Author’s response to reviews

**Title:** A systematic review identifying common data items in neonatal trials and assessing their completeness in routinely recorded United Kingdom national neonatal data

**Authors:**

Sena Jawad (s.jawad@imperial.ac.uk)

Neena Modi (n.modi@imperial.ac.uk)

A Prevost (a.prevost@imperial.ac.uk)

Chris Gale (christopher.gale@imperial.ac.uk)

**Version:** 1 **Date:** 28 Sep 2019

**Author’s response to reviews:**

Response to reviewers: A systematic review identifying common data items in neonatal trials and assessing their completeness in routinely recorded United Kingdom national neonatal data

**Editor comments**

Comment 1: The potential of using routine data for neonatal trials to make them more efficient and less costly trials deserves more attention. Please elaborate and better discuss this - is there empirical evidence supporting this claim?

Response: We have added the following text to the introduction to give more detail on trial costs and the evidence for more efficiency using simpler data collection.

“…the median cost of phase III trials of therapeutic agents in 2015-2016 was $19 million and even publicly funded pragmatic neonatal trials cost £1.5-2 million…”

“– mean trial data collection costs using conventional Case Record Forms have been estimated to be €1,135 per participant…”

We have elaborated with examples and references in the sentence: "More efficient collection, for example using electronic case record forms and routinely available clinical data, provide opportunities to reduce costs, and facilitate neonatal trials to improve the limited evidence base upon which much of neonatal care currently relies."

We also added the following text to the introduction with empirical evidence referenced: “An increasing proportion of neonatal Cochrane reviews are inconclusive because of insufficient high-quality data from randomised trials.”

We have added further detail in the introduction supporting the assertion that using routine data for neonatal trials may make them more efficient and less costly: “The effectiveness and efficiency of using routinely recorded clinical data held in the NNRD for data-enabled neonatal trials, are currently being investigated.”
Comment 2: Please carefully address issues related to the completeness of the data and to accuracy (in particular the comments raised by reviewer #2 in this regard). Misclassification of data used for adjustment may introduce different or more bias than missclassified outcome data, this needs to be highlighted and discussed.
Response: Please see our response to comments 47, 48 and 49 raised by reviewer 2.

Comment 3: Please better describe the ascertainment of e.g. infection and more importantly gestational age (GA) as the most commonly reported baseline variable and second most common stratification variable in the clinical trials assessed. Are there tendencies to over or underestimate GA, potential biases related to this?
Response: we have added the following text to the discussion to expand on this important point: “Furthermore, included trials used different approaches to ascertain commonly reported data items, for example the most commonly reported data item – gestational age – may be derived from maternal reported data, ultrasound measurement or clinical evaluation. Data held within the NNRD are extracted from routine clinical information used to inform clinical care, these clinically relevant data may be more appropriate for pragmatic trials than more granular data items reported in trials. Differences between trials and routinely recorded data sources in how data items are ascertained and synthesised have the potential to introduce biases into clinical trials seeking to use such routinely recorded data.”

Comment 4: Are there important baseline information that may not be included in the published trials or in the NNRD, for example because it is too difficult to obtain the data or too sensitive to be captured in the database?
Response: We did not identify any baseline information that are too sensitive to be captured in routinely recorded data held in the NNRD among included trials; this suggests that baseline data as reported in high impact neonatal clinical trials are able to be captured in the database. We have expanded the discussion around differences in ascertainment and data definitions for similar data items in the discussion, and discussed that some data items reported in trials may be more granular than those recorded in routine clinical data. This is described above in response to comment 3 and in the discussion section preceding this: “An additional limitation stems from the fact that some data items collected in clinical trials did align with data items in the NNRD, therefore there may be a loss of information from aggregating several data items into a common data item held by the NNRD to assess data quality.”

