PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	Study protocol: A modified Delphi survey for the evidence summarization of patient decision aids
AUTHORS	Saunders, Catherine; Durand, Marie-Anne; Dannenberg, Michelle; Giguere, Anik; Alper, Brian; Hoffmann, Tammy; Perestelo-Pérez, L; Campbell, Stephen; Elwyn, Glyn

VERSION 1 - REVIEW

REVIEWER	Purva Abhyankar University of Stirling, UK
REVIEW RETURNED	01-Nov-2018

GENERAL COMMENTS	Thank you for the opportunity to review this study protocol. I believe this is an important area where guidance is needed to ensure transparency and rigour in this aspect of the patient decision aid development. The paper is well written. I have some comments/questions about the manuscript. Introduction 1. The introduction is adequate and sets the scene for the study. The authors have explained the rationale for the need for procedural guidance on evidence summarisation in PDAs. They state that although there are agreed-upon approaches and methods for evidence summarisation in other areas such as clinical guidelines, there is no agreed process for the same in relation to PDAs. While I can understand that the requirements of this process are slightly different for PDAs compared to clinical guidelines, it would be useful if the authors clarify in the introduction why the same methods and approaches cannot or may not be applied for PDAs. 2. The authors state that currently there is no guidance for evidence summarisation in PDAs, even in the IPDAS criteria, minimum standards or the PDA certification efforts. While I agree with this in general, I feel that there is some evidence from IPDAS that is not referred to/incorporated. E.g. the IPDAS chapters and associated papers that reviewed the evidence underpinning each of the quality standards. The chapter: Based on Scientific Evidence: Basing information on comprehensive, critically appraised, and up-to-date syntheses of the scientific evidence: a quality dimension of the International Patient Decision Aid Standards, seems particularly relevant to this study, yet it is not referred to in the study. I would encourage the authors to incorporate the learning from this chapter or indicate how this has been done.

	Methods 3. Delphi method – could the authors clarify if they will be using the original Delphi method or a modified Delphi approach. Over the years, Delphi method has become a collective name, rather than a singular method, with multiple modifications of the approach. It would be useful if this could be briefly clarified. 4. The study will be managed by a study steering group, who seem to be key in making final decisions about the guidance steps, phases and criteria following the Delphi survey. While the authors have described the membership of this group, it would be useful to know the size of the group (how many members) and the geographical spread of the group members (how international is this group). It is also not clear if this is a virtual group or whether the group will meet in person following the Delphi survey to discuss, refine and finalise the criteria/steps. 5. In Participants – could the authors provide rationale for why the developers of PDAs must have developed or updated a tool within the last five years? 6. In the first survey – demographic questions – why is ethnicity data required (rather than country of residence e.g.)? If it is required for valid reasons, I would suggest that the authors use some standard/global classification system that is used in international research. The current categories seem applicable mainly to the USA based populations. Also, why is gender data required? 7. How will the decision about a third round of Delphi survey be made? Are there any criteria that the steering group will apply? What factors might influence this decision (e.g. what is meant by level of consensus)? 8. Would the members of the steering group also be participants in the Delphi survey? 9. In data analysis, the authors state that any items that are rated by 80% of participants will be removed or retained (as per the direction of the rating). What would happen if the steering group members disagree with any of the retentions/removals or there may be strong reasons/evidence suggesting otherwise? I think it is important to anticipate the various scenarios and flesh out the process for dealing with these upfront, to avoid bias. Data collection tools 10. Information on proposed process – Point 1 - The definition of decision aids seems inaccurate. PDAs do not just provide information about risks and benefits of health treatments and tests but also include methods to help people evaluate this information in relation to their values to make the choice that is right for them. While I see that the information part is the most relevant one to this study, I still feel that the definition should be one that is accurate and universally agreed. E.g. IPDAS defines PDAs as: Patient decision aids are tools designed to help people participate in decision making about health care options. They provide information on the options and help patients clarify and communicate the personal value they associate with different features of the options. Why not use the same definition? The next point 'Accurate and clear information is critical' progresses to the relevant part for this study, so that should be adequate. 11. Information on proposed process – Point 2 – 'It's important for decision aids to have accurate and trustworthy information from research evidence about the risks and benefits of health treatments and tests.' Does this not include evidence about what options may be available too? Is there a step/criteria that ensures

	that the process for identifying options is systematic, unbiased and comprehensive? E.g. how should the options be identified, who decides whether they are reasonable or not etc.? 12. Information on proposed process – Point 3 – We are trying to make evidence summarisation easier – Surely making the process easier is not the only purpose. I would suggest that this reflects the purpose of making the process transparent, rigorous and free from bias. 13. Information on proposed process – Point 5 - We sketched out a proposed process – would be useful to add 'of evidence summarisation' after 'process', to clarify.

