PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	Three-year Follow-up of a Randomized Clinical Trial of Intravenous versus Oral Iron for Anaemia in Pregnancy
AUTHORS	Khalafallah, Alhossain ; Dennis, Amanda; Ogden, Kath; Robertson, Iain; Charlton, Ruth; Bellette, Jackie Bellette; Shady, Jessica; Blesingk, Nep; Ball, Madeleine

VERSION 1 - REVIEW

REVIEWER	JJ Duvekot MD PhD Erasmus MC – University Medical Centre Rotterdam Division of Obstetrics and Prenatal Medicine Department of Obstetrics and Gynecology
REVIEW RETURNED	23-Mar-2012

GENERAL COMMENTS	The present study describes the follow-up of 200 participants of a study on iron-suppletion during pregnancy. Of the 196 patients in that study, 126 (64 %) completed HrQoL questionnaires (modified SF-36) 26-42 months postpartum. The response rate is not very high for this kind of study. Most studies on QoL try to reach a response rate of 80%. How representative is it to measure QoL so long after one intravenous iron infusion? Was this study open label, did the participants know which treatment was given? The study population is small, there is no control group. It is important in studies where QoL is measured to define a minimal clinical important difference and one should not only look at significancy. Title: Not accurate. It concerns a follow-up of the first three years postpartum. This should be stated more clearly in the title. Introduction: Adequate, but state more clearly when the women are going to be assessed. Patients and methods: How many questionnaires were filled in by the participants? Only one during pregnancy and one postpartum? How representative is it to fill in the questionnaire by telephone interview? I do not think filling in a questionnaire retrospectively for four time-points is very accurate. Iron status was not checked after delivery and after the follow up phase. Results: The OR's for the results of the SF-36 are quite small, 1,36 – 2,05. What is the clinical value of these small differences. Page 10, line 11: where are figures 2a and 2b. Discussion: Compare the results better with other studies (eg. Jansen et al.). The strong and weak points of the study are not very good discussed. The retrospective nature of the study, the telephone interview, the small SF-36 differences, the fact that no controls were

	used…. Figures: There are two figures 1. Mention in the second figure 1 the number of patients. Why is sometimes hemoglobin and sometimes ferritin used? Tables: From Table 1 it is not clear to me, which questions were left over in the modified SF-36. In Table 2 the HR's should be corrected for maternal age. This is a large confounder. Both the results and the discussion should be definitely written more clearly.

REVIEWER	Professor Stephen Walters University of Sheffield ,UK
REVIEW RETURNED	02-May-2012

THE STUDY	Since the study is an RCT: the abstract should follow the CONSORT guidelines for abstract reporting of RCTs. For example how many women were randomised to the oral and IV iron groups, confidence intervals for the main results. The censoring mechanism for the time to finishing breast feeding should be made clear. For the sub groups analysis looking at socioeconomic status and sex of the baby the apprioprate test of association is to test for a randomised group * sub group interaction.
RESULTS & CONCLUSIONS	There is no table of the baseline characteristics of the women in the two randomised groups. The study should report the timepoints post-randomisation that the data was collected not 28 and 34 weeks gestation or post delivery. The report should follow the CONSORT guidelines. Percentages should report the numerator and denominator that the percentage is based on. The results section should recap the results of the main study and quote confidence intervals and not just a P-value. The study should concentrate on comparing outcomes between the IV and oral Iron group women and not iron status. The developers of the SF-36 and most other researchers tend to treat the QoL scales as continuous and use statistical methods appropriate for continous data. See for example: Fairclough, D.L. (2002) Design and analysis of quality of life studies in clinical trials. New York, Chapman & Hall. Fayers, P.M., Hays, R.D. (editors) (2005) Assessing quality of life in clinical trials: methods and practice. 2nd edition. Oxford, Oxford University Press.

