PEER REVIEW HISTORY

BMJ Paediatrics Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below.

ARTICLE DETAILS

TITLE (PROVISIONAL)	Development of a core outcome set for clinical trials in childhood constipation: a study using a Delphi technique
AUTHORS	Steutel, Nina; Kuizenga-Wessel, Sophie; Benninga, Marc; Devreker, Thierry; Scarpato, Elena; Staiano, Annamaria; Szajewska, Hania; Vandenplas, Yvan; Tabbers, Merit

VERSION 1 - REVIEW

REVIEWER	Leoni Maffei, Helga Verena Retired Professor of Pediatric Gastroenterology, Department of Pediatrics, Botucatu Medical School - São Paulo State University (UNESP). Botucatu, SP, Brazil Clinical Researcher, since 2012, at the Children's Hospital 'Menino Jesus'. São Paulo, SP – Brazil Competing interests: None declared
REVIEW RETURNED	03-May-2017

GENERAL COMMENTS	This is an extremely important study in which the Delphi technique was applied to childhood functional constipation (FC). There are some biases, most of which were discussed by the authors, but in addition, some small corrections / explanations are needed before publication, in order to make it a little bit more 'reader friendly'. Title: Perhaps the title could be "Development of a core outcome set for clinical trials in child-hood constipation: a study using a Delphi technique" Abstract: - Design and setting: 'primary, secondary and tertiary settings' induces a false impression at the outset, whereas in the discussion, at the very end, it gets clear that tertiary is the predominant setting - the first step in the Abstract does not conform to step 1 in the text. Thus, the numbers of the steps must be carefully reviewed in order to be concordant. - Key words: why not include parents reported outcomes? Methods: A very meticulous description of employed methods is presented. However, there are some questions / suggestions: - were the 143 HCPs of step 2 (of the text) somehow selected? Which are the characteristics of the 34 who did not respond? (Most of the 109 respondents are pediatric gastroenterologists and – should the information be available - it would be interesting to know whether there are differences between answers of HCPs/physicians who follow the Rome criteria vs. those who do not). - the predefined domains presented in the text and in Table 2 must

	have the same denominations and appear explicitly in the Table. - the participation of HCPs/physicians in the various steps is a little bit confusing. It is not clear whether the physicians of step 3b (of the text) are partially the same who participated in step 2 (of the text); also, it is not clear whether they attended in primary, secondary or tertiary settings Some of these questions show up only in the discussion. On the other hand, the participating adolescents ≥12y are those being attended in a single tertiary setting; this latter bias has to be discussed. - children, whose parents answered about outcomes in step 2 (of the text), were selected for FC according to Rome III; this must appear in METHODS, instead of in the RESULTS section. - side effects (of treatment?) must be defined. Results: - why are only 3 outcomes presented in Table 1, instead of 5 with ≥ 10% answers, as shown in Table 2 of supplementary files? - straining and pain at defecation can occur separately. Therefore, the suggestion is to put them together (straining and/or pain at defecation), instead of disregard straining. - Fig 2a or the corresponding text could be eliminated, they are redundant. Discussion: The following aspects merit discussion - Outcome should rely on which symptoms children present before treatment; in this sense, 8% of blood in stools is an important indicator not to be overlooked. - items of the Rome criteria, like 'withholding impression/behaviour' and 'large diameter stools that may obstruct the toilet', were cited by only 2% of HCPs and by any one, respectively. - Table 1: it calls attention that defecation frequency was the most frequent outcome reported by HCPs in 4/4 possibilities, whereas among parents and adolescents it only appears as the most frequent outcome in 1/6 possibilities. - Table 3a: 18-22% HCPs rated faecal incontinence for 0-1year infants as an outcome measure; this is a very interesting aspect and should be approached References: Rome III for children <4y has to be cited

REVIEWER	Saxena, Romit Great Ormond street Hospital for children, London Competing interests: None
REVIEW RETURNED	20-May-2017