REVIEWER	Ilya Ivlev Oregon Health & Science University, USA
REVIEW RETURNED	15-Nov-2018

GENERAL COMMENTS	The results of the proposed study should help to ensure that new or updated patients decision aids present systemically collected evidence. Please consider my friendly peer-review comments as an attempt to consider additional approaches to conducting this study. Comment #1. Making final consensus-based decisions It's possible that after all three rounds of surveys were completed, the consensus won't be achieved for some items (processes, steps, or criteria). The authors proposed that "If no consensus is achieved, the steering group will decide whether or not to retain a criterion ..." Do you think that this approach for making final decisions may violate the proposed study aim – "to generate consensus"? Could you consider providing more flexible suggestions when the consensus was not achieved, but the committee thinks that these items (processes, steps, or criteria) are crucial? For example, these items (processes, steps, or criteria) may be still stated in your recommendations but following them will be optional. Comment #2. Measuring consensus You have proposed "If at least 80% of participants rate the item in the lower two categories (omit, possible) or in the higher two categories (desirable, essential), we will consider consensus to be achieved..." Why do you think that the items "omit" and "possible" are semantically close enough to be merged and indicate that the proposed process or criterion should be excluded from further consideration? In my opinion, the answer 'possible' is semantically closer to the answer "desirable;" however, I might be wrong. I think that by using the proposed approach–collapsing four answers in two groups–you may lose valuable information about participates' preferences. Also, the measure of 80% agreement on the collapsed answers doesn't seem to reflect respondents' views. To retain heterogeneity in answers, you may consider keeping these four alternative answers separately and use Kendall's dispersive coefficient of concordance (W) to establish a measure of the consistency of the participants' opinions. Strong conformity

	within the participants' preferences can be confirmed at W≥0.6 ($p<0.05$). I have a few additional minor questions/comments that you might consider addressing in your protocol: Comment #3. How many participants in each group (developers of PDAs, patient representatives) do you consider enough to complete each Delphi step? Comment #4. What is your strategy for retaining participants? Comment #5. What is your strategy for missing data (unit or item nonresponse) for each Delphi round? Comment #6. What are the eligibility criteria? Comment #7. The authors might consider adding a description of the participants' data management and safety. Comment #8. Are you planning on any additional statistical analyses (e.g., subgroup analyses)? Comment #9. It remains unclear for me how the authors will make final decisions about omitting or including processes, criteria, or steps. Comment #10. "PDA" – consider spelling out when used for the first time (see line 5)

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1

Reviewer Name: Purva Abhyankar

Institution and Country: University of Stirling, UK

Please state any competing interests or state 'None declared': None declared

Please leave your comments for the authors below

Thank you for the opportunity to review this study protocol. I believe this is an important area where guidance is needed to ensure transparency and rigour in this aspect of the patient decision aid development. The paper is well written. I have some comments/questions about the manuscript.

 Thank you for the positive feedback.

Introduction

1. The introduction is adequate and sets the scene for the study. The authors have explained the rationale for the need for procedural guidance on evidence summarisation in PDAs. They state that although there are agreed-upon approaches and methods for evidence summarisation in other areas such as clinical guidelines, there is no agreed process for the same in relation to PDAs. While I can understand that the requirements of this process are slightly different for PDAs compared to clinical guidelines, it would be useful if the authors clarify in the introduction why the same methods and approaches cannot or may not be applied for PDAs.

 Thank you for this helpful remark. We have made this important distinction clearer in the last paragraph of the introduction, page 4 of the revised manuscript (revised sections in blue):

"Evidence synthesis in other medical contexts is increasingly standardized, such as the selection and summarization of evidence for clinical practice guidelines and systematic reviews. This process minimizes the risk of bias in the end product [7-16]. The same level of scrutiny is justified when developing PDAs, as they may directly influence patient care and decision making. Tasks such as the

selection and identification of patient-relevant outcomes, analysis of patient concerns and priorities, description of the quality of evidence, and communication of uncertainty in ways that patients understand warrants the development of an agreed process and related steps and criteria that are specific to PDAs. For those reasons, it would not be appropriate to apply evidence summarization processes developed for clinical guidelines without integrating the evidence summarization steps and components that are specific to the development of interventions that target patients. The target group, scope and content differ significantly enough from clinical practice guidelines development to warrant a tailored evidence summarization process. Additionally, the IPDAS standards impose some prerequisites on the evidence summarization process on which the decision aid will be based. For example, IPDAS requires that the decision aid summarizes the evidence regarding all health options available to a patient facing a specific health problem, and that decision aids present positive and negative features of each option with an equal amount of details, among other specificities [18]. Efforts to develop an agreed evidence summarization process for PDAs should incorporate the substantial body of related evidence summarization guidance previously developed by other groups, and notably for clinical practice guidelines previously mentioned [9]."