	Fayers, P.M., Machin, D. (2007) Quality of Life: the assessment, analysis & interpretation of patient-reported outcomes. 2nd edition. Chichester, Wiley. Walters S.J. Quality of life outcomes in clinical trials and health care evaluation: a practical guide to analysis and interpretation. Chichester: Wiley 2009.
REPORTING & ETHICS	The study is an RCT. The CONSORT checklist for reporting RCTs should be included. It is not clear whether or not all the CONSORT statements have been answered in the manuscript. the abstract should follow the CONSORT guidelines for abstract reporting of RCTs. For example how many women were randomised to the oral and IV iron groups, confidence intervals for the main results. A fundamental flaw of the manuscript is the retrospective collection of QoL data from the mothers by there recall. I think the study should concentrate on the breast feeding outcomes and the baby growth data and just use the final QoL assessment - which I assume was collected prospectively and not retrospectively.

VERSION 1 – AUTHOR RESPONSE

General comments of the reviewers:

The present study describes the follow-up of 200 participants of a study on iron-suppletion during pregnancy. Of the 196 patients in that study, 126 (64 %) completed HrQoL questionnaires (modified SF-36) 26-42 months postpartum. The response rate is not very high for this kind of study. Most studies on QoL try to reach a response rate of 80%. How representative is it to measure QoL so long after one intravenous iron infusion? Was this study open label, did the participants know which treatment was given? The study population is small, there is no control group.

Response:
Reference the relatively small number of patients, there is no data elsewhere in the literature reporting specifically on perceived HRQoL during and after pregnancy in correlation with postnatal depression and breastfeeding. Therefore we consider that 126 patients is a relatively sufficient number to represent our preliminary data and to demonstrate these interesting results that may be a basis for a larger trial and further research in view of serious lack of data in this field.

Furthermore, a recent comprehensive meta-analysis and review by Reveiz et al in Cochrane Database Syst Rev; 2012 analysing the literature between 1970 till current on different treatments for iron deficiency anaemia (IDA) of pregnancy showed paucity of prospective randomised trials assessing clinical effects of iron administration on pregnant women with IDA in spite of the high incidence and burden of disease associated with IDA. This review showed only two prospective randomized trials were conducted in this period; one is our first published trial of this series (Khalafallah et al 2010) and the second was a follow up trial of the effect of iron therapy on the newborn. Both trials did fulfil the stringent independent reviewer quality criteria.

Due to the long-period of follow up (32 months), there are many patients moved to other states or moved houses and lost the contact due to change of their details. The study was open labelled and randomisation was performed by a different department (Pharmacy) to avoid bias. The QoL was

estimated prospectively at 4 points and at one point retrospectively in the same cohort of patients. We compared the participant's previous response to the prospective part with the similar questions in the retrospective part. This is explained in details in the response to specific point No 3. We believe that the data should reflect adequately HRQoL in this cohort of patients. The non-blindness should not alter the results as QoL questionaries, which showed great consistency of the prospective data in correlation with the retrospective data.

Specific comments:
1. Title: Not accurate. It concerns a follow-up of the first three years postpartum. This should be stated more clearly in the title.

Response:
The title has been changed as suggested to: A follow up study.

2. Introduction: Adequate, but state more clearly when the women are going to be assessed.

Response:
We have added the time point when the women were assessed in the introduction as suggested by the reviewer. This has been highlighted for easy tracking.

3. Patients and methods: How many questionnaires were filled in by the participants? Only one during pregnancy and one postpartum? How representative is it to fill in the questionnaire by telephone interview? I do not think filling in a questionnaire retrospectively for four time-points is very accurate. Iron status was not checked after delivery and after the follow up phase.

Response:
We analysed HRQoL for our cohort of pregnant women prospectively during the original study at the baseline; prior to treatment in the second trimester, 4 weeks after initiation of treatment and in the third trimester pre delivery, as well as at 6-8 weeks post delivery. In the follow-up study, HRQoL questionnaire is conducted incorporating the original questionnaire in addition to additional parameters such as length of breastfeeding period and occurrence of postnatal depression as well as child growth data. This was performed at a median of 32 months post intervention in order to assess the long-term effect of both iron therapies on mothers' HRQoL in correlation to previous prospective data. This questionnaire, although performed prospectively, it has a retrospective component by asking the participated mothers the same questions that they have previously answered prospectively about their QoL during and after pregnancy compared to the current questionnaire. These data were analysed against the mothers' original prospective QoL data for validation purposes.