GENERAL COMMENTS	1. Good study , but it has limited external validity to developing countries, as the correspondents, parents and children are from developed world predominantly . The eating habits and lifestyles in some developing countries, which in turn effects the causation of constipation , is not truly reflected, as they may have ranked the symptomatology differently . Could be mentioned as a shortcoming. 2. Functional diseases are usually effected by the socio-cultural backgrounds of the children, parents, birth order etc. And having a heterogeneous group , interferes with the interpretation of the same. Were the necessary adjustments to sample size done, to take this into account. 3. With regards to the statistical analysis, the inter-rater reliability and variability , is always a concern, in such studies. What

	statistical test were chosen for the same. Was Kappa analysis(Fleiss kappa)/interclass correlation used, to represent the inter-rater reliability(using HCP, parent , children as one group each) 4. To assess for internal consistency , since ranked measures were used to come towards COS, would use of cronbach's alpha be relevant , to assess for internal consistency among the various parameters in the COS. Since there would be overlap between the parameters used. 5. Since the parameters in the outcomes set, are subjective as is true for most functional disorders, would it help to objectify the same, as painful defecation (likert scale), stool consistency (Bristol chart).As parental/child's interpretation of stool consistency among other things, would not be uniform, unlike in conditions like diarrhoea. Do the authors feel , that may help in improving the external validity of the study

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1

Comments to the Author
This is an extremely important study in which the Delphi technique was applied to childhood functional constipation (FC). There are some biases, most of which were discussed by the authors, but in addition, some small corrections / explanations are needed before publication, in order to make it a little bit more 'reader friendly'.

Title: Perhaps the title could be "Development of a core outcome set for clinical trials in child-hood constipation: a study using a Delphi technique"

→ Thank you for this suggestion, we have altered the title accordingly.

Abstract:
- Design and setting: 'primary, secondary and tertiary settings' induces a false impression at the outset, whereas in the discussion, at the very end, it gets clear that tertiary is the predominant setting

→ After carefully discussing your remark in our group, we have concluded to leave the setting unaltered for two reasons:

1. **'Reducing' the setting to 'predominantly tertiary setting' would not do justice to the way healthcare is organised in several of the participating countries. In these countries the healthcare system is organised in such a way that (specialised) paediatricians represent primary care (and secondary) as well.**
2. **We prefer to be consistent with our previously published study (Steutel et al, Developing a core outcome set for infant colic for primary, secondary and tertiary care settings: a prospective study. BMJ Open, 2017). This study had the same design.**

- the first step in the Abstract does not conform to step 1 in the text. Thus, the numbers of the steps must be carefully reviewed in order to be concordant.

→ Thank you for pointing this out. We have adjusted the text in the abstract, please see page 2.

- Key words: why not include parents reported outcomes?

→ Thank you for this suggestion, we have added this as a key word, please see page 3.

Methods: A very meticulous description of employed methods is presented. However, there are some questions / suggestions:
- were the 143 HCPs of step 2 (of the text) somehow selected?

→ To clarify how participants were included we have made the following alteration to the text, please see page 7(clean copy): "Participants were randomly approached and asked to list up to five harmful/ beneficial treatment outcomes."

Which are the characteristics of the 34 who did not respond?
(Most of the 109 respondents are pediatric gastroenterologists and –should the information be available - it would be interesting to know whether there are differences between answers of HCPs/physicians who follow the Rome criteria vs. those who do not).

→ We agree it would be interesting to know if there are differences between respondents and non-respondents. However, we collected the data in the second questionnaire (rating and prioritising) anonymously, therefore we are unable to analyse any differences between responders and non-responders. We have added the following sentence to the limitations section of the Discussion (please see page 18, clean copy): "Furthermore, since answers were collected anonymously in both Delphi rounds, we were unable to assess any potential differences between respondents and non-respondents."

- the predefined domains presented in the text and in Table 2 must have the same denominations and appear explicitly in the Table.

→ Thank you for this suggestion, we have added the corresponding domains to this table. We also decided to add the domains to the other tables.
Please note that upon editorial request the numbers of the tables were altered.

