



Manual for ESHRE guideline development

Authors: N. Vermeulen, N. Le Clef, S. Mcheik, A. D'Angelo, K. Tilleman, Z. Veleva, W. Nelen

Table of contents

| | | | |
|---|----|---|----|
| Background information | 4 | Role of qualitative research | 18 |
| Clinical practice guidelines | 4 | 6. EVIDENCE SYNTHESIS | 20 |
| Aim and structure of this manual | 4 | Relevance and Quality check | 20 |
| ESHRE guidelines | 4 | Evidence tables | 20 |
| 1. TOPIC SELECTION | 8 | GRADE Evidence profiles and Summary of Findings Tables | 20 |
| Selection procedure | 8 | Quality of evidence for each outcome | 21 |
| Application procedure | 8 | Factors determining the quality of evidence | 22 |
| 2. FORMING GUIDELINE DEVELOPMENT GROUP | 9 | 7. DEVELOPING RECOMMENDATIONS | 26 |
| Composition guideline development group | 9 | Strong or weak recommendations | 26 |
| Chair of the Guideline Development Group | 9 | Wording of recommendations | 28 |
| GDG selection procedure | 9 | Research recommendations | 29 |
| Responsibilities of guideline development group members | 10 | 8. WRITING THE GUIDELINE DRAFT | 31 |
| Patient participation | 10 | Principles for writing | 31 |
| Handling Conflicts of Interest | 10 | Guideline structure | 31 |
| Consensus | 10 | 9. STAKEHOLDER CONSULTATION | 33 |
| 3. SCOPING THE GUIDELINE | 12 | Review procedure | 33 |
| Scoping procedure | 12 | 10. APPROVAL | 35 |
| 4. FORMULATING KEY QUESTIONS | 13 | Final version & authorization | 35 |
| Developing and selecting key questions | 13 | 11. PUBLICATION, DISSEMINATION, TRANSLATION AND IMPLEMENTATION | 36 |
| Defining key questions as PICO questions | 13 | Publications | 36 |
| 5. SEARCHING EVIDENCE | 17 | Dissemination | 36 |
| Literature search | 17 | Guideline translation and endorsement | 36 |
| Selection of evidence | 18 | ESHRE® copyright statement | 39 |

| | |
|--|----|
| ESHRE® disclaimer | 39 |
| Implementation and evaluation | 40 |
| 12. UPDATING THE GUIDELINE | 42 |
| REFERENCE LIST | 44 |
| FORMS AND CHECKLISTS | 46 |
| A. APPLICATION FORM | 47 |
| B. DISCLOSURE FORM | 48 |
| C. CONFIDENTIALITY FORM | 49 |
| D. SCOPING CHECKLIST | 50 |
| E. PICO CHECKLIST | 51 |
| F. RELEVANCE AND QUALITY CHECK | 52 |
| G. QUALITY ASSESSMENT CHECKLIST | 53 |
| H. EVIDENCE TABLES | 70 |
| I. FRAMEWORK FOR RECOMMENDATIONS | 73 |
| J. REVIEWER COMMENTS FORM | 75 |
| K. DOCUMENT ASSESSMENT TOOL FOR UPDATING GUIDELINES | 77 |

Background information

Clinical practice guidelines

Clinical practice guidelines are defined as “systematically developed statements to assist care providers and patient decisions about appropriate health care for specific clinical circumstances” [1]. Clinical guidelines are intended as neither cookbook nor textbook but should be helpful in everyday clinical medical decision-making about appropriate and effective care. Therefore, clinical guidelines should be presented in formats easy to interpret.

Aim and structure of this manual

The principal aim of this manual is to provide stepwise advice to individual members of ESHRE guideline development groups (GDG).

The manual is based on the ESHRE manual for guideline development 2017 and draws on the most up-to-date evidence on international guideline development methodology and resources available, including Grading of Recommendations Assessment, Development, and Evaluation (GRADE).

GRADE: Key Points

- GRADE offers a transparent and structured process for developing and presenting summaries of evidence, including its quality, for systematic reviews and recommendations in health care.
- GRADE provides guideline developers with a comprehensive and transparent framework for carrying out the steps involved in developing recommendations.
- GRADE's use is appropriate and helpful irrespective of the quality of the evidence: whether high or very low.
- Although the GRADE system makes judgments about quality of evidence and strength of recommendations in a systematic and transparent manner, it does not eliminate the inevitable need for judgments

In addition, the manual is based on internationally acceptable criteria of methodological quality, as articulated by the Appraisal of Guidelines for Research and Evaluation in Europe (AGREE) instrument [2,3]. All 23 items of the AGREE Reporting checklist were incorporated in the manual and listed as tips at the end of each chapter.

The structure of this manual follows guideline development from its proposal through to publication and beyond.

This ESHRE manual is intended to be a “living” publication and it will be updated regularly based on new developments in guideline development and experiences in the guideline groups. Comments on either content or presentation are welcome and should be sent

to guidelines@eshre.eu. At the time of change, GDG members will be notified.

Previous versions

- Manual for ESHRE guideline development v1. 2007
- Manual for ESHRE guideline development v2. 2014
- Manual for ESHRE guideline development v3. 2017

Details on the update 2019

In addition to some minor adaptations and corrections, 2 major adaptations were made in the current manual:

- Adaptation of the methodology for forming a guideline development group, with more stringent rules and approval by the Executive Committee.
- Addition of a chapter on translation and adaptation of the ESHRE Guidelines, outlining the different policies.

ESHRE guidelines

The main goal of ESHRE guideline development is the provision of clinical recommendations to improve the quality of health care delivery within the European field of human reproduction and embryology. (For more information on ESHRE visit www.eshre.eu).

ESHRE guideline development methodology is similar to the methodology of other societies [5–7,4] and complies with the criteria used by the AGREE instrument for good quality guidelines [2,3]. Furthermore, all relevant items of the Guidelines International network (GIN) Guideline Development Checklist were included [8,9].

ESHRE clinical guidelines contain recommendations on a particular clinical issue. These guidelines are based on the best available evidence (most relevant and highest level of evidence) and not on all evidence available. There is an explicit link between recommendations and their available evidence. Furthermore, scientific and clinical evidence take precedence over expert judgement.

ESHRE Guidelines will not include a formal analysis of cost effectiveness of recommended as compared to established practice, as this is not the main aim, and is sometimes impossible because of the obvious differences in current European economic and healthcare systems. The clinical and organizational impact of costs on recommendations will be considered in GDG meetings, and if relevant, described in the justification section. The economic feasibility of recommendations will not be covered.

ESHRE guidelines can be adapted and translated by National Societies ensuring more efficient use of resources and improvement of patient outcomes throughout Europe. ESHRE guidelines should therefore be flexible and adaptable such that individual circumstances can be taken into consideration. ESHRE has established a

policy for translation of its guidelines to ensure quality and validity of translated documents (see also chapter 11).

Medico-legal implications of ESHRE guidelines

Potential medico-legal implications of clinical guidelines have been of ongoing concern to medical practitioners [10]. However, clinical guidelines are intended as an aid to clinical judgement, not to replace it. The ultimate decision about a particular clinical procedure or treatment will always depend on each individual patient's condition, circumstances and wishes, and the clinical judgement of the healthcare team as is represented within the disclaimer in the beginning of each guideline. Clinical guidelines are not intended to deprive clinicians of their medical freedom to treat, nor relieve them of their responsibility to make appropriate decisions based on their own knowledge and experience.

To clarify the legal perspective all ESHRE guidelines carry the following statement in the disclaimer:

The European Society of Human Reproduction and Embryology (hereinafter referred to as 'ESHRE') developed the current clinical practice guideline, to provide clinical recommendations to improve the quality of healthcare delivery within the European field of human reproduction and embryology. This guideline represents the views of ESHRE, which were achieved after careful consideration of the scientific evidence available at the time of preparation. In the absence of scientific evidence on certain aspects, a consensus between the relevant ESHRE stakeholders has been obtained.

The aim of clinical practice guidelines is to aid healthcare professionals in everyday clinical decisions about appropriate and effective care of their patients.

However, adherence to these clinical practice guidelines does not guarantee a successful or specific outcome, nor does it establish a standard of care. Clinical practice guidelines do not replace the need for application of clinical judgment to each individual presentation, nor variations based on locality and facility type. Ultimately, healthcare professionals must make their own clinical decisions on a case-by-case basis, using their clinical judgment, knowledge, and expertise, and taking into account the condition, circumstances, and wishes of the individual patient, in consultation with that patient and/or the guardian or carer.

ESHRE makes no warranty, express or implied, regarding the clinical practice guidelines and specifically excludes any warranties of merchantability and fitness for a particular use or purpose. ESHRE shall not be liable for direct, indirect, special, incidental, or consequential damages related to the use of the information contained herein. While ESHRE makes every effort to compile accurate information and to keep it up to date, it

cannot, however, guarantee the correctness, completeness, and accuracy of the guideline in every respect. In any event, these clinical practice guidelines do not necessarily represent the views of all clinicians that are member of ESHRE.

