PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	Protocol of an Interdisciplinary Consensus Project Aiming to Develop an AGREE II Extension for Guidelines in Surgery
AUTHORS	Antoniou, George A.; Mavridis, Dimitris; Tsokani, Sofia; Lopez-Cano, Manuel; Florez, Ivan; Brouwers, Melissa; Markar, Sheraz R.; Silecchia, Gianfranco; Francis, Nader; Antoniou, Stavros

VERSION 1 – REVIEW

REVIEWER	Kurinchi Gurusamy University College London, United Kingdom I am a guideline developer and am currently working on a project to find the relationship between correct strength of recommendation and AGREE-II criteria.
REVIEW RETURNED	19-Feb-2020

GENERAL COMMENTS	Lines 132-136: “For example, the item “The potential resource implications of applying the recommendations have been considered“ may be difficult to be universally addressed, because cost-effectiveness studies are scarce in the surgical literature and relevant evidence typically varies in different settings” This does not refer to only cost-effectiveness. This has implications on whether there is adequate expertise available to perform a procedure. If a complex procedure is shown to be effective, it requires adequate expertise to perform the procedure. The AGREE-II tool explanation clearly states: “For example, there may be a need for more specialized staff, new equipment, and expensive drug treatment” and this is not just about cost-effectiveness. Lines 174 -177: “We employed a series of statistical methods to explore reliability…” Improving the reliability of AGREE-II tool is a good justification, but it is not clear why this is done for surgical guidelines alone. Lines 233/234: “The second round will include closed-ended questions in a numeric Likert scale to assess participants”. The Likert scale range and interpretation should be provided. Lines 236/237: "We will discard low-scoring items and use the shortlisted items in a third Delphi round". What does ‘low scoring items’ mean in second round Delphi – will these be removed only based on the scores or will there be a certain level of agreement? Lines 238/239: “We will repeat the process until new information reaches saturation and consensus with an alpha level of 0.8”. Is there any reference for this approach in Delphi? This sounds more like focus group meetings where you continue the focus groups until

	saturation is reached. There is no new information introduced after round 1 – as this is the only open-ended questionnaire. After this, it is only change of scores/agreement that occurs. What does alpha level mean here – alpha error? Clearly, never seen an alpha error of 0.8 being used. Is this agreement level? At what level of agreement is an item included or excluded? Lines 244-251: "Qualitative synthesis" More details about qualitative research synthesis (this is what is meant here and is different from qualitative synthesis which may refer to narrative synthesis of quantitative data or qualitative data) should be provided. Currently, it lacks any details. Line 289: "Publication and dissemination" The intellectual property content should be clarified. Including a lead member of AGREE-II tool, on its own, does not mean that modifications/extensions for surgery can be done.

REVIEWER	Paschalis Gavriilidis University Hospitals Birmingham NHS Trust Queen Elizabeth Hospital Department of Vascular Access and Renal Transplantation, Birmingham, UK
REVIEW RETURNED	12-Mar-2020

GENERAL COMMENTS	The protocol is well structured and written

REVIEWER	Mary E Brindle Cumming School of Medicine, University of Calgary, Calgary Canada
REVIEW RETURNED	04-May-2020

GENERAL COMMENTS	This protocol paper proposes to develop an extension to the AGREE II framework to focus specifically on a guideline for developing guidelines in surgery. The protocol is very well-written and the methods are excellent, thorough, well-considered and generally very well-developed. The eventual extension is important. The authorship is international in its scope with broad representation of scientific and clinical backgrounds including authorship from AGREE and the Health Research Methods Evidence and Impact program at McMaster University. The broad authorship reflects the necessary diversity required to develop the guideline. The introduction appropriately provides a rationale for the AGREE II extension broad scope of literature based on the wide sources of literature that are currently available to aid surgeons in decision-making and the recognition that the AGREE II framework may be inadequate to address the specific needs of surgical guidelines. Overall the methods are rigorous. Appropriate patient representation is considered within the Delphi process. The use of pilot testing and assessment of internal validity are a great strength, as well as the plans to broadly disseminate, acquire feedback and update the extension as required. There are a few minor areas where improvements could be considered.