Comment 5: Please be more careful with the statements and conclusion on the feasibility of large studies using this approach, as this is not directly supported by the data presented.
Response: We have redrafted the conclusion of the abstract and the conclusion section accordingly, as described more fully in responses to comments 47 and 48 raised by reviewer 2.

Comment 6: Minor: Please report the search strategy separately (not included in the flowchart)
Response: We have removed the search strategy from the flowchart and added it as a supplemental data file 1. This is described in paragraph 2 of the Methods section.

Reviewer 1 comments
The authors describe a study with 2 aims. First in a systematic review of 4 journals, they identified randomized trials with neonates as the target population. From these trials, they
identified a set of common variables used for baseline, stratification, and adjusting outcomes. They then looked at a research registry to report on the completeness of these variables. This effort sets the stage for such data elements to be used both in trial design and in off-setting some of the primary data collection efforts needed in trial conduct.

Comment 7: Introduction: Authors need to make a more compelling case for why this exercise to use real-world data elements is necessary. Comment on costs and efficiency. Can you cite literature and actual dollars. If I do not have to collect 9 data elements as part of a trial…is that going to really be more efficient (how much), more cost effective (how much)…
Response: Please see our response to Comment 1.

Comment 8: Also, are there references you can add from the pragmatic literature that help frame what level of completeness is desired to make a variable useful e.g. is 90% good enough?
Response: The importance of baseline variable completeness depends on its role in the trial. We have added further discussion of this in the fourth paragraph of the discussion: “It is important to note that some NNIRD data items had between 10 and 30% missing data. The implications of such degrees of missingness depend on the role of the data item in the trial, but are likely to lead to a loss of precision. Baseline variables have a role in pre-specified statistical analyses of outcomes in order that treatment effects can be estimated more precisely. Where the baseline is missing, there are methods which do allow incomplete baseline variables to be included without removing the patients with missing baselines, and to achieve some increase in precision. This is relevant to individually randomised trials, whereas incomplete baseline may have a greater impact in trials randomising centre clusters when baseline completeness varies by centre. Baseline variables are also used to describe the trial population, for example to allow readers to judge generalisability, and a high level of baseline completeness may be important for this purpose. Finally, baseline variables are important for subgroup analyses and missing data may limit such analyses. The results presented here will allow the impact that different degrees of missingness have in neonatal trials to be further explored and modelled to better understand which trials are most suitable to use routinely recorded data. The more widespread use of routinely collected data for clinical trials also has the potential to improve the recording of such data”

Comment 9: Discussion needs added commentary on how this study adds significance despite no study using all 14 of most common variables and most variables were not common. This might suggest unique data from primary data collection are actually needed for studies, an idea counter to the suggestion that a common data core will be used and thus drive efficiency in study conduct.
Response: We have discussed this in the discussion as follows: “We recognise that the trials included in the systematic review also reported a wide range of additional non-outcome data items that were not included in the common set identified here. In planning future pragmatic neonatal trials, the completeness and accuracy of additional data items critical to the integrity of a planned trial can be evaluated using the approach applied here. Furthermore, the finding that reported data items were variable even between similar trials (supplemental figure 2) suggests that some reported data items may not have been critical to trial integrity, and that harmonisation of non-outcome data items may improve the consistency and efficiency of future neonatal trials.”
Comment 10: Lastly - Intro framed that using common data core would allow more efficient, cost-effective study conduct, but this point is absent from discussion. Authors need to frame how this would happen, and look at other non-neonatal areas where this has successfully been done, framing their results and suggestions in such context.
Response: We have added the following text to the discussion: “The clinical and economic efficiency of using routinely recorded common data items has been demonstrated by trials that have used common registries such as SWEDHEART (24, 25). Common data items, as identified here and in core outcome sets (26), can be used to ensure existing primary data capture systems such as EPR systems and registries capture appropriate data for trials, and in planning such trials. High accuracy and completeness of data are critical for trials; it may however, not be feasible to evaluate such metrics for all data items within a database or registry – common data items and core outcome sets can be used to target quality assessment of data items most critical to a range of clinical trials. Ongoing data-enabled pilot trials that use routinely recorded data held in the NNRD will provide prospective data regarding the feasibility of such an approach.”

