PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	Early indicators of disease progression in Fabry disease that may indicate the need for disease-specific treatment initiation: findings from the opinion-based PREDICT-FD modified Delphi consensus initiative
AUTHORS	Hughes, Derralynn; Aguiar, Patricio; Deegan, Patrick; Ezgu, Fatih; Frustaci, Andrea; Lidove, Olivier; Linhart, Aleš; Lubanda, Jean-Claude; Moon, James; Nicholls, Kathleen; Niu, Dau-Ming; Nowak, Albina; Ramaswami, Uma; Reisin, Ricardo; Rozenfeld, Paula; Schiffmann, Raphael; Svarstad, Einar; Thomas, Mark; Torra, Roser; Vujkovac, Bojan; Warnock, David; West, Michael; Johnson, Jack; Rolfe, Mark; Feriozzi, Sandro

VERSION 1 – REVIEW

REVIEWER	Dr. Ignacio Marín-León CIBERESP. IBIS. Rocio University Hospital. ENEBRO Foundation. Seville- Spain
REVIEW RETURNED	22-Jan-2020

GENERAL COMMENTS	2: ABSTRACT and TITLE: The Title is some confuse. I Think it could be clear for the readers to included the word "opinion -based" consensus: ie: Finding.... FD Delphy opinion-based consensus initiative. Because NOT all the consensus process are opinion-based; the ones that use the RAND Delphy consensus methods are evidence-based. 10. RESULTS Table 2: need to add some footnotes, to remark that some early indicators as "elevated troponin"; "microalbuminuria" or "NT-ProBNP" need a causal relationship between the indicator and Fabry disease. Table 3: need a footnote to state that the white cells Shadow means that consensus were not met 12 LIMITATIONS. I miss two relevant weaknesses in the discussion of limitations. First, we don't know the reproducibility of this consensus, What could happens with another composition of the panel, let say if instead of 8 nephrologist and 1 neurologist, there were 1 dermatologist, 3, neurologist, 1 Rhinolaringologist and only 2 nephrologist: In that case would be the results the same? The second one is the absent of some contrast between the consensus results and the evidence wisdom that back the "rightness" of the results. 13 SUPPLEMENTARY Part of the method information provided in the supplementary is NEEDLESS. All the related to "Literature review" (page 30, lines 23-59, and page 31, l 3-16. The literature review is not USED to

	support the consensus reached, neither to contrast the rightness of the indicators selected. Thus some reader could be confused, thinking that the article include an evidence review that back the consensus. And that is not true.

REVIEWER	Annabel Griffiths Costello Medical, United Kingdom
REVIEW RETURNED	16-Feb-2020

GENERAL COMMENTS	BMJ Review – Author Comments A really interesting study that clearly addresses an unmet need in the Fabry disease field. Application of a Delphi-like technique used appropriately with a wide geographical spread of participants, and an incredibly high response rate, which shows a high level of engagement and appetite in the clinical community to see these results published. Abstract The abstract did not clearly describe the following: - The number of rounds and what exercise was carried out at each round of the Delphi, as such it was unclear how the importance and consensus agreement rounds linked together - "Classified provisionally as important" was hard to interpret and from the abstract alone it was unclear what the agreement was determining if not importance - The number of participants, the fact they were all clinicians (as opposed to other healthcare professionals) and the number of countries represented should be included to provide additional context on the perspective this study provides - What the n numbers represent, could be interpreted as the number of panellists suggesting indicators of this type - How many met importance criteria and the role the co-chairs played in revising these prior to the initiation of round 2 - Results of indicators that met consensus differs from those presented in the main body of the manuscript – I assume this is in part due to recategorisation (see comment below) but abstract seems to be presenting final results (including completion of Round 4) and therefore expect to align with Table 2: o Abstract states 5 other early indicators of FD (in agreement with Table S4) whilst Table 2 includes 6 o Abstract states 6 patient-reported indicators (in agreement with Table S5) but Table 2 has 3 - The other activities conducted beyond the importance and consensus exercise (i.e. the chronology of manifestation of indicators during the disease course, and the initiation and cessation of FD-specific treatment via analysis of different scenarios) Description of Methods It is acknowledged that there are no formal, universally agreed guidelines on Delphi methods, however, it is largely accepted that a Delphi must have the following characteristics: be an iterative process, involve controlled feedback (whereby responses are summarised between rounds and communicated back to the participants), be consensus gathering and ensure participant (pseudo) anonymity.1, 2 It seems that no result summaries were

	provided to participants between rounds meaning that the current status of the groups' collective opinion was not repeatedly fed back. All other aspects of this Delphi criteria, however, were met. Furthermore, only some aspects of a classical Delphi were included, for example, an open first round was included, but a separate importance setting round was included before a single consensus gathering round (with Round 4 conducted on a subset of indicators). I would therefore propose using the term "modified Delphi" or "Delphi-like". The distinction between rounds and exactly what steps were taken to collate and interpret results in between rounds is unclear. For example, it should be made clearer how the initiation or cessation of FD-specific treatment in different scenarios fitted in – it seems a subset of the Round 2 questionnaire, but then unclear how these results translated into Round 3. Further explanation of Round 4 should be added, as it was unclear why results were inadvertently omitted and what "just met the importance criteria" was defined as – it seems the round was corrective and decided upon following the realisation of a mistake in the analysis of Round 3 (as opposed to statements being taken through that did not meet consensus as part of an a priori decision on how to proceed/terminate the Delphi process). Additional details are reported in the results (in "Refinements to consensus indicators") however further clarity is required and should be described in the methods. Would suggest the Delphi process figure S1 is moved into the main body and the following revisions made: - It to be made clearer that a data analysis/curation step by the Co-Chairs fits in between Round 1 and Round 2 - Separation of Round 3 and Round 4 in the diagram so it is clearer how they link together and what indicators were taken through - Clearer presentation of the importance rating and consensus gathering aspects of the respective rounds i.e. the figure indicates Round 2 purely assessed importance however in the methods it is stated that Round 2 also included agreement questions about initiation or cessation of FD-specific treatment in different scenarios - Inclusion of the chronology of signs and symptoms so it is clear at which stage of the process this was explored The other following additions to the methods would help aid clarity: - The dates over which the Delphi was conducted so that a) results can be interpreted in light of any changes to the treatment landscape, and b) for future interpretation of these findings - Mention of protocol development in the methods including stating who was involved in this step, in addition to the review of the protocol by Jack Johnson - Clarification over how the median score for starting or stopping FD-specific treatment was ≥7.5 was generated (i.e. unclear which scale was used here) - It to be made clear that all questions were compulsory - Justification of chosen consensus thresholds - Whether the number of rounds was determined a priori and/or based on a required threshold for terminating the Delphi process - Suggest explanation of the term "pivoted" Likert scale is added or the term removed - It to be made clear how the results of the literature review were used in the questionnaire development process