The information provided in this document does not constitute business, medical or other professional advice, and is subject to change.

Guideline development in 12 steps

Guideline development, implementation, and evaluation is no linear process, but a cycle of interdependent activities. Key steps within this process are topic selection, synthesis of evidence, formulation of recommendations, consultation and review, dissemination and implementation, evaluation and updating.

Timelines

The time taken to develop an ESHRE guideline varies according to the scope of the topic, the volume of relevant literature, the amount of feedback received, and the time needed to reach consensus about some topics. In general, it is recommended to keep the guideline to a reasonable size to ensure its development within an 18–24-month period.

Budget

In an effort to cut costs and time, it is strongly recommended to organize GDG meetings in conjunction with other meetings/congresses. The use of e-mails and teleconference for communication is also strongly encouraged to increase efficiency and avoid unnecessary meetings and travel.

A fixed budget is set to cover the costs of necessary meetings of a GDG. These expenses cover meeting costs, including travel (economy class tickets), accommodation, food and meeting facilities. Costs are reimbursed upon request within four weeks, on presentation of original receipts, invoices, bills, tickets etc., together with a provided ESHRE expense claim form.

ESHRE GUIDELINE DEVELOPMENT: 12 STEPS

1. Topic selection

2. GDG Formation

3. Scoping

4. Key questions

5. Evidence search

6. Evidence synthesis

7. Recommendations

8. Writing draft

9. Stakeholder review

10. Approval

11. Publication

12. Updating / Revising

Summary of meetings and timelines



* The development should start as soon as a proposal is approved, but can be postponed due to other projects or workload.

01. Topic Selection

Selection procedure

The Coordinators of all Special Interest Groups (SIGs) are regularly invited to propose new guideline topics. These proposals are made on an application form (Appendix A), and subjects chosen are within the field of reproductive medicine and embryology with the aim of assisting physicians and laboratory staff in diagnosis and/or clinical management.

Individual ESHRE members wanting to present a guideline topic are encouraged to contact the relevant SIG coordinator, who will be responsible for submitting the application form (by emailing guidelines@eshre.eu).

ESHRE uses a 2-step selection procedure to decide which proposals for guideline development will be accepted. In a first step, all proposals will be checked by the ESHRE research specialist and, if relevant, a representative of the SIG SQART based on the following criteria:

- Is the proposal complete?
- Is the proposal clear and well organized?
- Are the details in the proposal correct?
- Is the topic appropriate for an evidence-based “Clinical” guideline?

The research specialist will also add information on existing guidelines and overlap with other ESHRE guidelines. To check the presence of existing guidelines the following websites will be screened: Guidelines International Network (www.g-i-n.net), National Guideline Clearinghouse (www.guideline.gov), American Society for Reproductive Medicine (www.asrm.org), the National Institute for Clinical Excellence (www.nice.org.uk) and the Royal College of Obstetricians and Gynecologists (www.rcog.org.uk). All guidelines found must be evaluated by using the AGREE checklist. Adaptation of existing guidelines (rather than developing new guidelines) can be considered; methodologies for adaptation are available [11,12]. The Cochrane Library should be consulted to estimate the available existing evidence.

If necessary, additional information is requested from the applicant to complete the proposal before submission to the ESHRE executive committee.

In a second step the ESHRE executive committee evaluates the application for guideline development and decides whether the proposal is acceptable for ESHRE guideline development. The ESHRE executive committee may suggest revisions to the application. If not acceptable, the applicant will be informed of the reason. The decision of the executive committee and any comments will be communicated to the applicant by the research specialist.

Appropriate topic selection is important to ensure that an ESHRE guideline is relevant and addresses priority issues for the improvement of European reproductive medicine.

Within the selection procedure priority is given on topics with:

- high volume
- high costs
- major patient impact (e.g. health burden or high risks)
- high practice variation (within Europe)
- high ethical/legal impact
- high improvement potential.

Application procedure

The guideline application form (Form **A**) can be requested via email (nathalie@eshre.eu). Completed application forms should be sent to the ESHRE research specialist.

Proposals can be added at any time and will generally be evaluated at the next meeting of the Executive committee.



Available forms

Application form **A**

02. Forming guideline development group

Convening an effective guideline development group (GDG) is a crucial stage in producing a guideline; the GDG agrees on the key questions, considers the evidence and has considerable influence on the final guideline recommendations [13]. Therefore, it is strongly recommended that representatives of all key groups and disciplines affected by a guideline topic participate.

Composition guideline development group

Diversity is an essential feature of a GDG, and its exact composition should be tailored to the guideline topic (and scope) and reflect the range of stakeholders involved. A GDG should comprise at least:

- content expert(s)
- non-expert clinician(s)
- a patient or their representatives
- allied health care provider(s) and an
- ESHRE research specialist.

Industry representatives are excluded from membership.

A maximum of 10 to 15 GDG members are recommended in addition to the chairperson. Simultaneous membership of more than one active GDG or GPR WG is generally not recommended. Simultaneous membership of more than 2 active¹ GDG or GPR WGs is not allowed.

In composing a GDG, the following points should be considered:

- balance in geographical location; representatives from all parts of Europe
- balance in gender
- balance in expertise (academic, non-academic, senior, junior, ...)

Depending on the guideline topic, a representative from a related society might be considered for membership of the GDG. In the case of a joint guideline development with partner organizations, the Executive Committee must approve the collaboration (preferably at the same time as the application).

Chair of the Guideline Development Group

The chairperson of the GDG is either the applicant, the responsible SIG coordinator, or any GDG member with appropriate expertise, and team-working skills. A GDG chair is appointed for a period of

four years and should be a respected content expert, experienced in group facilitation, maintaining constructive dynamics, identifying and resolving conflicts, remaining neutral and objective, and having methodological expertise.

GDG selection procedure

When a topic is accepted for guideline development, the applicant/responsible SIG coordinator is invited to propose GDG members.

First, the applicant/responsible SIG coordinator should consider inviting one representative of each of the relevant ESHRE Special Interest Groups. Experts in the topic of interest can also be invited to join the GDG. Finally, an application process ("open call") can be set up by the research specialist where ESHRE members are asked to apply for a position in the GDG.

Independent of how they were recruited, everyone with an interest of joining the GDG will be asked to send a short cv, a motivation on why s/he should be included in the GDG and the completed COI form (form **B**). Based on the provided information, the profiles to be included (as above) and considering the balance in gender, geography and expertise, the applicant/responsible SIG coordinator prepares a proposal for the GDG composition. This proposal is to be discussed and ratified by ExCo before the GDG can be formalized.

At the start of the guideline development, all GDG members, except for patient representatives and invited experts, should be a member of ESHRE.

Once all members have agreed to participate, the GDG can become functional.

New members should usually not be added to the GDG once the development process has started. Additional needed expertise or the replacement of a GDG member should be discussed within the GDG group and approved by ExCo. The research specialist should ensure that new GDG members have all information on the previous steps in the guideline development and receive training similar to the rest of the GDG.

¹ Active meaning from the first GDG meeting (scope) to the last GDG meeting (after stakeholder review)

Responsibilities of guideline development group members

To ensure that the GDG functions effectively and achieves its aims, all GDG members should engage to the following responsibilities:

- Attend all GDG meetings ²
- Sign a statement of confidentiality at the start of the project
- Declare of any conflict of interest (in case of changes, and at least annually)
- Contribute to the formulation of clinical questions (PICO questions)
- Assess and summarize papers for a specific section of the guideline
- Write a summary of evidence and draft recommendations for a specific section of the guideline
- Participate in discussion and decision making, with acceptance and tolerance of varying viewpoints
- Approve of the final recommendations

The GDG will be supported by an ESHRE research specialist who will be responsible for overall project management and organizing the meetings in collaboration with the chair of the guideline group. In addition, the research specialist will provide specific training on the different steps in guideline development during the guideline meetings. The aim of such training is to increase and equalize the level of guideline development expertise within a GDG. Finally, the research specialist will perform the literature searches, and collect all input in one guideline document.

Patient participation

Patient involvement in guideline development is important to ensure reflection of their needs, concerns and preferences, as they may have different perspectives on healthcare processes, priorities, and outcomes from those of health care professionals. Ideally, patients are involved starting from the scoping process [14]. Patient needs and preferences should be for each guideline at least be considered with respect to:

- information
- communication
- health care organization
- financial constraints

→ shared decision making, and

→ self-management.

For the identification of patients' views other methodologies can also be applied, including literature search, patient (organization) consultation e.g. by (focus group) interviews, and/or guideline review by patients or their representatives.

Handling Conflicts of Interest

Because ESHRE aims to ensure objectivity and independence in its European guidelines, the guidelines are developed without external funding. All GDG chairpersons and members, have to provide disclosure statements of all potential conflicts of interest and confidentiality (see forms [B](#) and [C](#)). To ensure objectivity of the guideline, group members with conflicts of interest in specific topics can be excluded from performing evidence selection on one or more key questions. In addition, it can be relevant to ask a second GDG member to check and evaluate a certain key question.