Comment 11: Page 4 Line 12: add ref and more detail around cost.
Response: Please see our response to Comment 1.

Comment 12: Page 4 Line 19: Sentence is redundant.
Response: We have reworded this sentence; please see our response to Comment 1.

Comment 13: Page 4 Line 42-44: This area is redundant to prior.
Response: We have removed the redundant section.

Comment 14: Add sentences or paragraph with references defining pragmatic research, which is where this study's results could be most applicable.
Response: We have added the sentence “These approaches are most likely to be applicable to pragmatic trials” to the second paragraph of the introduction with an additional reference.

Comment 15: Page 4 Line 55. Comment on approach to identify non-outcome related variables. Are you defining the approach of 'how' to define here as well as defining a set for neonates? IF the intent is to highlight an approach (steps to take)...this needs to be defined clearly throughout.
Response: We are not aware of any established method for the identification of common baseline data items; we therefore undertook a systematic review of neonatal trials published in high impact journals. We have explained this in the final paragraph of the introduction: “...as there is no established approach for the identification of common baseline data items we undertook a systematic review to identify baseline data items reported in neonatal trials.”
We have also discussed other published approaches to determining core data items in paragraph 3 of the discussion. We have added text to this paragraph to discuss the strengths and limitations of the approach we used and wider implications:
“The approach that we used was a more limited systematic review of trials published in high impact journals. This approach was chosen a-priori to focus on data-items reported in trials that influence neonatal practice. This was a pragmatic decision and there are limitations to this approach: by limiting our review to general medical journals we may have missed influential trials published in specialty journals, and have not sampled the range of outcomes reported in smaller trials. Furthermore, no approach to date has sought parent or patient views on the
importance of different non-outcome data items; this may be important given the different priorities identified by these groups compared to health professionals and researchers (22). The examples cited here demonstrate the interest in, and potential value of, common sets of non-outcome data items, across different specialties. The development of an established methodological approach, analogous to that developed by the COMET initiative (12) would increase the consistency, robustness and comparability of such endeavours in future.”

Comment 16: Word choice --Could you refer to 'data item' as "variable" throughout?
Response: We accept that this is a personal choice, but would prefer to use data item as we feel this clearer in the context of routinely recorded data. We have changed variable to data item throughout for consistency.

Comment 17: Page 5 Line 52: In checking your online prospective registry, it says inclusion population &lt;34 wks receiving special, high-dependency, or intensive care. Was there a change from what was proposed? Need to add that descriptor here if changed for a reason.
Response: We have clarified the change to the study inclusion criteria in the second paragraph of the methods: “Prior to data extraction we changed the inclusion criteria for studies from less than 34 gestational weeks, as registered, to include trials that enrolled infants of any gestation on the neonatal unit so that the results would be more generalisable to neonatal trials.”

Comment 18: Page 7. Line 7. Aggregating many different types of support and counting presence of any one variable as complete for a very important clinical trial variable like respiratory support is a bit of a leap that should be discussed in limitations.
Response: Addressed in the discussion as follows “There may also be a loss of information by aggregating several data items to assess data quality. Further exploration is needed to understand how to accurately assess the data quality of similar data items.”