- the participation of HCPs/physicians in the various steps is a little bit confusing. It is not clear whether the physicians of step 3b (of the text) are partially the same who participated in step 2 (of the text); also, it is not clear whether they attended in primary, secondary or tertiary settings. Some of these questions show up only in the discussion.

→ In order to clarify the flow of participating HCPs, we have made the following alterations to the Methods section, please see page 8: "Shortlists for HCPs were sent to those that participated in the first survey and agreed to participate in the second survey as well."
→ Since we collected data in the second survey anonymously, we are unable to give any details about the work setting of the respondents.

On the other hand, the participating adolescents ≥12y are those being attended in a single tertiary setting; this latter bias has to be discussed.

→ As mentioned in the Methods section, this was considered to be a pilot study. We have added the following sentence to the limitations section of the Discussion (see page 18): "Also, patients ≥ 12 years were recruited in a tertiary care setting as a pilot study. Only recruiting patients in such a specialised setting might have led to biased results. "

- children, whose parents answered about outcomes in step 2 (of the text), were selected for FC according to Rome III; this must appear in METHODS, instead of in the RESULTS section.

→ Thank you for pointing this out, we have added this to the Methods section (please see page 7).

- side effects (of treatment?) must be defined.

→ We agree that this needed to be specified, we have therefore altered this into side-effects *of treatment*.

Results:
- why are only 3 outcomes presented in Table 1, instead of 5 with ≥ 10% answers, as shown in Table 2 of supplementary files?

→ Thank you for your question. If we would include all outcomes mentioned by > 10% of respondents we feel this table would be redundant since this information is already depicted in Tables 4 – 6 (all outcomes mentioned by ≥ 10% of respondents were included in the shortlists). Therefore we chose to present the 3 outcomes which are mentioned most often by respondents.

- straining and pain at defecation can occur separately. Therefore, the suggestion is to put them together (straining and/or pain at defecation), instead of disregard straining.

→ Thank you for your suggestion. Although 'straining' was included in the draft COS, it was excluded in the consensus meeting of the Working Group because it was mentioned in only 1/10 shortlists. Except for school attendance (2/10 shortlists) all remaining core outcomes appeared in 6 – 10/10 shortlists. An explanation for the fact that 'straining' was excluded in this final step might be the commonly held dogma that 'straining' might occur together with 'painful defecation' as described in the Rome IV criteria (Benninga et al and Hyams et al, Gastroenterology, 2016).

- Fig 2a or the corresponding text could be eliminated, they are redundant.

→ Thank you for pointing this out, we have eliminated figure 2a.

Discussion: The following aspects merit discussion
- Outcome should rely on which symptoms children present before treatment; in this sense, 8% of blood in stools is an important indicator not to be overlooked.

→ As mentioned in the Discussion, this core outcome set should function as a minimum of outcomes to be measured. Researchers are of course free to measure more outcomes in their study.

- items of the Rome criteria, like 'withholding impression/behaviour' and 'large diameter stools that may obstruct the toilet', were cited by only 2% of HCPs and by any one, respectively.

→ We agree that it is an interesting finding that these Rome criteria are mentioned as important outcome measures by 2% and 0% of the HCPs respectively. This might be explained by the fact that we did not ask our respondents which criteria they use to diagnose functional constipation, but we asked them to mention the 5 outcome measures that they find most important during the treatment. Apparently these 2 specific Rome criteria were considered to be less important than others.
We have added the following sentence to the Discussion, please see page 17: "Interestingly, most Rome IV-criteria are represented in our COS, except for 'presence of a large faecal mass in the rectum', 'retentive posturing' and 'large diameter stools' which were mentioned by 3%, 2% and 0% of HCPs respectively."

- Table 1: it calls attention that defecation frequency was the most frequent outcome reported by HCPs in 4/4 possibilities, whereas among parents and adolescents it only appears as the most frequent outcome in 1/6 possibilities.