	Presentation of Results A number of changes to Table 2 are suggested: - The inclusion of the importance and agreement results from the Delphi should be presented so it is transparent how each of the indicators for which consensus on importance was achieved compared - Recommend a single column format rather than two columns, unless split between columns is reflective of the results/interpretation in which case this should be made clear - The table includes the heading “Early indicators of PNS damage in FD” however the supplementary table S3 stats CNS. Furthermore, the wording of the gastrointestinal symptoms indicator differs between the two - As described above, seems to be discrepancies between the abstract and supplementary tables and what is reported here - There is a strikethrough in the table for neuro-otologic abnormalities, which looks as if left in error, would suggest removing from the table and including a fuller description in the footnote o On a related point the heading of Table S5 (and presumably the other supplementary tables) should be amended to make clear results of Round 3 only - It is unclear if the following footnotes are based on feedback from the participants or based on literature, if the latter, please can references be added: o *The prognostic significance of this indicator is different in male and female patients o ||A causal relationship between this indicator and FD is required to justify treatment initiation - The table footnotes make reference to recategorisation however the justification for this and the stage at which this step occurred is unclear, please can further details be added The description of the chronology exercise does not make clear how panellist responses were collected and analysed, and as noted above, it is also unclear at which stage of the Delphi process this occurred. Study Limitations The authors suggest there was no group-interaction bias owing to the anonymous consensus process, however, it should be acknowledged that in a small field it is likely that these participants are known to one another, even though they cannot attribute responses to individuals. This limitation should be described, furthermore the term quasi- or pseudo-anonymity may be more appropriate.2, 3 In addition, as it seems all participants were ultimately included as authors it would be good to clarify at which stage the author group was determined and made aware of one another. The small number of participants is understandable in a rare disease, however, this context should be described. The authors cite Mehta 2019 in support of the number of participants, however, this reference states “based on published estimates that Delphi studies typically enrol 15–20 participants” supported by a Sandford reference. As such the reference does not support the authors claim that 15-22 participants is necessary for robust consensus, rather just describes what is typically done.

	There are questions around the generalisability of this study, as the authors claim to be presenting a global representative panel that is reflective of the real-world view of clinicians. There is an average of only 1.4 clinicians per country, so context such as (if available) the number of specialist treatment centres would help put this into perspective. The breakdown in Table 1 demonstrates the high level of expertise of the panel and the additional description of which specialities are represented is informative. It would be helpful to include details of which medical specialities typically treat Fabry disease and any other healthcare professionals that are involved in the management of these patients to a) put into context how representative these experts are of the treating clinical community, and b) explain why only clinicians were consulted versus the inclusion of other healthcare professions. Particularly with the emphasis on early indicators, it would be helpful to understand the role of primary care in these initial diagnoses. The lack of a neutral or do not wish to answer option in the importance Likert scale could result in bias in the results, or participants answering despite not having sufficient experience to do so especially as all questions were compulsory. Of note, a neutral option was provided in the pivoted Likert scale. Consensus disagreement was also not explored in this study. With any opinion gathering exercise it should be acknowledged that these results are opinion of the panellists only. The discussion includes several definitive statements such as “female patients and male patients with non-classical disease should be treated based on existing guideline recommendations” which should be framed in this context and/or additional references added if supported by other studies. Other - The high response rate is remarkable for a Delphi panel, although technically if one individual did not complete Round 1 then this response rate is not 100% in all rounds - In the initial strength listed by the authors: “a globally representative panel of experts in FD was recruited”, it should be made clear that all experts were clinicians - Suggest rephrasing of the final strength and limitation of this study “importance and agreement rating steps in a Delphi consensus are opinion based” as Delphi techniques do not require importance and agreement steps to both be conducted - The reference supporting “high costs of FD-specific treatment” is not robust, and particularly as a global perspective is being taken the wording should reflect any geographical differences in access and cost - Suggest avoiding the term “interviewed” (page 7) and replace with “consulted” - The term “group interaction” is used, which I would suggest is changed to “social interaction” as in a Delphi (that should provide controlled feedback) there are aspects of “group interaction”. It appears controlled feedback was not used in this case, however, so “group interaction” could be used as long as it is made clear that the controlled feedback element was not in place (see comment above)

	- Would review and align the use of “panel” and “participants”, I think these are being used interchangeably but can cause confusion with the term “Delphi panel” or imply an expert panel was employed in addition to the Co-Chairs - The footnotes in Table 3 and 4 for the green shading are not aligned in terms of description - The authors state that no statistical analyses were performed however the analysis of data and calculation of whether thresholds were met is a form of statistical analysis, albeit a simple one - The discrepancy between the results and the current guidelines could be explored more in the discussion References 1. Hasson F, Keeney S, McKenna H. Research guidelines for the Delphi survey technique. J Adv Nurs 2000;32:1008-15. 2. Keeney S. HF, McKenna H. The Delphi Technique in Nursing and Health Research: Wiley-Blackwell, 2011. 3. McKenna HP. The Delphi technique: a worthwhile research approach for nursing? J Adv Nurs 1994;19:1221-5.

VERSION 1 – AUTHOR RESPONSE

Reviewer 1	
2: ABSTRACT and TITLE: The Title is some confuse. I Think it could be clear for the readers to include the word "opinion -based" consensus: ie: Finding.... FD Delphi opinion-based consensus initiative. Because NOT all the consensus process are opinion-based; the ones that use the RAND Delphi consensus methods are evidence-based.	The title (p.1) has been amended accordingly.
10. RESULTS Table 2: need to add some footnotes, to remark that some early indicators as "elevated troponin"; "microalbuminuria" or "NT-ProBNP" need a causal relationship between the indicator and Fabry disease. Table 3: need a footnote to state that the white cells shadow means that consensus were not met	The footnotes to Tables 2 (pp.15) and 3 (p.20) have been amended.
12 LIMITATIONS. I miss two relevant weaknesses in the discussion of limitations. First, we don't know the reproducibility of this consensus, What could happens with another composition of the panel, let say if instead of 8	The ‘Strengths and weaknesses’ section on pp.25–26 has been amended to note the possible effect of having different medical specialties represented among the panel.

