Surgical Endoscopy
https://doi.org/10.1007/s00464-022-09354-z

CONSENSUS STATEMENT

AGREE-S: AGREE II extension for surgical interventions: appraisal instrument

Stavros A. Antoniou[1,2] · Ivan D. Florez[3,4] · Sheraz Markar[5,6] · Patricia Logullo[7] · Manuel López-Cano[8] · Gianfranco Silecchia[9] · George A. Antoniou[10,11] · Sofia Tsokani[12] · Dimitrios Mavridis[12,13] · Melissa Brouwers[3] · The GAP Consortium

Received: 24 March 2022 / Accepted: 18 May 2022

Abstract

Background The Appraisal of Guidelines Research and Evaluation (AGREE) II instrument was developed to evaluate the quality of clinical practice guidelines. Evidence suggests that development, reporting, and appraisal of guidelines on surgical interventions may be better informed by modification of the instrument.

Objective We aimed to develop an AGREE II extension specifically designed for appraisal of guidelines of surgical interventions.

Methods In a three-part project funded by the United European Gastroenterology and the European Association for Endoscopic Surgery, (i) we identified factors that were associated with higher quality of surgical guidelines, (ii) we statistically calibrated the AGREE II instrument in the context of surgical guidelines using correlation, reliability, and factor analysis, and (iii) we undertook a Delphi consensus process of stakeholders to inform the development of an AGREE II extension instrument for surgical interventions.

Results Several features were prioritized by stakeholders as of particular importance for guidelines of surgical interventions, including development of a guideline protocol, consideration of practice variability and surgical expertise in different settings, and specification of infrastructures required to implement the recommendations. The AGREE-S—AGREE II extension instrument for surgical interventions has 25 items, compared to the 23 items of the original AGREE II instrument, organized into the following 6 domains: *Scope and purpose*, *Stakeholders*, *Evidence synthesis*, *Development of recommendations*, *Editorial independence*, and *Implementation and update*. As the original instrument, it concludes with an overall appraisal of the quality of the guideline and a judgement on whether the guideline is recommended for use. Several items were amended and rearranged among domains, and an item was deleted. The *Rigor of Development* domain of the original AGREE II was divided into *Evidence Synthesis* and *Development of Recommendations*. Items of the AGREE II domain *Clarity of Presentation* were incorporated in the new domain *Development of Recommendations*. Three new items were introduced, addressing the development of a guideline protocol, support by a guideline methodologist, and consideration of surgical experience/expertise.

Conclusion The AGREE-S appraisal instrument has been developed to be used for assessment of the methodological and reporting quality of guidelines on surgical interventions.

Keywords Clinical practice guideline · Surgery · Guideline quality · Quality appraisal · AGREE-S · AGREE II

The Appraisal of Guidelines Research and Evaluation (AGREE) is an international initiative aiming to improve the quality of patient care and health system performance through advances in the science and practice of clinical practice guidelines [1]. The AGREE instrument, first published in 2003 and updated as AGREE II in 2010, is an instrument for the systematic assessment of the methodological and the reporting quality of clinical practice guidelines [2]. The instrument assesses the methodological rigor and transparency with which a guideline has been developed. AGREE II can be used by health care providers and policy makers to assess a guideline before adopting its recommendations into practice [3].

The members of the GAP Consortium are listed in the Acknowledgements section.

✉ Stavros A. Antoniou
stavros.antoniou@hotmail.com

Extended author information available on the last page of the article

Published online: 15 June 2022

AGREE II is, however, a generic instrument, intended to be applied to guidelines in any field of medicine and public health, for disease diagnosis, prevention or treatment [4]. Surgery is a unique field, where outcomes of interventions are often operator-dependent, with experience, expertise, and infrastructure playing an important role. For instance, the external validity of a recommendation on a complex procedure may be affected by the experience and the expertise of the surgical team, and the availability of instruments and devices. Furthermore, details of the interventions, such as surgical instrumentation and types of prosthetic materials, may affect the intervention outcome. The research ecosystem of surgical interventions is also different from that of pharmacological interventions, with generally limited randomized evidence to inform clinical practice [5]. We hypothesized that an AGREE II extension specifically designed for guidelines on surgical interventions might better inform their development, reporting, and appraisal.

The design and development of the AGREE-S extension instrument to permit appraisal of guidelines for surgical interventions is reported in this article.

Methods

We developed a protocol for this tripartite project named *Guideline Assessment Project* (GAP) [6] which was composed of three stages that are described below.