The issue of conflicts of interest and how these are handled should be discussed within the guideline development group before the evidence is selected and evaluated. The strategy and all disclosed conflicts of interest will be mentioned in the appendix of the guideline. The disclosure form must be updated if any individual changes occur during the guideline development process and will be updated at least annually and at the end of the guideline development process.

Consensus

GDG members need to make collective decisions throughout the entire development of a guideline. Such consensus includes generating key questions, agreeing the best evidence to answer them, and formulating recommendations. There are many different approaches to making group decisions and reaching consensus – but there is no blueprint about which approach should be used. Resources for consensus development methods can be found in the systematic review by Murphy and others [13]. The most commonly used consensus development methods are the: nominal group technique, Delphi survey, and RAND/UCLA appropriateness method.

² If a GDG member cannot attend two meetings in a row, he/she may be asked to stand down by the chairperson.

Summary



Tips

- Decide on methods for recruitment and enrollment of member for the GDG.
- Consider the optimum group size for the guideline development group, particularly the guideline panel (e.g. too small of a group may lack sufficient experience, content expertise and wide representation, too large of group may lack cohesiveness and effective group interaction).
- Set expectations and awareness of the group process through an introduction, training, and support for the GDG members (e.g. setting ideal conditions for group discussion and decision-making).
- Set a quorum for meetings (e.g. 75% of group must be present to formulate guideline recommendations) but expect that all group members attend all meetings as far as possible.
- Document the guideline group member selection process and roles to ensure transparency.
- Record the composition of the GDG (names, professions, represented organizations, geographical location) within the guideline.
- Record competing interests of the GDG within the guideline, particularly where the conflicts bear on specific recommendations.
- Record within the guideline that its development was without external funding.

Available forms/checklists:

Disclosure form (B)

Confidentiality form (C)

03. Scoping the guideline

The aim of the scoping process is to define the overall objectives of the guideline (e.g. potential impact and benefits), the patients and target users to whom the guideline is meant to apply and its relation to other (ESHRE) documents.

Scoping procedure

In general, a scoping procedure will start with a kick-off meeting of the GDG. A preliminary literature search, or a survey of target users and patients can be performed to provide input for the scope of the guideline.

The scoping checklist (form ①) is completed to document the consensus of the GDG on what is within and outside the scope of the guideline.

The scope should be accepted by the entire group before the GDG begins to formulate the key questions as the basis for literature searching. After scoping, the timelines for guideline development should be set.

Summary



Tips

- Guideline scoping provides the opportunity for patient consultation.
- Set timelines for the whole guideline development process.
- Establish a method and criteria to generate and prioritize a candidate list of topics to be addressed within the guideline (e.g. where evidence is most confusing or controversial, where there is currently uncertainty or inconsistency in practice, questions about screening, diagnosis, and treatment, etc.).
- Consult appropriate stakeholders to ensure all relevant topics for the guideline have been identified and will meet the needs of the target audience(s).
- Record the overall objectives of the guideline, and the specific health benefits.
- Describe the population to whom the guideline is meant to apply.
- Record the target users of the guideline.

Available forms/checklists:

Scoping checklist ①

04. Formulating key questions

Effective and efficient guideline development involves asking and answering key questions. Key questions should be clear, focused and closely define the boundaries of the topic. They are important both as the starting point for the subsequent systematic literature review and as a guide for the development of recommendations.

Developing and selecting key questions

The key questions are developed from the guideline's scope. The scope is divided into different clinical stages (e.g. diagnosis, prognosis, treatment) and for each stage key questions are defined. It is generally not acceptable to define key questions on topics that have not been covered in the scope. Generally, a list of key questions is proposed by the chair or after consulting the GDG members, which are further defined and structured at the first GDG meeting.

Around 15–20 questions would be a reasonable number of key questions for guidelines taking 18–24 months to develop, but this depends largely on the complexity of the topic and the questions. It may be necessary to divide a guideline topic requiring more questions into subtopics or more guidelines.

During the final selection of key questions within a guideline the overall guideline outline should be kept in mind; each step of a clinical scenario needs to be addressed in a logical sequence. For example: diagnostics, treatment options, monitoring options, potential benefits/risks, outcome, prevention, information provision. The GDG selects the definitive key questions by consensus.

A significant proportion of the key questions will focus on interventions: these questions should be formulated in a structured format, based on the defined PICO components (see below):

→ Should [intervention] vs. [comparison] be used for [health problem] / [population]?

For intervention questions, PICO components should be defined, a formal evidence synthesis should be carried out, and GRADE evidence profiles should be developed.

A similar approach is suggested for questions on diagnosis, although it is not yet relevant to create GRADE evidence profiles for diagnostic questions. The proposed structured question is:

→ Should [intervention] vs. [comparison] be used to diagnose [target condition] in [health problem and/or population]?

In addition to interventions and diagnosis, other types of questions may arise. Some of these questions will fit the PICO format (although some components may be non-applicable), and a systematic assessment of the available evidence can be relevant, and recommendations can be formulated. For other questions, a formal systematic assessment of evidence synthesis is often irrelevant. These questions are often answered in a narrative format and conclusions

or good practice points, rather than recommendations, are formulated by the GDG. When defining these questions, the GDG should define whether a systematic or narrative review is relevant. Examples of questions, and how they can be handled are:

| | | |
|--------------------------------------|-------------------------|-------------|
| WHAT CAUSES THE PROBLEM? | AETIOLOGY, RISK FACTORS | PICO |
| WHAT IS THE FREQUENCY OF THE PROBLEM | FREQUENCY, PREVALENCE | PO |
| WHO WILL GET THE PROBLEM? | PROGNOSIS, PREDICTION | PICO |
| WHAT IS THE DEFINITION? | DEFINITION | (narrative) |
| WHAT IS THE CLINICAL PRESENTATION | DEFINITION | (narrative) |

For these questions (not on diagnosis or interventions), it is not relevant to create GRADE evidence profiles.

Defining key questions as PICO questions

The PICO framework is a well-accepted methodology for framing clinical questions [15]. This framework divides each question into four components (see also template for PICO questions (E)): Patients/population, Interventions, Comparisons and Outcomes

Table 4.1 Definition of PICO components and factors to consider [16,4] ³

| | | EVIDENCE-BASED GUIDELINES |
|----------|---------------------------|--|
| P | PATIENT POPULATION | <p>The patients or population to whom the recommendations are meant to apply</p> <ul style="list-style-type: none"> → How is the disease/condition defined? → What are the most important characteristics that describe the people? → Are there any relevant demographic factors (e.g. age, sex, ethnicity)? → What is the setting (e.g. hospital, community)? → Who should make the diagnosis? → Are there any other types of people who should be excluded from the review (because they are likely to react to the intervention in a different way)? → How will studies involving only a subset of relevant participants be handled? → <i>Consider the prevalence of multiple comorbidities in the population</i> |
| I | INTERVENTION | <p>The therapeutic, diagnostic, or other intervention under investigation (e.g. the experimental intervention, or in observational studies the exposure factor)</p> <ul style="list-style-type: none"> → What are the experimental and control (comparator) interventions of interest? → Does the intervention have variations (e.g. dosage/intensity, mode of delivery, personnel who deliver it, frequency of delivery, duration of delivery, timing of delivery)? → Are all variations to be included (for example is there a critical dose below which the intervention may not be clinically appropriate)? |
| C | COMPARISON | <p>The alternative intervention; intervention in the control group</p> <ul style="list-style-type: none"> → How will trials including only part of the intervention be handled? → How will trials including the intervention of interest combined with another intervention (co intervention) be handled? → <i>Identify whether or not multiple (treatment) comparisons should be included.</i> |
| O | OUTCOME | <p>The outcome(s) of interest</p> <ul style="list-style-type: none"> → Main outcomes, for inclusion in the 'Summary of findings' table, are those that are essential for decision making, and emphasis should be on patient important outcomes. → Primary outcomes are the two or three outcomes among the main outcomes that the review would be likely to be able to address if sufficient studies are identified, in order to reach a conclusion about the effects (beneficial and adverse) of the intervention(s). → Secondary outcomes include the remaining main outcomes (other than primary outcomes) plus additional outcomes useful for explaining effects. → Ensure that outcomes cover potential as well as actual adverse effects. → Consider outcomes relevant to all potential decision makers, including economic data. → Consider the type and timing of outcome measurements. → <i>Include both desirable (e.g. benefits, less burden, savings) and undesirable effects (e.g. harm, burden, costs, and decrease in patient autonomy).</i> → <i>Do not ignore important outcomes for which evidence may be lacking.</i> |

³ Additions in Italic based on GIN-McMaster Guideline Development Checklist [9]

Defining the patient population and intervention for guideline questions is challenging and should consider the underlying biology. Defining a broad patient population or intervention may be okay if across the range of patients and interventions it is plausible that the magnitude of effect on the key outcomes is more or less the same. If that is not the case the review will generate misleading estimates for at least some subpopulations of patients and interventions, and the questions should be defined more narrow or split up. Also, and different to a systematic review calculating relative risks, recommendations may differ across subgroups of patients at different baseline risk of an outcome, despite there being a single relative risk that applies to all of them. Thus, guideline panels must often define separate questions (and produce separate evidence summaries) for high- and low-risk patients, and patients in whom quality of evidence differs.

