Author's response to reviews

Title: Guidelines on acute gastroenteritis in children: a critical appraisal of their quality and generalizability in primary care.

Authors:

Marjolein Y Berger (m.y.berger@med.umcg.nl)
José van den Berg (josevandenberg@yahoo.com)

Version: 2 Date: 3 September 2011

Author's response to reviews: see over
Dear Editor,

Thank you for your reply and for the opportunity provided to strengthen our paper. Below please find our point-by-point reply to the issues raised by the reviewers. The changes have been marked in the manuscript.

Reviewer: Joseph Watine

Reviewer's report:
A systematic review is a literature review focused on a research question that identifies, appraises, selects and synthesizes all the evidence relevant to that question. A systematic review aims to provide an exhaustive summary of the literature relevant to that question. This work can therefore be considered as a systematic review. However, unless I misunderstood what the authors did, I think that this title: "Guidelines on acute gastroenteritis in children lack of applicability in general practice: a critical appraisal" would even better reflect what this work is about.

Reply: Thank you for this suggestion which reflects the opinion of reviewer 2. We changed the title into:

“Guidelines on acute gastroenteritis in children: a critical appraisal of their quality and generalizability in primary care”

The authors write "An up-to-date high-quality guideline generally means that potential biases of guideline development have been addressed, and that the recommendations are valid, practical and evidence based [8-10]". This is not clear enough. What does valid mean here? What sort of evidence are the authors referring to? Is it the evidence about benefits and harms? Recommendations also need to consider the evidence about patients values, and also about justice issues (mostly resource use) [see: Watine J. What sort of bioethical values are the evidence-based medicine and the GRADE approaches willing to deal with? J Med Ethics 2011 Mar;37(3):184-6].

Reply: We agree that, although terms as valid, practical and evidence based are frequently used and explained in the literature on evidence based medicine, these terms should be explicitly explained in the context of guideline recommendations. Based on the references mentioned by the reviewer we adapted the text into:

“An up-to-date high-quality guideline generally means that potential biases of guideline development have been addressed, and that the recommendations are practicable, reflect the balance between desirable and undesirable effects, incorporate quality of evidence, and consider variability in values of benefits, risks, inconveniences, and costs [10-12]”.

Regarding justice by the way, maybe the authors could also explain why they have excluded guidelines originating from low income countries.

Reply: Adherence to the guidelines for management of gastroenteritis in low income countries is good, in contrast to adherence in the high income countries. It should not surprise this reviewer, that heterogeneity in
practicability and values of benefits and risks between low and high income countries will be large and that therefore, recommendations for low income countries may differ from that for high income countries. We intended to investigate reasons for non-adherence in high income countries. You can’t explain non-adherence in high income countries by evaluating recommendations for low income countries.

As the authors critically appraised guidelines systematically, I am also surprised that they do not quote "Qual Saf Health Care. 2010 Dec;19(6):e58. The quality of clinical practice guidelines over the last two decades: a systematic review of guideline appraisal studies. Alonso-Coello P, Irfan A, Solà I, Gich I, Delgado-Noguera M, Rigau D, Tort S, Bonfill X, Burgers J, Schunemann H", which is probably the most comprehensive systematic review available on the subject, and which also shows that lack of applicability is generally the main shortcoming of guidelines. One of the authors' finding is therefore not completely original, and this should be acknowledged, and discussed.

Reply: we systematically searched for guidelines for acute gastroenteritis in children. We did not systematically searched the literature for articles on the evaluation of the quality of guidelines. Nevertheless we are grateful for mentioning this paper and have included it to the discussion section:

“Lowest scores were obtained for the items ‘stakeholder involvement’ and ‘applicability’. This is in accordance with the findings of Alonso-Coello et al [59], who performed a systematic review of guidelines in order to evaluate the quality of guidelines over time. They found that ‘stakeholder involvement’ and ‘applicability’, were two of the lowest scoring items of guidelines in general, and that little improvement was seen over two decades up to 2007. In contrast with their findings scores for ‘rigour of development’ were high for four of the guidelines we evaluated. Two of these were published after 2007(37,33). Alonso-Coello et al. argue, that this domain may be the strongest indicator for quality, therefore our finding might indicate that some improvement on guideline development can be seen. Further improvement on ‘stakeholder involvement’ and ‘applicability’ as well as ‘rigour of development’ might make guidelines for management of acute gastroenteritis in children, more trustworthy to the user, and could improve guideline adherence.”