nephrologist and 1 neurologist, there were 1 dermatologist, 3, neurologist, 1 Rhinolaringologist and only 2 nephrologist: In that case would be the results the same? The second one is the absent of some contrast between the consensus results and the evidence wisdom that back the "rightness" of the results.	We may have misunderstood what the reviewer is asking for. If it is evidence that supports the utility of the indicators that achieved consensus, then this is discussed in paras. 2–4 on p.24. Please advise if there is another aspect of this that we need to address.
13 SUPPLEMENTARY Part of the method information provided in the supplementary is NEEDLESS. All the related to "Literature review" (page 30, lines 23-59, and page 31, l 3-16. The literature review is not USED to support the consensus reached, neither to contrast the rightness of the indicators selected. Thus some reader could be confused, thinking that the article include an evidence review that back the consensus. And that is not true.	Reviewer 2 asked that we explain what role the Literature review section had in development of the protocol. We have amended the text in the supplement to explain this and hopefully have therefore justified keeping the details of how the literature search was conducted. We are also willing to delete the section if this solution is unacceptable to Reviewer 1.
Reviewer 2	
General A really interesting study that clearly addresses an unmet need in the Fabry disease field. Application of a Delphi-like technique used appropriately with a wide geographical spread of participants, and an incredibly high response rate, which shows a high level of engagement and appetite in the clinical community to see these results published	We thank the reviewer for recognizing the relevance and utility of this research.
Abstract The abstract did not clearly describe the following: - The number of rounds and what exercise was carried out at each round of the Delphi, as such it was unclear how the importance and consensus agreement rounds linked together - "Classified provisionally as important" was hard to interpret and from the abstract alone it was unclear what the agreement was determining if not importance - The number of participants, the fact they were all clinicians (as opposed to other healthcare professionals) and the number of countries represented should be included to provide additional context on the perspective this study provides - What the numbers represent, could be interpreted as the number of panellists suggesting indicators of this type	The 'Design and setting' section has been reworded to clarify the sequence of the consensus process. The phrase "classified as provisionally important" has been removed. The text has been reworded to clarify this point.

- How many met importance criteria and the role the co-chairs played in revising these prior to the initiation of round 2	
- Results of indicators that met consensus differs from those presented in the main body of the manuscript – I assume this is in part due to recategorisation (see comment below) but abstract seems to be presenting final results (including completion of Round 4) and therefore expect to align with Table 2: o Abstract states 5 other early indicators of FD (in agreement with Table S4) whilst Table 2 includes 6 o Abstract states 6 patient-reported indicators (in agreement with Table S5) but Table 2 has 3	The text has been reworded to clarify this point. The total number of indicators meeting importance criteria has been included. The abstract now states that the indicators rated for importance were compiled by the Co-Chairs and administrator.
- The other activities conducted beyond the importance and consensus exercise (i.e. the chronology of manifestation of indicators during the disease course, and the initiation and cessation of FD-specific treatment via analysis of different scenarios)	The numbers in the abstract now agree with those in Table 2 and the Method has been extended to describe the final round of the consensus in which indicators were refined or grouped. The chronology is now included but description of the various scenarios for initiating and stopping treatment and the consensus results obtained is probably beyond the scope of the abstract.
Description of Methods It is acknowledged that there are no formal, universally agreed guidelines on Delphi methods, however, it is largely accepted that a Delphi must have the following characteristics: be an iterative process, involve controlled feedback (whereby responses are summarised between rounds and communicated back to the participants), be consensus gathering and ensure participant (pseudo) anonymity.[1, 2] It	We have made changes to use the phrase "modified Delphi" in the Title (p.1), Running head (p.2), Abstract (p.3), Introduction (p.6), Method (p.8), Results (pp.11, 12), Discussion (pp.25, 26), and for expediency have deleted the term 'Delphi' elsewhere if it is unnecessary or clearly implied. We acknowledge the

seems that no result summaries were provided to participants between rounds meaning that the current status of the groups' collective opinion was not repeatedly fed back. All other aspects of this Delphi criteria, however, were met. Furthermore, only some aspects of a classical Delphi were included, for example, an open first round was included, but a separate importance setting round was included before a single consensus gathering round (with Round 4 conducted on a subset of indicators). I would therefore propose using the term "modified Delphi" or "Delphi-like".	absence of a consensus round before the importance-setting round, but would like to point out that the structure of this consensus protocol meant that the panel were effectively apprised of the results of the previous round by the terms included in each subsequent round.
The distinction between rounds and exactly what steps were taken to collate and interpret results in between rounds is unclear. For example, it should be made clearer how the initiation or cessation of FD-specific treatment in different scenarios fitted in – it seems a subset of the Round 2 questionnaire, but then unclear how these results translated into Round 3.	Regarding initiation and cessation of treatment in different scenarios, only two panel rounds were needed because the clinical scenarios tested were selected by the Co-Chairs and not by consensus. Methods p.8 para. 4 describes how in Round 1 panellists were asked to "*rate…the likelihood that they would start or stop FD-specific treatment in different patient groups*". In round 2 panellists rated their agreement with the initiation/cessation findings collated from round 1 (Methods p.9, paragraph 1 "*Regarding initiation or cessation…*". The text on p.8–9 paras. 4 & 5 and p.9 para. 1 has been amended to try to clarify this. While addressing this point we noticed that the likelihood scores were missing that informed the agreement round shown in Table 3. Table 3 has been adapted to include these scores and Table 4 has been adapted to match the format of Table 3.
Further explanation of Round 4 should be added, as it was unclear why results were inadvertently omitted and what "just met the importance criteria" was defined as – it seems the round was corrective and decided upon following the realisation of a mistake in the analysis of Round 3 (as opposed to statements being taken through that did not meet consensus as part of an a priori decision on how to proceed/terminate the Delphi process). Additional details are reported in the results (in "Refinements to consensus indicators") however further clarity is required and should be described in the methods.	
Would suggest the Delphi process figure S1 is moved into the main body and the following revisions made: - It to be made clearer that a data analysis/curation step by the Co-Chairs fits in between Round 1 and Round 2 - Separation of Round 3 and Round 4 in the diagram so it is clearer how they link together and what indicators were taken through - Clearer presentation of the importance rating and consensus gathering aspects of the respective rounds i.e. the figure indicates Round	There was an administrative error at the end of round 2. Twenty of the 21 responses had been received and analysed correctly but when the 21st response was included in the data table the proportion of panellists who awarded an importance rating ≥3 did not update correctly. The error was detected after round 3 had been initiated, so 6 factors were circulated among the panel as part of round 4. The phrase "just met the importance criteria" reflects the fact that these factors had not reached the threshold as