Another challenge is defining the comparators. Mostly, guideline groups will be strict in defining the intervention, but will define the comparator as “all other interventions”. Clarity in choice of the comparator makes for interpretable guidelines, and lack of clarity can cause confusion.

In order to make sensible recommendations all relevant outcomes that are important or critical to patients for decision making must be considered and included. Recommendations cannot be made on the basis of information about single outcomes and decision-making always involves a balance between health benefits and harms. GDGs must base the choice of outcomes on what is important, not on what outcomes are measured and for which evidence is available. If evidence is lacking for an important outcome, this should be acknowledged, rather than ignoring the outcome. Most systematic reviews do not summarize the evidence for all important outcomes, and evidence from other sources should be included.

In GRADE, outcomes should be classified on importance for decision-making in 3 categories; critical, important but not critical, and of limited importance. Ranking outcomes by their relative importance can help to focus attention on those outcomes that are considered most important and help to resolve or clarify disagreements. Practically, to generate a list of relevant outcomes, one can use the following type of scales [4]:

It is important to realize that the importance of outcomes is likely to vary within and across cultures or when considered from the perspective of the target population (e.g. patients or the public), clinicians or policymakers. The perspective would generally be that of the patient, and a literature search can be conducted on patients’ values and preferences about the intervention in question in order to inform the rating of the importance of outcomes. Reviewing the evidence may provide the panel with insight about the variability in patients’ values, the patient experience of burden or side effects, and the weighing of desirable versus undesirable outcomes. However, often such evidence is not available and panel members should use their prior experiences with the target population to assume the relevant values and preferences.

Modifications to the key questions

In general, after consensus by the GDG, the key questions are final, and modifications should be minimized. However, once the evidence has been searched, the key questions may need refining. In any case, the entire GDG should be informed of and agree with any changes to the key questions.

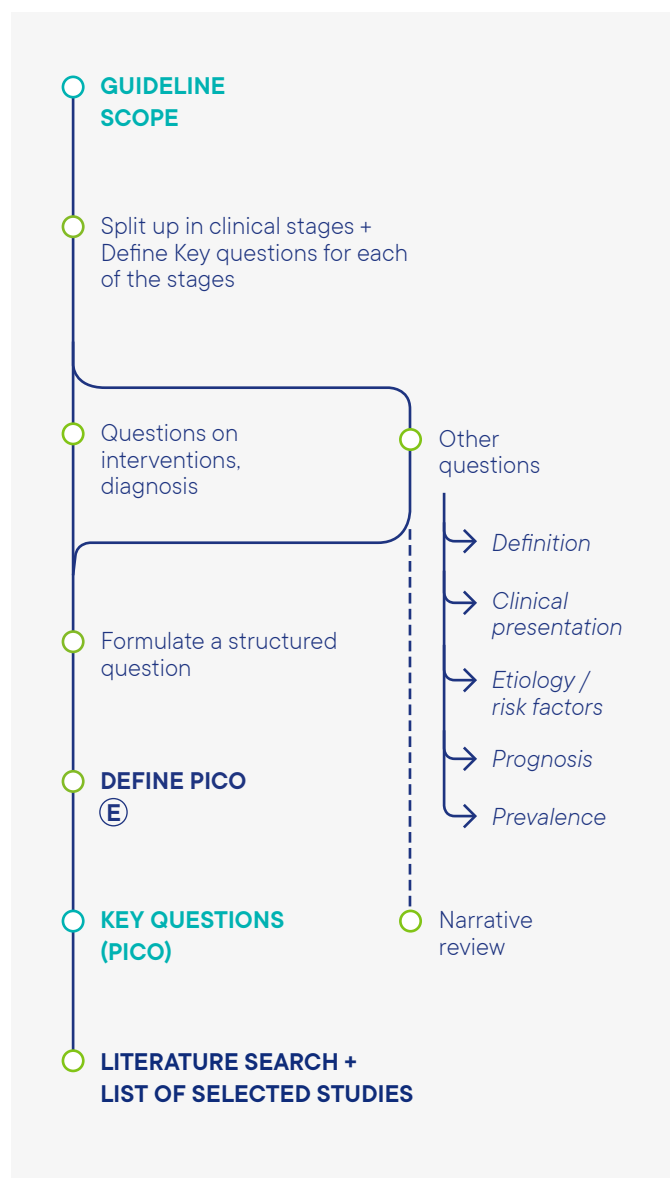
Changes to the key questions could include:

- Reassessment of the importance of the outcomes
- Addition of an outcome: for instance, the importance of an outcome (e.g. a serious adverse effect) may only become known after the evidence synthesis
- Addition of an intervention that is relevant for the flow and consistence of the guideline
- Specific search on an important subgroup, not defined as such in the PICO questions
- Merging of key questions based on significant overlap of the evidence synthesis.

As changes to the key questions imply additional work for the research specialist and the GDG members, these should be avoided as much as possible.

| RATING SCALE | | | | | | | | |
|--|---|---|------------------------------------|---|---|---------------------------------------|---|---|
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| Of least importance | | | | | | Of most importance | | |
| OF LIMITED IMPORTANCE FOR MAKING A DECISION | | | IMPORTANT, BUT NOT CRITICAL | | | CRITICAL FOR MAKING A DECISION | | |
| Not included in evidence profile | | | Included in evidence profile | | | Included in evidence profile | | |

Summary



Tips

- Generate and document the key questions (e.g. clinical, health, policy) to be answered in the guideline using a standard format (e.g. PICO) and determine the criteria by which the questions generated will be prioritized if it is not feasible to answer all questions (e.g. survey guideline panel members, survey stakeholders).
- Select no more than 15–20 key questions.
- Define key questions in such a way that answering the question gives the opportunity to make a recommendation.
- Think of formulating key questions in addition to health benefits - and on side effects and risks.
- List all key questions in the guideline, at the start of each guideline section.

Available forms/checklists:

Template PICO Question (E)

05. Searching evidence

The identification and selection of evidence is an essential step towards answering the key questions. Secondly, in order to perform an evidence search the key questions should be translated into key words or search terms. The evidence search itself should be gathered in a systematic process to avoid or minimize bias. Finally, from the identified literature the relevant evidence should be selected for summary and evaluation.

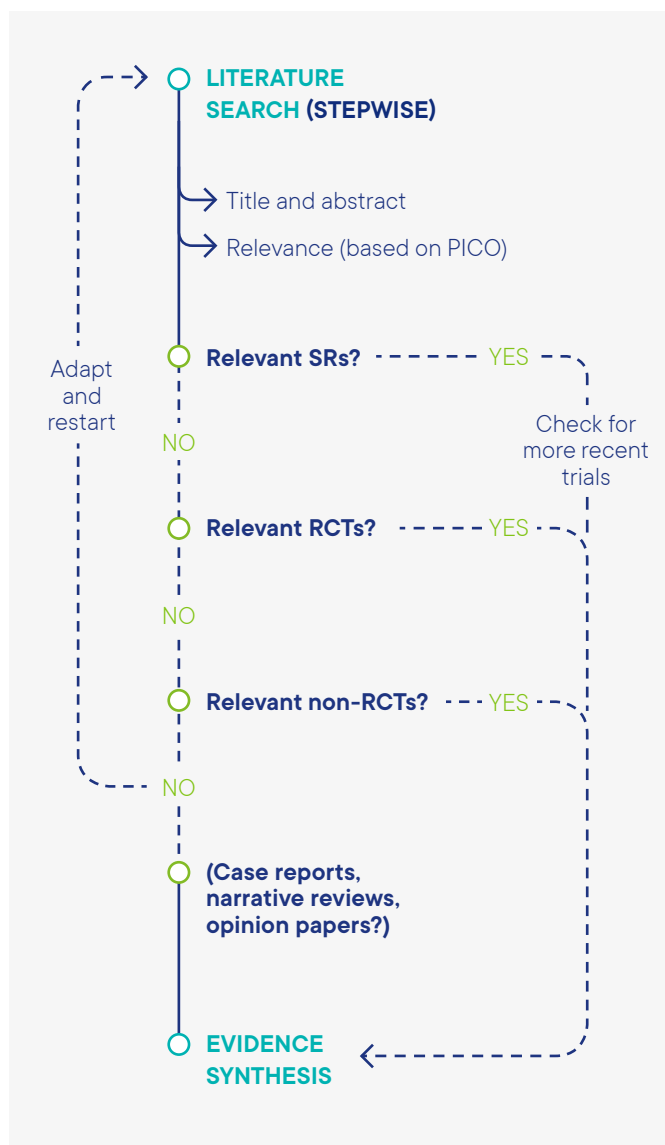
Literature search

According to the Institute of Medicine, literature searching is the key step in developing valid guidelines, because incomplete or biased literature evaluation can lead to inappropriate recommendations. The search for relevant research should be comprehensive, based on explicit criteria, and the validity of the results should be judged in a rigorous and reproducible fashion [17].