Comment 19: Page 7 line 10. 2 errors. We = was. Or = nor…ie ….neither…nor
Response: changed to “was not”

Comment 20: Page 7: Need definition for implausible, how was determination made?
Response: We have clarified this with the additional text: “according to the neonatal data set data dictionary definition”

Comment 21: Page 7 Line 36. Avoid 'nearly' in results…put actual number
Response: “Nearly” replaced with “42”

Comment 22: Page 7 Line 45 - need to clean up stratification variable paragraph…too easy to get lost in numbers in current state. Can you try using the same cadence? 35 trials reported n=x stratification items with y and z being most common. 33 trials reported n=y adjustment items. Then keep same cadence/terminology as baseline data paragraph….of these x# defined as common (&gt;20% of trials) and y# were only present in single studies.
Response: Addressed by rephrasing paragraph: “Sixteen stratification items were reported by 35 trials. Neonatal unit identifier (57%) and gestational age (39%) were the most common items used for stratification during randomisation. Two (13%) of these stratification items were reported by more than 20% of trials and 9 (56%) were reported by 1 study only (supplementary
tables). Twenty-four items were reported by 33 trials to adjust the primary outcome. Of these, 3 (13%) were reported by more than 20% of all trials and 12 (50%) were reported by 1 study only (supplementary tables).”

Comment 23: It would be helpful for reader to walk away from these 2 paragraphs with clear understanding of how many variables meet definition of 'common', broken out by baseline, stratification, and adjusting categories.
Response: We have added the following text to summarise this: “Two (13%) of these stratification items were reported by more than 20% of trials” as well as “Of these, 3 (13%) were reported by more than 20% of all trials”.

Comment 24: Page 8 Line 2. Would be helpful to know how demographics and other characteristics of NNRD population lines up to trials…but this may be too difficult to obtain.
Response: We have referenced descriptive data for infants included in the NNRD in the Data completeness section of the methods.

Response: We changed this to include the following: “Data completeness in the NNRD is 99.9% for gestational age at birth, 99.9% for sex, 100% for birth weight, 99.7% for multiple birth and 100% for respiratory support on day 1”

Comment 26: Page 8 Line 22. Can you put in avg for completeness by term/preterm here?
Response: Mean completeness added as follows “Completeness was higher for all data items for preterm (mean completeness 94.4%) compared to term babies (mean completeness 89.2%) (table 3).”

Comment 27: Page 8 Line 37: Needs to be tempered. I'm not so sure that the assertion is supported by your finding…rather the NNRD may serve as a reliable source for some important variables in neonatal clinical trials.
Response: This has been changed to specify that items with high completeness in NNRD helps support the idea of conducting trials using NNRD as follows: “That a common set of non-outcome data items can be identified across the range of disease areas and interventions found in neonatal clinical trials with a high level of completeness in the NNRD supports the assertion that multiple large, efficient neonatal trials are feasible using the NNRD”

Comment 28: Page 8 Line 43: How would using the common data elements help one determine suitability and feasibility of NNRD for a specific study…more so I think you are suggesting that given the completeness of data element in the NNRD, it is suitable for investigators to use in planning trials?
Response: The word “planning” has been added

Comment 29: Page 8 Line 50: First mention that these are 'simple' trials? Do you mean this?
Response: We have removed the term “simple” as this was not previously defined

Comment 30: Page 8 Line 64: Are researchers really going to have much sway in changing the culture of routinely collected data across multiple countries.
Response: There are ongoing research initiatives aiming to improve the accuracy and completeness of routinely recorded data.

Comment 31: Page 9 Line 5: recorded at baseline? Some of the adjustment variables are not baseline e.g respiratory support.
Response: We have add “or used as explanatory data items” to clarify this.

Comment 32: Page 9 Line 17: 'avoiding error in trials'? what is meant here. Error in design, bias, misinterpretation of results?
Response: We have clarified this by changing to “avoiding misinterpretation of results”.

Comment 33: Page 9 Line 19: second reference to existing methodology for outcomes variables. Might be helpful to compare/contrast that methodology to what you have done in discussion. You do a little bit of this in line 50…but bring them together and add to.
Response: We have expanded on this in the discussion as described in our response to comment 15.

Comment 34: Limitations: would move to last paragraph before conclusion.
Response: We have reordered the discussion so that strengths and limitations are discussed immediately before the conclusion.