2 purely assessed importance however in the methods it is stated that Round 2 also included agreement questions about initiation or cessation of FD-specific treatment in different scenarios - Inclusion of the chronology of signs and symptoms so it is clear at which stage of the process this was explored	a proportion of 20 votes but passed it (by approximately 1%) when the 21st vote was included. However, "just" is unnecessary; the factors met the importance criteria and the point to be conveyed is that they were omitted from round 3 by accident. The Method has been amended on p.8 para.2.
The other following additions to the methods would help aid clarity: - The dates over which the Delphi was conducted so that a) results can be interpreted in light of any changes to the treatment landscape, and b) for future interpretation of these findings	Figure S1 has been amended to address the reviewer's points and is now Figure 1 in the main manuscript. Other figures in the main text and supplement have been renumbered.
- Mention of protocol development in the methods including stating who was involved in this step, in addition to the review of the protocol by Jack Johnson	
- Clarification over how the median score for starting or stopping FD-specific treatment was ≥7.5 was generated (i.e. unclear which scale was used here)	
- It to be made clear that all questions were compulsory	
- Justification of chosen consensus thresholds	
- Whether the number of rounds was determined a priori and/or based on a required threshold for terminating the Delphi process	
- Suggest explanation of the term "pivoted" Likert scale is added or the term removed	Dates have been added on p.8 para.3 and p.9 para.3.
- It to be made clear how the results of the literature review were used in the questionnaire development process	Unless we have misunderstood the reviewer's point, this is reported on p.8 para.3 where the Delphi process is described.

	The text on p.8–9 para.5 has been expanded to explain this. The text has been amended on p.8 para.3. As far as we know the choice of thresholds in Delphi initiatives is arbitrary but must be made *a priori*. We used ones that were used in another published initiative and have noted this in the additional information provided in the Supplement p.2 para 2. Three rounds were specified a priori, which has been noted on p.8 para. 4, and the fourth was conducted *post hoc* as noted on p.9 para.2. It means that the extreme ratings of the scale pivot around the middle of the scale rather than increasing continuously from the lowest to the highest score. However, if the term is unknown to the reviewer then it is better deleted. It has been removed from the abstract on p.3, and from the method on p.9. The results of the literature search were summarized and shared with the Co-Chairs before the modified Delphi initiative commenced. It was helpful in informing questions about starting and stopping treatment in different patient groups and scenarios and provided a resource for subsequent production of a study report and publications. The text in the Supplement (p.2, para. 3) has been expanded to capture this. **Please note**: Reviewer 1 requested that this section be removed completely; accordingly, we have alerted Reviewer 1 to the comments made here.
Presentation of Results A number of changes to Table 2 are suggested:	

- The inclusion of the importance and agreement results from the Delphi should be presented so it is transparent how each of the indicators for which consensus on importance was achieved compared - Recommend a single column format rather than two columns, unless split between columns is reflective of the results/interpretation in which case this should be made clear	We respectfully disagree. The importance and agreement values for each indicator are provided in the supplementary tables but our understanding of the consensus process is that having achieved consensus it is inappropriate to assert that some consensus parameters are more important than others by dwelling on the scores awarded. This speaks to a point raised by Reviewer 1 that changing the panel composition might change the outcome of the consensus. Of course, this may occur, but setting importance and agreement thresholds that are less than 100% and ensuring that a certain number of panel members participate should accommodate some variation in the absolute scores without major changes to the empirical result.
- The table includes the heading "Early indicators of PNS damage in FD" however the supplementary table S3 states CNS. Furthermore, the wording of the gastrointestinal symptoms indicator differs between the two	The layout does not imply anything about the interpretation of the Results and the format was chosen pragmatically to avoid presenting a 40-line table. We have proposed a different layout which preserves the order in which the indicators appear in the relevant supplementary tables and would be pleased to re-format the table again if the journal has a preferred layout
- As described above, seems to be discrepancies between the abstract and supplementary tables and what is reported here - There is a strikethrough in the table for neuro-otologic abnormalities, which looks as if left in error, would suggest removing from the table and including a fuller description in the footnote o On a related point the heading of Table S5 (and presumably the other supplementary tables) should be amended to make clear results of Round 3 only	Table 2 is the final consensus and the accompanying footnotes endeavour to explain what refinements have been applied (noted in Results p.17, end of para 3 "*explanatory footnotes*…"). We have added a note in the text before Table 2 (p.12, para. 2) to bring the reader's attention to the section on p.17 that describes refinements to the consensus. The dilemma we face is that we would like the focus of the manuscript to be the refined consensus result not the results achieved by the end of round 3, hence the latter are provided in the supplement. The discrepancy between "CNS" in Table S3 and "PNS" in Table 2 arose as follows: the round 1 questionnaire solicited feedback about neurological signs, referred to as "indicators of CNS damage", but as the consensus process progressed, one of the panel members commented that none of the remaining neurological indicators were CNS-

	related, so we had to 'recategorize as PNS' as reported in the footnote.
- It is unclear if the following footnotes are based on feedback from the participants or based on literature, if the latter, please can references be added: o *The prognostic significance of this indicator is different in male and female patients o ||A causal relationship between this indicator and FD is required to justify treatment initiation - The table footnotes make reference to recategorisation however the justification for this and the stage at which this step occurred is unclear, please can further details be added The description of the chronology exercise does not make clear how panellist responses were collected and analysed, and as noted above, it is also unclear at which stage of the Delphi process this occurred.	We alert the reader to the fact that the consensus results in round 3 were further refined in round 4 (p.12, para.2) and have amended the text in the same paragraph (and preceding Table 2) to direct the reader to the "Refinements to consensus indicators" section. We hope that this, in conjunction with other changes to the supplementary tables (see below) and various footnotes, now makes it clear to the reader why Table 2 and the results in the supplementary tables are ostensibly discrepant. The text strikethrough was intentional but should not have extended to the footnote symbol. This indicator reached consensus but the cluster of indicators that it represents did not do so individually. Rather than remove it we have removed the strikethrough from the footnote symbol and have expanded the footnote to explain why the strikethrough is shown (pp.14–15). The reviewer is correct that Tables S1–S6 (Supplement pp.65–76) only show data from rounds 2 and 3 (apart from the indicators that were inadvertently omitted in round 3 and submitted for consensus in round 4, which are now marked with a footnote). The supplementary table headings have been amended. The footnotes are all based on feedback from the panel at the refinement stage and reported in Table S7 (Supplement pp.77–78). The reviewer's point about recategorization of 'CNS' as 'PNS' is discussed above, was instigated following round 4, and is documented in Table S7. The footnote has been amended to note this, as have other footnotes referring to category changes.