ESHRE applies a stepwise methodology, common to guideline development, focusing on the best available evidence to address each key question [5]. A set of standard search filters is used for identification in the following order:

- Systematic reviews/meta-analyses
- Randomized controlled trials
- Non-randomized studies / observational studies
- Case reports/opinion documents.

The literature search will be performed in this stepwise approach, but all studies will be available for selection. Where adequate published systematic reviews exist, it may be appropriate to select the review and additional studies from the time period since the review was conducted. If no systematic review exists, the next type of studies to be assessed are RCTs (at least for intervention questions), followed by non-randomized and observational studies.



The GDG should establish in advance a set of basic selection criteria (e.g. duration of a follow-up period, the primary outcome measure, age limits). The process for evidence identification should also be repeatable and transparent. The search strategy, including search terms, should therefore be documented and stored. This also simplifies running the search strategies to check the validity of a guideline.

The ESHRE research specialist will conduct the literature searches, based on a list of search terms for each of the PICO questions defined by the GDG members. Literature searching includes at least MEDLINE/PubMed and the Cochrane Library, but additional sources can be covered (e.g. NHS Economic Evaluations Database (NEED), PsycInfo and Embase) specific to the topic under review.

The searches are limited to:

- peer reviewed published literature
 - the use of abstracts should be avoided except in very rare instances (and always combined with a search for the full paper)
 - unpublished clinical trials should be avoided to support any recommendation.
- English language
- human subjects
- defined time frame; searches in a guideline update are limited to the period following the last publication of the guideline; if a suitable systematic review is identified, an update search is limited to the time period following the reported search cut-off date.

Although the research specialist performs a preliminary level of selection based on title and abstract, the clinical expertise of GDG members is necessary to carry out the definitive selection of the search output.

Different questions may be best answered by different databases or may rely on different levels of evidence. Following evaluation of the first search results the key questions may be redefined, and subsequent searches focused on the most appropriate sources and study types. As a result, the assembly of evidence is a stepwise and iterative process.

Selection of evidence

Papers are initially pre-selected according to title and abstract by the ESHRE research specialist and the final selection is made by the GDG member.

First, the titles of the retrieved citations are scanned and those that fall outside the topic of the guideline are eliminated. Next, a quick check of the remaining abstracts identifies further papers not relevant to the key questions, and these are also excluded.

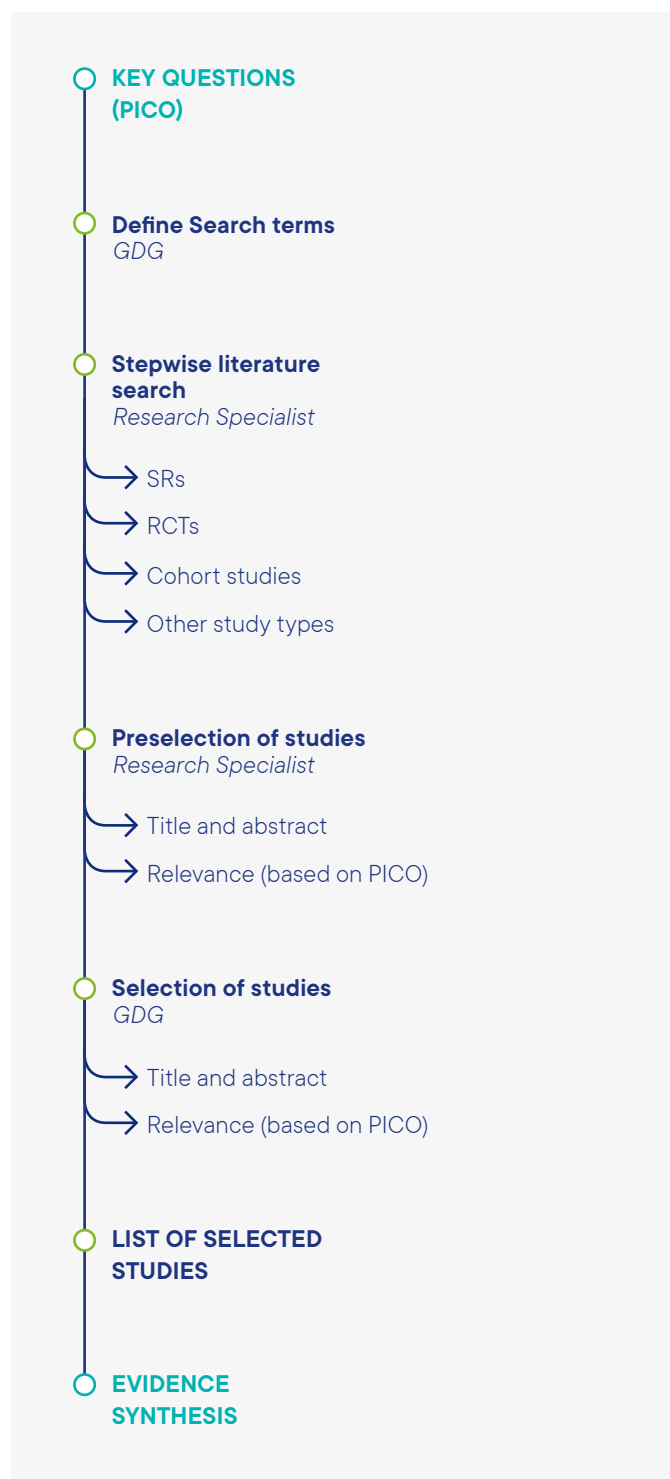
The remaining abstracts are investigated if they fulfil the selection criteria agreed by the GDG. If no or incomplete information is available in the abstract, the reference is selected and in the next step, the full text is assessed for relevance and quality to ensure its validity and applicability. The study selection process is clearly documented and details the applied inclusion criteria.

Role of qualitative research

At present there is no established mechanism for incorporating qualitative studies in evidence-based guideline development. Nevertheless, the use of qualitative studies can help identify issues of concern to patients. A qualitative approach to complement trial data in the collection of information on patient preferences and the values placed on outcomes would perhaps help bridge the gap between scientific evidence and clinical practice. In case qualitative studies are used to support recommendations, an appropriate quality assessment checklist should be used to validate the quality of the studies.

Narrative or descriptive review are generally not selected during evidence synthesis, but they may be helpful as background information. These papers represent an interpretation of evidence in the context of experts' experiences and knowledge. Expert opinion is not evidence per se and should not be used as evidence; rather, experience or observations that support expert opinions should be described, identified and, if possible, appraised in a systematic and transparent way.

Summary





Tips

- Follow systematic review methods (either full systematic reviews or rapid systematic reviews depending on the topic and organization's framework) or provide a rationale for why this is not done.
- Develop a protocol for locating, selecting, and synthesizing the evidence (e.g. conduct a search for existing systematic reviews, new systematic review and grey literature search) and determine the types of evidence to include (e.g. data-bases searched, types of studies, inclusion and exclusion criteria, searching for specific studies on adverse effects or deciding to abstract information on adverse effects from studies on benefit).
- Found evidence gaps can be used for future research goals.
- Document and store the search strategies used.
- Record how patients' perspectives are included within the evidence search.

06. Evidence synthesis

Studies identified during the stepwise literature search should be reviewed to identify the most appropriate data for answering the key questions and ensure that recommendations are based on the best available evidence. This process should be explicit and transparent and should be carried out through a systematic review process. This involves selecting relevant studies (step 6), assessing relevance and quality, summarizing the results, and grading the evidence.

Relevance and Quality check

Relevance and quality assessment (template ) of the selected evidence is necessary to ensure that recommendations are based on the highest quality evidence available. Quality assessment is performed on each individual study. However, if the study is a meta-analysis or systematic review quality assessment should be performed on the meta-analysis or review itself and not on the studies included. Depending on the type of study, different checklist should be used (Checklist ). For systematic reviews and meta-analysis, the AMSTAR quality assessment checklist is recommended.

A study should be rejected if its quality is assessed as low. If no better evidence can be found, the study might be considered as low-level evidence, comparable with expert opinions.

In addition to the inherent quality of a paper, the applicability of findings (relevance) should also be assessed.

Applicability or relevance is related to the definition of the components (PICO) of the formulated key questions. Comparison of the available articles with the defined PICO components guides the selection of papers with the relevant evidence.

The validity of a study is the extent to which its design and conduct are likely to prevent systematic errors, or bias. There are four potential sources of systematic bias in healthcare trials:

- Selection bias – randomization (Patients/population)
- Performance bias – blinding (Intervention)
- Attrition bias – handling participant loss (Comparison) and
- Detection bias – outcome assessment (Outcome).