Comment 35: Page 9 Line 22: main limitation…this sentence is choppy and needs to be reworked. Also I would not say this is the main limitation, actually seems rather minor that studies may have not reported all variables collected. Main limitation is probably that there is no way in current study to evaluate accuracy of data in addition to completeness.
Response: “Main limitation” changed to “a limitation”. Added “Current studies lack a feasible method to evaluate both data accuracy as well as completeness.”

Comment 36: Page 9 Line 33 - would expound on this limitation. This is actually a limitation to the whole concept of using something like NNRD in lieu of primary data collection…there will frequently be some data elements that are difficult to capture in real-world data.
Response: We

Comment 37: Also in limitations, you need to comment on the aggregation of data elements, which could have biased your results towards completeness.
Response: We have added the following text to the second paragraph of the discussion to address this: “Additionally, some data items collected in clinical trials did align with data items in the NNRD, and there may be a loss of information from aggregating several data items into a common data item held by the NNRD to assess data quality. Data held within the NNRD is extracted from routine clinical information used to inform clinical care, these clinically relevant data may be more appropriate for pragmatic trials than more granular data items reported in trials. Further exploration is needed to understand how to accurately assess and synthesis similar data items.”

Comment 38: Additionally, recognizing that some of the NNRD data elements, may not align exactly with those used in trials.
Response: Please our response to comment 37 above.

Comment 39: Lastly -commentary on completeness is all positive, but somewhere in limitations or discussion, authors should comment on the implications of some of these variables having 10-20% missingness...that is actually a pretty big deal for analyses.
Response: Please see our response to comment 8.

Comment 40: Page 9 line 50: Not sure I follow why doing this for laboratory tests is not relevant to clinical trial conduct, the authors just chose not to look at labs…but those might be very important data elements for stratification and adjusting.
Response: We modified the text to clarify this section as follows: “This study identified a wide range of laboratory tests for feasibility studies. Diagnostic test data were not identified in our systematic review of large neonatal trials as commonly reported non-outcome data items, indicating that such data items are not as relevant to pragmatic neonatal trials that are the focus of this work.”

Comment 41: Page 9 line 52: Not sure I see relevance of Sheehan ref here?
Response: The Sheehan reference has been included as an example of other approaches aimed at identifying common data elements. We have clarified this.

Comment 42: Page 9 line 59: add more on chari comparing what you did.
Response: Please see our response to comment 15

Comment 43: Page 11 Line 35. First mention of accuracy here…which is actually not appropriate as your study did not look at accuracy at all. Remove and discuss in limitations.
Response: We have redrafted the conclusion to remove this mention of accuracy, and have discussed this in the limitations as follows: “Another limitation is that we did not evaluate the accuracy of common non-outcome data-items in the NNRD in this study. Completeness and accuracy are key factors in determining the suitability of using routinely recorded clinical data for clinical trials and should be evaluated for all data items deemed critical to any trial seeking to use such data.”

Comment 44: Page 11 Line 38: Reference to suggestion for lower impact studies seems random and not supported by study.
Response: We have removed this sentence.

Comment 45: Conclusion needs to be crisper. Is conclusion focused on NNRD can be used for trial design only or are you also suggesting that trials should use NNRD for conduct of trial instead of primary data collection.
Response: We have extensively redrafted the conclusion to ensure it is more focused.

Comment 46: Page 11 Line 45: point is not really when planning EMR systems…but rather that care providers/health systems continue to work towards a culture of completeness in primary data collection because the completeness and accuracy of such data are important for research.
Response: We changed the final sentence of the conclusion to reflect this valuable point: “We suggest that when planning primary data collection systems such as EPR systems, registries or
clinical databases, consideration is given to fostering a culture of completeness and ensuring that important items are accurately and completely captured.”

