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 a modified version of Baby-Led Weaning on iron intake and status: a randomised controlled trial
AUTHORS	Daniels, Lisa; Taylor, Rachael; Williams, Sheila; Gibson, R; Fleming, Elizabeth; Wheeler, Benjamin; Taylor, Barry; Haszard, Jillian; Heath, Anne-Louise

VERSION 1 – REVIEW

REVIEWER	Mary Fewtrell UCL GOS Institute of Child Health, London, UK
REVIEW RETURNED	01-Sep-2017

GENERAL COMMENTS	This manuscript reports data from an important and much-awaited RCT comparing a modified version of baby-led weaning (BLISS) with conventional weaning. The authors are to be congratulated for undertaking this challenging research. The manuscript is clearly written with respect to the iron intake/status outcomes and I found it very interesting to read. The conclusions reflect the data and strengths and limitations are appropriately presented. My main comments and suggestions are as follows: 1. The primary and secondary trial outcomes are not clearly stated in the methods. From the protocol, it seems that the primary outcomes were BMI z-score at 12 and 24 months of age, and the rate of suboptimal iron status at 12 months of age. The growth data are already published, and this should be clearly explained here. I think it is important to be absolutely clear on these methodological issues not least because it will be relevant for the inclusion and quality assessment of this study in subsequent systematic reviews and meta-analyses. 2. The planned sample size for the study was 500 according to the protocol. Although the authors mention that this was not achieved, it would be relevant to explain the reasons for this as they may be relevant to the subsequent implementation of the BLISS approach. For example, did parents not like the idea of having the method of weaning 'imposed' or did they not like the idea of the BLISS approach? 3. As the main outcome for this paper (iron status) was only available for a sub-group (58%) of the infants, it is relevant to consider whether there is selection bias; are these infants representative of the whole study cohort? Please present the characteristics of the subjects with and without this outcome. 4. Although compliance with the BLISS approach was very good, the parents received a number of additional contacts with healthcare professionals or the study team. It would be relevant to discuss whether this level of contact is considered necessary to ensure

	compliance with the intervention, and the practical and resource implications of this. Do the authors think that some or all of these additional visits could be replaced by written/online information or telephone advice, for example? Minor comments: 1. Please comment on whether delayed cord clamping is practiced in this population. This may influence the infants' iron status and this information will provide context for comparison with any future studies. 2. Why are iron variables adjusted for maternal education and parity (which are not imbalanced between groups at baseline)? 3. Please clarify the heading for table 1 – it implies that there are 3 different sample sizes, but all the data are at baseline when presumably all recruited subjects provided data?

REVIEWER	Nicholas Embleton Newcastle Hospitals NHS Trust & Newcastle University, UK
REVIEW RETURNED	08-Sep-2017

GENERAL COMMENTS	This is a well conducted RCT of a complex intervention. There are very few RCTs in this area, and there are important concerns about the risk of iron deficiency in infancy so overall this type of study is welcomed. The study will reassure practitioners that given the right type of advice, baby led weaning is 'safe' from this perspective. The authors rightly point out that BLW without appropriate advice is an un-tested and potentially harmful approach. Over 50% of the study recruits were university educated, and whilst there was information on deprivation indices, I did not get a good sense of whether that might have impacted on adherence or outcomes. Was that tested at all? I am not sure it needs changing but the authors may care to note that in the UK BLISS is the acronym for a well know advocacy organisation supporting newborn babies so this might create a degree of confusion Background - appropriate and concise Methods Well described Biochemical assessment The use of 82% which was 58% of total study participtants is a bit tricky to read. I would say the % in whom blood tests were obtained, and then describe the reasons for not getting a result (which includes consent given but no blood obtained) p.14 2nd paragraph – I thought breast milk intakes were estimated from existing data on age, so how do they compare differences in estimates? 74% of both groups had inadequate iron intake at 7m age – was that expected or of concern? Other causes of anemia were not described but seemed prevalent – it would be interesting to know what these were

	There are lots of tables, but they are well laid out and they all contain useful information so I do not see a specific reason to limit References appear appropriate