2. The authors state that currently there is no guidance for evidence summarisation in PDAs, even in the IPDAS criteria, minimum standards or the PDA certification efforts. While I agree with this in general, I feel that there is some evidence from IPDAS that is not referred to/incorporated. E.g. the IPDAS chapters and associated papers that reviewed the evidence underpinning each of the quality standards. The chapter: Based on Scientific Evidence: Basing information on comprehensive, critically appraised, and up-to-date syntheses of the scientific evidence: a quality dimension of the International Patient Decision Aid Standards, seems particularly relevant to this study, yet it is not referred to in the study. I would encourage the authors to incorporate the learning from this chapter or indicate how this has been done.

 Thank you for this remark. We were aware of this chapter but had omitted to cite it. We have revised the following paragraph on page 4 of the revised manuscript and added the citation:

"A 2013 review of the literature conducted by the IPDAS working group on the synthesis of scientific evidence highlighted the importance of rigorously selecting and summarizing evidence used to populate a patient decision aid. They did not provide clear practical guidance on how to conduct evidence summarization for the development of patient decision aids except recommending that developers apply the GRADE methodology."

Methods

3. Delphi method – could the authors clarify if they will be using the original Delphi method or a modified Delphi approach. Over the years, Delphi method has become a collective name, rather than a singular method, with multiple modifications of the approach. It would be useful if this could be briefly clarified.

 As specified in the manuscript title, abstract and methods, we are using a modified Delphi approach.

4. The study will be managed by a study steering group, who seem to be key in making final decisions about the guidance steps, phases and criteria following the Delphi survey. While the authors have described the membership of this group, it would be useful to know the size of the group (how many members) and the geographical spread of the group members (how international is this group). It is also not clear if this is a virtual group or whether the group will meet in person following the Delphi survey to discuss, refine and finalise the criteria/steps.

 We have clarified on page five of the revised manuscript the size and geographical spread of the steering group. See revised content below:

“The study steering group includes nine international experts in PDA development, evaluation and implementation, evidence summarization and clinical practice guidelines, and one patient representative. Six steering group members are based in the US, one in Canada, one in Australia and one in Spain. Google drive and video-conferencing facilities will be used to facilitate the exchange and review of information and documents, virtual meetings, as well as real-time collaboration and version-control.”

5. In Participants – could the authors provide rationale for why the developers of PDAs must have developed or updated a tool within the last five years?

 As stated in the manuscript, we have previously developed an inventory of established patient decision aid developers who are specialized in developing those interventions and produce and maintain at least five patient decision aids. We feel that it would be difficult to have established robust development and evidence summarization processes for developers who have developed one or two patient decision aids on an ad hoc basis and may not have maintained the interventions. This approach has been used in previously published studies on related topics. Although it seemed fair and logical to target established patient decision aid developers with the Delphi survey (thus sending the Delphi invitation to developers already listed on our inventory), we did not exclude smaller decision aid developers from completing the survey, who may have come across the Delphi survey invitation through included listservs.

6. In the first survey – demographic questions – why is ethnicity data required (rather than country of residence e.g.)? If it is required for valid reasons, I would suggest that the authors use some standard/global classification system that is used in international research. The current categories seem applicable mainly to the USA based populations. Also, why is gender data required?

 We systematically collect basic demographic information, including ethnicity data and gender when conducting online surveys in the US. It is important to provide a basic description of the sample’s characteristics, particularly given patients and patient representatives were invited to participate. Given the majority of steering group members and the main study authors and coordinators were based in the US, and the survey hosted in the US, we used survey responses that were appropriate for this country. We did not want to complexify the survey even more by adding skip logic and multiple ethnicity categories depending on country of completion. Based on previous studies of this type, and selected listservs (many of which are hosted in the US), we knew that the majority of participants were likely to be based in the US. We will mention this as a potential, yet small, limitation in the published Delphi survey results manuscript. We feel that it is not necessary to mention it in the protocol.

7. How will the decision about a third round of Delphi survey be made? Are there any criteria that the steering group will apply? What factors might influence this decision (e.g. what is meant by level of consensus)?

 As described in the manuscript, the steering group will make this decision collectively based on consensus ratings. We have also clarified this section on page 7 of the revised manuscript to address the reviewer’s comment. Further, it is worth noting that the members of the steering group, collectively, have conducted and published many modified Delphi survey and have always been able to reach consensus within 3 rounds.

“Depending on the level of consensus (see data analysis section), a third round may be conducted. This will be determined by the steering group after round 2 data analysis is completed.”

8. Would the members of the steering group also be participants in the Delphi survey?

 Thank you for this helpful comment. No, the steering group unanimously decided not to participate in the survey. This was added on page 6 of the revised manuscript, in the participants section.

“In order not to contaminate the Delphi survey results and express their views twice (in developing the original items and taking the surveys), the steering group members have unanimously decided not to complete the Delphi surveys.”