One of the most important factors leading to bias and distorted treatment comparisons is patient assembly (selection bias). An appropriate method for preventing foreknowledge of treatment assignment is crucial for any study. True randomization is administered by someone who is not responsible for the recruitment of study subjects. Thus, studies are for the selection bias judged on the quality of the used allocation concealment methodology.

Performance bias refers to systematic differences in the provision of care to the participants in the intervention and control group. Those providing and receiving care can be 'blinded' to protect

against unintended differences in care.

Attrition bias, also known as exclusion bias, alludes to systematic differences in the approach to handling the loss of participants (e.g. withdrawals, dropouts, protocol deviations) in the two study groups. This may have great potential for biasing results.



Detection bias is a systematic difference between two study groups in outcome assessment. Trials that blind those assessing outcomes are logically less likely to be biased than trials that do not.

Risk of bias assessment categorizes studies as low, moderate or high-risk bias based on the 4 sources of potential bias. The assessment can then be used as a (1) a threshold for study inclusion (e.g. for studies judged at high risk of bias, this assessment constitutes grounds for study exclusion), or (2) a possible explanation for found differences in study results.

Risk of bias assessment provides a structured evaluation of the possible sources of bias. However, it is important to go back to the question and assess how important the study flaws are in the interpretation of the overall results.

The study selection procedure and results of the risk of bias assessment and relevance should be documented and will be published as an annex to the guideline. At this point the available evidence is ready for summary.

Evidence tables

Evidence tables help to identify similarities and differences between studies. Data for inclusion within an evidence table should be extracted according to a standard template (checklists ). Here, key characteristics of the study population (e.g. sample size, age), intervention (e.g. follow-up period, kind of intervention), comparison (e.g. IUI versus timed intercourse) and outcome measures (e.g. effect size) are important. The evidence table was updated to the recommendations of the evidence table working group of the Guidelines International Network (<http://www.g-i-n.net/>). The completed evidence tables will be published as an appendix to the guideline (Template evidence table 

GRADE Evidence profiles and Summary of Findings Tables

A GRADE evidence profile allows presentation of key information about all relevant outcomes for a given health care question. It presents information about the body of evidence (e.g. number of studies), the judgments about the underlying quality of evidence, key statistical results, and the quality of evidence rating for each outcome.

A GRADE evidence profile is particularly useful for presentation of evidence supporting a recommendation in clinical practice

guidelines. It includes:

- A list of outcomes (those considered critical and important; classification of the importance can be added)
- The number of studies and study design(s)
- Judgements about each of the quality of evidence factors assessed; risk of bias, inconsistency, indirectness, imprecision, other considerations (including publication bias and factors that increase the quality of evidence)
- The assumed risk; a measure of the typical burden of the outcomes, i.e. illustrative risk or also called baseline risk, baseline score, or control group risk
- The corresponding risk; a measure of the burden of the outcomes after the intervention is applied, i.e. the risk of an outcome in treated/exposed people based on the relative magnitude of an effect and assumed (baseline) risk
- The relative effect; for dichotomous outcomes the table will usually provide risk ratio, odds ratio, or hazard ratio
- The absolute effect; for dichotomous outcomes the number of fewer or more events in treated/exposed group as compared to the control group
- Rating of the overall quality of evidence for each outcome (which may vary by outcome)
- Footnotes, if needed, to provide explanations about information in the table such as elaboration on judgements about the quality of evidence

GRADE evidence profiles are used for discussion of recommendations during guideline meetings. Summary of findings tables provide similar information in a more accessible format. Where relevant, and at least for all intervention questions with more than 1 RCT available, summary of findings tables will be added in the guideline

(either in the body text, or as an annex). The corresponding GRADE evidence profiles will be used for discussion and are available upon request. Any studies summarized in GRADE evidence profiles should not be necessarily included in the evidence table.

Quality of evidence for each outcome

The quality of evidence reflects the extent to which our confidence in an estimate of the effect is adequate to support a particular recommendation. It gives the reader a quick impression of the quality of the supporting evidence, which is not necessarily related to the importance of the recommendation.

Guideline panels must make judgments about the quality of evidence relative to the specific context for which they are using the evidence.

The GRADE approach involves separate grading of quality of evidence for each patient-important outcome (across studies) followed by determining an overall quality of evidence across outcomes. Although the quality of evidence represents a continuum, the GRADE approach results in an assessment of the quality of a body of evidence in one of four grades:

Table 6.1: Quality of Evidence Grades [18]

| GRADE | DEFINITION |
|----------|---|
| High | We are very confident that the true effect lies close to that of the estimate of the effect. |
| Moderate | We are moderately confident in the effect estimate: The true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different. |
| Low | Our confidence in the effect estimate is limited: The true effect may be substantially different from the estimate of the effect. |
| Very low | We have very little confidence in the effect estimate: The true effect is likely to be substantially different from the estimate of effect. |

Factors determining the quality of evidence

The GRADE approach to rating the quality of evidence begins with the study design (trials or observational studies) and then addresses five reasons to possibly rate down the quality of evidence and three to possibly rate up the quality [18].

For intervention studies, randomized trials provide, in general, far stronger evidence than observational studies, and rigorous observational studies provide stronger evidence than uncontrolled case series. As such, RCTs without important limitations provide high quality evidence, while observational studies without special strengths or important limitations provide low quality evidence.

In case of RCTs, 5 factors should be assessed to detect limitations and reduce the quality of the evidence (for a certain outcome). In observational studies, 3 factors should be assessed to detect strengths and increase the quality of the evidence (for a certain outcome). If one or more of these factors is met (and there is no reason for downgrading), it is possible to rate up the quality.

In the end, the overall quality of evidence for an intervention across outcomes, is the lowest quality of evidence for the critical outcomes, as the overall confidence in effect estimates cannot be higher than the lowest confidence in effect estimates for any outcome that is critical for a decision.

Non-randomized experimental trials (quasi-RCT) without important limitations also provide high quality evidence but will automatically be downgraded for limitations in design (risk of bias) – such as lack of concealment of allocation and tie with a provider (e.g. chart number).

Although the GRADE approach focusses on RCTs, **large observational studies**, specifically multivariate regression analyses, can provide moderate quality evidence, and can answer questions that are impossible to be answered by RCTs. Quality assessment is essential and should focus on whether confounding factors are accounted for and data screened is sufficiently large.

Case series and case reports are observational studies that investigate only patients exposed to the intervention. Source of control group results is implicit or unclear, thus, they will usually warrant downgrading from low to very low-quality evidence.

Table 6.2: Factors of upgrading and downgrading quality of evidence [18]

| SOURCE OF THE BODY OF EVIDENCE | INITIAL QUALITY OF THE BODY OF EVIDENCE | FACTORS THAT MAY DECREASE THE QUALITY | FACTORS THAT MAY INCREASE THE QUALITY | QUALITY OF A BODY OF EVIDENCE |
|--------------------------------|---|---------------------------------------|---|-------------------------------|
| Randomized trials | High | 1. Risk of bias | 1. Large effect | High ⊕⊕⊕⊕ |
| | | 2. Inconsistency | 2. Dose-response | Moderate ⊕⊕⊕⊖ |
| Observational studies | Low | 3. Indirectness | 3. All plausible residual confounding would reduce the demonstrated effect or would suggest a spurious effect if no effect was observed | |
| | | 4. Imprecision | | Very low ⊕⊖⊖⊖ |
| | | 5. Publication bias | | |

Summary



Tips

- Document and publish the search and selection of evidence, judging eligibility, range of evidence included, and search strategies used to ensure the methods are explicit and transparent.
- Summarize the evidence using a concise summary (e.g. evidence table, evidence profile or summary of findings table) of the best available evidence for each important outcome, including diagnostic test accuracy, anticipated benefits, harms, resources (costs), the quality of evidence rating, and a summary of the relative and absolute results/estimate of effect for each outcome.
- Assess the quality of evidence for each important outcome.
- Assess and report the overall quality of evidence (e.g. lowest quality of evidence from outcomes rated as most important or critical, or highest quality of evidence when all outcomes point in the same direction).
- Document the judgements made in appraising the quality of evidence to ensure they are transparent and explicit.
- Record the set of evidence selection criteria.
- Record the strengths and limitations of the evidence.

Available forms/checklists:

Template Relevance and Quality check (F)

Quality assessment checklists (G)

Template evidence table (H)

Background information

Factors for downgrading the quality of evidence

1. Risk of bias

The risk of bias of the included studies should be assessed in relation to the effect on the outcome [19]. In assessing the studies, the weight of the studies in the meta-analysis should be considered as small studies with high risk of bias may not necessarily impact on the estimate of effect if combined with a very large study at low risk of bias.

Guidance to assess risk of bias and corresponding downgrading for limitations in study design is listed in table 6.3.

2. Inconsistency of results

Estimates of treatment effect across studies can differ because of clinical heterogeneity (P, I, O: for instance larger effect with higher dose, or shorter time of follow-up), or methodological heterogeneity (differences in study design) [20].

In case of (unexplained) inconsistency, the quality can be downgraded, or subgroups can be presented.