REVIEWER	Kathryn Barger Tufts University, USA
REVIEW RETURNED	29-Nov-2017

GENERAL COMMENTS	Summary The authors report on a secondary outcome from the BLISS study, which was originally aimed to determine if the intervention prevents young children from becoming overweight. Iron intake and status is one of the secondary outcomes from the BLISS study, and it is important to evaluate since a previous smaller study (N=51) showed evidence that infants following BLW had mean dietary iron intake less than half that of infants following traditional spoon-feeding. This manuscript contributes scientific knowledge by comparing iron markers between BLISS subjects and a control group, as well as evaluate dietary patterns of subjects including factors such as age breastfed and age high-iron foods were introduced, and contributes evidence that a baby-led approach to complementary feeding does not appear to increase the risk of iron deficiency. Issues with this manuscript are detailed below, but I applaud the authors for including an interpretation of the confidence interval results on plasma ferritin (page 17 line 55 thru page 18 line 10). Major Issues Similar to a safety study, iron intake and status is being investigated in order to show that iron deficiency under BLISS is no different then under Well Child usual care. This scenario calls for statistical equivalence testing [Hoenig, John M., and Dennis M. Heisey. "The abuse of power: the pervasive fallacy of power calculations for data analysis." The American Statistician 55.1 (2001): 19-24.]. The authors should make the following changes. 1. Recalculate the power analysis stated in reference #12 for plasma ferritin. The mean, SD, and detectable difference in geometric mean (5.0 ug/L) are reported in Table 2 (in reference #12), thus this conversion to the equivalence testing scenario is possible and post-experimental power analysis paradoxes can be avoided. The result will be estimated power of the equivalence test, or probability that the significance test will conclude that the difference in iron status is small, when assuming the difference in iron status is small (equal or within a small margin of equivalence). This alternative sample size calculation should be reported in the Statistical Analysis section. 2. Report equivalence testing results with p-values for the three main results: iron intake at 7 and 12 months, and iron status at 12 months. Or alternatively, use the method evaluating if the confidence interval fits within the margin of equivalence [see Hoenig et al. section 4 Equivalence Testing] along with the reported power analysis suggested above, and you can use your already calculated confidence intervals.

	The results on breast milk intake are confusing. Page 10, lines 50-55: "It was not possible to directly measure breast milk intake so it was assumed to be 750 g per day at 7 months and 448 g per day at 12 months…" In Table 2 total energy and dietary iron are reported for subjects at age 7 and 12 months. However, it cannot be assumed that the breast milk intake is independent of intake from complementary foods. How can you justify your method here? It appears to me that only energy and dietary iron from complementary foods can be assessed properly. Page 14, lines 10-21: Differences in estimated breast milk intake at 7 and 12 months are reported with estimated differences, confidence intervals, and p-values. Why does it make sense to report these statistics, and where did the variability come from if you are assuming all subjects at 7 months have the same breast milk intake? I see that the profile of milk consumers is different in each group (eTable 1), however, reporting the differences on page 14 do not make sense. Iron status is listed as a primary objective in the 29 August 2011 protocol (Lower South Regional Ethics Committee - Protocol). However, iron status is listed as a secondary objective in reference #12 (numbered in the manuscript – Daniels et al 2015). Please explain this inconsistency. This is a major issue because the identification of an outcome as primary or secondary influences how the outcome is treated and the statistical analysis that follows, and so must be treated carefully. Minor Issues Page 2, lines 22-24: "The funders provided funds only and played no role in carrying out the study or the analysis or interpretation of results." Authors should also state who developed the study protocol, especially in regard to developing study aims. Did the funders have input in defining the study aims? Please state here. Page 4, lines 12-14: "This randomized controlled trial included 206 participants assigned to Control (n=101) or BLISS (n=105) groups." Sample size reported should be number of completers relevant to the outcomes under study here (iron intake and status). Page 8, lines 39-43: please restate in this section Well-Child care recommendations on breastfeeding and complementary feeding, to contrast with recommendations under BLISS, "to encourage and support exclusive milk feeding (ideally breastfeeding) and delay the introduction of complementary foods until 6 months of age" (page 9, lines 14-18). Table 1 (page 26 – 27): The title describes N=162 at 7 months, yet Control N=81 and BLISS n=88. These sample sizes appear inconsistent. Table 2 (page 28 – 29): Table footnote (b) Difference adjusted for infant age and sex, and maternal education and parity. How are you adjusting for age if all infants are 7 months, for example? Are you adjusting for age in days, if so, this needs to be explicit. This footnote appears in other tables as well.