GAP I

In the first part of this project, we assessed the quality of a sample of 67 clinical practice guidelines in the field of Surgery. Assessment of these guidelines using the generic AGREE II appraisal instrument, suggested that 40% were of below-average quality and may therefore not be recommended for use. In exploratory analyses, we found that guidelines produced by organizations with high guideline output (at least 1 guideline per year) and organizations with a guidelines committee were more likely to be recommended for use. Furthermore, using the GRADE methodology increased the odds of a guideline to be recommended for use by 8 times (odds ratio 8.2, 95% confidence interval 2.5–26.3). We coded these findings into items, which were nominated for inclusion in the extension document [7].

GAP II

In GAP II, we employed correlation, reliability, and factor analyses, and the item response theory to the same sample of guidelines using the AGREE II scores obtained from GAP I. Statistical modeling suggested removing or rearranging some items across domains and reducing the number of domains to 4 or 5 [8].

GAP III

In GAP III, the GAP Consortium convened a one-day meeting to discuss the findings of GAP I and GAP II, and to identify stakeholders for the Delphi exercise. The rationale behind the choice of panelists was to compose an inclusive and multidisciplinary group that would allow representativeness from various stakeholders' groups, professional, geographic, and ethnic backgrounds, and including patient advocates.

We summarized the research output from GAP I and GAP II, coded into candidate items for inclusion in or exclusion from an AGREE II extension for surgical interventions. We invited stakeholders from across different surgical specialties and different continents (SE, NL, LB, GdB, RG, AS, JJC, BWE, JLM); representatives from organizations advocating evidence-informed medicine and guideline development, and transparency in research (GRADE, Guidelines International Network—GIN, Enhancing the QUAlity and Transparency Of health Research—EQUATOR) (EA, PAC, PL); a surgical journal editor (NS); a guideline implementer (AS); a national authority representative (AM); and a patient representative (KI), to participate in an anonymous Delphi survey.

In the first round, panel members were asked to nominate candidate items to be included in an AGREE II extension for surgical interventions. Their responses were thematically synthesized and presented in a second round. In this round, Delphi participants were asked whether they agreed with inclusion or exclusion of the candidate items summarized from GAP I, GAP II, and the previous round. Agreement was documented using a 5-point Likert scale, 1 indicating disagreement and 5 indicating agreement, with an option of a response indicating no opinion. They were also asked to suggest modification of the proposed items.

Consensus was defined, as at least 80% of participants providing a score of 4 or 5. When consensus on an item was reached, the item was shortlisted for inclusion. When consensus was not reached in two Delphi rounds, the item was shortlisted for exclusion. After two rounds of Delphi, 5 new items and 1 modified AGREE II item were shortlisted for inclusion and 4 items were shortlisted for exclusion.

The GAP Consortium had an in-person consensus meeting, where findings of the three parts of the project were discussed and critically overviewed. Such findings informed the development of the AGREE-S instrument. Additional details on the methodology and the development process are provided in the Online Appendix.

AGREE-S appraisal instrument

The instrument comprises of 25 core items organized into 6 domains (Table 1).

1. Domain *Scope and purpose* refers to the objectives, the health question(s) and the protocol of the guideline.
2. Domain *Stakeholders* refers to the availability of methodological expertise, the composition of the guideline panel, and the definition of the target users.
3. Domain *Evidence synthesis* refers to the methods used to search, select, and appraise the evidence.
4. Domain *Development of recommendations* refers to what is broadly known as the evidence to decision framework. Relevant information pertains to the link between evi-

Table 1 AGREE-S: AGREE II extension for surgical interventions appraisal instrument

Domains	Items	Assessment*
1. Scope and purpose	1. The guideline has been developed according to a pre-established protocol and the link to it is provided	1–7
	2. The overall objective(s) of the guideline is (are) specifically described	1–7
	3. The health question(s) covered by the guideline [patient, interventions/procedures, outcomes] are specifically described	1–7
2. Stakeholders	4. The guideline was supported by a guideline development committee, including a guideline methodologist	1–7
	5. The guideline development group includes individuals from all relevant professional groups and patients	1–7
	6. The target users of the guideline are specifically described	1–7
3. Evidence synthesis	7. Systematic methods were used to search for evidence	1–7
	8. The criteria for selecting the evidence are clearly described	1–7
	9. The strengths and limitations of the body of evidence are clearly described	1–7
4. Development of recommendations	10. The views and preferences of the target population (patients, public, etc.) have been sought	1–7
	11. The methods for formulating the recommendations are clearly described	1–7
	12. The health benefits, side effects, and risks have been considered in formulating the recommendations	1–7
	13. There is an explicit link between the recommendations and the supporting evidence	1–7
	14. The recommendations are specific and unambiguous	1–7
	15. The different options for management of the condition or health issue are clearly presented	1–7
	16. Key recommendations are easily identifiable	1–7
	17. Potential resource implications of applying the recommendations have been considered	1–7
	18. The guideline considers potential variability in surgical expertise of those performing the interventions/procedures	1–7
5. Editorial independence	19. The views of the funding body have not influenced the content of the guideline	1–7
	20. Competing interests of guideline development group members have been recorded and addressed	1–7
6. Implementation and update	21. The guideline describes facilitators and barriers to its application	1–7
	22. The guideline has been externally reviewed by clinical and methodological experts prior to its publication	1–7
	23. A procedure for updating the guideline is provided	1–7
	24. The guideline provides advice and/or tools on how the recommendations can be put into practice	1–7
	25. The guideline presents monitoring and/or auditing criteria	1–7
Overall guideline assessment	26. Rate the overall quality of this guideline	1–7
	27. I would recommend this guideline for use	•Yes •Yes, with modifications •No