9. In data analysis, the authors state that any items that are rated by 80% of participants will be removed or retained (as per the direction of the rating). What would happen if the steering group members disagree with any of the retentions/removals or there may be strong reasons/evidence suggesting otherwise? I think it is important to anticipate the various scenarios and flesh out the process for dealing with these upfront, to avoid bias.

 Thank you for this important remark. This is described on page 7 and 8 of the revised manuscript. We have tried to clarify further (see below). In brief, data analysis and decision to retain or exclude items will primarily be guided by consensus. However, if the steering group felt that it was important to remove or modify items that have reached consensus but are contradicted by recurring open text comments (for example), they will retain editorial control and will make the changes that they collectively deem suitable, guided by open text comments. We have successfully used this approach before.

“If at least 80% of participants rate the importance of the item in the lower two categories, or in the higher two categories, we will consider consensus to be achieved and the item will be removed or retained, respectively. If no consensus is achieved or the consensus ratings are contradicted by recurring open text comments, the steering group will decide whether or not to retain a criterion, basing this decision on qualitative feedback from the participants where possible, and the steering group’s views. We have successfully used this approach before.”

10. Information on proposed process – Point 1 - The definition of decision aids seems inaccurate. PDAs do not just provide information about risks and benefits of health treatments and tests but also include methods to help people evaluate this information in relation to their values to make the choice that is right for them. While I see that the information part is the most relevant one to this study, I still feel that the definition should be one that is accurate and universally agreed. E.g. IPDAS defines PDAs as: Patient decision aids are tools designed to help people participate in decision making about health care options. They provide information on the options and help patients clarify and communicate the personal value they associate with different features of the options. Why not use the same definition? The next point ‘Accurate and clear information is critical’ progresses to the relevant part for this study, so that should be adequate.

 Thank you for this pertinent suggestion. We have revised the definition on page 4 of the revised manuscript to match IPDAS and Stacey et al.’s Cochrane review published in 2017. We have added the IPDAS citations.

“They are typically defined as: “evidence-based tools designed to help patients make specific and deliberated choices among healthcare options. Patient decision aids supplement (rather than replace) clinicians' counselling about options”.

11. Information on proposed process – Point 2 – ‘It's important for decision aids to have accurate and trustworthy information from research evidence about the risks and benefits of health treatments and tests.’ Does this not include evidence about what options may be available too? Is there a step/criteria that ensures that the process for identifying options is systematic, unbiased and comprehensive? E.g. how should the options be identified, who decides whether they are reasonable or not etc.?

 Yes, this criterion is included in phase 2, step 1 (search for evidence):

“There is a systematic search for evidence that relates to the options included in the PDA.”

12. Information on proposed process – Point 3 – We are trying to make evidence summarisation easier – Surely making the process easier is not the only purpose. I would suggest that this reflects the purpose of making the process transparent, rigorous and free from bias.

 We are not sure exactly which section of the manuscript or supplementary files the reviewer is referring to. We feel that we already make those arguments in several paragraphs of the introduction. However, we have tried to improve selected sentences to address the reviewer's comments:

"This process promotes transparency, rigor, and minimizes the risk of bias in the end product [7-16]."page 4

"This will in turn improve transparency, rigor and minimize the risk of bias of the evidence summarization processes leading to the development of patient decision aids." Page 5

13. Information on proposed process – Point 5 - We sketched out a proposed process – would be useful to add 'of evidence summarisation' after 'process', to clarify.

 Thank you for this suggestion. The Delphi survey round 1 has already been sent out. We will add what the reviewer suggests in rounds 2 and 3 of the Delphi survey.

Reviewer: 2

Reviewer Name: Ilya Ivlev

Institution and Country: Oregon Health & Science University, USA

Please state any competing interests or state 'None declared': None declared

Please leave your comments for the authors below

The results of the proposed study should help to ensure that new or updated patients decision aids present systemically collected evidence. Please consider my friendly peer-review comments as an attempt to consider additional approaches to conducting this study.

 Thank you for your positive feedback and helpful comments.

Comment #1. Making final consensus-based decisions

It's possible that after all three rounds of surveys were completed, the consensus won't be achieved for some items (processes, steps, or criteria). The authors proposed that "If no consensus is achieved, the steering group will decide whether or not to retain a criterion ..." Do you think that this approach for making final decisions may violate the proposed study aim – "to generate consensus"? Could you consider providing more flexible suggestions when the consensus was not achieved, but the committee thinks that these items (processes, steps, or criteria) are crucial? For example, these items (processes, steps, or criteria) may be still stated in your recommendations but following them will be optional.

 See answer to reviewer 1 comments (comment 9). In addition, we can confirm that this approach does not violate the modified Delphi survey methodology and has been successfully used in previously published Delphi surveys. A citation has been added on page 8.

Comment #2. Measuring consensus

You have proposed “If at least 80% of participants rate the item in the lower two categories (omit, possible) or in the higher two categories (desirable, essential), we will consider consensus to be achieved...”