Reviewer #2
The authors, in this study did a systematic review of neonatal clinical trials published in four high impact medical journals over 10 years (2006-2015) and extracted baseline characteristics items, stratification items, and potential confounders-items used to adjust primary outcomes. Then they also examined the availability of the most common items identified (in &gt;20% of the neonatal trials) in the NNRD database of routinely collected data from the National Health Service (NHS) Neonatal Units in England, Wales and Scotland. In the 44 analyzed neonatal clinical trials the authors identified 126 such data items; 14 of those items were reported by more than 20% of analyzed trials and these 14 items were identified in &gt;90% of records in the NNRD database.

Comment 47: In the Discussion section the authors state: (a) "We have identified a common set of data items reported in high impact neonatal trials. That a common set of non-outcome data items can be identified across the range of disease areas and interventions found in neonatal clinical trials supports the assertion that multiple large, efficient neonatal trials are feasible using the NNRD. The common non-outcome data items we identified can be used to assess the suitability and feasibility of using the NNRD and other similar routinely recorded data sources for such trials." (b) Also in their Conclusion section the authors state "High impact neonatal trials report a common set of non-outcome data items in their primary publications. This indicates that large neonatal trials using existing data sources are feasible where such data items are recorded to a high degree of accuracy and completeness." (c) and in the Abstract-conclusion section the authors state: "The efficiency of neonatal clinical trials could be increased by using high quality, routinely recorded EPR data such as that held in the NNRD rather than collecting these items anew": These conclusions about the "feasibility" of large neonatal trials perusing data from the NNRD are not supported from this study's findings and are misleading; the above concluding statements should be deleted. The only conclusion that can be made from this study (and this should be clearly stated) is that "a limited set of 14 baseline non-outcome items (2 of which were also very vague and clinically non-specific) can be found in routinely collected EPR data and could be used to "inform" about the study design and provide some general information about the neonates that could be potentially considered for study eligibility. However, the fact that these specific 14 items (gestational age at birth, sex, birth weight, antenatal steroids, maternal ethnicity, multiple births, mode of delivery, Apgar score at 5 min, maternal age, inborn, "drug treatment" in day 1 and "respiratory support" in day 1) were identified with good completeness in the NNRD, cannot -even remotely- ascertain the feasibility of large neonatal trials. The accuracy and specificity of recording in the NNRD of clinically-relevant information that is needed to identify neonates who fulfill certain study-eligibility/inclusion criteria - and thus assess the feasibility of a clinical trial- were not explored in this study (to support the above conclusions). The authors need to acknowledge that in their study-limitations section.
Response: We acknowledge the reviewer’s comments. We have added additional description of, and reference to, data reporting the accuracy of the common data items identified in this paper. We have extensively redrafted the conclusion section of the abstract to read:
“High impact neonatal clinical trials share common data items. In the United Kingdom, these items can be obtained at a high level of completeness from routinely recorded data held in the
NNRD. The feasibility and efficiency using routinely recorded EPR data such as that held in the NNRD for neonatal clinical trials, rather than collecting these items anew, should be further examined.”

And we have extensively redrafted the first paragraph of the discussion to read:

“We have identified a common set of non-outcome data items reported in high impact neonatal trials. We find that these 14 data items can be obtained from the NNRD with a high level of completeness for most items. The 14 common data items identified here have previously been validated against independently collected trial data where they were shown to be highly accurate and complete in the NNRD. This supports the assertion that non-outcome data held in the NNRD can be used to support large, efficient neonatal trials. We recognise that the trials included in the systematic review also reported a wide range of additional non-outcome data items that were not included in the common set identified here. In planning future pragmatic neonatal trials, the completeness and accuracy of additional data items can be evaluated using the approach applied here where these are critical to the integrity of a planned trial. Furthermore, the finding that reported data items were variable even between similar trials (supplemental figure 2) suggests that many reported data items may not have been critical to trial integrity, and that harmonisation of non-outcome data items may improve the consistency and efficiency of future neonatal trials. The common non-outcome data items we identify here, and their completeness and accuracy in the NNRD, can be used to assess the suitability and feasibility of using the NNRD and other similar routinely recorded data sources for neonatal trials.”