*7 corresponds to the highest possible quality

dence and recommendations, stakeholders' input, formulation and presentation of the recommendations, alternative options, and resource considerations, among others.
5. Domain *Editorial independence* refers to the role of the funding body, and reporting and management of conflicts of the guideline development group.
6. Domain *Implementation and update* refers to external review, facilitators and barriers, tools for implementation, update, and monitoring.

Each AGREE-S core item and the global rating item (final AGREE-S item for an overall assessment of the quality of the guideline) are rated on a scale between 1 and 7, 1 indicating the lowest possible quality, and 7 indicating the highest possible quality.

Calculating domain scores

Domain scores are calculated by summing up all the scores of the individual items in a domain and by scaling the total as a percentage of the maximum possible score for that domain.

Maximum possible score : 7 (strongly agree) × X (number of items in the domain) × Y (number of appraisers) = 7XY

Minimum possible score : 1(strongly disagree) × X (number of items in the domain) × Y (number of appraisers) = XY

The domain score is calculated with the formula:

(Obtained score − minimum possible score) / (Maximum possible score − minimum possible score)

An example is provided in Table 2.

Table 2 Example of calculation of domain score

	Item 1	Item 2	Item 3
Appraiser 1	5	6	1
Appraiser 2	4	5	2
Appraiser 3	5	7	1
Total	14	18	4

Calculation of domain score: (Obtained score − minimum possible score)/(Maximum possible score − minimum possible score)

Maximum possible score: 7 (strongly agree)×3 (items)×3 (appraisers)=63

Minimum possible score: 1 (strongly disagree)×3 (items)×3 (appraisers)=9

[(14+18+4) − (1×3×3)]/(63 − 9)=50%

Below we elaborate on each item in a number of bullet points to highlight the main issues to consider for the appraisal. They are not intended to be numerical subitems and the appraiser(s) need to consider them globally, and to decide what score would be the best for each item, according to their judgement of the elements considered. Summary differences between AGREE II and AGREE-S are provided in Table 3.

Domain 1. Scope and purpose

Item 1: The guideline has been developed according to a pre-established protocol and the link to it is provided.

- A protocol for development of the guideline was developed.
- The protocol was developed before commencing the guideline development process.
- The authors provide the protocol (e.g., as supplementary file), reference a protocol publication, or provide a functional link to the protocol.
- Amendments to the protocol are reported and justified.

Table 3 Summary differences between AGREE II and AGREE-S

SCOPE AND PURPOSE	No change in domain name
	New item 1. *The guideline has been developed according to a pre-established protocol and the link to it is provided*
1. The overall objective(s) of the guideline is (are) specifically described	No change in item. Renumbered to 2
2. The health question(s) covered by the guideline is (are) specifically described	Changed to *The health question(s) covered by the guideline [patient, interventions/procedures, outcomes] are specifically described* AND renumbered to 3
STAKEHOLDER INVOLVEMENT	Renamed to *STAKEHOLDERS*
	New item 4. *The guideline was supported by a guideline development committee, including a guideline methodologist*
3. The population (patients, public, etc.) to whom the guideline is meant to apply is specifically described	Deleted item
4. The guideline development group includes individuals from all relevant professional groups	Changed to *The guideline development group includes individuals from all relevant professional groups and patients* AND renumbered to 5
5. The views and preferences of the target population (patients, public, etc.) have been sought	Moved to Domain 4 *Development of recommendations* AND renumbered to 10
6. The target users of the guideline are clearly defined	Changed to *The target users of the guideline are specifically described* AND renumbered to 6
RIGOR OF DEVELOPMENT	Split to *Evidence synthesis* and *Development of recommendations*
	New domain *EVIDENCE SYNTHESIS*
7. Systematic methods were used to search for evidence	No change
8. The criteria for selecting the evidence are clearly described	No change
9. The strengths and limitations of the body of evidence are clearly described	No change
10. The methods for formulating the recommendations are clearly described	Moved to Domain 4 *Development of recommendations* AND renumbered to 11
	New domain *DEVELOPMENT OF RECOMMENDATIONS*
	The views and preferences of the target population (*patients*, *public*, etc.) *have been sought* Moved from *Stakeholder involvement* AND renumbered to 10
	The methods for formulating the recommendations are clearly described Moved from *Rigor of development* AND renumbered to 11
11. The health benefits, side effects, and risks have been considered in formulating the recommendations	Renumbered to 12
12. There is an explicit link between the recommendations and the supporting evidence	Renumbered to 13
13. The guideline has been externally reviewed by experts prior to its publication	Changed to *The guideline has been externally reviewed by clinical and methodological experts prior to its publication* AND moved to Domain 6 *Implementation and update* AND renumbered to 22
14. A procedure for updating the guideline is provided	Moved to Domain 6 *Implementation and update* AND renumbered to 23
15. The recommendations are specific and unambiguous	Renumbered to 14
16. The different options for management of the condition or health issue are clearly presented	Renumbered to 15
17. Key recommendations are easily identifiable	Renumbered to 16
	The potential resource implications of applying the recommendations have been considered Moved from *Applicability* AND renumbered to 17
	New item 18. *The guideline considers potential variability in surgical expertise of those performing the interventions/procedures*
APPLICABILITY	Deleted domain
18. The guideline describes facilitators and barriers to its application	Moved to Domain 6 *Implementation and update* AND renumbered to 21
19. The guideline provides advice and/or tools on how the recommendations can be put into practice	Moved to Domain 6 *Implementation and update* AND renumbered to 24