Inconsistency can be detected by assessing confidence intervals and direction of effect for the included studies in a forest plot, or by means of the outcome of statistical tests for heterogeneity (I^2 statistic; $>60\%$ = substantial, p -value).

3. Indirectness of evidence

Downgrading for indirectness can be considered if the evidence from the studies is different from the PICO question [21]. Examples could be:

- Indirect comparison between 2 interventions (A vs placebo and B vs placebo instead of drug A vs drug B)
- Population : evidence in menopausal women instead of POI
- Comparator
- Outcomes: surrogate outcomes: bone density instead of fractures

Table 6.3. Risk of bias and impact on GRADE of evidence

| RISK OF BIAS | ACROSS STUDIES | INTERPRETATION | CONSIDERATIONS | GRADE |
|-----------------------------|---|---|--|--|
| LOW RISK OF BIAS | Most information is from studies at low risk of bias | Plausible bias unlikely to seriously alter the results | No apparent limitation | No serious limitations, do not downgrade |
| UNCLEAR RISK OF BIAS | Most information is from studies at low or unclear risk of bias | Plausible bias that raises some doubt about the results | Potential limitations are unlikely to lower confidence in the estimate of effect | No serious limitations, do not downgrade |
| | | | Potential limitations are likely to lower confidence in the estimate of effect | Serious limitations, downgrade 1 level |
| HIGH RISK OF BIAS | The proportion of information from studies at high risk of bias is sufficient to affect the interpretation of results | Plausible bias that seriously weakens the confidence in the results | Crucial limitation for one criterion or some limitations for multiple criteria, sufficient to lower confidence in the estimate of effect | Serious limitations, downgrade 1 level |
| | | | Crucial limitation for one or more criteria, sufficient to substantially lower confidence in the estimate of effect | Serious limitations, downgrade 2 levels |

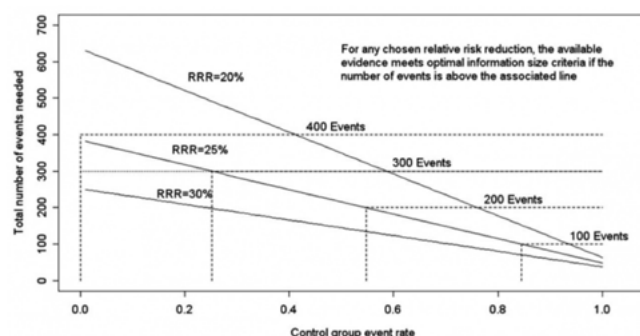
Background information

4. Imprecision

In general, results are imprecise when studies include relatively few patients and few events and thus have a wide confidence interval (CI) around the estimate of the effect [22]. In this case, one may judge the quality of the evidence lower than it otherwise would be considered because of resulting uncertainty about the results.

In general, downgrading for imprecision can be considered in cases of small sample size (optimal information size not met), and/or wide confidence intervals.

The optimal information size is the threshold of number of events that needs to be included. The criteria for the optimal information size can be estimate by plotting the background risk against a chosen relative risk reduction in the following graph.



If the total number of patients included in a systematic review is less than the number of patients generated by a conventional sample size calculation for a single adequately powered trial, consider rating down for imprecision.

Regarding the wide confidence intervals, downgrading is appropriate when the confidence interval crosses the clinical decision threshold between recommending and not recommending treatment. This clinical decision threshold should be set based on the intervention. For example, for an intervention with limited adverse events, inconvenience and cost, the threshold for an absolute reduction in pregnancy loss may be set at 0.5%, which means that even a small benefit of the intervention would lead to recommending it. Alternatively, for an intervention with significant toxicity, the clinical decision threshold could be set at at least 1%. A theoretical intervention with an absolute reduction of pregnancy loss of 1.3% (95% CI 0.6% to 2.0%), would be rated down for imprecision with the clinical decision threshold of 1%, but not with a threshold of 0.5%.

5. Publication bias

Publication bias is a systematic under-estimation or an over-estimation of the underlying beneficial or harmful effect due to the selective publication of studies [23]. Confidence in the combined estimates of effects from a systematic review can be reduced when publication bias is suspected, even when the included studies

themselves have a low risk of bias. Funnel plots can be used as a means to detect publication bias.

Factors for upgrading the quality of evidence [24]

1. Large magnitude of effect

When the body of evidence from observational studies yield large or very large estimates of the magnitude of an intervention effect, then we may be more confident about the results. Decisions to rate up quality of evidence because of large ($RR > 2$ or $RR < 0.5$) or very large effects ($RR > 5$ or $RR < 0.2$) should consider not only the point estimate but also the precision (width of the CI) around that effect. Furthermore, upgrading should only be considered in absence of any problems with risk of bias.

2. Dose-response gradient

The presence of a dose-response gradient has long been recognized as an important criterion for believing a cause-effect relationship. The presence of a dose-response gradient may increase our confidence in the findings of observational studies and thereby increase the quality of evidence.

3. Opposing plausible residual confounding

The reason that in most instances we consider observational studies as providing only low-quality evidence is that unmeasured or unknown determinants of outcome unaccounted for in the adjusted analysis are likely to be distributed unequally between intervention and control groups, referred to as "residual confounding" or "residual biases". On occasion, all plausible confounders (biases) from observational studies unaccounted for in the adjusted analysis (i.e. residual confounders) of a rigorous observational study would result in an underestimate of an apparent treatment effect. If, for instance, only sicker patients receive an experimental intervention or exposure, yet they still fare better, it is likely that the actual intervention or exposure effect is even larger than the data suggest. A parallel situation exists when observational studies have failed to demonstrate an association.

07. Developing recommendations

Once the selection and summary of evidence is complete, the available evidence must be combined and condensed into recommendations.

Recommendations are statements mostly proposing a course of action. An ESHRE guideline recommendation should be a stand-alone text written in a complete sentence.

Based on the available evidence, each GDG member prepares specific recommendations and presents them to the other GDG members at the GDG meeting. In addition to the evidence summary, the full systematic review(s) and the original studies and other sources of evidence will be available for the entire GDG during the process and prior to the meetings to inform deliberations (through a collaborative website and/or via electronic communication).

When the GDG has reached consensus on the recommendations, the draft version of the guideline can be written.

Strong or weak recommendations

The strength of a recommendation reflects the extent to which a guideline panel is confident that desirable effects of an intervention outweigh undesirable effects, or vice versa, across the range of patients for whom the recommendation is intended.

According to GRADE, recommendations are classified as “strong” or “weak”. The strength of a recommendation may not be directly correlated with its priority for implementation [25].

When the GDG formulates a strong recommendation, they have to be certain about the various factors that influence the strength of a recommendation. The GDG also should have the relevant information at hand that supports a clear balance towards either the desirable effects of an intervention (to recommend an action) or undesirable effects (to recommend against an action). When the GDG is uncertain whether the balance is clear or when the relevant information about the various factors that influence the strength of a recommendation is not available, a guideline panel should be more cautious and in most instances, they would opt to make a weak recommendation. Alternatively to weak recommendations, the terms conditional (depending on patient values, resources available or setting) or discretionary (based on opinion of patient or practitioner) can be used.

Table 7.1: The implications of the two grades of strength of recommendations in the GRADE approach [25], with the addition of the implications of a “research only” recommendation and a GPP:

| TARGET GROUP | STRONG RECOMMENDATIONS* | CONDITIONAL (WEAK) RECOMMENDATIONS | RESEARCH ONLY RECOMMENDATIONS | GOOD PRACTICE POINTS (GPP)** |
|---------------|--|---|---|--|
| PATIENTS | Most people in your situation would want the recommended course of action and only a small proportion would not. | The majority of people in your situation would want the recommended course of action, but many would not. | The test or intervention should only be considered by patients and clinicians within the setting of a research trial for which appropriate approvals and safety precautions have been established | Clinicians, patients and policy makers are informed of the advice of the GDG regarding a certain recommendation. |
| CLINICIANS | Most patients should receive the recommended course of action. | Recognize that different choices will be appropriate for different patients and that you must make greater effort with helping each patient to arrive at a management decision consistent with his or her values and preferences. Decision aids and shared decision making are particularly useful. | | |
| POLICY MAKERS | The recommendation can be adopted as a policy in most situations. | Policy making will require substantial debate and involvement of many stakeholders. | NA | |

* Strong recommendations based on high quality evidence will apply to most patients for whom these recommendations are made, but they may not apply to all patients in all conditions; no recommendation can take into account all of the often compelling unique features of individual patients and clinical circumstances.

** A good practice point or GPP is written by the GDG to support the recommendations. Advice can for instance be provided on how to establish shared decision making, and on factors to be taken into account for a specific test or intervention.

Table 7.1 will be provided in the methodology section of the guidelines to provide clear direction on the implication of the strength of recommendation for clinicians, patients, policy makers, and any other target audience groups.

The decision on a strong or a weak recommendation depends on 5 key factors [26,4]. Judgement on these factors will be documented in a framework and summarized (narratively), with information on the explicit link between the recommendation and evidence supporting the recommendation in a justification statement in the guideline.