	We note in the method (p.9 para. 3) that the chronology was generated after the refined list of consensus indicators had been generated and that its generation did not involve Delphi techniques. We have amended the text to note that chronology development was conducted under the direction of the Co-Chairs, and that collation of the panel's comments was coordinated by the administrator.
Study Limitations The authors suggest there was no group-interaction bias owing to the anonymous consensus process, however, it should be acknowledged that in a small field it is likely that these participants are known to one another, even though they cannot attribute responses to individuals. This limitation should be described, furthermore the term quasi- or pseudo-anonymity may be more appropriate.[2, 3] In addition, as it seems all participants were ultimately included as authors it would be good to clarify at which stage the author group was determined and made aware of one another. The small number of participants is understandable in a rare disease, however, this context should be described.	Regarding consensus bias, we have expanded the 'Strengths and weaknesses' section of the Discussion (pp.25–26) to acknowledge the possibility of bias described by the reviewer and have noted when the panel members became aware of the other identities of other participants. We have also revised the related "Strengths and limitations" bullet on p.5.
The authors cite Mehta 2019 in support of the number of participants, however, this reference states "based on published estimates that Delphi studies typically enrol 15–20 participants" supported by a Sandford reference. As such the reference does not support the authors claim that 15-22 participants is necessary for robust consensus, rather just describes what is typically done.	We have rephrased the first paragraph of the supplement (p.2) to make readers aware that panel size selection was based only on precedent, and we have included a reference by Hsu & Sandford 2007 which reviews this point.
There are questions around the generalisability of this study, as the authors claim to be presenting a global representative panel that is reflective of the real-world view of clinicians. There is an average of only 1.4 clinicians per country, so context such as (if available) the number of specialist treatment centres would help put this into perspective.	The organization and structure of health care systems vary substantially in different countries with respect to management of rare diseases. Some, like the UK, have a few specialist centres that each support a relatively large number of patients, whereas countries like Italy

The breakdown in Table 1 demonstrates the high level of expertise of the panel and the additional description of which specialities are represented is informative. It would be helpful to include details of which medical specialities typically treat Fabry disease and any other healthcare professionals that are involved in the management of these patients to a) put into context how representative these experts are of the treating clinical community, and b) explain why only clinicians were consulted versus the inclusion of other healthcare professions. Particularly with the emphasis on early indicators, it would be helpful to understand the role of primary care in these initial diagnoses.	and France have many specialist centres, sometimes supporting a relatively small number of patients. We feel there is no succinct way to address this point so have noted in the 'Strengths and weaknesses' section of the Discussion (p.26, para.1) that the generalisability of findings is influenced both by the composition of the panel and how well a panellist's views align with those of colleagues in their own country. We have included a list of further specialties covered by the members of our panel (p.12, para. 1) and believe that the key ones involved in the management of patients with FD are represented. Having acknowledged this, it is important to note that the different opinions expressed depend only partly on the specialties of the Delphi panel and are mostly drawn from professional experiences. The initiative focused on specialist physicians because they have responsibility for treating patients and determine when to initiate FD-specific therapies. By contrast, primary care physicians would in general neither diagnose FD, manage patients with FD, nor prescribe FD-specific therapies, and would usually have very limited knowledge and experience of the disease (for example, a general practitioner in the UK would be unlikely even to see a patient with FD as, on average, there may be one such patient per 5–50 primary care medical centres).
The lack of a neutral or do not wish to answer option in the importance Likert scale could result in bias in the results, or participants answering despite not having sufficient experience to do so especially as all questions were compulsory. Of note, a neutral option was provided in the pivoted Likert scale. Consensus disagreement was also not explored in this study. With any opinion gathering exercise it should be acknowledged that these results are opinion of the panellists only. The discussion includes several definitive statements such as "female patients and male patients with non-classical disease should be treated based on existing guideline recommendations" which should be framed in this context and/or additional references added if supported by other studies.	The absence of a neutral option during importance rating has been noted in the 'Strengths and weaknesses' section (p.26, para.1). We have tried to infer what the reviewer had in mind regarding the application of consensus disagreement in our study – was it to stratify indicators that did not meet consensus into a group that could be discounted completely and a group whose utility is unproven or unknown? We have drafted text to this effect but please advise if this is not what was intended.

	The text has been revised to note that this is the panellists' collective opinion, in the Discussion p.24 para. 1, p.25 para. 1,
Other - The high response rate is remarkable for a Delphi panel, although technically if one individual did not complete Round 1 then this response rate is not 100% in all rounds - In the initial strength listed by the authors: "a globally representative panel of experts in FD was recruited", it should be made clear that all experts were clinicians - Suggest rephrasing of the final strength and limitation of this study "importance and agreement rating steps in a Delphi consensus are opinion based" as Delphi techniques do not require importance and agreement steps to both be conducted - The reference supporting "high costs of FD-specific treatment" is not robust, and particularly as a global perspective is being taken the wording should reflect any geographical differences in access and cost - Suggest avoiding the term "interviewed" (page 7) and replace with "consulted" - The term "group interaction" is used, which I would suggest is changed to "social interaction" as in a Delphi (that should provide controlled feedback) there are aspects of "group interaction". It appears controlled feedback was not used in this case, however, so "group interaction" could be used as long as it is made clear that the controlled feedback element was not in place (see comment above) - Would review and align the use of "panel" and "participants", I think these are being used interchangeably but can cause confusion with the term "Delphi panel" or imply an expert panel was employed in addition to the Co-Chairs - The footnotes in Table 3 and 4 for the green shading are not aligned in terms of description - The authors state that no statistical analyses were performed however the analysis	The abstract (p.3), strengths and limitations (p.5), and results (p.12) have been modified to note that the response rate was >95% (21/22 = 95.5%) The text (p.5) has been amended. The sentence (p.5) has been deleted because the preceding bullet notes that scoring is based on clinicians' real-world views (which must include their opinions), so further qualification of this now seems redundant. Our apologies. The wrong reference by Rombach was cited here. It should have been Rombach SM, Hollak CE, Linthorst GE, Dijkgraaf MG. Cost-effectiveness of enzyme replacement therapy for Fabry disease. Orphanet J Rare Dis 2013;8:29. This study reports costs in the Dutch healthcare system and cites estimates from the UK and US. The correct reference has been included and the text modified. The text has been amended.

of data and calculation of whether thresholds were met is a form of statistical analysis, albeit a simple one - The discrepancy between the results and the current guidelines could be explored more in the discussion	The text has in the Introduction (p.7) has been rephrased to highlight the rationale given in the supporting citation by Hsu and Sandford (2007). We have changed “participant” to “panellist” The text (pp.20 & 22) has been amended. The Statistical analysis text has been rephrased on p.10. This is the focus of a manuscript in preparation.
References 1. Hasson F, Keeney S, McKenna H. Research guidelines for the Delphi survey technique. J Adv Nurs 2000;32:1008-15. 2. Keeney S. HF, McKenna H. The Delphi Technique in Nursing and Health Research: Wiley-Blackwell, 2011. 3. McKenna HP. The Delphi technique: a worthwhile research approach for nursing? J Adv Nurs 1994;19:1221-5.	