I am not sure to understand these sentences: "For all consistent recommendations based on consensus, a GP was present in guideline development. They are therefore generalizable to patients in general practice". Could the authors clarify what they mean? Do they believe that the simple fact that one GP is present guaranties the generalizability to patients in general practice?

Reply: If a recommendation was based on evidence from published studies generalizability to general practice was based on whether the setting in which the study was performed reflected the setting of general practice. If a recommendation was based on consensus because evidence was lacking, the presence of a GP gave the best available indication that applicability, and balance between benefit and harm had been evaluated from the perspective of the GP. We added this remark as a limitation of the review in the discussion section.
Could the authors also explain why their "Judgment on applicability in daily practice was influenced by the items ‘clarity and presentation’, and ‘editorial independence’"… whereas "Low scores on the item... ‘applicability’ did not influence this outcome". This does not sound logical at all to me.

Reply: If all guidelines score low on one item, the item has no discriminative power anymore. Given the reference of Alonso-Coello et al and the comment of reviewer 2 we adjusted this paragraph as indicated above.

It is also not clear to me why "Uniformity of LofE systems is needed to make LofE easy to use for guideline-users and -producers". Wouldn't having a good LofE system be more important than having an uniform LofE system? The same applies to this sentence in the conclusion: "Finally, consensus is needed on the grading systems currently used to assess the level of evidence of research". By the way, have the authors any idea about what is a good LofE system (as opposed to a not so good one)? Maybe they could propose a best candidate? Maybe the GRADE system which is being increasingly adopted worldwide?

By the way, I advise the authors to read this paper, authored by the GRADE working group: http://www.ncbi.nlm.nih.gov/pubmed/18467413, as well as a few others which are freely available on the Internet, where they could find a number of points that might inspire and enrich their discussion a little bit deeper (see other reference to GRADE above).

Reply: We are grateful for mentioning this paper and have used it to enrich our discussion.

In the conclusion, the author also write "For 43% of the recommendations there is a lack of evidence". Did the authors rather mean "43% of the recommendations are not based on LofE"?

Reply: We clarified the sentence in the text. "In 43% the quality of the evidence could not be interpreted because no level of evidence was stated"

**Reviewer:** Athina Tatsioni  
**Reviewer's report:**  
Van den Berg and Berger have conducted an interesting systematic review on the quality and applicability of guidelines on the diagnosis and management of acute gastroenteritis in children. The paper revealed important issues related to the development, and the generalizability of these guidelines in general practice. In addition, it stresses that guidelines often did not corroborate on their recommendations. The article is well written; however, additional clarifications would be requested on certain points.

**Major Compulsory Revisions**  
**Title**  
1) Page 1, “Guidelines on acute gastroenteritis in children: do they apply in..."
general practice? Systematic review”. The authors may consider rephrasing the title so that they capture the main purpose of their study. Based on the submitted article, the authors assessed both quality and applicability of the guidelines.

Reply: We changed the title into:
‘Guidelines on acute gastroenteritis in children: a critical appraisal of their quality and applicability in primary care’

Background

2) Page 3, “In high-income countries, implementation of oral rehydration therapy (ORT) has not been as successful as in low-income countries.” The authors may consider adding a reference if available.

Reply: We added a reference

3) Page 4, “The present study identifies international practice guidelines for the management of acute diarrhea in children in developed countries, assesses their quality, investigates the recommendations, evaluates the level of evidence, as well as the generalizability of these recommendations to general practice, where most children with AGE initially present.” The authors may consider rephrase this sentence so that they first report what is the main purpose of their study. In a following sentence they may add secondary purposes such as whether guidelines corroborated with each other on the recommendations they suggested. “Identify the guidelines” is probably part of the methodological process of the study and may not be included as one of its aims.