Why do you think that the items “omit” and “possible” are semantically close enough to be merged and indicate that the proposed process or criterion should be excluded from further consideration? In my opinion, the answer ‘possible’ is semantically closer to the answer “desirable;” however, I might be wrong.

I think that by using the proposed approach–collapsing four answers in two groups–you may lose valuable information about participates’ preferences. Also, the measure of 80% agreement on the collapsed answers doesn’t seem to reflect respondents’ views. To retain heterogeneity in answers, you may consider keeping these four alternative answers separately and use Kendall’s dispersive coefficient of concordance (W) to establish a measure of the consistency of the participants’ opinions. Strong conformity within the participants’ preferences can be confirmed at $W \geq 0.6$ ($p < 0.05$).

 Thank you for this comment. Grouping the higher two or the lower two categories is a standard approach. We do not feel that there is a semantic issue here. In order to address the reviewer’s comment and clarify our analytic approach, we have reworded the following paragraph on pages 7 and 8:

“Following round one, the ratings will be summarized using percentages and the views of all participants will be given equal weight. If at least 80% of participants rate the item in the lower two categories (omit, possible) or in the higher two categories (desirable, essential), we will consider consensus to be achieved and the item will be removed or retained, respectively. Items where ratings do not meet the consensus threshold and conflict with open text comments will be grouped together and explained to round 2 participants. They will be asked to re-rate those items taking the qualitative feedback into account. Following the first survey round, a consensus meeting involving the steering group will be held. The steering group will review and discuss the ratings and qualitative feedback received, including rewording suggestions per criterion, suggestions to add new phases, steps or criteria and more general comments or questions. The wording or order of the phases, steps or criteria will be revised if two or more respondents suggest it or if the steering group members agree that the phase, step or criterion would benefit from rewording, reordering or merging.

Following the second survey round, a second consensus meeting will be held. Decisions on whether to conduct a third round and retain items in the scale will be made based on the ratings in the survey rounds and feedback/comments from participants. The ratings will be summarized using percentages and the views of all participants will be given equal weight. If at least 80% of participants rate the importance of the item in the lower two categories, or in the higher two categories, we will consider consensus to be achieved and the item will be removed or retained, respectively. If no consensus is achieved or the consensus ratings are contradicted by recurring open text comments, the steering group will decide whether or not to retain a criterion, basing this decision on qualitative feedback from the participants where possible, and the steering group’s views. We have successfully used this approach before [21].

Only complete surveys will be included in the analysis. We will report the amount of missing data in the manuscript reporting the results of the Delphi survey.”

I have a few additional minor questions/comments that you might consider addressing in your protocol:

Comment #3. How many participants in each group (developers of PDAs, patient representatives) do you consider enough to complete each Delphi step?

 We have chosen not to pre-define a minimum number of participants (particularly for rounds 2 and 3). We will keep the survey opened for 3 weeks and analyze answers received during this time.

Comment #4. What is your strategy for retaining participants?

 As stated in the manuscript, we will use email addresses provided in round 1 and email those participants again (rounds 2 and 3) and use up to 2 reminders per round.

Comment #5. What is your strategy for missing data (unit or item nonresponse) for each Delphi round?

 Thank you for this helpful remark, we omitted to provide this information. We have added it on page 8 of the revised manuscript.

“Only complete surveys will be included in the analysis. We will report the amount of missing data in the manuscript reporting the results of the Delphi survey.”

Comment #6. What are the eligibility criteria?

 The eligibility criteria are listed on page 5 of the manuscript. We have no other eligibility criteria and have added this sentence:

To maximize the generalizability and applicability of the criteria, we plan to invite participation in the survey from the following groups: 1) all known developers of PDAs who created or updated a tool within last five calendar years (using existing inventory), 2) all members of the of the IPDAS group, 3) the Shared Decision Making listserv; 4) the Society for Participatory Medicine listserv ; 5) an overdiagnosis google group ; 6) the evidence-based healthcare listserv ; 7) the Society for Medical Decision Making ; the 8) the Society of Behavioral Medicine (Health Decision Making Interest Group) , 9) HTAi-ISG Patient Involvement listserv, 10) GRADE Working group, 11) the Guidelines International Network, 12) convenience sample of policy makers with interest and expertise in PDA certification; 13) the BMJ patient group; 14) the ProPublica Patient Safety Community. We have no other eligibility criteria.

Comment #7. The authors might consider adding a description of the participants’ data management and safety.

 Thank you for this excellent suggestion. We have added the following section on page 8 of the revised manuscript:

Data Management and Safety

“Data to be collected include information about the participant’s role as it relates to patient decision aids, general demographics, and their opinion of what to add/change/include in an evidence summarization process. We are careful to protect the identity of all study participants. We will store the data securely in accordance with standard human subject research protocols. All data will be retained for three years, per the Dartmouth College data retention policy (or for the period specified by journals in which arising manuscripts are published, if longer) and then destroyed securely.”