This has been added and highlighted for clarification in the introduction.

4. Results: The OR's for the results of the SF-36 are quite small, 1,36 – 2,05. What is the clinical value of these small differences. Page 10, line 11: where are figures 2a and 2b.

Response:
Thanks for highlighting the point with Figure 2, which should be split to figure a and b. We have removed the separated figures 2a and b in an earlier version and amalgamated them in one figure. However for clarification, we take the point of the reviewer and we added the following: figure 2a on the left and figure 2b on the right.

The variation in the QoL of the women is likely to be predominantly determined by a large range of

different physical, mental and social issues entirely independent of iron status and its management. The effect of improved iron status is likely to be incremental in this context.

In response to a comment from the reviewer, we have estimated the absolute effect sizes (in direct QoL scale scores), which we include in the tables, figures and text. As the reviewer has shown, they may have some illustrative usefulness.

What these new estimates show is that an OR of about 2.00 is equivalent of a QoL score difference of about 10 (out of a score range of 100). For the association between haemoglobin and PF scores, a 1 SD rise in Hb (10.9 g/L is associated with an 11 point rise in PF score. Since 4 SDs (in theory) represent about 95% of the range of Hb values, then the range of Hb is associated with a notional range of 45% of the possible values of PF scores. Since 0 to 100 represents the range from no functioning to perfect functioning the effect of Hb change cannot be said to be clinically insignificant.

The presentation of numbers arose from the need to keep within space limits dictated by paper-format constraints. A more expansive format would allow exploration of the ambiguities of the statistical analytic process, whilst assisting the interpretation of the clinical relevance of the subject.

Therefore, Figure 2 has been reformatted to show absolute QoL score, rather than odds ratios. The supplementary tables show both absolute QoL scores and odds ratios, representing the two forms of analysis.

5. Discussion: Compare the results better with other studies (eg. Jansen et al.). The strong and weak points of the study are not very good discussed. The retrospective nature of the study, the telephone interview, the small SF-36 differences, the fact that no controls were used…

Response:
We added a comparison between other studies and also we highlighted in the methods section the prospective/retrospective validation component of the study.

Due to paucity of data regarding HRQoL during and after pregnancy, there are only very few literatures available. Jansen et al studied the effect of delivery and postpartum on the HRQoL. A cohort of 141 pregnant women were included in this study. HRQoL questionnaires were measuring the immediate effect of delivery on HRQoL. The were conducted less than 1 day after vaginal delivery and less than two days after caesarean sections in a comparison to 3-6 weeks post delivery questionnaires for both groups. The study focused on patients HRQoL recovery after both delivery interventions. In this study, the different time-points of conduction of the questionnaire may not necessary reflect the HRQoL during pregnancy and also after the postpartum period. Furthermore, the immediate questionnaire after delivery and 3-6 weeks time during the post-partum period may be at least, in theory, influenced by the acute event of delivery, in particular if complications occur, as well as the associated emotional and hormonal fluctuations during this period. It is worthwhile noting that the same study did not correlate between association of iron status and perceived HRQoL in conjunction with breastfeeding.
6. Figures: There are two figures 1. Mention in the second figure 1 the number of patients. Why is sometimes hemoglobin and sometimes ferritin used?
Response:
Many thanks for highlighting the typographic error with two times figure 1. The second figure “1” actually is figure 2a and b. This has been corrected.

Regarding Haemoglobin and ferritin, the iron status variables used in the multivariate regression models were selected by stepwise regression, with the haemoglobin being selected sometimes and ferritin selected at other times. Substitution of one for the other tended to produce a marginal

reduction in effect size and a substantial increase in variance resulting in higher P-values, which we would judge is likely to be misinterpreted by some readers as indicating absence of effect. Therefore we have simply shown what we found.