VERSION 2 – REVIEW

REVIEWER	Ignacio Marín-León. CIBERESP. Rocio University Hospital. IBIS. Seville, SPAIN Fundación Enebro

	Seville, SPAIN
REVIEW RETURNED	22-Apr-2020

GENERAL COMMENTS	I agree with the answers given by the authors to the previous review, and the amended included in the new manuscript. It is a paper that merit to be published, because addresses an unmet need in a moving knowledge field as Fabry disease. Given the great uncertainty and strong implications with the disease management, a well developed consensus process could be very helpful for doctors treating Fabry patients in real world practice

REVIEWER	Annabel Griffiths Costello Medical, UK Employee of Costello Medical
REVIEW RETURNED	04-May-2020

GENERAL COMMENTS	Presentation of the methods much improved, thank you! Details of remaining/outstanding comments below and detailed on the annotated PDF, which also includes a few minor suggested track changes. Abstract - Last sentence of the “Design and setting” section of the abstract is unclear. Does this mean finally consensus indicators with an agreement of >75% (i.e. higher than the consensus threshold of 67% employed initially) were then refined or grouped? If so, it is unclear why this threshold would be applied, and why those that had achieved consensus would need further modifications and why. - Within the results section, please make clear if these results relate to current and/or future indicators so it is easier to cross-reference to the results in the main body of the manuscript. - Re the last sentence of the results section of the abstract, would suggest either mentioning in the methods how/when/why the chronologies were assessed, and then adding in a result to describe the outcome after these chronologies were proposed to participants (i.e. did they agree? if not, why not?), or removing this sentence from the abstract results. - The exercise to establish consensus on when to initiate and stop FD-specific treatment seems really valuable. Appreciate word count limits, but suggest mentioning in the abstract, including presenting any key conclusions. Also current methods in the abstract don't make clear how consensus was achieved with respect to this aspect of the study (from the main body methods, seems like questions on this topic included from round 1), please make clear or avoid the term "consensus" here if a formal consensus gathering exercise wasn't conducted. Description of Methods - Please explicitly state that no controlled feedback was provided to participants between rounds, as this is a key characteristic of the Delphi process that has not been carried out. This should also be mentioned in the discussion of study limitations. o As per my previous feedback, it is acknowledged that there are no formal, universally agreed guidelines on Delphi methods, however, it is largely accepted that a Delphi must have the following characteristics: be an iterative process, involve controlled feedback (whereby responses are summarised between rounds

	and communicated back to the participants), be consensus gathering and ensure participant (pseudo) anonymity.1, 2 - Suggest saying explicitly that consensus on treatment recommendations were then not taken forward beyond round 2. - It is unclear from the abstract and methods description whether importance was re-assessed in parallel to consensus in round 3, or importance stopped at round 2 (currently the abstract implies the latter, but the methods in the main body suggest the former). - Please make clearer in the methods how, when and for what purpose the distinction was made between current and future indicators - e.g. were future indicators treated differently in subsequent rounds? Presumably the current versus future distinction was made by the Co-Chairs? If just more of an observation made after study or round completion to aid interpretation, then the phrasing should be softened so this is clearer (currently it is presented as a formal distinction made during the study). - Other minor points: o In the “Modified Delphi process” section would make clear that the “outcome of voting” means (presumably) median likelihood score o Suggest these sentences starting with "Agreement was sought..." is moved to the end of the previous paragraph and adjusted to read "Agreement was subsequently sought in round 2 for those scenarios where.... " o The sentence "If the score was..." is unclear, please rephrase - something like "In contract, participants were asked whether they agreed with not starting or stopping treatment for..." (if I've understood correctly). o Re the sentence starting “Agreement scores were compiled by the administrator…” - I think you are using the term "agreement score" to mean output of the consensus gathering exercise (rather than the importance scoring), as such, if this sentence stays in this position I think it needs to say "agreement and importance scores"? o As previously suggested, the dates over which the Delphi was conducted so that a) results can be interpreted in light of any changes to the treatment landscape, and b) for future interpretation of these findings (currently dates only included for the chronology of signs and symptoms exercise) o As previously suggested, mention of protocol development in the methods including stating who was involved in this step, in addition to the review of the protocol by Jack Johnson o As previously suggested, it to be made clear how the results of the literature review were used in the questionnaire development process Presentation of Results - The following changes to Table 2 are recommended: o Appreciate the inclusion of the importance and agreement results from the Delphi (as previously suggested) may not have been presented due to space limitations, however, would suggest including in Table 2 at least the consensus results. o Would suggest rephrasing the title as in the rest of the manuscript you seem to be distinguishing the importance rating and consensus agreement steps, and the term “consensus on importance” is confusing. o Would either change the main body titles or subtitles within the table so the "Additional indicators" or "Other" titles are consistently used.

	□ Related to the above, table title currently says “PNS” where as main body title says “CNS/PNS” – would align how you refer to this category throughout. - Re “Indicators of cardiac damage”: o Please make clear if this consolidation step was completed in round 4 - in which case the results presented are adjusted for the consolidation completed in round 4. Also applies to the “Indicators of CNS/PNS damage”. o Unclear how the 3 presented here relate to the abstract results (6 cardiac - that met the importance criteria and reached consensus) - please clarify wording here and/or in the abstract. o When describing “The other current indicators”, please clarify if these were considered important as well or if this is regardless of importance (if so, suggest not presenting those that didn't reach the importance threshold in the main body results). - Re the sentence “Results by organ system and category are described below…” - this sentence makes it hard to understand how the scenarios on when to start and stop FD-specific treatment (which were asked in parallel to the indicator questions, and therefore can't have been based on the final list of indicators) feed into the recommendations. Please amend for clarity. - Within “Patient-reported indicators” would split out patient-reported signs and symptoms relevant to FD-specific treatment initiation so clearer how these results relate to the table and abstract. - Please make clear if the chronologies presented are before or after panel feedback, presumably the latter. - Within the “Initiation and cessation of FD-specific…” - please make clear which round the results relate to - it is understood the consensus gathering round was round 2 only (applies throughout this subsection) - Other minor points: o Suggest rephrasing "substantial proportions" - e.g. leaving as "managing both patients with classical and those with non-classical FD", or saying "approximately equal numbers of ..." o "co-chairs" is capitalised elsewhere in the manuscript but not within “Indicators of renal damage” o As previously noted, re the strikethrough in the table for neuro-otologic abnormalities, which looks as if left in error, would suggest presenting in () instead o Suggest presenting early indicators within the respective columns (e.g. early cardiac indicators within the cardiac column but with a subheading) or not having in the table and just listing in the main text o Suggest moving additional indicators and patient-reported indicators above the table - It is unclear if the following footnotes are based on feedback from the participants Discussion - As noted above, please mention the absence of controlled feedback as a limitation. Other minor points: - Would suggest aligning with the 95% response rate you've used elsewhere to avoid confusion (rather than introducing the concept of "voting" and "non-voting" rounds at this stage in the manuscript). References 1. Hasson F, Keeney S, McKenna H. Research guidelines for the Delphi survey technique. J Adv Nurs 2000;32:1008-15.

