Reviewer's report

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

Version: 0 Date: 20 Jul 2019

Reviewer: Despina Contopoulos-Ioannidis

Reviewer's report:

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 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 >90% of records in the NNRD database.

1. General comment: Although one might find of interest the first part of this study (the mapping of the types of baseline items commonly reported in neonatal clinical trials), I have serious reservations about the clinical relevance and the conclusions made in the second part of this study. 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.

2. 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.

3. 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.

4. Additional minor comments:
   a. 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?)
   b. 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.
      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?).
      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
   c. Re Supplementary Table 2:
      i. Please expand the heading for some baseline items: e.g. Respiratory (should be respiratory, support in the first 24 hrs, correct?); Inborn (?)
Level of interest
Please indicate how interesting you found the manuscript:

An article of limited interest

Quality of written English
Please indicate the quality of language in the manuscript:

Not suitable for publication unless extensively edited

Quality of figures

All images and figures within the manuscript should be genuine i.e. without evidence of manipulation. No specific feature within an image may be enhanced, obscured, moved, removed, or introduced. If you have concerns about the veracity of the figures you should choose the first option below.

Statistical review
Is it essential that this manuscript is seen by an expert statistician? If so, please give your reasons in your report.

Declaration of competing interests
Please complete a declaration of competing interests, considering the following questions:

1. Have you in the past five years received reimbursements, fees, funding, or salary from an organisation that may in any way gain or lose financially from the publication of this manuscript, either now or in the future?

2. Do you hold any stocks or shares in an organisation that may in any way gain or lose financially from the publication of this manuscript, either now or in the future?

3. Do you hold or are you currently applying for any patents relating to the content of the manuscript?

4. Have you received reimbursements, fees, funding, or salary from an organization that holds or has applied for patents relating to the content of the manuscript?

5. Do you have any other financial competing interests?

6. Do you have any non-financial competing interests in relation to this paper?

If you can answer no to all of the above, write 'I declare that I have no competing interests' below. If your reply is yes to any, please give details below.
No Conflicts of Interest

I agree to the open peer review policy of the journal. I understand that my name will be included on my report to the authors and, if the manuscript is accepted for publication, my named report including any attachments I upload will be posted on the website along with the authors' responses. I agree for my report to be made available under an Open Access Creative Commons CC-BY license (http://creativecommons.org/licenses/by/4.0/). I understand that any comments which I do not wish to be included in my named report can be included as confidential comments to the editors, which will not be published.

I agree to the open peer review policy of the journal

Were you mentored through this peer review?

No