VERSION 1 – AUTHOR RESPONSE

"Impact of a modified version of Baby-Led Weaning on iron intake and status: a randomised controlled trial" (bmjopen-2017-019036)

Response to reviewers (please find attached a copy of this response with highlighted text)

Editor comments:

1. The authors don't make clear to the reader in both the abstract and the body of the paper is that this is a "slice" of the trial in that they are just reporting what seem to be post-hoc outcomes. The trial registry /https://www.anzctr.org.au/Trial/Registration/TrialReview.aspx?ACTRN=ACTRN12612001133820) does mention as secondary outcome number 5 "dietary intake" (7,12 and 24 months), so authors need to clarify whether these are indeed post-hoc outcomes or prespecified secondary outcomes (it is confusing at this stage). Please clarify whether these outcomes are pre-specified or not and make all this clear in the paper.

Thank you, some additions have been made to clarify this.

Abstract (Page 4):
"The primary outcome of the BLISS study (growth) has been previously reported. This paper reports the key pre-specified secondary outcomes iron intake and iron status."

Introduction (Last paragraph, Page 7):
"In this paper we report the key pre-specified secondary outcomes iron intake (at 7 and 12 months of age) and iron status (at 12 months) of infants following BLISS compared with traditional spoon-feeding."

Reviewer 1 comments:

This manuscript reports data from an important and much-awaited RCT comparing a modified version of baby-led weaning (BLISS) with conventional weaning. The authors are to be congratulated for undertaking this challenging research. The manuscript is clearly written with respect to the iron intake/status outcomes and I found it very interesting to read. The conclusions reflect the data and strengths and limitations are appropriately presented.

Thank you.

1. The primary and secondary trial outcomes are not clearly stated in the methods. From the protocol, it seems that the primary outcomes were BMI z-score at 12 and 24 months of age, and the rate of suboptimal iron status at 12 months of age. The growth data are already published, and this should be clearly explained here. I think it is important to be absolutely clear on these methodological issues not least because it will be relevant for the inclusion and quality assessment of this study in subsequent systematic reviews and meta-analyses.

We apologise for the lack of clarity. The original ethics application (submitted with the paper) does not reflect the changes we made subsequently to our primary and secondary outcomes in order to have a single primary objective. We have attached the published protocol paper which clearly states the iron outcome as a secondary objective (Daniels et al., 2015). This is consistent with the trial registry (http://www.anzctr.org.au/TrialSearch.aspx?searchTxt=ACTRN12612001133820&isBasic=True) and our published main outcomes paper by Taylor et al. (2017).

A sentence explaining that the primary outcome (i.e. growth) has been published has been added to the abstract (Page 4):

"The primary outcome of the BLISS study (growth) has been previously reported. This paper reports the key pre-specified secondary outcomes iron intake and iron status."

2. The planned sample size for the study was 500 according to the protocol. Although the authors mention that this was not achieved, it would be relevant to explain the reasons for this as they may be relevant to the subsequent implementation of the BLISS approach. For example, did parents not like the idea of having the method of weaning 'imposed' or did they not like the idea of the BLISS approach?

The initial sample size was based on recruitment rates for our previous study (assessing diet and sleep) which achieved a 59% response rate. Both studies used an "opt out" method of recruitment where all individuals who were eligible to participate were given the option to opt out of participation, rather than being expected to actively contact the study in order to participate. At the time the previous study was carried out it was the only study recruiting using this opt out method. However, by the time the BLISS study rolled out there were 14 other studies using the same recruitment method. We did not collect any data on the reasons why some prospective participants chose not to take part in the BLISS study, however, we speculate that this was due to the fact that 14 other studies were also recruiting participants, which reduced the pool of potential participants. We reduced the sample size because we did not have sufficient funding to be able to extend the recruitment period enough to account for the reduced rate of recruitment.

3. As the main outcome for this paper (iron status) was only available for a sub-group (58%) of the infants, it is relevant to consider whether there is selection bias; are these infants representative of the whole study cohort? Please present the characteristics of the subjects with and without this outcome.

Thank you for this suggestion. We have added these data as a supplementary table (eTable 1) and described them in the Results section (Page 14):
"There were no differences in the characteristics of participants who were included in this analysis (i.e. provided either intake or status data) compared with those were not included (i.e. provided neither intake nor status data) with the exception of maternal age at birth, which was lower for those who did not provide data (eTable 1)."