We have redrafted the conclusion to read:

“High impact neonatal trials report a common set of non-outcome data items in their primary publications. In the UK, our study indicates that these core non-outcome data can be obtained from the NNRD; the feasibility and efficiency using routinely recorded EPR data such as that held in the NNRD for neonatal clinical trials, rather than collecting these items anew, should be examined. We suggest that when planning EPR systems, registries or clinical databases”

Comment 48: The authors need to acknowledge also in the study-limitations’ section that even though certain clinical items can be identified to be recorded in routinely collected data in EPRs, this does not provide any ascertainment that these items were also recorded correctly and accurately.

Response: We have added for following text in the third paragraph of the discussion section to make this limitation clearer: “Another limitation is that we did not evaluate the accuracy of common non-outcome data-items in the NNRD in this study. Completeness and accuracy are key factors in determining the suitability of using routinely recorded clinical data for clinical trials and should be evaluated for all data items deemed critical to any trial seeking to use such data.”

Comment 49: The clinical usefulness/relevance of several of those recorded items in EPR is unclear particularly when information in routinely collected data is recorded under very non-specific terms. For example: a) a recorded item "respiratory support in day 1" is very unclear what it means (just some O2 supplementation via nasal canulla for a short time or intubation with high respiratory support requirements). b) The same also for the item "infection"; does it mean "culture positive-confirmed -infection"? "possible infection"? or "rule out infection"? Moreover, it does not provide any information about the severity of the infection. c) Also items like "drug treatment during the first 24 hrs"; what does this mean? (how many drugs? which drugs?, what doses?). This should also be acknowledged in the study-limitations section.
Response: Different trials defined individual data items differently. We have described this in the methods section as follows: “A comprehensive list of reported data items and frequencies was extracted. Items were combined where appropriate, for example administration of different medications was combined into the item “medications”.”

We have added the following text to the discussion section to highlight this potential limitation: “Additionally, some data items collected in clinical trials did align with data items in the NNRD, and there may be a loss of information from aggregating several data items into a common data item held by the NNRD to assess data quality.”

Comment 50: Re Table 3: Unclear item-terms are used and should have additional annotation provided for the EPR fields that are covered under these items: (a) drug treatment in first day of life; (2) inborn (born in hospital?)
Response: We have added this additional information to the table.

Comment 51: Re Supplementary Table 1: i. The maternal clinical and maternal socioeconomic baseline characteristics should be reported separately from the neonatal baseline clinical and neonatal baseline laboratory/imaging characteristics.
Response: We have reported these data items as they were reported in the reviewed trials, rather than introducing additional categorisation we have ordered them by the proportion of all included studies that reported them as we feel this is most relevant to this paper.

Comment 52: Re Supplementary Table 1: ii. For some of the listed baseline characteristics it is unclear what exactly they mean: e.g. Clinical complications (maternal or neonatal?); Diagnostic group (?); Fluid or normally sterile body fluid (?); infection (maternal or neonatal?).
Response: These data items are included as they were reported in the reviewed trials and refer to infant characteristics unless otherwise stated. We have clarified this in the table.

Comment 53: Re Supplementary Table 1: iii. Some of the recorded items could have been grouped together to make more clinically meaningful item-categories (e.g. for the surgery related items: surgical stress, surgical procedures, bowel perforation of definite NEC). It is also unclear why some of them were listed separately, e.g.: umbilical cord blood tests, umbilical cord hemoglobin, umbilical arterial pH.
Response: Please see our response to comment 51.

Comment 54: Re Supplementary Table 2 Please expand the heading for some baseline items: e.g. Respiratory (should be respiratory, support in the first 24 hrs, correct?); Inborn (?)
Response: These data items are included as they were reported in the reviewed trials. We have clarified the respiratory item.