Comment #8. Are you planning on any additional statistical analyses (e.g., subgroup analyses)?

 We are not. All planned analyses are described in the protocol.

Comment #9. It remains unclear for me how the authors will make final decisions about omitting or including processes, criteria, or steps.

 See data analysis section and answer to reviewer 1 (comment 9).

Comment #10. “PDA” – consider spelling out when used for the first time (see line 5)

 Thank you for your attention to detail. We had already spelt out PDA in the first sentence of the introduction:

“Patient Decision Aids (PDAs) are tools that help patients and their clinicians make preference-sensitive decisions together.”

However, we notice that we had not spelt it in the abstract. Thank you for the reminder, we have made this change.

VERSION 2 – REVIEW

REVIEWER	Purva Abhyankar University of Stirling, UK
REVIEW RETURNED	14-Jan-2019

GENERAL COMMENTS	Thank you for submitting a revised version of the protocol paper and an opportunity for reviewing it. I feel the manuscript has improved in transparency. Below are my comments/views (in orange text) on the authors’ responses to the reviewer comments, where I do not feel the responses are satisfactory enough: 1. The introduction is adequate and sets the scene for the study. The authors have explained the rationale for the need for procedural guidance on evidence summarisation in PDAs. They state that although there are agreed-upon approaches and methods for evidence summarisation in other areas such as clinical guidelines, there is no agreed process for the same in relation to PDAs. While I can understand that the requirements of this process are slightly different for PDAs compared to clinical guidelines, it would be useful if the authors clarify in the introduction why the same methods and approaches cannot or may not be applied for PDAs. Authors’ reply to comment 1 (of reviewer 1): Thank you for this helpful remark. We have made this important distinction clearer in the last paragraph of the introduction, page 4 of the revised manuscript (revised sections in blue): “Evidence synthesis in other medical contexts is increasingly standardized, such as the selection and summarization of evidence for clinical practice guidelines and systematic reviews. This process minimizes the risk of bias in the end product [7-16]. The same level of scrutiny is justified when developing PDAs, as they may directly influence patient care and decision making. Tasks such as the selection and identification of patient-relevant outcomes, analysis of patient concerns and priorities, description of the quality of evidence, and communication of uncertainty in ways that patients understand warrants the development of an agreed process and related steps and criteria that are specific to PDAs. For those reasons, it would not be appropriate to apply evidence summarization processes developed for clinical guidelines without integrating the evidence summarization steps

	and components that are specific to the development of interventions that target patients. The target group, scope and content differ significantly enough from clinical practice guidelines development to warrant a tailored evidence summarization process. Additionally, the IPDAS standards impose some prerequisites on the evidence summarization process on which the decision aid will be based. For example, IPDAS requires that the decision aid summarizes the evidence regarding all health options available to a patient facing a specific health problem, and that decision aids present positive and negative features of each option with an equal amount of details, among other specificities [18]. Efforts to develop an agreed evidence summarization process for PDAs should incorporate the substantial body of related evidence summarization guidance previously developed by other groups, and notably for clinical practice guidelines previously mentioned [9]." Comments on author response: Where the changes appear (in blue), I don't think the reader is aware that the previous sentence contains the reasons/the differences between evidence summarization tasks for clinical guidelines vs PDAs. So when you say 'for those reasons..' it's confusing. Additionally, the reasons are still implicit/unclear e.g. why to the listed tasks require different standards/processes for evidence summarisation? The following sentence (highlighted in grey) does seem to represent one of the reasons, but other reasons are not clear/apparent. I could not understand the point about IPDAS standards imposing some pre-requisites on evidence summarisation process for PDAs (which are listed thereafter). However, what implications do these have for requiring a different summarization process? 7. How will the decision about a third round of Delphi survey be made? Are there any criteria that the steering group will apply? What factors might influence this decision (e.g. what is meant by level of consensus)? Authors' reply on comment 7 of reviewer 1: □ As described in the manuscript, the steering group will make this decision collectively based on consensus ratings. We have also clarified this section on page 7 of the revised manuscript to address the reviewer's comment. Further, it is worth noting that the members of the steering group, collectively, have conducted and published many modified Delphi survey and have always been able to reach consensus within 3 rounds. Comments on author response While the clarification on page 7 is satisfactory, I do not think 'having been able to reach a consensus within 3 rounds' is a valid argument in academic and scientific practice. This may be due to the (power) dynamics of the steering group, rather than open critical debate and discussion followed by a democratic consensus process. 10, 11, 12, 13. Information on proposed process: These comments were meant for the supplementary files - Overall Proposed