4. Although compliance with the BLISS approach was very good, the parents received a number of additional contacts with healthcare professionals or the study team. It would be relevant to discuss whether this level of contact is considered necessary to ensure compliance with the intervention, and the practical and resource implications of this. Do the authors think that some or all of these additional visits could be replaced by written/online information or telephone advice, for example?

Unfortunately, we can only speculate on this. There may be some capacity for some aspects of the additional contacts to be provided as written or online information, however, these additional contacts provided individualised as well as generic advice. This individualisation could possibly be provided over the telephone rather than face to face. Our lactation consultant commented that, at least in terms of the breastfeeding component of the study, it was important that the initial appointment was face to face, but that telephone advice may be appropriate after this. It is not known whether the advice would be less effective if given over the telephone rather than in person.

Minor comments:
1. Please comment on whether delayed cord clamping is practiced in this population. This may influence the infants' iron status and this information will provide context for comparison with any future studies.

A sentence has been included on this in the Methods section after the description of the hospital from which the participants were recruited (Page 8):
"Delayed cord clamping was infrequently practiced in Queen Mary Maternity Hospital at the time of the study."

2. Why are iron variables adjusted for maternal education and parity (which are not imbalanced between groups at baseline)?

Parity and maternal education were used as stratification variables in the design of the trial so we followed the advice of the European Medicines Agency (2013) to include them in the analysis.

3. Please clarify the heading for table 1 – it implies that there are 3 different sample sizes, but all the data are at baseline when presumably all recruited subjects provided data?

The data presented in Table 1 are for participants who provided data at 7 and 12 months of age as well as at baseline. The numbers of participants providing data at these time points have been removed to prevent any confusion regarding the sample size.

Reviewer 2 comments:

This is a well conducted RCT of a complex intervention. There are very few RCTs in this area, and there are important concerns about the risk of iron deficiency in infancy so overall this type of study is welcomed.

The study will reassure practitioners that given the right type of advice, baby led weaning is 'safe' from this perspective. The authors rightly point out that BLW without appropriate advice is an un-tested and potentially harmful approach.

Thank you.

1. Over 50% of the study recruits were university educated, and whilst there was information on deprivation indices, I did not get a good sense of whether that might have impacted on adherence or outcomes. Was that tested at all?

This was not tested, however, we did adjust for maternal education in our outcome analyses (Tables 2, 3 and 4).

2. I am not sure it needs changing but the authors may care to note that in the UK BLISS is the acronym for a well know advocacy organisation supporting newborn babies so this might create a degree of confusion

Thank you for bringing this to our attention. We have had a look into this and can acknowledge that this could potentially create some confusion for some UK readers. We have added a sentence at the end of the introduction to the Methods section (Page 8):
"The BLISS study, and intervention, are not in any way related to the UK-based Bliss charity for "babies born premature or sick" (www.bliss.org.uk)."

Biochemical assessment
3. The use of 82% which was 58% of total study participtants is a bit tricky to read. I would say the % in whom blood tests were obtained, and then describe the reasons for not getting a result (which includes consent given but no blood obtained)

The wording under 'Biochemical Assessment' in the Methods section (Page 11) has been edited for clarity:
"A non-fasting venous blood sample was obtained from 119 infants at 12 months of age (58% of total study participants). Of those who did not provide a blood sample, 26 blood draws were unsuccessful, 22 had withdrawn from the study by 12 months of age, 13 could not be contacted or were living out of town, and 26 parents did not provide consent for the blood test."

4. p.14 2nd paragraph – I thought breast milk intakes were estimated from existing data on age, so how do they compare differences in estimates?

The reviewer is correct, breast milk estimates were used. However:
(a) Mean intakes of breast milk were calculated across all the participants in each group, and not all infants consumed breast milk; and
(b) Not all breastfed infants were being given breast milk as their only infant milk (i.e. many were being both breastfed and infant formula fed [eTable 2]). For these mixed fed infants, their intake of breast milk was estimated as the full estimate (i.e. 750 or 448 g per day) minus the amount of formula given (as described in paragraph 1 on Page 11).

5. 74% of both groups had inadequate iron intake at 7m age – was that expected or of concern?