→ This is certainly an interesting observation and we have added the following to the Discussion, please see page 15: "….depending on the age group, 46 – 55% of HCPs found this the most important treatment outcome, whereas parents and patients mentioned this less often as an important outcome."

- Table 3a: 18-22% HCPs rated faecal incontinence for 0-1year infants as an outcome measure; this is a very interesting aspect and should be approached

→ This is indeed a peculiar finding considering the age of the infants. We can think of two potential explanations for this. First, 'faecal incontinence' was one of the Rome III criteria for children aged 0 – 4 years, referring to 'faecal incontinence' in combination with 'large diameter

stools', which can occur in not toilet trained children as well. Since this is very difficult to assess in infants and toddlers wearing diapers, this Rome criterion was altered when the Rome IV-criteria were developed. Now, 'faecal incontinence' is only used as an additional criterion to diagnose FC in toilet trained children.
The Rome IV criteria were not published yet when we conducted our survey, which might explain why respondents mentioned 'faecal incontinence' in infants as well.
Another explanation might be that they did not read the age group properly. Please see page 17-18 of the Discussion.

References: Rome III for children <4y has to be cited
→ **We agree that the Rome criteria needed to be cited for both age groups. Therefore, we have added both the Rome III and Rome IV criteria for neonates and toddlers.**

Reviewer: 2

Comments to the Author
1. Good study , but it has limited external validity to developing countries, as the correspondents, parents and children are from developed world predominantly . The eating habits and lifestyles in some developing countries, which in turn effects the causation of constipation , is not truly reflected, as they may have ranked the symptomatology differently . Could be mentioned as a shortcoming.

→ **Thank you for this suggestion. We have added the following sentence to the limitations section of the Discussion (please see page 18, clean copy): "Although we included participants from around the globe, developing countries were underrepresented which may limit the external validity of this COS in a developing country setting."**

2. Functional diseases are usually effected by the socio-cultural backgrounds of the children, parents, birth order etc. And having a heterogeneous group , interferes with the interpretation of the same. Were the necessary adjustments to sample size done, to take this into account.

→ **We did not perform a formal sample size adjustment but we made sure stakeholders represented a broad range of backgrounds. As depicted in Table 2, HCPs originated from 28 countries across 5 continents. Parents were recruited in 4 different countries. Physicians recruiting parents and patients were instructed to recruit a heterogeneous group of respondents that would represent their patient population.**

3. With regards to the statistical analysis, the inter-rater reliability and variability , is always a concern, in such studies. What statistical test were chosen for the same. Was Kappa analysis(Fleiss kappa)/interclass correlation used, to represent the inter-rater reliability(using HCP, parent , children as one group each)

→ **Thank you for raising this point. However, we did not aim to assess the degree of consensus among the different groups of respondents and therefore only used descriptive analysis to describe our results.**

4. To assess for internal consistency , since ranked measures were used to come towards COS, would use of cronbach's alpha be relevant , to assess for internal consistency among the various parameters in the COS. Since there would be overlap between the parameters used.

→ **We did not assess internal consistency for the following reason: Ranked measures were purely used to assemble a top 5 of outcome measures. We asked respondents which outcome measure they found most important in the treatment of functional constipation. We did not ask them which outcome measure best measures the presence of (i.e. diagnoses) functional constipation. In the latter case, determining the internal consistency would be relevant.**

→ Additionally, except for quality of life which is likely to depend on the other outcome measures in the COS, we do not expect there is much overlap between the other core outcome measures.

5. Since the parameters in the outcomes set, are subjective as is true for most functional disorders, would it help to objectify the same, as painful defecation (likert scale), stool consistency (Bristol chart).As parental/child's interpretation of stool consistency among other things, would not be uniform, unlike in conditions like diarrhoea. Do the authors feel , that may help in improving the external validity of the study

→ Thank you for your suggestion. We will take this issue into account in our next step. We will search for validated measurement instruments to measure the core outcomes in the core outcome set.