	2. Keeney S. HF, McKenna H. The Delphi Technique in Nursing and Health Research: Wiley-Blackwell, 2011.

VERSION 2 – AUTHOR RESPONSE

Reviewer 1	
I agree with the answers given by the authors to the previous review, and the amended included in the new manuscript. It is a paper that merit to be published, because addresses an unmet need in a moving knowledge field as Fabry disease. Given the great uncertainty and strong implications with the disease management, a well-developed consensus process could be very helpful for doctors treating Fabry patients in real world practice	We thank the reviewer for supporting our endeavour
Reviewer 2	
General Presentation of the methods much improved, thank you! Details of remaining/outstanding comments below and detailed on the annotated PDF, which also includes a few minor suggested track changes.	**N.B.** Page numbers included for reference pertain to the manuscript version showing tracked changes.
Abstract - Last sentence of the "Design and setting" section of the abstract is unclear. Does this mean finally consensus indicators with an agreement of >75% (i.e. higher than the consensus threshold of 67% employed initially) were then refined or grouped? If so, it is unclear why this threshold would be applied, and why those that had achieved consensus would need further modifications and why. - Within the results section, please make clear if these results relate to current and/or future indicators so it is easier to cross-reference to the results in the main body of the manuscript. - Re the last sentence of the results section of the abstract, would suggest either mentioning in the methods how/when/why the chronologies were assessed, and then adding in a result to describe the outcome after these chronologies were proposed to participants (i.e. did they agree? if not, why not?), or removing this sentence from the abstract results. - The exercise to establish consensus on when to initiate and stop FD-specific treatment seems really valuable. Appreciate word count limits, but suggest mentioning in the abstract, including presenting any key conclusions….	The ">75%" figure applies to agreement with refinements in round 4, not to a more stringent threshold applied in round 3. We have revised the last sentence of "Design and setting" to try to make this clear. We realised that no mention of the ">75%" threshold was made in the main text, so have addressed this on p.8 We decided to exclude data relating to future indicators from the abstract (there were only three and these coincide with current ones) because describing them added little and was costly in terms of wordcount. In terms of clarity, please note that we state "83 possible **current** indicators" in the opening sentence of the Results. In response to this, and the reviewer's next comment, we included mention of the chronologies at last review following the reviewer's earlier request that more

….Also current methods in the abstract don't make clear how consensus was achieved with respect to this aspect of the study (from the main body methods, seems like questions on this topic included from round 1), please make clear or avoid the term "consensus" here if a formal consensus gathering exercise wasn't conducted.	information about the study be detailed in the abstract "- *The other activities conducted beyond the importance and consensus exercise (i.e. the chronology of manifestation of indicators during the disease course, and the initiation and cessation of FD-specific treatment via analysis of different scenarios)*". We considered reporting information about starting/stopping treatment but that would have required explanation of the rounds in which the relevant data were collected. Clearly, choices have to be made about what is reported in a 300-word abstract – mentioning that chronologies were developed is a self-contained statement that really needs no qualification, whereas a description of how agreement was reached on treatment initiation/ cessation demands methodological detail that cannot be accommodated without sacrificing information elsewhere. We will delete the statement about the chronologies if the reviewer insists but would prefer to keep it. We have amended the wording slightly to clarify panellists' involvement in their development. Text revised, pp.6, 8, 15.
Description of Methods Please explicitly state that no controlled feedback was provided to participants between rounds, as this is a key characteristic of the Delphi process that has not been carried out. This should also be mentioned in the discussion of study limitations. o As per my previous feedback, it is acknowledged that there are no formal, universally agreed guidelines on Delphi methods, however, it is largely accepted that a Delphi must have the following characteristics: be an iterative process, involve controlled feedback (whereby responses are summarised between rounds and communicated back to the participants), be consensus gathering and ensure participant (pseudo) anonymity.1, 2 - Suggest saying explicitly that consensus on treatment recommendations were then not taken forward beyond round 2. - It is unclear from the abstract and methods description whether importance was re-assessed in parallel to consensus in round 3, or importance stopped at round 2 (currently the abstract implies the latter, but the methods in the main body suggest the former).	Text revised, pp.7,21

Please make clearer in the methods how, when and for what purpose the distinction was made between current and future indicators - e.g. were future indicators treated differently in subsequent rounds? Presumably the current versus future distinction was made by the Co-Chairs? If just more of an observation made after study or round completion to aid interpretation, then the phrasing should be softened so this is clearer (currently it is presented as a formal distinction made during the study).	Text revised, p.8. Importance rating stopped at round 2. Agreement whether an indicator was important was surveyed in round 3 (and round 4 for the few that had to be revisited). Unfortunately, we are struggling to see what remains unclear. We have made an amendment to the text (p.8) and hope that the revision addresses the issue.
Other minor points: o In the "Modified Delphi process" section would make clear that the "outcome of voting" means (presumably) median likelihood score o Suggest these sentences starting with "Agreement was sought..." is moved to the end of the previous paragraph and adjusted to read "Agreement was subsequently sought in round 2 for those scenarios where.... " o The sentence "If the score was..." is unclear, please rephrase - something like "In contrast, participants were asked whether they agreed with not starting or stopping treatment for..." (if I've understood correctly). o Re the sentence starting "Agreement scores were compiled by the administrator…" - I think you are using the term "agreement score" to mean output of the consensus gathering exercise (rather than the importance scoring), as such, if this sentence stays in this position I think it needs to say "agreement and importance scores"? o As previously suggested, the dates over which the Delphi was conducted so that a) results can be interpreted in light of any changes to the treatment landscape, and b) for future interpretation of these findings (currently dates only included for the chronology of signs and symptoms exercise) o As previously suggested, mention of protocol development in the methods including stating who was involved in this step, in addition to the review of the protocol by Jack Johnson [*To clarify, I meant that the methods should include a mention of development of a pre-agreed protocol (presumably after "Selection of the Chairs and expert panel"), and who reviewed/signed this off (i.e. not just mentioned within the "Patient and*	In round 1, panellists were surveyed separately regarding which indicators they regarded as relevant to current practice and which they foresaw as of potential future relevance. The questionnaires in the appendix can be reviewed to confirm this (e.g. appendix pp. 8, 12). The two categories (current indicators, future indicators) were separately subjected to the same consensus building methodology. In the methods on p.7 we state "Panellists provided free-text responses to open questions about early indicators of renal, cardiac and CNS damage <u>that can be assessed in current routine clinical practice, or which are not assessed routinely at present, but might be in the future</u>". We have amended the text in case the existing phrasing implies that indicators were suggested to the panel. They were not. All indicators were suggested by the panel. Text revised, p.8 We have re-ordered the text, moving this segment to an earlier point in the same paragraph, p.8