However, it is worthwhile to note that we have measured peripheral blood values of “iron status”. These measures were designed to show the likely effect on red cell and haemoglobin production, and have not been specifically designed for more subtle effects, such as psychological effects. In order to examine and explain associations between “iron status” and psychological state of participants, it might be necessary to measure iron levels and iron-related metabolic effects in specific tissues, organs and systems that are much less accessible to measurement and whose functional effect on quality of life we do not understand.

The relationship between peripheral blood haemoglobin and ferritin and the relevant iron effects in the tissue responsible for these putative iron status benefits may not be direct, and is certainly still in the area of unknown. Thus we regard this ambiguity as an accurate statement of our level of understanding, and that any spurious precision would be inappropriate at this time.

7. The number of patients being analysed is included in the number-at-risk table shown in the figure. The total numbers from the overall study are shown in the patient flow diagram (Figure 1).

Response:
The number of patients being analysed is included in the total numbers from the overall study and are shown in the patient flow diagram (Figure 1).

8. Tables: From Table 1 it is not clear to me, which questions were left over in the modified SF-36.

Response:
We accept the reviewer point. Due to limitation of the space, we did not add all 36 points questions, however as response to the reviewer’s comment we added an expanded explanation of the questionnaire has been provided in the footnotes of Table 2 for a further clarification.

Perhaps we can send the actual questionnaire as supplementary material to the BMJ-Open, in addition to the modified table. This table will be required by the terms of use stipulation by the original producers of the SF-36 questionnaire.

9. In Table 2 the HR’s should be corrected for maternal age. This is a large confounder.

Response:
Thanks for this important point; we have already corrected the HR’s for maternal age. An expanded explanation of the analysis has been provided in the footnotes of the Table.

10. Both the results and the discussion should be definitely written more clearly.

Response:
We went through the discussion and results and clarified any possible points that may create confusion for the readers.

11. The censoring mechanism for the time to finishing breast feeding should be made clear.

Response:
Thanks for raising this important point; Censoring was taken as the time reported in the baby growth booklet, which was available and complete for all women/babies (i.e. had been recorded routinely). All

women had ceased breast-feeding by the time of the data collection. A sentence was added to Statistical methods section: “The time to cessation of breast-feeding was taken from the subject’s baby growth booklet.”

12. For the sub groups analysis looking at socioeconomic status and sex of the baby the appropriate test of association is to test for a randomised group/sub group interaction

Response:
Many thanks for highlighting this issue. As a matter of fact as part of the stepwise regression process, such interactions were included in the covariates at the start of the stepwise process. No associations were found, and these were excluded from the final model as per the predetermine criteria.

We have added as response a note in the methods section to this effect.

13. There is no table of the baseline characteristics of the women in the two randomised groups.

This has been added as requested. We included number of women who underwent vaginal or caesarean deliveries as this may influence the outcomes.

14. The study should report the time points post-randomisation that the data was collected not 28 and 34 weeks gestation or post delivery.

Response:
Thanks for raising this issue as this has been corrected.

15. Percentages should report the numerator and denominator that the percentage is based on.

Response:
We accept the valid point of the reviewer and we have added all numerator and denominator as requested.

16. The results section should recap the results of the main study and quote confidence intervals and not just a P-value.

Response:
We accept the point of the reviewer and we have added confidence intervals in addition to P-values as requested.

17. The study should concentrate on comparing outcomes between the IV and oral Iron group women and not iron status.

Response:
Please see our thorough response on this issue in point 6

18. The developers of the SF-36 and most other researchers tend to treat the QoL scales as continuous and use statistical methods appropriate for continuous data.

Response:
The early versions of the analysis of the data included both ordered logistic regression and generalised estimating equations: this produced a large set of numbers and tables, and so the latter was dropped during the production of the final draft.

We are aware that many people apply statistical methods appropriate for continuous interval data. The alternative is the one adopted by us in this paper of using a rank-order form of analysis. There are advantages and disadvantages to both approaches, but it is our opinion that there is no satisfactory accessible individual statistical method of comparing the effects and describing the clinical meaning of those effects.