Reply: Thank you for this suggestion. We changed the text as follows:
‘The aim of the present study is to assess the quality of international practice guidelines for the management of acute diarrhea in children in high income countries with the Appraisal of Guidelines, Research and Evaluation (AGREE) instrument. We aim to investigate the consistency of the recommendations, evaluate the quality of evidence and the practicability of these recommendations to general practice, where most children with AGE initially present’.

Methods

Identification of guidelines

4) Page 5, “…searches were made in websites of professional societies in the clinical fields of paediatrics, gastroenterology and general practice, and websites of institutes involved in guideline development.” Did the authors use certain criteria to select the specific societies and institutions? If yes, authors may report these criteria.

Reply: we more clearly described this selection criterium: ‘…..articles were selected…… were produced by a professional organisation (i.e. an organisation that is the official representative of an important user group (paediatricians, gastro-enterologists or general practitioners)’.

Selection of guidelines

5) Page 5, “Guidelines were excluded if they were: … a systematic review or not an original publication,….” Could authors clarify more on that?

Reply: adding these criteria is not useful because we gave a clear definition of a guideline that precludes systematic reviews and non-original publications. We therefore deleted these criteria.
Consistency of recommendations

6) Is the use of the 50% cut-off for consistency arbitrary? Authors may also use the percentage of the studies that corroborate for each recommendation as a continuous variable. This might give more information
    Reply: The 50% cut-off point for consistency is indeed arbitrary. More information upon the number and the percentage of studies that corroborate for each recommendation, however, is given in tables 3 and 5.

Results

7) Page 9, “The three guidelines that would not be recommended by the reviewers scored below 40% on all items, except clarity and presentation.” It is not clear in the Methods whether the authors would intend to recommend or not a guideline for use. If this had been part of the initial protocol, the criteria for recommendation should have been clearly reported in Methods
    Reply: Recommending a guideline or not is part of the AGREE instrument. We agree with the reviewer that this was not explained well enough in the methods. We added the following paragraph to the Methods:
    ‘The AGREE instrument contains 23 items grouped in six domains - 1/ scope and purpose, 2/stakeholder involvement 3/ rigour of development; 4/ clarity and presentation; 5/applicability 6/editorial independence - and one overall assessment item, judging whether the guideline ought to be recommended for its use in clinical practice. To evaluate each item within the domains a four-point Likert scale is used, ranging from strongly disagree to strongly agree (1 to 4). For the overall judgment a three-point scale is used ranging from not recommended to strongly recommended’.

8) Page 10, “Table 3 presents recommendations on diagnosis (signs and symptoms) and therapy for dehydration. The consistency of each recommendation is stated, as well as the setting of the research that was used to form the recommendation.” The authors may consider deleting this paragraph
    Reply: we think some explanation about the contents of table 3 is necessary.

Discussion

9) Page 15, “Based on predefined quality criteria...” Which are the criteria the authors refer to?
    Reply: The criteria are now more clearly described in the methods (see above)

Quality of the guidelines

10) Most of the information included in this paragraph has already been stated in the Results. Authors may comment on the possibility that the difference in the year of publication for the guidelines might have influence their applicability in daily practice.
    Reply: we revised this paragraph and included a comment on change of quality over time. (see also reviewer 1)

Levels of evidence

11) Authors may comment on specific differences that ranking systems used for the classification of the LoF may have so that they can give an idea of the variability
12) Limitations of the study: the authors may consider reporting potential limitations of their study in one paragraph before the final conclusion.

Reply: Thank you for this suggestion. We added the following to the discussion section:

‘Limitations of the review

Appraisal of quality was performed by two reviewers, where preferably four appraisers are needed. Because we specifically wanted to review the guidelines from the perspective from the GP and both appraisers were GPs, and because the ranges of scores were small, we feel another two GPs would not give more precise or valid estimates of quality.

Even though a thorough search of possible guidelines was performed, the possibility remains that we missed a guideline. We assumed that if a recommendation was based on consensus, because evidence was lacking, the presence of a GP gave the best available indication that applicability, and balance between benefit and harm had been evaluated from the perspective of the GP. We realise that this assumption is disputable.’