	Phases, Steps and Criteria - INFORMATION ON PROPOSED PROCESS. This is the actual survey material – following the demographic data collection. However, now that the round 1 has already been sent out, I don't know how much of this can be incorporated. Authors' reply to comment #6 from reviewer 2: What are the eligibility criteria? The eligibility criteria are listed on page 5 of the manuscript. We have no other eligibility criteria and have added this sentence: To maximize the generalizability and applicability of the criteria, we plan to invite participation in the survey from the following groups: 1) all known developers of PDAs who created or updated a tool within last five calendar years (using existing inventory), 2) all members of the of the IPDAS group, 3) the Shared Decision Making listserv; 4) the Society for Participatory Medicine listserv ; 5) an overdiagnosis google group ; 6) the evidence-based healthcare listserv ; 7) the Society for Medical Decision Making ; the 8) the Society of Behavioral Medicine (Health Decision Making Interest Group) , 9) HTAi-ISG Patient Involvement listserv, 10) GRADE Working group, 11) the Guidelines International Network, 12) convenience sample of policy makers with interest and expertise in PDA certification; 13) the BMJ patient group; 14) the ProPublica Patient Safety Community. We have no other eligibility criteria. If I understand it correctly, there are no criteria for participation as such apart from being a member of the listed groups. The list given in the paragraph above are not criteria, but a list of sources/groups from where the participants will be recruited. Could the authors clarify or make it explicit that anyone who is a member of the listed groups is eligible to participate in the Delphi?

REVIEWER	Ilya Ivlev Oregon Health & Science University, USA
REVIEW RETURNED	02-Jan-2019

GENERAL COMMENTS	This protocol was improved. I do not have additional comments or suggestions.

VERSION 2 – AUTHOR RESPONSE

Reviewer 2

Reviewer Name: Ilya Ivlev

Institution and Country: Oregon Health & Science University, USA

Please state any competing interests or state 'None declared': None declared

Please leave your comments for the authors below

This protocol was improved. I do not have additional comments or suggestions.

Thank you for submitting a revised version of the protocol paper and an opportunity for reviewing it. I feel the manuscript has improved in transparency. Below are my comments/views (in orange text) on the authors' responses to the reviewer comments, where I do not feel the responses are satisfactory enough:

⇒ AUTHOR RESPONSE: Thank you, we are glad all comments have been satisfactorily addressed.

Reviewer: 1

Reviewer Name: Purva Abhyankar

Institution and Country: University of Stirling, UK

Please state any competing interests or state 'None declared': None declared

Please leave your comments for the authors below

Thanks you for the opprtunity to review the revised version of this paper. I have attached a document detailing my comments on the authors' response to reviewer comments.

1. The introduction is adequate and sets the scene for the study. The authors have explained the rationale for the need for procedural guidance on evidence summarisation in PDAs. They state that although there are agreed-upon approaches and methods for evidence summarisation in other areas such as clinical guidelines, there is no agreed process for the same in relation to PDAs. While I can understand that the requirements of this process are slightly different for PDAs compared to clinical guidelines, it would be useful if the authors clarify in the introduction why the same methods and approaches cannot or may not be applied for PDAs.

Authors' reply to comment 1 (of reviewer 1):

 Thank you for this helpful remark. We have made this important distinction clearer in the last paragraph of the introduction, page 4 of the revised manuscript (revised sections in blue):

"Evidence synthesis in other medical contexts is increasingly standardized, such as the selection and summarization of evidence for clinical practice guidelines and systematic reviews. This process minimizes the risk of bias in the end product [7-16]. The same level of scrutiny is justified when developing PDAs, as they may directly influence patient care and decision making. Tasks such as the selection and identification of patient-relevant outcomes, analysis of patient concerns and priorities, description of the quality of evidence, and communication of uncertainty in ways that patients understand warrants the development of an agreed process and related steps and criteria that are specific to PDAs. For those reasons, it would not be appropriate to apply evidence summarization processes developed for clinical guidelines without integrating the evidence summarization steps and components that are specific to the development of interventions that target patients. The target group, scope and content differ significantly enough from clinical practice guidelines development to warrant a tailored evidence summarization process. Additionally, the IPDAS standards impose some prerequisites on the evidence summarization process on which the decision aid will be based. For example, IPDAS requires that the decision aid summarizes the evidence regarding all health options available to a patient facing a specific health problem, and that decision aids present positive and negative features of each option with an equal amount of details, among other specificities [18].

Efforts to develop an agreed evidence summarization process for PDAs should incorporate the substantial body of related evidence summarization guidance previously developed by other groups, and notably for clinical practice guidelines previously mentioned [9]."

Comments on author response:

Where the changes appear (in blue), I don't think the reader is aware that the previous sentence contains the reasons/the differences between evidence summarization tasks for clinical guidelines vs PDAs. So when you say 'for those reasons..' it's confusing.

Additionally, the reasons are still implicit/unclear e.g. why to the listed tasks require different standards/processes for evidence summarisation?

The following sentence (highlighted in grey) does seem to represent one of the reasons, but other reasons are not clear/apparent.