We have added some discussion of this point to the Discussion section (Page 18):
"There was a high proportion (74% of both groups) of infants at risk of inadequate iron intakes at 7 months of age. Unfortunately, we do not have a measure of iron status at 7 months to determine whether this high prevalence of inadequate intakes is reflected in poor iron status. However, at 12 months of age the risk of inadequate intakes had decreased (23% of Controls, 26% of BLISS). It is possible that this high prevalence at 7 months of age may be due to the cut offs available for determining the risk of inadequate iron intakes - currently, there is no specific cut off for infants less than 8 months of age that has Institute of Medicine probabilities of inadequacy that are needed in order to calculate the prevalence of inadequacy.[35]"

6. Other causes of anemia were not described but seemed prevalent – it would be interesting to know what these were

Thank you for this suggestion. A paragraph regarding other types of anemia has been included in the discussion (Page 19):
"The BLISS study focused on iron deficiency anaemia, but 10% of Control infants and 13% of BLISS infants were diagnosed as having anaemia that was not concurrent with iron deficiency. Non-iron-deficient anaemia can be caused by a wide range of conditions, including infection (e.g., with malaria, HIV, or hookworm), folate or Vitamin B12 deficiency, and genetic disorders such as thalassemia and sickle cell anaemia.[36] We took care to minimise rates of infection in our study design, and malaria, HIV and hookworm are extremely rare in this age group in New Zealand. Similarly, no participant had a mean cell volume >86 fL which would be indicative of the megaloblastic anaemia of folate or Vitamin B12 deficiency.[37] We cannot rule out haemoglobinopathies as a cause of anaemia for some of the infants, but these would be fairly rare in this population. An alternative explanation for the high proportion of other anaemia could be the cut off used for defining anaemia (<110 g/L).[36,37] This value has been extrapolated from older age groups,[38] and there has been some discussion as to whether a lower cut off may be more appropriate in this age group.[39]"

Reviewer 3 comments:

… Issues with this manuscript are detailed below, but I applaud the authors for including an interpretation of the confidence interval results on plasma ferritin (page 17 line 55 thru page 18 line 10).

Thank you.

1. Similar to a safety study, iron intake and status is being investigated in order to show that iron deficiency under BLISS is no different then under Well Child usual care. This scenario calls for statistical equivalence testing [Hoenig, John M., and Dennis M. Heisey. "The abuse of power: the pervasive fallacy of power calculations for data analysis." The American Statistician 55.1 (2001): 19-24.]. The authors should make the following changes.

a) Recalculate the power analysis stated in reference #12 for plasma ferritin. The mean, SD, and detectable difference in geometric mean (5.0 ug/L) are reported in Table 2 (in reference #12), thus this conversion to the equivalence testing scenario is possible and post-experimental power analysis paradoxes can be avoided. The result will be estimated power of the equivalence test, or probability that the significance test will conclude that the difference in iron status is small, when assuming the difference in iron status is small (equal or within a small margin of equivalence). This alternative sample size calculation should be reported in the Statistical Analysis section.

A number of authors, including Hoenig and Heisey (2001), Greenland et al. (2016) and Senn (2002) argue that power after the event is irrelevant. Although Hoenig and Heisey argue that once confidence intervals are constructed power calculations yield no additional information they go on to suggest equivalence testing. Using their “seat-of-pants” argument it seems that our study demonstrates that the differences between the treatments for iron intake at 7 and 12 months are small and unlikely to be of clinical importance. Although this seems rather unsatisfactory, there doesn’t appear to be an alternative as in this case nothing, as far as we know, is known about what sized difference would have a clinical impact and in turn what range of values would be considered equivalent for the iron intake of children in this particular age group. The confidence intervals suggest that quite large differences are plausible for iron status at 12 months. Again it appears nothing is known about what difference would be regarded as equivalent.

b) Report equivalence testing results with p-values for the three main results: iron intake at 7 and 12 months, and iron status at 12 months. Or alternatively, use the method evaluating if the confidence interval fits within the margin of equivalence [see Hoenig et al. section 4 Equivalence Testing] along with the reported power analysis suggested above, and you can use your already calculated confidence intervals.

We have presented the differences and their confidence for iron intake at 7 and 12 months and iron status at 12 months more explicitly in the Results section and commented on them in the Discussion.

