PEER REVIEW HISTORY

BMJ 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)

A consensus definition of a radiologically healed fistula on MRI in perianal Crohn's disease: An international Delphi study.

Authors

Anand, Easan; Devi, Jalpa; Joshi, Shivani; Antoniou, Anna; Doering, Michelle; Stoker, Jaap; Lung, Phillip; Hart, Ailsa L.; Ballard, David H; Deepak, Parakkal; Tozer, Phil

---

VERSION 1 - REVIEW

Reviewer	**1**
Name	**Nakase, Hiroshi**
Affiliation	**The Japanese Society of Gastroenterological Surgery**
Date	**12-Jul-2024**
COI	**None**

This paper is a protocol for developing a consensus definition for assessing perianal fistula healing in patients with Crohn's disease by MRI. The content is acceptable. It would be interesting to know the consensus results using the Delphi method.

Reviewer	**2**
Name	**West, Charlie**
Affiliation	
Date	**15-Jul-2024**
COI	**No competing interests**

Very interesting and relevant topic - excellent that patients are involved even though this is a very technical area where hearing the patient voice is difficult, also great strategy to define and check that you only have experts involved. This is important work and I only have minor comments that can be addressed with expansion of a few areas.

It is important to be very precise on the a priori definition of consensus when you come to report the final results - it is not quite clear currently if statements will be dropped between Delphi rounds. You say >80% agreement in your definition of consensus (which I assume is a score of 4 or 5 on the Likert scale), in the description on round 2 it sounds that only statements achieving >80% in round 1 will go into the second round but then there is a sentence saying all items where there is lack of consensus will be identified, following round 2 it is clear that only those with >80% agreement will be retained. A third Delphi round is very reasonable but 'insufficient consensus' is not defined, this will no doubt become clearer after piloting so could remain a little speculative.

How attrition bias will be addressed is unclear, and missing data is not mentioned -there should be brief sentences on these. Although not a core outcome set, looking through COS-STAP could be helpful (https://trialsjournal.biomedcentral.com/articles/10.1186/s13063-019-3230-x/tables/1).

ACCORD, CREDES and COSMIN acronyms appear without being defined - which may make reading the manuscript more difficult for someone not familiar with consensus methods. The TOpCLASS acronym is not defined, and it is not clear what this consortium is.

VERSION 1 - AUTHOR RESPONSE

Response to Dr. Nakase:

Thank you for your encouraging feedback. We appreciate your interest in seeing the Delphi consensus results, and we are confident that the inclusion of expert input will result in a robust definition that can contribute meaningfully to clinical practice.

Response to Dr. West:

We greatly appreciate your constructive comments and detailed suggestions. In response, we have provided further clarification on several points:

1. **A Priori Definition of Consensus and Dropping Statements:**

   o   We have now clarified that an a priori level of ≥80% agreement is required for a statement to reach consensus. As noted, this agreement corresponds to a score of 4 or 5 on the Likert scale (p.14/15)

   o   We have added an additional statement clarifying that statements will only be dropped between Delphi rounds if they clearly fail to reach consensus, as defined by agreement below 80%. This decision will be made in consultation with the study management group to ensure a transparent and rigorous process. Any statement considered for exclusion will be carefully reviewed, and the final decision will be based

on both the level of agreement and expert input from the management group (p18).

- o We have expanded our explanation to define "insufficient consensus" as agreement below 80%, and noted that any statements that fail to reach consensus across rounds will be excluded from the final set.(p.18)

2. **Attrition Bias and Missing Data Handling:**

- o We have now explicitly addressed how attrition bias and missing data will be handled. Automated reminders via the Qualtrics platform will be used to ensure participants who have started but not completed the survey are reminded to finish their responses. Only participants who provide complete responses will be eligible for inclusion in subsequent rounds. For cases of incomplete data, consensus percentages will be calculated based on those participants who completed the relevant questions (p.16)

3. **Clarification of Acronyms:**

- o We have added definitions for the acronyms ACCORD (p.20), CREDES (p.8), COSMIN (p.3), and TOpCLASS (p.8) to ensure that readers unfamiliar with these terms can easily follow the manuscript.