGAR scale of the newborns.

We have no information about APGAR scores as indicated above.

3.Page 5, line 15: Did they receive corticosteroids during pregnancy?

No information is available about corticosteroids or any other treatment women received in pregnancy. Evidence suggest no effect on maternal depression following multiple versus single doses of corticosteroids (Murphy, Hannah, Willan et al 2011) and levels of depressive symptoms comparable with our population.

The chapter ‘Discussion’ is the weakest part of the manuscript. It includes numerous repetitions from the chapter ‘Results’. The discussion is a minor contribution to decent research work. It only includes references to other results. No bold attempt at interpretation has been made, which leaves the reader unsatisfied and disappointed.

At the beginning of the discussion it is usual to recap the results and put them in the context of the literature.

Additional material has been added to the discussion as suggested.

The title of the journal, British Medical Journal, suggests that it focuses, among other issues, on the problems of a widely understood medical mechanism. Therefore, a reference should have been made to such challenging issues and the interesting problems presented in the manuscript ought to have been discussed in such a context. Then the article would make sense, would be more interesting and recognisable, and it would gain the opportunity to be cited. If it were up to me, I would accept the publication on condition that the content of the chapter 'Discussion' was edited and the amendments suggested above were made.

Text reflecting possible mechanisms and material on the impact of preterm birth and implications for care more broadly has been added to the discussion.

Reviewer 2 Antje Horsch, University of Lausanne, Switzerland

Thank you for inviting me to review this interesting study reporting data from the UK Millennium Cohort Study, which makes an important contribution to the existing literature. Important strengths of this study are the large sample size (which allows to control for several confounding factors) of a nationally representative cohort.

Thank you for your positive comment.

This paper reports results from the National Maternity Survey 2014, not the UK Millennium Cohort Study.

Important weaknesses of the study are the use of a non-validated questionnaire to measure symptoms of anxiety, depression, fatigue, three selective PTSD symptoms, four questions relating to early contact with the baby and development of the mother-infant relationship and asking mothers to retrospectively respond whether they had experienced symptoms at 10 days, 1 month, and 3 months.

These points are addressed individually below:
(1) Given the design of the data reported here is cross-sectional, any causal language should be avoided, e.g. "impact of x on y" or "x affects y".

We have minimised the use of causal language and acknowledged the weakness of the survey design. 'Impact' is used only in the title, in discussion of other literature and at the beginning of the results. 'Affect' is used only in the Objectives. Although causality cannot be inferred from cross-sectional studies, the fact that preterm birth precedes postnatal outcomes lends weight to that interpretation.

(2) Outcomes were measured at 3 months rather than 3 months corrected age, as is common practice in preterm populations. This means that for the mothers, the time since the birth of their baby will vary greatly (with a range of approximately 4 months between the extremely premature babies at the one end and the post-term babies at the other end of the spectrum). Evidence shows that the time parents had to adjust to the birth of their (premature) baby plays an important factor with regard to their psychological adjustment. This needs to be discussed.

The time since birth was the same for all participants. We consider this a strength of the study. If outcomes were measured at 3 months corrected age, as the reviewer points out, there would be a wide variation in time since the actual birth. We clarified this point in the manuscript (p15 3rd para).

(3) If I understand correctly, then mothers received one questionnaire at 3 months but they had to retrospectively respond whether they had experienced symptoms at 10 days, 1 month, and 3 months. The validity of responses is questionable given the possibility of a recall bias.

The issue of recall is referred to in the study limitations, nevertheless individual women do answer differently for the different time points as is reflected in the data presented.

(4) The description of the non-validated symptom checklist lacks detail and no rationale for the selection of items is given. How can anxiety and depression be assessed with one item, when the only validated screening questions for depression (Whooley questions) consist of 2 items?

The checklist of symptoms used had been employed in previous National Maternity Surveys. A standard instrument, the Edinburgh Postnatal Depression Scale (EPDS) was also used, as described in the methods. More detail has been added including the rationale for the symptoms listed (p5 2nd para).
Anxiety and depression were separate items as is clear in the text and tables.

How can one question when mothers “first felt that their baby really belonged to them” measure mother-infant relationship?

Several questions were asked that reflected the mother-infant relationship, including the adjective checklist (see Table 3).

(5) How was the “number of positive/negative adjectives used by mother about baby” measured? Was it a pre-defined list or were mothers ask to generate their own list of adjectives? In case of the latter, was the number of possible adjectives restricted? Again, sufficient detail and a rationale are needed.

The adjective checklist was a pre-defined list of 16 items. This has been clarified in the methods section (p5 para 2).

(6) An a-priori power analysis is not reported.

This study was based on secondary analysis of existing data and thus an a-priori power analysis was not appropriate or possible.

(7) For the results in Table 2, it is not clear to me which groups were compared here (please clarify in the data analysis section). I would also suggest changing the way the results in Table 2 are presented in order to make clear which group differences were significantly different.

The comparisons in tables 1-4 are between the 3 gestational ages groups and thus a global test for association across categories was used. This has been clarified in the data analysis section (p5 last para).

(8) The authors had no access to objective medical record data and therefore have to rely on the accuracy of self-reported health problems related to the mother herself and her baby. This needs to be discussed as a limitation.

This has been acknowledged in the Discussion

Minor issues:
(1) I would suggest quoting the number of preterm babies in 2014 (the year in which the data was collected) rather than presenting a figure from 2010 (abstract).

This has been changed as suggested.

(2) In the abstract, “statistically significant differences” are reported without explaining which groups were compared.

This has been clarified.

(3) The word “closely” is repeated twice in the second sentence of “Strengths and limitations”.

This has been amended.

(4) Please give test statistic and p-value for the comparison of responders and non-responders.

This has been added to the 1st para of the Results section.

(5) In the discussion it is stated that mothers of preterm and very preterm babies had more depressive symptoms, although this was not significant. I suggest taking this out (as the difference was not significant).
P<0.05 is an arbitrary cut-off for deciding what is important and relevant.

We consider that it is part of the whole pattern of poorer maternal health following preterm birth and, thus, should remain.

(6) In the discussion (p. 15) the authors mention the term “attachment” for the first time without defining it.

This has been clarified.

VERSION 2 – REVIEW

REVIEWER	Antje Horsch University of Lausanne, Switzerland
REVIEW RETURNED	15-Aug-2016

GENERAL COMMENTS	Thank you for inviting me to review this revised version. I am satisfied that the authors have responded to all of my comments, except for one: I don't agree with the authors' response to my comment (5) under minor issues. The results regarding the difference of EPDS depressive symptoms between the different groups were not significant and should thus not be discussed as being different. This is also in line with what the authors wrote in their abstract: "Comparing the three gestational age groups, no statistically significant différences in rates of depressive symptoms measured on the EPDS were found." If the authors want to make the case that this lack of statistical significance was probably due to a lack of power, then I suggest they say so directly and chose an alternative method to present these results, such as by reporting confidence intervalls and effect sizes.

VERSION 2 – AUTHOR RESPONSE

We have made a minor revision addressing this point as the reviewer suggests on page 14 of the manuscript.