Results (Page 14):
“The differences in iron intake between the BLISS group and the Control group at 7 and 12 months were 0.6 mg/day (95% C: -1.0, 2.3) at 7 months and -0.1 mg/day (-1.6, 1.4) at 12 months. In both cases the differences were small and the confidence intervals exclude clinically interesting differences. The same applies to intakes of the iron absorption modifiers that were measured,..”

Results (Page 16):
“The difference between the BLISS and Control groups for plasma ferritin was -2.6 µg/L (-10.9, 5.8), and not statistically significant, although the lower confidence limit does suggest it is plausible that BLISS infants as a population could have plasma ferritin concentrations that are 11 µg/L lower than those of Control infants. Differences between the groups for the other biochemical indicators of iron

status were small and not statistically significant (all p>0.55) (Table 4)."

Discussion (Pages 19-20):
"Although we did not reach our planned sample size, it is important to note the most extreme difference in plasma ferritin concentration consistent with the data was -10.9 µg/L (i.e. the lower confidence limit for the difference). This suggests that, in response to a BLISS intervention, the Control group's median plasma ferritin concentration might, at most, fall to 18.0 µg/L – a value above the cutoffs usually associated with deficiency (i.e. 12 or 15 µg/L). The data are also consistent with plasma ferritin rising to 34.7 µg/L (applying the upper confidence limit). Similarly, the confidence limits for the differences in dietary iron intake at 7 and 12 months of age suggest that any differences may be too small to be of clinical interest with plausible ranges of -1 to 2.3 at 7 months and -1.6 to 1.4 at 12 months."

2. The results on breast milk intake are confusing. Page 10, lines 50-55: "It was not possible to directly measure breast milk intake so it was assumed to be 750 g per day at 7 months and 448 g per day at 12 months…" In Table 2 total energy and dietary iron are reported for subjects at age 7 and 12 months. However, it cannot be assumed that the breast milk intake is independent of intake from complementary foods. How can you justify your method here? It appears to me that only energy and dietary iron from complementary foods can be assessed properly. Page 14, lines 10-21: Differences in estimated breast milk intake at 7 and 12 months are reported with estimated differences, confidence intervals, and p-values. Why does it make sense to report these statistics, and where did the variability come from if you are assuming all subjects at 7 months have the same breast milk intake? I see that the profile of milk consumers is different in each group (eTable 1), however, reporting the differences on page 14 do not make sense.

The differences arise because not all participants were breastfed at 7 and 12 months (eTable 2); in fact only half of participants in each group were having breast milk as their only infant milk, and 24 to 46% were having no breast milk at all. In addition, there were a number of infants who were mixed formula and breastfed. For these mixed fed infants, their intake of breast milk was estimated as the full estimate (i.e. 750 or 448 g per day) minus the amount of formula given (as described in paragraph 1 on Page 11). As a result, there is variability in estimated breast milk intakes.

Although we agree that estimated breast milk volumes are not ideal, particularly at the individual level, infant milks make an important, and differential, impact on total iron intake at this age. We were able to calculate infant formula intake using the weighed diet records. Infant formula is an important contributor to total iron intake, so we consider that it is important for the reader to have information on total iron intake that includes infant formula. This then requires that we also report a best estimate of breast milk intake. We have added a comment in the limitations paragraph on the use of estimated breast milk volumes to acknowledge this (Page 20):
"Also, estimated breast milk volumes were used. This approach is commonly used when other methods are not feasible[32,40-44] but does mean that we do not have specific intake values for individuals."

3. Iron status is listed as a primary objective in the 29 August 2011 protocol (Lower South Regional Ethics Committee - Protocol). However, iron status is listed as a secondary objective in reference #12 (numbered in the manuscript – Daniels et al 2015). Please explain this inconsistency. This is a major issue because the identification of an outcome as primary or secondary influences how the outcome is treated and the statistical analysis that follows, and so must be treated carefully.

Please see response to reviewer 1, comment number 1.

Minor issues:

1. Page 2, lines 22-24: “The funders provided funds only and played no role in carrying out the study or the analysis or interpretation of results.” Authors should also state who developed the study protocol, especially in regard to developing study aims. Did the funders have input in defining the study aims? Please state here.