⇒ AUTHOR RESPONSE: We do not fully understand the reviewer comment. We have clearly outlined the tasks specific to patient decision aid development , and different from clinical guideline development (Tasks such as the selection and identification of patient-relevant outcomes, analysis of patient concerns and priorities, description of the quality of evidence, and communication of uncertainty in ways that patients understand warrants the development of an agreed process and related steps and criteria that are specific to PDAs) and that therefore justify developing an agreed evidence summarization process for PDAs

I could not understand the point about IPDAS standards imposing some pre-requisites on evidence summarisation process for PDAs (which are listed thereafter). However, what implications do these have for requiring a different summarization process?

⇒ AUTHOR RESPONSE: We fail to understand the reviewer's confusion and feel that the information provided is clear. As stated, IPDAS requires that the decision aid summarizes the evidence regarding all health options available to a patient facing a specific health problem, and that decision aids present positive and negative features of each option with an equal amount of details, among other specificities. This is one more argument, or reason (as stated above) to justify why are developing an agreed process and related steps and criteria that are specific to PDAs

7. How will the decision about a third round of Delphi survey be made? Are there any criteria that the steering group will apply? What factors might influence this decision (e.g. what is meant by level of consensus)?

Authors' reply on comment 7 of reviewer 1:

 As described in the manuscript, the steering group will make this decision collectively based on consensus ratings. We have also clarified this section on page 7 of the revised manuscript to address the reviewer's comment. Further, it is worth noting that the members of the steering group, collectively, have conducted and published many modified Delphi survey and have always been able to reach consensus within 3 rounds.

Comments on author response

While the clarification on page 7 is satisfactory, I do not think 'having been able to reach a consensus within 3 rounds' is a valid argument in academic and scientific practice. This may be due to the (power) dynamics of the steering group, rather than open critical debate and discussion followed by a democratic consensus process.

⇒ AUTHOR RESPONSE: To address the reviewer's concern, we have reworded this section as follows, see page 7 of the revised manuscript.

"This will be determined by the steering group after round 2 data analysis is completed. We will use open debate and discussion followed by a democratic consensus."

10, 11, 12, 13. Information on proposed process: These comments were meant for the supplementary files - Overall Proposed Phases, Steps and Criteria - INFORMATION ON PROPOSED PROCESS. This is the actual survey material – following the demographic data collection. However, now that the round 1 has already been sent out, I don't know how much of this can be incorporated.

⇒ AUTHOR RESPONSE: Thank you for clarifying. We agree with the reviewer. Given that round 1 has already been sent out, we are not able to make changes other than those that were mentioned in the initial revision.

Authors' reply to comment #6 from reviewer 2:

What are the eligibility criteria?

 The eligibility criteria are listed on page 5 of the manuscript. We have no other eligibility criteria and have added this sentence:

To maximize the generalizability and applicability of the criteria, we plan to invite participation in the survey from the following groups: 1) all known developers of PDAs who created or updated a tool within last five calendar years (using existing inventory), 2) all members of the of the IPDAS group, 3) the Shared Decision Making listserv; 4) the Society for Participatory Medicine listserv ;

5) an overdiagnosis google group ; 6) the evidence-based healthcare listserv ; 7) the Society for Medical Decision Making ; the 8) the Society of Behavioral Medicine (Health Decision Making Interest Group) , 9) HTAi-ISG Patient Involvement listserv, 10) GRADE Working group, 11) the Guidelines International Network, 12) convenience sample of policy makers with interest and expertise in PDA certification; 13) the BMJ patient group; 14) the ProPublica Patient Safety Community. We have no other eligibility criteria.

If I understand it correctly, there are no criteria for participation as such apart from being a member of the listed groups. The list given in the paragraph above are not criteria, but a list of sources/groups from where the participants will be recruited. Could the authors clarify or make it explicit that anyone who is a member of the listed groups is eligible to participate in the Delphi

⇒ AUTHOR RESPONSE: This is correct. We have tried to clarify this paragraph as follows, on page 5 of the revised manuscript.

"To maximize the generalizability and applicability of the criteria, we plan to invite participation in the survey from members of the following groups: 1) all known developers of PDAs who created or updated a tool within last five calendar years (using existing inventory), 2) all members of the of the IPDAS group, 3) the Shared Decision Making listserv; 4) the Society for Participatory Medicine listserv ; 5) an overdiagnosis google group ; 6) the evidence-based healthcare listserv ; 7) the Society for Medical Decision Making ; the 8) the Society of Behavioral Medicine (Health Decision Making Interest Group) , 9) HTAi-ISG Patient Involvement listserv, 10) GRADE Working group, 11) the Guidelines International Network, 12) convenience sample of policy makers with interest and expertise in PDA certification; 13) the BMJ patient group; 14) the ProPublica Patient Safety Community. We have no other eligibility criteria (except for membership to one of the above listed groups)."