Table 3 (continued)

20. The potential resource implications of applying the recommendations have been considered	Moved to Domain 4 *Development of recommendations* AND renumbered to 17
21. The guideline presents monitoring and/or auditing criteria	Moved to Domain 6 *Implementation and update* AND renumbered to 25
EDITORIAL INDEPENDENCE	No change in domain name
22. The views of the funding body have not influenced the content of the guideline	Renumbered to 19
23. Competing interests of guideline development group members have been recorded and addressed	Renumbered to 20
	New domain I*MPLEMENTATION AND UPDATE*
	The guideline describes facilitators and barriers to its application Moved from *Applicability* AND renumbered to 21
	The guideline has been externally reviewed by clinical and methodological experts prior to its publication changed from *The guideline has been externally reviewed by experts prior to its publication* AND moved from *Rigor of development* AND renumbered to 22
	A procedure for updating the guideline is provided Moved from *Rigor of development* AND renumbered to 23
	The guideline provides advice and/or tools on how the recommendations can be put into practice Moved from *Applicability* AND renumbered to 24
	The guideline presents monitoring and/or auditing criteria Moved from *Applicability* AND renumbered to 25

Item 2: The overall objective(s) of the guideline is (are) specifically described.

- There are specific health intents (i.e., treatment, prevention, screening, diagnosis).
- There is/are (a) specific scope (i.e., to inform clinical decision making, to inform policy decisions).
- There are specific expected benefits or outcomes.
- There is a specific target population (e.g., patients, people at risk, public)

Item 3: The health question(s) covered by the guideline [patient, interventions/procedures, outcomes] are specifically described.

- There is (a) specific question framework(s); e.g., PICO format (patients, intervention(s), comparator(s), outcomes). For example, *"Should open versus endovascular repair be preferred in patients with abdominal aortic aneurysm?"*
- There is (are) (a) specific target population(s) (e.g., patients with disease, people at high risk; specific characteristics, such as gender, age, disease stage).
- There is (are) specific intervention(s), exposure(s) and/or diagnostic tests.
- There is (are) specific comparator intervention(s), exposure(s) and/or diagnostic test(s).
- There are specific outcomes, or a structured procedure was followed to specify outcomes of importance (e.g., systematic review of patients' views, survey of panel members, interviews, and/or focus groups).
- There is a specific health care setting or context (e.g., healthcare systems, centers of expertise).

Domain 2. Stakeholders

Item 4: The guideline was supported by a guideline development committee, including a guideline methodologist.

- The guideline development process was steered or overseen by a guideline development committee.
- The guideline methodology was developed by or in collaboration with a guideline methodologist.
- A qualified guideline methodologist (e.g., certified or with experience in guideline development) directed the guideline development process.

Item 5: The guideline development group includes individuals from all relevant professional groups and patients.

- All groups of stakeholders were represented in the guideline panel (e.g., surgeons, obstetricians, gynecologists, midwifes, patient advocates, directorate managers in a guideline panel for non-obstetric surgery in pregnancy).
- At least one patient representative or a patient advocate was a member of the guideline panel.
- All stakeholder groups and patient representatives or advocates had equal opportunities to contribute to the development of the guideline and to vote on the direction and strength of recommendation(s).