The data produced by the individual SF-36 scales are inherently discontinuous rank-order scores whose steps are irregular to an unpredictable extent in any particular subject group and sample. Furthermore, those individual scales are subject to ceiling and floor effects, which difficulties are illustrated by the calculated standard deviations of some of the subgroups that approach or exceed the entire range of the scales. Therefore, a rank-order repeated-measures multivariate analysis method such as ordered logistic regression is appropriate. However, this method has the major limitation of producing an effect size estimate in the form of an odds ratio, which is very difficult to interpret in the clinical context (see Reviewer comment above and our response). It is also subject to distortion due to missing data and random effects.

The alternative of a random effects continuous interval methodology (the generalized estimating equations model used by us and shown in this revision) gives clinically-understandable effect size estimates, but post-estimation testing for assumptions of linear regression (the assumptions on which the estimates are based) shows major violations of those assumptions (various combinations of heteroskedasticity, skewness and kurtosis of the regression residuals in different QoL score models).

It seems that the above arguments are recognized by researchers who have thought about the issue, but that there has been a collective decision to ignore the difficulties in order to come to some sort of workable process for handling the data being produced by these QoL scales. (I regard QoL issues as very important in determination treatment effectiveness and related health-service decision-making issues.) We not necessary assume that our analysis method is incorrect, just that others do it differently. This is all fine for the researchers familiar with the uncertainties of handling QoL measurements, but it has the potential to leave the impression of excessive precision in the minds of those not familiar with such handling.

We believe that the results of the continuous interval and rank-order methodologies should be considered together, and an impressionistic judgment be made about the magnitude of the effect sizes and the certainty of those effect sizes in the light of the correspondence of the two sets of the results, taking into account the possibilities of false negative and false positive results in both sets.

As stated above, the results have been expanded to include both absolute QoL scale values and odds ratios, representing the two possible ways of analysing the data (both have their own problems), and the reader may draw their own conclusions from the results and any ambiguities that any difference in those results.

19. A fundamental flaw of the manuscript is the retrospective collection of QoL data from the mothers by there recall. I think the study should concentrate on the breast feeding outcomes and the baby growth data and just use the final QoL assessment - which I assume was collected prospectively and not retrospectively

This point has been explained in details previously in the response to questions 3 and 5. However, it is worth noting that there is a paucity of data recording HRQoL in pregnancy. Our study would be the first trial in the literature that looked on various aspects of HRQoL in conjunction with Haemoglobin and iron studies, postnatal depression and breastfeeding adding another important angle and

dimension for further research and discussions.

Furthermore, reanalysis of the data showed a diminished association between male babies and various mental component scales, although this variable has continued to feature as an adjustment covariate. We therefore withdraw our comments on this issue. Therefore, the last sentence of the Results section has been removed.

VERSION 2 – REVIEW

REVIEWER	Dr. J.J. Duvekot, MD, PhD Gynaecologist/Perinatologist Dept Ob/Gyn Erasmus MC, University Medical Centre Rotterdam Rotterdam The Netherlands. I declare that I have no competing interests with the present study.
REVIEW RETURNED	04-Jul-2012

THE STUDY	The revised version of this paper gives only answers on few of my earlier posed questions. Also the methods and results sections should be more systematically described. Title: Not accurate enough still. It still not clear that a randomized study was performed with iv iron vs. only oral iron medication. Also is not clear that the iron therapy was given during pregnancy. Abstract: Design section: the trial is intravenous and oral versus only oral iron therapy. In the results section it should be noted when the HRQoL exactly were evaluated. Introduction: Can be shortened, expecially the details on oral iron therapy should be made shorter. The initial trial is not a reference and pops up out of the blue. The initial results should be referred to (PMID: 20546462). The rationale and objectives section should not be put here but should be part of the methods section or should be shortened as last part of the introduction. Patients and methods: This still is not clear enough for me. Many of the former questions were not answered. My major problem is that patients were asked to answer questions at one time moment about four time points. This is not very accurate. Results: This section is still not very adequately described. It should be more systematic. The results section could be longer. Discussion: Page 14, lines 1-7, the english is not proper. Figures: Figure 2a and 2b should be separate dor combined. Later is not defined.
RESULTS & CONCLUSIONS	The revised version of this paper gives only answers on few of my earlier posed questions. Also the methods and results sections should be more systematically described. Title: Not accurate enough still. It still not clear that a randomized