	1. argument for the inadequacy of the current AGREE II framework to specifically guide surgical guidelines could be strengthened. The example given that the current AGREE framework may not be appropriate specifically related to considering resource implications is concerning, as resource considerations will likely continue to be important in guideline development even in the absence of high quality cost effectiveness data. Surgery is not the only complex and multifaceted aspect of care that benefits from guidelines. Stronger support for why surgery, specifically, will benefit from this extension would strengthen the introduction. 2. A brief description of the findings from GAP I are provided. A short description of the overarching findings from GAP II would also be of benefit. 3. There are a few areas of the methods that would benefit from further details. The investigators report that they will perform qualitative evidence synthesis to identify factors of conceptual importance of the quality of evidence in surgery. A few details on the type of QES undertaken (approach) would be of benefit As well, the investigators report that they will survey users of social media to nominate parameters of importance in the development and reporting of guidelines in surgery. More details on who these users are, how they are identified, and how they would be surveyed will help the readers better understand this aspect of the proposal.

VERSION 1 – AUTHOR RESPONSE

Reviewer(s)' Comments to Author:

Reviewer: 1

Reviewer Name

Kurinchi Gurusamy

Institution and Country

University College London, United Kingdom

Please state any competing interests or state ‘None declared’:
I am a guideline developer and am currently working on a project to find the relationship between correct strength of recommendation and AGREE-II criteria.

Please leave your comments for the authors below
Lines 132-136: “For example, the item “The potential resource implications of applying the recommendations have been considered“ may be difficult to be universally addressed, because cost-effectiveness studies are scarce in the surgical literature and relevant evidence typically varies in different settings”
This does not refer to only cost-effectiveness. This has implications on whether there is adequate expertise available to perform a procedure. If a complex procedure is shown to be effective, it requires adequate expertise to perform the procedure. The AGREE-II tool explanation clearly states: “For

example, there may be a need for more specialized staff, new equipment, and expensive drug treatment" and this is not just about cost-effectiveness.

Response:

This is a valid point. We have added the following: "Since surgical expertise varies across countries and institutions, the is a need for the instrument to consistently apply to different healthcare settings." L137-139

Lines 174 -177: "We employed a series of statistical methods to explore reliability…"
Improving the reliability of AGREE-II tool is a good justification, but it is not clear why this is done for surgical guidelines alone.

Thank you for the opportunity to clarify this. The entire GAP project is focused on surgical guidelines. GAP I dealt with appraisal of surgical guidelines. GAP II used material from GAP I and applied statistical methods to calibrate the AGREE II instrument specifically for surgical guidelines. "We have used quality appraisal data from GAP I and employed a series of statistical methods to explore reliability, internal consistency and unidimensionality of the AGREE II instrument when it is applied in surgical guidelines." (L194-199)

Lines 233/234: "The second round will include closed-ended questions in a numeric Likert scale to assess participants".
The Likert scale range and interpretation should be provided.

Thank you for pointing this out. We have added: "As per protocol, 1/2 indicates strong/moderate disagreement, 3 indicates no opinion, and 4/5 indicates moderate/strong agreement." (L258-260)

Lines 236/237: "We will discard low-scoring items and use the shortlisted items in a third Delphi round".
What does 'low scoring items' mean in second round Delphi – will these be removed only based on the scores or will there be a certain level of agreement?

We have clarified this as follows: "We will discard low-scoring items (i.e. those with a median score of 1/2 on the Likert scale) and use the shortlisted items in a third Delphi round." (L261-263)

Lines 238/239: "We will repeat the process until new information reaches saturation and consensus with an alpha level of 0.8".
Is there any reference for this approach in Delphi? This sounds more like focus group meetings where you continue the focus groups until saturation is reached. There is no new information introduced after round 1 – as this is the only open-ended questionnaire. After this, it is only change of scores/agreement that occurs. What does alpha level mean here – alpha error? Clearly, never seen an alpha error of 0.8 being used. Is this agreement level? At what level of agreement is an item included or excluded?

Thank you for this comment. The term saturation is indeed confusing. We have changed the statement as follows: “We will repeat the process until an agreement of 80% (4/5 on the Likert scale) is reached among Delphi participants.” (L263-264)

Lines 244-251: “Qualitative synthesis”
More details about qualitative research synthesis (this is what is meant here and is different from qualitative synthesis which may refer to narrative synthesis of quantitative data or qualitative data) should be provided. Currently, it lacks any details.