Item 6: The target users of the guideline are specifically described.

- The guideline was developed and written to be used by specific target users (e.g., surgeons, primary care physicians, patients, public).
- Target users are represented in the guideline development process (e.g., as panel members).
- Different versions of the guideline report were developed for specific target users (e.g., patient versions), if applicable.

Domain 3. Evidence synthesis

Item 7: Systematic methods were used to search for evidence.

- Electronic bibliographic databases or evidence sources were searched (e.g., MEDLINE, EMBASE, PsychINFO, CINAHL)
- Appropriate sources were searched (e.g., preprint databases for a novel intervention).
- Search terms, search syntaxes, and search limits were appropriate and provided in the guideline document or as an appendix/supplementary file.
- Search strategies were developed by experts in outreach, knowledge, and evidence search.
- Record screening was appropriately performed (e.g., parallel blinded).
- If one or more existing systematic reviews were used, their quality was assessed and was judged to be sufficient.

Item 8: The criteria for selecting the evidence are clearly described.

- There are specific and appropriate inclusion and exclusion criteria on study design.
- There are specific and appropriate inclusion and exclusion criteria on target population(s) (e.g., age, stage of disease).
- There are specific and appropriate inclusion and exclusion criteria on intervention(s), exposure(s) and/or diagnostic method(s), and comparator(s).
- Outcome selection and rating of their importance are explicitly and clearly described.
- Data extraction was appropriately performed (e.g., single reviewer extraction and cross checked by a second reviewer).

Item 9: The strengths and limitations of the body of evidence are clearly described.

- The evidence was assessed for risk of bias using appropriate methods (e.g., RoB-2 for randomized trials and ROBINS-I for non-randomized studies of interventions) on the outcome level (i.e., separate risk of bias assessment for each outcome, or summary risk of bias assessment for groups of outcomes, with justification).
- The certainty of the evidence was assessed using specific criteria at the outcome level (i.e., separate certainty of evidence assessment for each outcome).
- The certainty of the evidence was assessed using appropriate methods (e.g., GRADE methodology, which considers risk of bias, heterogeneity, imprecision, indirectness, publication bias, effect magnitude).
- The overall certainty of the evidence (across outcomes) was appropriately assessed (e.g., the lowest level of evidence certainty on a critical outcome defined the overall certainty of the evidence).

Domain 4. Development of recommendations

Item 10: The views and preferences of the target population (patients, public, etc.) have been sought.

- Patients'/public's views and preferences were sought after (e.g., participation in the guideline development group, evidence from the literature, surveys, interviews and/or focus groups)
- The methods by which patients'/public's preferences and views were considered in the development of recommendations were appropriate [e.g., contribution to the evidence to

decision framework discussion as an equal panel member, voting on direction and strength of recommendation(s)]
- Sociocultural acceptability, societal implications, human rights, health equity, equality and non-discrimination have been considered in guideline development.

Item 11: The methods for formulating the recommendations are clearly described.

- The criteria that were used to decide on the strength and the direction of the recommendation(s) were appropriate (e.g., GRADE evidence to decision framework).
- The process that was followed to decide on the strength and the direction of the recommendation(s) were appropriate (e.g., Delphi technique, voting procedures, level of consensus).

Item 12: The health benefits, side effects, and risks have been considered in formulating the recommendations.

- Health benefits with supporting data were considered when formulating the recommendations.
- Harms and/or side effects with supporting data were considered when formulating the recommendations.
- The balance between health benefits and harms/side effects was considered when developing the recommendation(s).

Item 13: There is an explicit link between the recommendations and the supporting evidence.

- The process through which the guideline development group used the evidence to inform recommendations was appropriate (e.g., evidence tables, evidence to decision framework with consideration of the balance between benefits and harms; systematic observation form to retrieve expert-based evidence; evidence on cost, patients' values and preferences, acceptability and feasibility).

Item 14: The recommendations are specific and unambiguous.

- There is (are) (a) clearly recommended course(s) of action.
- The wording of the recommendation(s) reflects its (their) strength (e.g., *'We recommend'* for strong recommendations; *'We suggest'* for weak/conditional recommendations).
- The direction of the recommendation(s) is clear (e.g., *'We suggest laparoscopic over open cholecystectomy'*).
- Good practice statements are explicitly reported as such and are not graded.
- Caveats or qualifying statements, if relevant, are provided (i.e., patients or conditions or settings where the recommendations would not apply; e.g., *"We suggest transanal total mesorectal excision over laparoscopic total mesorectal excision if surgical expertise is available"*).

Item 15: The different options for management of the condition or health issue are clearly presented.

- Different options and alternatives for treatment, management, diagnosis, prevention, or screening have been considered.
- Patients, populations or clinical settings most appropriate to each option or alternative are specified.