	study was performed with iv iron vs. only oral iron medication. Also is not clear that the iron therapy was given during pregnancy. Abstract: Design section: the trial is intravenous and oral versus only oral iron therapy. In the results section it should be noted when the HRQoL exactly were evaluated. Introduction: Can be shortened, expecially the details on oral iron therapy should be made shorter. The initial trial is not a reference and pops up out of the blue. The initial results should be referred to (PMID: 20546462). The rationale and objectives section should not be put here but should be part of the methods section or should be shortened as last part of the introduction. Patients and methods: This still is not clear enough for me. Many of the former questions were not answered. My major problem is that patients were asked to answer questions at one time moment about four time points. This is not very accurate. Results: This section is still not very adequately described. It should be more systematic. The results section could be longer. Discussion: Page 14, lines 1-7, the english is not proper. Figures: Figure 2a and 2b should be separate dor combined. Later is not defined.

VERSION 2 – AUTHOR RESPONSE

1. The methods and results sections should be more systematically described.

This has been systematically described as suggested.

2. Title: Not accurate enough still. It still not clear that a randomized study was performed with iv iron vs. only oral iron medication. Also is not clear that the iron therapy was given during pregnancy.

We have changed the title to address the reviewer recommendation. The new title is:

“Three-year Follow-up of a Randomized-Controlled Study of Intravenous versus Oral Iron Therapy for Pregnancy Anaemia demonstrating that Intravenous Iron is Associated with Improved Maternal Quality of Life, Less Postnatal Depression and Longer Breastfeeding”

3. Abstract: Design section: the trial is intravenous and oral versus only oral iron therapy. In the results section it should be noted when the HRQoL exactly were evaluated.

We changed the abstract as suggested.

4. Introduction: Can be shortened, especially the details on oral iron therapy should be made shorter. The initial trial is not a reference and pops up out of the blue. The initial results should be referred to (PMID: 20546462).

We shortened the introduction as suggested and referred to the initial paper as suggested.

5. The rationale and objectives section should not be put here but should be part of the methods section or should be shortened as last part of the introduction.

We moved the rationale and objectives to the methods section as suggested.

6. Patients and methods: This still is not clear enough for me. Many of the former questions were not answered. My major problem is that patients were asked to answer questions at one time moment about four time points. This is not very accurate.

This has been clarified in the methods section.

7. Results: This section is still not very adequately described. It should be more systematic. The results section could be longer.

We elaborated our findings systematically as we can in the results section as suggested. We also expanded the results section as suggested.

8. Discussion: Page 14, lines 1-7, the english is not proper.

We reviewed the English in the discussion and rewrote the above mentioned paragraph.

9. Figures: Figure 2a and 2b should be separate dor combined. Later is not defined.

We separated the Figures 2a and b to be below each other instead of side by side and we defined later as suggested.

Finally, many thanks for the reviewer's criticism. We believe it has strengthened this article in its current form.

VERSION 3 – REVIEW

REVIEWER	Dr. J.J. Duvekot, gynaecologist Erasmus MC, University Medical Center Rotterdam Dept. of Obstetrics and Gynaecology Rotterdam, The Netherlands. I have no competing interests with this study.
REVIEW RETURNED	12-Aug-2012

THE STUDY	This paper is not very clearly written even after this second revision. Furthermore, the design of the study is very weak, the numbers of participants and the recall rate is small. I don't think this is a very good study, although it is always difficult to perform clinical studies like these. With a third revision I do not expect that this manuscript and especially the design will improve.
GENERAL COMMENTS	In general, it is still difficult to read this article. The design of the study is still not very well described. Also the text could be compromised to a certain extent.