We have now provided details on the methodology that we applied. (P271-299).

Line 289: “Publication and dissemination”
The intellectual property content should be clarified. Including a lead member of AGREE-II tool, on its own, does not mean that modifications/extensions for surgery can be done.

This project is the result of a partnership between an international team of surgical research experts and two of the AGREE research team leads (IF and MB), who are co-authors of this manuscript. The AGREE research team is currently under a membership renovation process, and, therefore, neither of the authors can speak on behalf of the entire AGREE group. However, both AGREE research team leads state that AGREE has supported the project from its inception. Furthermore, they have agreed to support dissemination activities by making this new tool available in the AGREE website (https://agreetrust.org). (L161-168)

Reviewer: 2

Reviewer Name

Paschalis Gavriilidis

Institution and Country

University Hospitals Birmingham NHS Trust Queen Elizabeth Hospital Department of Vascular Access and Renal Transplantation, Birmingham, UK

Please state any competing interests or state ‘None declared’:
None

Please leave your comments for the authors below
The protocol is well structured and written

Thank you for your comments.

Reviewer: 3

Reviewer Name

Mary E Brindle

Institution and Country

Cumming School of Medicine, University of Calgary, Calgary Canada

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

Please leave your comments for the authors below
This protocol paper proposes to develop an extension to the AGREE II framework to focus specifically on a guideline for developing guidelines in surgery.
The protocol is very well-written and the methods are excellent, thorough, well-considered and generally very well-developed. The eventual extension is important.

The authorship is international in its scope with broad representation of scientific and clinical backgrounds including authorship from AGREE and the Health Research Methods Evidence and Impact program at McMaster University. The broad authorship reflects the necessary diversity required to develop the guideline.

The introduction appropriately provides a rationale for the AGREE II extension broad scope of literature based on the wide sources of literature that are currently available to aid surgeons in decision-making and the recognition that the AGREE II framework may be inadequate to address the specific needs of surgical guidelines.

Overall the methods are rigorous. Appropriate patient representation is considered within the Delphi process. The use of pilot testing and assessment of internal validity are a great strength, as well as the plans to broadly disseminate, acquire feedback and update the extension as required.

There are a few minor areas where improvements could be considered.

1. argument for the inadequacy of the current AGREE II framework to specifically guide surgical guidelines could be strengthened. The example given that the current AGREE framework may not be appropriate specifically related to considering resource implications is concerning, as resource considerations will likely continue to be important in guideline development even in the absence of high quality cost effectiveness data. Surgery is not the only complex and multifaceted aspect of care that benefits from guidelines. Stronger support for why surgery, specifically, will benefit from this extension would strengthen the introduction.

Thank you. In response to comments from Reviewer 1 and Reviewer 3, we have added the following: "Since surgical expertise varies across countries and institutions, the is a need for the instrument to consistently apply to different healthcare settings. Surgical interventions are complex and detailed information on interventions/comparators are imperative for the target users to be able to assess the external validity of guidelines. Specialists from different specialties and allied health professionals with a wide range of expertise are involved in the treatment of surgical patients, which makes their involvement in guideline development paramount." (L137-144)

2. A brief description of the findings from GAP I are provided. A short description of the overarching findings from GAP II would also be of benefit.

Thank you for this suggestion. We have added the following: “Statistical modeling showed that excluding 5 items from the original tool (items 1, 2, 5, 7 and 8) and re-arranging the remaining items into 4 domains instead of 6 would enhance the instrument.“ (L203-205)

3. There are a few areas of the methods that would benefit from further details.
The investigators report that they will perform qualitative evidence synthesis to identify factors of conceptual importance of the quality of evidence in surgery. A few details on the type of QES undertaken (approach) would be of benefit

Thank you for this suggestion. Further details on the methodology of qualitative synthesis have been added (P271-299).

As well, the investigators report that they will survey users of social media to nominate parameters of importance in the development and reporting of guidelines in surgery. More details on who these users are, how they are identified, and how they would be surveyed will help the readers better understand this aspect of the proposal.

The term ’survey’ may be not inaccurate. We have modified the statement as follows: "Furthermore, we will invite users of social media through the project account on Twitter (@GAProject2) and through communication streams of the sponsoring bodies (Facebook, Twitter and email newsletters) to nominate parameters of importance..." (L300-302)