Item 16: Key recommendations are easily identifiable.

- Recommendations are provided in a summarized box, typed in bold, underlined, or presented as flow charts or algorithms.
- Specific recommendations are grouped together in one section.

Item 17: Potential resource implications of applying the recommendations have been considered.

- Costs and resources utilization (e.g., up-to-date costs of the different alternative interventions) were obtained and presented to the panel, or cost-effectiveness analyses (own-developed or available in literature) were used to inform the panel.
- There was an appropriate consideration of cost to inform the guideline development process and/or formulation

of the recommendations (e.g., as part of the evidence to decision framework).
- Methods by which cost information was sought was appropriate (e.g., only costs description of the interventions, economic evaluations, or cost–benefit analyses, and if no such information was available, input from an expert in economic analyses, or from the experience from the guideline panel).
- Human resources, health and care infrastructure and settings, and organization of health services have been considered in the guideline development process.

Item 18: The guideline considers potential variability in surgical expertise of those performing the interventions/procedures.

- Experience and expertise of those performing the interventions, diagnostic procedures etc. was considered when developing the recommendation(s) (e.g., previous courses, hands-on training, previous experience with specific number of procedures/interventions).
- Experience and expertise of those performing the interventions, diagnostic procedures etc. in the source studies was documented and considered when developing the recommendations.

Domain 5. Editorial independence

Item 19: The views of the funding body have not influenced the content of the guideline.

- An explicit statement from the authors that the funding body did not influence the content of the guideline is provided.
- The name of the funding body or source of funding, or statement of no funding is provided.

Item 20: Competing interests of guideline development group members have been recorded and addressed.

- Potential financial and non-financial competing interests of all members of the guideline development group (e.g., steering group, evidence outreach group, statisticians' group, panel members, external advisors etc.) were sought at the outset and upon completion of the guideline development process.
- Potential competing interests are provided in detail in the guideline manuscript, supplementary file, or a valid link.
- The steering group managed appropriately both financial and non-financial (intellectual, professional etc.) competing interests (e.g., no steering group members with financial or non-financial conflicts; members of the guideline development group with competing interests not allowed to vote on strength, direction, and formulation of recommendations).

Domain 6. Implementation and update

Item 21: The guideline describes facilitators and barriers to its application.

- The guideline identifies the types of facilitators and barriers that were considered when formulating the recommendation(s) and those that are expected to be encountered when implementing the recommendation(s) (e.g., lack of expertise, resistance to change, organizational culture).
- Appropriate methods were used to collect information regarding facilitators and barriers to implementing recommendations (e.g., feedback from key stakeholders, pilot testing of guidelines before widespread implementation).
- The guideline describes how the collected information influenced the guideline development process and/or formation of the recommendations (e.g., part of the evidence to decision framework).

Item 22: The guideline has been externally reviewed by clinical and methodological experts prior to its publication.

- If submitted for journal publication, review by both content and methodological experts were proposed to the editorial board.
- If not published in a peer-reviewed journal (e.g., website, newsletter), a peer review by both content and methodological experts was performed.
- If not published in a peer-reviewed journal, the purpose and intent of the external review are reported (e.g., to improve quality, gather feedback on draft recommendations, assess-applicability and feasibility, disseminate evidence).

- If not published in a peer-reviewed journal, the methods followed to undertake the external review (e.g., rating scale, open-ended questions, or the AGREE-S appraisal instrument) are reported.
- If not published in a peer-reviewed journal, information on external reviewers is provided (e.g., number, type of reviewers, qualifications, expertise, affiliations).
- If not published in a peer-reviewed journal, the outcomes/information gathered from the external review (e.g., summary of key findings) and description of how the information gathered was used to inform the guideline development process and/or formation of the recommendations are provided (e.g., guideline panel considered results of review in forming final recommendations).

Item 23: A procedure for updating the guideline is provided.

- A plan for updating the guideline is provided.
- Specific time interval or explicit criteria to guide guideline updates are provided (e.g., expected publication of new studies that will inform the evidence, identified through a scoping review of trial registries).

Item 24: The guideline provides advice and/or tools on how the recommendations can be put into practice.

- The guideline provides tools and resources to facilitate application, such as guideline summary documents, patient, or public version of the guideline or lay summaries, treatment/management algorithms, how-to manuals, solutions linked to barrier analysis, tools to capitalize on guideline facilitators, outcome of pilot test and lessons learned, decision aids).

Item 25: The guideline presents monitoring and/or auditing criteria.