	Title: Improved, but I still think the language could be better. Article summary: It is not always clear whether improvement of HRQoL is during, just after pregnancy or during longstanding follow-up. Make this more clear. This goes also for the time point where iron status was assessed. Abstract: see previous comments. In the methods section the postdelivery status is assessed twice: 6-8 wks and 32 months. The conclusion should state that repletion of iron stores should be done during pregnancy and that this is related to longstanding HRQoL. Introduction: State here clear that this is a second study as a follow-up to the initial study: In this follow-up study… Patients and methods: It became clear now, that initially a clinical questionnaire and later a modified SF-36 questionnaire was administered. It is doubtful whether these two questionnaires can be compared. This makes the validation doubtful. Results: How large are the scales of the HRQoL measurements? What is the clinical value of the small differences. Discussion: Method of delivery is not discussed and analysed except in case of cessation of breast feeding. Figures: Figure 2a and 2b: Explain what is meant by later. It is still not clear why haemoglobin and ferritin is used separately. Tables: Table 4 is a doubtful comparison.

VERSION 3 – AUTHOR RESPONSE

1. Title: Improved, but I still think the language could be better.

We have changed the title to: “Three-year Follow-up of a Randomized Clinical Trial of Intravenous versus Oral Iron for Anaemia in Pregnancy demonstrates that Intravenous Iron Therapy is Associated with Improved Maternal Quality of Life, Less Postnatal Depression and Longer Breastfeeding”

2. Article summary: It is not always clear whether improvement of HRQoL is during, just after pregnancy or during longstanding follow-up. Make this more clear. This goes also for the time point where iron status was assessed.

We have clarified this point and stated that the HRQoL improvement was during and after pregnancy.

3. Abstract: see previous comments. In the methods section the post-delivery status is assessed twice: 6-8 wks and 32 months. The conclusion should state that repletion of iron stores should be done during pregnancy and that this is related to longstanding HRQoL.

We changed the abstract methods and conclusion sections as suggested.

4. Introduction: State here clear that this is a second study as a follow-up to the initial study: In this follow-up study…

We stated clearly that this is a second follow-up study of the original cohort of patients.

5. Results: How large are the scales of the HRQoL measurements? What is the clinical value of the small differences.

The scale of measurement is from 1-5. The clinical value between small differences needs to be investigated in further research and is beyond the scope of our study. The variation of the QoL in pregnancy is likely to be predominantly determined by a large range of different physical, mental and social issues entirely independent of iron status and its management. The effect of improved iron status is likely to be incremental in this context.

6. Discussion: Method of delivery is not discussed and analysed except in case of cessation of breast feeding.

There were no statistical differences in terms of HRQoL assessment regarding method of delivery between women who delivered normally and those who had caesarean section. This has been clarified in the results section.

7. Figures: Figure 2a and 2b: Explain what is meant by later. It is still not clear why haemoglobin and ferritin is used separately.

Later is referring to post delivery follow up assessment. We have added this to the figure.

The iron status variables used in the multivariate regression models were selected by stepwise regression. Substitution of one for the other tended to produce a marginal reduction in effect size. There is a well established link between iron status and haemoglobin. Most measures are designed to show the likely effect on red cell and haemoglobin production, and have not been specifically designed for more subtle effects, such as psychological effects. In order to examine and explain associations between “iron status” and psychological state of people, it might be necessary to measure iron levels and iron-related metabolic effects in specific tissues, organs and systems that are much less accessible to measurement and whose functional effect on quality of life we do not understand. The relationship between peripheral blood haemoglobin and ferritin and the relevant iron effects in the tissue responsible for these putative iron status benefits may not be direct, and is certainly unknown by us. Thus we regard this ambiguity as an accurate statement that reflects our current level of understanding.

8. Tables: Table 4 is a doubtful comparison.

Table 4 shows an important correlation and statistical significance between both prospective and retrospective assessment of HRQoL at different timepoints. We believe it highlights the association between both conducted questionnaires to the readers.

Many thanks for the reviewer’s criticism. We believe it has strengthened this article in its final form.