The following has been added to the funding statement (Page 2):
“The funders provided funds only and played no role in the study design, development of the study aims, carrying out the study or the analysis, or interpretation of the results.”

2. Page 4, lines 12-14: “This randomized controlled trial included 206 participants assigned to Control (n=101) or BLISS (n=105) groups.” Sample size reported should be number of completers relevant to the outcomes under study here (iron intake and status).

The number of completers relevant to the outcomes presented in this paper have been added to the results section (Page 13), as we felt this was a more appropriate place to explain the sample presented in this paper:
“A total of 214 mother-infant pairs were randomised, of whom eight were excluded after birth (n=5 Control, n=3 BLISS), providing a final sample size of 206 participants (Figure 1). Of these 206 participants, 81 Control and 88 BLISS participants provided data for this secondary analysis (Table 1).”

3. Page 8, lines 39-43: please restate in this section Well-Child care recommendations on breastfeeding and complementary feeding, to contrast with recommendations under BLISS, “to encourage and support exclusive milk feeding (ideally breastfeeding) and delay the introduction of complementary foods until 6 months of age” (page 9, lines 14-18).

Thank you for pointing out that we had not expressed ourselves clearly here. We have edited the paragraph to clarify the difference in the recommendations regarding the timing of complementary food introduction: around 6 months of age for “Well Child” compared with from 6 months of age for BLISS (Pages 8-9):
“‘Well Child Tamariki Ora’ is a nationally funded program to support and educate families with children under 5 years of age. The program recommends exclusive breastfeeding until around 6 months of age with the introduction of complementary foods at around 6 months.[13]
Participants in the BLISS group received routine midwifery and ‘Well Child’ care, and BLISS support and education from before birth (approximately 34-35 weeks gestation) until 9 months of age. The BLISS approach was based on three key principles of BLW: exclusive milk feeding until 6 months of age, infant self-feeding from the start of complementary feeding (i.e. baby-led from 6 months of age), and offering family foods as finger foods so they can be picked up by the infant.”

4. Table 1 (page 26 – 27): The title describes N=162 at 7 months, yet Control N=81 and BLISS n=88. These sample sizes appear inconsistent.

Please see reviewer 1 minor comment number 3 for a response.

5. Table 2 (page 28 – 29): Table footnote (b) Difference adjusted for infant age and sex, and maternal education and parity. How are you adjusting for age if all infants are 7 months, for example? Are you adjusting for age in days, if so, this needs to be explicit. This footnote appears in other tables as well.

The reviewer is correct, we adjusted for infant age by day. All footnotes which include the statement regarding adjustment for age have now been amended:
“Difference adjusted for infant age (by day) and sex, and maternal education and parity”

References

Daniels L, Heath A-LM, Williams SM, et al. Baby-Led Introduction to SolidS (BLISS) study: a randomised controlled trial of a baby-led approach to complementary feeding. BMC Pediatr 2015;15:1–15.

European Medicines Agency. Guideline on adjustment for baseline covariates in clinical trials. 2015 (accessed 12th December 2017). http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2015/03/WC500184923.pdf

Greenland S, Senn SJ, Rothman KJ, Carlin JB, Poole C, Goodman SN, Altman DG. Statistical tests, P values, confidence intervals, and power: a guide to misinterpretations. Eur J Epidemiol 2016;31:337-50. doi: 10.1007/s10654-016-0149-3.

Hoenig JM and Heisey DM. The abuse of power: the pervasive fallacy of power calculations for data analysis. The American Statistician 2001;55:19-24.

Senn SJ, citing Sir David Cox. Power is indeed irrelevant in interpreting completed studies. BMJ 2002;325. doi: https://doi.org/10.1136/bmj.325.7375.1304.

Taylor RW, Williams SM, Fangupo LJ, et al. Effect of a Baby-Led Approach to Complementary Feeding on Infant Growth and Overweight. JAMA Pediatr 2017;171:838-846.