- A plan for monitoring or measuring uptake of the guideline is provided through the development of indicators based on key recommendations (e.g., survey of target users 2 years after publication).
- Specific criteria to assess guideline implementation or adherence to key recommendations were set (e.g., adoption by at least 80% of target users).
- The guideline sets criteria for assessing the impact of implementing key recommendations (e.g., 50% reduction in surgical site infection).
- The guideline provides advice on the frequency and interval of measurement.

Overall score

Upon completing the assessment of the individual items within each domain, AGREE-S users are asked to provide 2 global assessments of the guideline. The overall assessment requires the user to make a judgment as to the quality of the guideline, considering the criteria in the assessment process. The user is also asked whether he/she would recommend use of the guideline.

How the instrument should be used

As in the original AGREE II instrument [4], the overall assessment requires the user to make a judgment as to the quality of the guideline, taking into account the criteria considered in the assessment process. Although users may place greater value, e.g., on the *Evidence synthesis* and the *Development of recommendations* domains, we do not recommend using cutoff domain scores that will inform judgements on global assessment. Furthermore, there is no overall score threshold suggesting that a guideline should be recommended for use.

Several items of this instrument might not be applicable to guidelines on surgical interventions that include question frameworks with non-interventional components. For example, for the assessment of a guideline on the duration of thrombosis prophylaxis after cancer surgery, the items on surgical expertise will not be applicable. Users are advised to adjust their scoring (i.e., the maximum score) when calculating domain scores.

A formal assessment of the validity of the AGREE-S appraisal instrument has not been undertaken as yet. Until results of the assessment will be published in the AGREE-S website, www.agree-s.org; the AGREE Trust website, www.agreetrust.org; and in peer-reviewed journal publications, we recommend that each guideline be assessed by at least 2 appraisers, preferably 4 [4]. The AGREE-S website www.agree-s.org provides an interactive interface for completing appraisals and downloading the results in a printable document.

Who can use the instrument

AGREE-S is intended to be used by the following stakeholder groups:

- health care providers (surgeons, primary care physicians, allied healthcare professionals) who wish to undertake their own assessment of a guideline before adopting its recommendations into their practice;
- guideline developers, to support them in following a structured and rigorous development methodology, conducting an internal assessment to ensure that their guidelines are sound, or evaluating guidelines from other groups for potential adaptation to their own context or de novo development;
- scientific societies, to assist in deciding which guidelines to endorse and in selecting which guidelines to adapt, thereby overcoming the need to develop further national or international guidelines;
- policy makers, to help them decide which guidelines could be recommended for use in practice or to inform policy decisions; and
- educators, to help enhance critical appraisal skills amongst health professionals and to teach core competencies in guideline development, reporting, and appraisal.

Limitations

The AGREE-S appraisal instrument has not been pilot tested nor tested for content validity or utility yet. Users are instructed to monitor project updates in the AGREE-S website, www.agree-s.org. Because of the multitude of stakeholder groups, only one representative from each group participated in the Delphi process. This has facilitated operationalization of the process; however, it may have narrowed the spectrum of input from each stakeholder group.

Conclusion

We used a protocol-based, transparent, and rigorous methodology to develop the AGREE-S instrument, which is an extension of the AGREE II instrument designed for developing, reporting, and appraising guidelines on surgical interventions.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s00464-022-09354-z.

Acknowledgements Collaborative authorship: *The GAP Consortium*: Yasser Sami Abdel Dayem, Luca Bertolaccini, Pablo Alonso-Coello, Elie Akl, Manish Chand, John J Como, Gert J. de Borst, Salomone Di Saverio, Sameh Emile, Bang Wool Eom, Ramon Gorter, George Hanna, Kaisa Immonen, Quirino Lai, Nicolaas Lumen, Joseph L. Mathew, Alessandro Montendori, Martin Moya, Gianluca Pellino, Alvaro Sanabria, Athanasios Saratzis, Neil Smart, Dimitrios Stefanidis, Giovanni Zaninotto.

Funding The Guideline Assessment Project (GAP) III received financial support from the *United European Gastroenterology* (UEG) and from the *European Association for Endoscopic Surgery and Other Interventional Techniques* (EAES), both non-profit organizations. The funders had no role in the design or development of this project.

Declarations

Disclosure Ivan Florez and Melissa Brouwers are leads of the AGREE Consortium. Stavros A. Antoniou, Ivan D. Florez, Sheraz Markar, Patricia Logullo, Manuel López-Cano, Gianfranco Silecchia, George A. Antoniou, Sofia Tsokani, Dimitrios Mavridis and Melissa Brouwers have no conflicts of interest or financial ties to disclose. Detailed conflicts of interest statements of the authors are provided in http://osf.io/fau4d.

Ethical approval The project was evaluated by the NHS/HSC Research Ethics Committee and ethics approval was waived.

Patient consent Not applicable.

Permission to reproduce material from other sources The authors obtained permission from the AGREE Consortium to reproduce parts of the AGREE II User's Manual and 23-item Instrument.