public involvement statement"). This is important for emphasising which aspects of the study design (e.g. thresholds) were agreed in advance]	Text revised, p.8
o As previously suggested, it to be made clear how the results of the literature review were used in the questionnaire development process [*To clarify, I think this should be made clear in the main body methods (not just adjusting the wording in the supplementary material)*]	Only agreement scores were compiled in round 3. No change made (p.8).
	The dates for the consensus were included on p.7 at last review ('Modified Delphi process' para. 1).
	Text revised, p.7
	Text revised, p.7

Presentation of Results - The following changes to Table 2 are recommended: o Appreciate the inclusion of the importance and agreement results from the Delphi (as previously suggested) may not have been presented due to space limitations, however, would suggest including in Table 2 at least the consensus results. [*Apologies, I didn't review your response initially, however still stand by the suggestion that consensus results should be included as a minimum (in the same way p values would be shown even though a significance threshold is applied). I'm not convinced by your rationale for not doing so. If you do, however, wish to not present in the main body, then equivalent results from Table 4 should also be removed for consistency.*]	We think that our decision and rationale are justified, and so have not added agreement scores in Table 2. Regarding Tables 3 and 4, we would have removed the agreement scores for consistency but the reviewer also pointed out that consensus-building methodology was applied only partially in gaining agreement on when, and in which patient groups, treatment should be started or stopped. Without the benchmark of consensus, we think that presentation of agreement scores for these data seems particularly relevant.
o Would suggest rephrasing the title as in the rest of the manuscript you seem to be distinguishing the importance rating and consensus agreement steps, and the term "consensus on importance" is confusing.	
o Would either change the main body titles or subtitles within the table so the "Additional indicators" or "Other" titles are consistently used.	Text revised, p.13
- Related to the above, table title currently says "PNS" whereas main body title says "CNS/PNS" – would align how you refer to this category throughout. [*Again, apologies for not reviewing your response initially, I can appreciate that ultimately this category makes more sense as PNS only (following recategorisation), however, would suggest you keep the title the same within the results section*]	Text revised, p.12
- Re "Indicators of cardiac damage": o Please make clear if this consolidation step was completed in round 4 - in which case the results presented are adjusted for the consolidation completed in round 4. Also applies to the "Indicators of CNS/PNS damage".	Text revised, p.11
o Unclear how the 3 presented here relate to the abstract results (6 cardiac - that met the importance criteria and reached consensus) - please clarify wording here and/or in the abstract.	
o When describing "The other current indicators", please clarify if these were considered important as well or if this is regardless of importance (if so, suggest not	Consolidation by the Co-chairs was at the end of round 1 in both cases. Text revised, p.11

presenting those that didn't reach the importance threshold in the main body results). - Re the sentence “Results by organ system and category are described below…” - this sentence makes it hard to understand how the scenarios on when to start and stop FD-specific treatment (which were asked in parallel to the indicator questions, and therefore can't have been based on the final list of indicators) feed into the recommendations. Please amend for clarity. - Within “Patient-reported indicators” would split out patient-reported signs and symptoms relevant to FD-specific treatment initiation so clearer how these results relate to the table and abstract. - Please make clear if the chronologies presented are before or after panel feedback, presumably the latter. - Within the “Initiation and cessation of FD-specific…” - please make clear which round the results relate to - it is understood the consensus gathering round was round 2 only (applies throughout this subsection) - Other minor points: o Suggest rephrasing "substantial proportions" - e.g. leaving as "managing both patients with classical and those with non-classical FD", or saying "approximately equal numbers of ..." o "co-chairs" is capitalised elsewhere in the manuscript but not within “Indicators of renal damage” o As previously noted, re the strikethrough in the table for neuro-otologic abnormalities, which looks as if left in error, would suggest presenting in () instead [*Thank you for removing it from the footnote, however, I still stand by the suggestion that the strikethrough would be better presented another way*] o Suggest presenting early indicators within the respective columns (e.g. early cardiac indicators within the cardiac column but with a subheading) or not having in the table and just listing in the main text o Suggest moving additional indicators and patient-reported indicators above the table - It is unclear if the following footnotes are based on feedback from the participants	There were 6 renal indicators and 10 cardiac indicators that reached consensus (abstract states: “*….27 indicators (kidney, 6; cardiac, 10; CNS/PNS, 2;….*). This subset of ‘3’ cardiac indicators achieved consensus as current and future indicators independently. On p.11 the preceding sentence states “*Consensus was reached for 10 current indicators, 3 of which also reached consensus as future indicators*…” Text revised, p.11 Text revised, p.11 We have amended the text to clarify how the final list of patient-reported indicators was reached (p.12). This is covered in the method on p.8 (and yes, chronologies show what was agreed after panel feedback and review) Text revised, p.15

	Text revised, p.10 Text revised, p.11 Text revised, p.13 These are “Early cardiac indicators of FD that may be used in **future**”. As explained earlier, these achieved consensus independently as future indicators and as current indicators. We prefer to include them in table 2 (pp.13–14). We have moved a larger block of text (not tracked – to leave other changes visible) above table 2 so that all information pertaining to the table precedes it (now, p.12). Text revised, p.12
Discussion - As noted above, please mention the absence of controlled feedback as a limitation. *Other minor points:* - Would suggest aligning with the 95% response rate you've used elsewhere to avoid confusion (rather than introducing the concept of "voting" and "non-voting" rounds at this stage in the manuscript).	Text revised, p.21 Text revised, pp.20–21