VERSION 2 – REVIEW

REVIEWER	Kathryn Barger Tufts University, USA
REVIEW RETURNED	15-Jan-2018

GENERAL COMMENTS	Comment 1a: Authors should and can compute power for a statistical equivalence test, completely independent of the reported data from this paper so as to avoid any paradoxes. The estimates of variation from the preliminary data, as well as effect size can be used to "convert" the original calculation to the equivalence testing scenario. The following statement is inaccurate, "Similarly, the confidence limits for the differences in dietary iron intake at 7 and 12 months of age suggest that any differences may be too small to be of clinical interest with plausible ranges of -1 to 2.3 at 7 months and -1.6 to 1.4 at 12 months." A large confidence interval, does not mean the effect size is not of clinical importance. Comment 1b: Addressed. Comment 2: Address in the limitations that differences in estimated breast milk intake at 7 and 12 months do not account for any reduced intake when part of the total daily calories are from complementary foods.

	Comment 3: Addressed. Minor issues 1-4: Addressed. Minor issue 5: Change “by day” to “in days”. “By day” implies a factorial design, while “in days” better describes units of measurement.

VERSION 2 – AUTHOR RESPONSE

"Impact of a modified version of Baby-Led Weaning on iron intake and status: a randomised controlled trial" (bmjopen-2017-019036 R1)
Response to reviewers

Reviewer 3 comments:

1a. Authors should and can compute power for a statistical equivalence test, completely independent of the reported data from this paper so as to avoid any paradoxes. The estimates of variation from the preliminary data, as well as effect size can be used to "convert" the original calculation to the equivalence testing scenario.

There are three reasons why we do not feel comfortable adding the proposed power calculation for a statistical equivalence test:
(a) The purpose of a power calculation is to inform the design of a study, in particular the sample size, so a power calculation should be carried out before the data are collected, and be informed by the a priori objectives of the study. Statistical equivalence testing was not an a priori objective of our study.
(b) Equivalence testing would require a definition of equivalence for plasma ferritin in this age group. To the best of our knowledge, no such value for equivalence has been proposed or agreed in the literature.
(c) The 95% CI for the group difference in plasma ferritin is wide and this may indicate that the variation estimate from the preliminary data we used to calculate the power for this test may have been insufficient. Given this, it would not make sense to do post-hoc power calculations using the preliminary data.
It is the case that we did not reach the desired sample size, so it could be argued that a true difference might remain undetected. However, we provide effect size estimates and 95% confidence intervals so that the reader can see the possible effect of the intervention, and we explicitly state this in the Results on page 16:
"The difference between the BLISS and Control groups for plasma ferritin was -2.6 µg/L (-10.9, 5.8), and not statistically significant, although the lower (and upper) confidence limits do not rule out clinically meaningful effects."
We also include a discussion of this in the limitations section.

The following statement is inaccurate, "Similarly, the confidence limits for the differences in dietary iron intake at 7 and 12 months of age suggest that any differences may be too small to be of clinical interest with plausible ranges of -1 to 2.3 at 7 months and -1.6 to 1.4 at 12 months." A large confidence interval, does not mean the effect size is not of clinical importance.

We apologise - the word "Similarly" confuses the statement. We have removed the word "Similarly" from the text on page 20:
"The confidence limits for the differences in dietary iron intake at 7 and 12 months of age suggest that

any differences may be too small to be of clinical interest with plausible ranges of -1 to 2.3 at 7 months and -1.6 to 1.4 at 12 months."

1b. Addressed.

Thank you.

2. Address in the limitations that differences in estimated breast milk intake at 7 and 12 months do not account for any reduced intake when part of the total daily calories are from complementary foods.

We have clarified our comment in the limitations (page 20):
"Also, estimated breast milk volumes were used. This approach is commonly used when other methods are not feasible[32,40-44] but does mean that we do not have specific intake values for individuals. In particular, although the estimated breast milk volumes were determined in infants who were consuming complementary foods (Dewey et al. 1991), we cannot rule out the possibility that BLISS had different effects on the amount of breast milk consumed. However, there was no evidence in the current study that BLISS impacted on the amount of infant formula consumed at 7 and 12 months of age."

3. Addressed.

Thank you.

Minor issues
1-4. Addressed.

Thank you.

5. Change "by day" to "in days". "By day" implies a factorial design, while "in days" better describes units of measurement.

Thank you for the recommendation. All footnotes have now been amended:
"Difference adjusted for infant age (in days) and sex, and maternal education and parity"

References

Dewey KG, Heinig MJ, Nommsen LA, et al. Adequacy of energy intake among breast-fed infants in the DARLING study: relationships to growth velocity, morbidity, and activity levels. J Pediatr 1991;119:538-47.