References

1. Makarski J, Brouwers MC, AGREE Enterprise (2014) The AGREE Enterprise: a decade of advancing clinical practice guidelines. Implement Sci 9:103. https://doi.org/10.1186/s13012-014-0103-2
2. Brouwers MC, Kho ME, Browman GP, Burgers JS, Cluzeau F, Feder G, Fervers B, Graham ID, Grimshaw J, Hanna SE, Littlejohns P, Makarski J, Zitzelsberger L, AGREE Next Steps Consortium (2010) AGREE II: advancing guideline development, reporting and evaluation in health care. J Clin Epidemiol 63(12):1308–11. https://doi.org/10.1016/j.jclinepi.2010.07.001
3. Brouwers MC, Kho ME, Browman GP, Burgers JS, Cluzeau F, Feder G, Fervers B, Graham ID, Grimshaw J, Hanna SE, Littlejohns P, Makarski J, Zitzelsberger L, AGREE Next Steps Consortium (2010) AGREE II: advancing guideline development, reporting and evaluation in health care. CMAJ 182(18):E839-42. https://doi.org/10.1503/cmaj.090449
4. AGREE Next Steps Consortium (2009). The AGREE II instrument [electronic version]. http://www.agreetrust.org. Accesssed 20 Dec 2021
5. McCulloch P, Taylor I, Sasako M, Lovett B, Griffin D (2002) Randomised trials in surgery: problems and possible solutions. BMJ 324(7351):1448–1451. https://doi.org/10.1136/bmj.324.7351.1448
6. Antoniou GA, Mavridis D, Tsokani S, López-Cano M, Flórez ID, Brouwers M, Markar SR, Silecchia G, Francis NK,

Antoniou SA (2020) Protocol of an interdisciplinary consensus project aiming to develop an AGREE II extension for guidelines in surgery. BMJ Open 10(8):e037107. https://doi.org/10.1136/bmjopen-2020-037107

7. Antoniou SA, Tsokani S, Mavridis D, López-Cano M, Antoniou GA, Stefanidis D, Francis NK, Smart N, Muysoms FE, Morales-Conde S, Bonjer HJ, Brouwers MC (2019) Guideline assessment project: filling the GAP in surgical guidelines: quality improvement initiative by an international working group. Ann Surg 269(4):642–651. https://doi.org/10.1097/SLA.0000000000003036
8. Tsokani S, Antoniou SA, Moustaki I, López-Cano M, Antoniou GA, Flórez ID, Silecchia G, Markar S, Stefanidis D, Zanninotto G, Francis NK, Hanna GH, Morales-Conde S, Bonjer HJ, Brouwers MC, Mavridis D (2021) Guideline assessment project II: statistical calibration informed the development of an AGREE II extension for surgical guidelines. Surg Endosc 35(8):4061–4068. https://doi.org/10.1007/s00464-021-08604-w

Authors and Affiliations

Stavros A. Antoniou[1,2] · Ivan D. Florez[3,4] · Sheraz Markar[5,6] · Patricia Logullo[7] · Manuel López-Cano[8] · Gianfranco Silecchia[9] · George A. Antoniou[10,11] · Sofia Tsokani[12] · Dimitrios Mavridis[12,13] · Melissa Brouwers[3] · The GAP Consortium

1 Department of Surgery, Mediterranean Hospital of Cyprus, Limassol, Cyprus

2 European University Cyprus, Nicosia, Cyprus

3 Department of Health Research Methods, Evidence and Impact, McMaster University, Hamilton, ON, Canada

4 Department of Pediatrics, University of Antioquia, Medellin, Colombia

5 Nuffield Department of Surgery, University of Oxford, Oxford, UK

6 Department of Molecular Medicine and Surgery, Karolinska Institute, Stockholm, Sweden

7 UK EQUATOR Centre, Centre for Statistics in Medicine, Nuffield Department of Orthopaedics, Rheumatology & Musculoskeletal Sciences, University of Oxford, Oxford, UK

8 Abdominal Wall Surgery Unit, Val d' Hebrón University Hospital, Universidad Autónoma de Barcelona, Barcelona, Spain

9 Department of Medico-Surgical Sciences and Translation Medicine, Faculty of Medicine and Psychology, Sapienza University of Rome, Rome, Italy

10 Department of Vascular and Endovascular Surgery, Manchester University NHS Foundation Trust, Manchester, UK

11 Division of Cardiovascular Sciences, School of Medical Sciences, The University of Manchester, Manchester, UK

12 Department of Primary Education, School of Education, University of Ioannina, Ioannina, Greece

13 Paris Descartes University, Sorbonne Paris Cité, Faculté de Médecine, Paris, France