Table 7.2: Key factors for deciding on a strong or a weak recommendation [26,4]:

| FACTORS | COMMENTS |
|--|--|
| Balance between desirable and undesirable outcomes (trade-offs) taking into account: → best estimates of the magnitude of effects on desirable and undesirable outcomes → importance of outcomes (estimated typical values and preferences) | The larger the differences between the desirable and undesirable consequences, the more likely a strong recommendation is warranted. The smaller the net benefit and the lower certainty for that benefit, the more likely a weak recommendation is warranted. |
| Confidence in the magnitude of estimates of effect of the interventions on important outcomes (overall quality of evidence for outcomes) | The higher the quality of evidence, the more likely a strong recommendation is warranted. |
| Confidence in values and preferences of patients and their variability | The greater the variability in values and preferences, or uncertainty about typical values and preferences, the more likely a weak recommendation is warranted. |
| Resource use | The higher the costs of an intervention (the more resources consumed), the less likely a strong recommendation is warranted. |
| Health system perspective, including → equity (what would be the impact on health inequities?), → acceptability (is the option acceptable to key stakeholders?) → feasibility (is the option feasible to implement?) | A higher impact on equity, acceptability and feasibility makes a strong recommendation more likely. |

The methods in which the additional information is to be incorporated with the synthesized evidence is documented in the annex of the guideline to ensure transparency (e.g. formal consensus on patient values, consensus on equity issues, formal economic analysis, consideration of disaggregated resource use data in a qualitative manner).

Regarding resource use, the guideline will not include a formal analysis of cost effectiveness of recommended practice versus current or established practice. The economic feasibility of recommendations will also not be covered. The clinical and organizational impact of recommendations on costs will be considered in the GDG meetings and, if relevant, described in the justification section of the guideline.

Wording of recommendations

ESHRE guideline recommendations could stand alone and contain enough detailed information to be understandable without

references to supporting material. Recommendations are written in complete sentences and should answer the key questions. In addition, the wording must be:

- unambiguous
- clearly defined
- actionable
- easy to translate into clinical practice and
- agreed by the complete GDG.

Indicate in the recommendation statements the population for which the recommendation is intended, the intervention being recommended, and the alternative approach(es) or intervention(s). A help to guarantee the formulation of such clear recommendations is the five 'W' rule: each recommendation should be a description about **who** does **what** for **whom**, **when** and in **which way**.

Possible benefits and harms should be quantified as much as possible. Any exceptions to the recommendations should be listed whenever possible.

Despite the lack of studies supporting this, a standardized wording is usually defined reflecting the strength of the recommendation. Standardized wording to use for recommendations provides structure for the GDG members and aids to ensure clarity and to maintain consistency throughout the guideline, avoiding wording that may be vague and nonspecific (see table 7.3).

Table 7.3: Recommended phrasing for recommendations in the ESHRE guidelines.

| RECOMMENDED PHRASING | |
|----------------------------------|---|
| Strong recommendation | <ul style="list-style-type: none"> — Clinicians should* — It is recommended* — It is indicated — Do* |
| Weak recommendation | <ul style="list-style-type: none"> — It is conditionally recommended* — It is suggested* — Clinicians might* — Clinicians could consider — Clinicians may/might consider |
| Good practice point (GPP) | <ul style="list-style-type: none"> — The GDG recommends |

* suggested by the GRADE working group

Good practice points (GPPs) are not an alternative to evidence-based recommendations; any evidence relating to a key question excludes the possibility of using a GPP to make a recommendation. GPPs might be used to emphasize the importance of patient participation in decision making about specific procedure, provide advice on the management of specific surgical procedures for which there is an evidence based recommendation, or advise caution where there is perceived risk of harm but no available direct evidence of such harms [5].

If the GDG group feels strongly that they want to make a recommendation even though there is **no significant evidence**, this should be done as a weak recommendation rather than a GPP. For such recommendation, the evidence can consist of opinion (from outside the GDG) supporting the recommendation. If no such evidence exists, formal methods should be used to develop a consensus-based recommendation which will be clearly identified as such within the guideline by a statement accompanying the recommendation. The methods used to reach consensus should be described in the methodology annex of the guideline.

For newer diagnostic tests and interventions, the GDG is

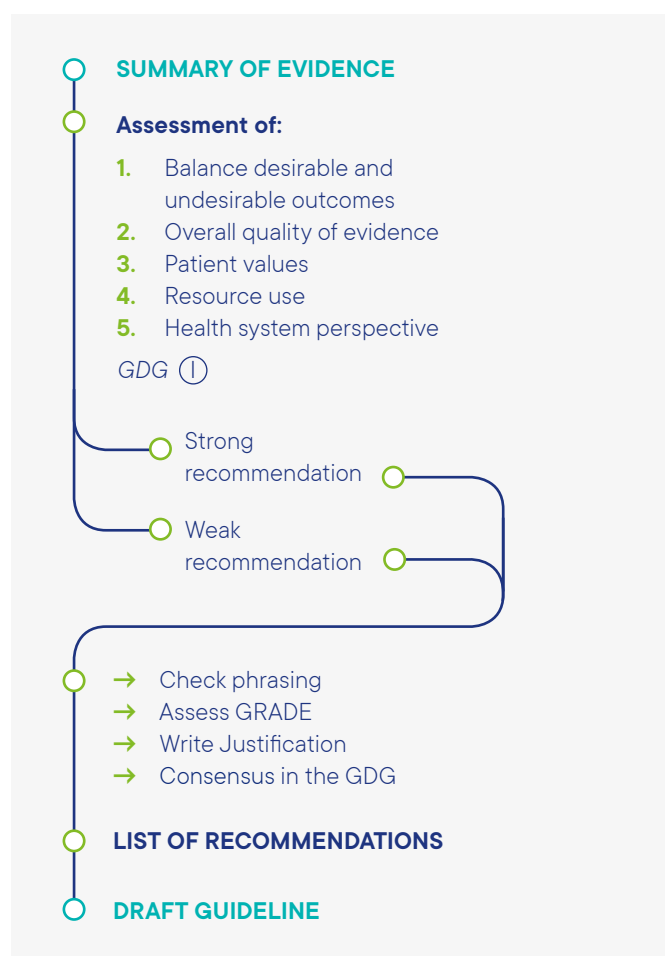
encouraged to clearly state whether the test/intervention is recommended in routine clinical practice or whether it can be used only in a research context. Further data could allow for a more comprehensive recommendation in the update of the guideline.

For some tests and interventions, the GDG may decide not to make a recommendation at all.

Research recommendations

In discussing the available evidence and its shortcomings, the GDG may feel it is important to formulate recommendations for future research. Where possible, research recommendations should be specific, detailing the necessary type of studies (RCTs, large multicenter studies), but also the (specific) interventions, or patient subgroups where further research is needed. The aim of formulating research recommendations is to stimulate research with a direct impact on future clinical recommendations. For this aim, the GDG will also be asked to reach consensus on a top-3 of research recommendations with the highest priority.

Summary



Tips

- If applicable, make provisions for formulating recommendations in situations where there is insufficient evidence or very low quality evidence (e.g. conditional recommendation with judgements laid out transparently, no recommendation if the guideline panel feels there is substantial risk that their decision may be wrong, recommend that the intervention be used in the context of research complemented by guidance for what are the best management options until further research becomes available).
- Provide suggestions about whether the recommendations are appropriate to serve as performance measures/quality criteria (e.g. management options associated with strong recommendations based on high- or moderate-quality evidence are particularly good candidates for quality criteria, when a recommendation is weak, discussing with patients the relative merits of the alternative management strategies and appropriate documentation of this interaction may become a quality criterion).
- Record or refer to the methodology used for recommendations' formulation.
 - If no consensus is reached, describe the different views and options.
 - Record benefits and harms considerations.
- Recommendations should be specific and unambiguous.

Available forms/checklists:

Framework for justification of recommendations ①

08. Writing the guideline draft

Principles for writing

Once key questions are answered and there is consensus about the guideline's recommendations, the first draft version can be written. However, writing in committee requires prior agreement about the consistent use of terminology and writing style. ESHRE guidelines should be written in English and within a European scope. Furthermore, they should be comprehensive and flexible in order to allow adaptation to diverse settings and circumstances of clinical practice.

The use of paragraphs and headings are recommended to facilitate readers' navigation. Moreover, the use of tables, illustrations, figures and algorithms is encouraged. For guideline uniformity an ESHRE guideline is written according to an established structure.

The research specialist is responsible for merging the input of the different GDG members and to adapt the content where needed to result in a consistent and well-structured guideline.

Guideline structure

In general, an ESHRE guideline consists of 3 main parts; the introduction section, the key question-related part and the annexes.

The introduction section is subdivided in the introduction to the guideline, and a clinical introduction.

The introduction to the guideline includes the scope of the document based on the scoping checklist, information on target users, and details on how the guideline was developed (referring to an annex with the full methodology). If relevant, previous versions of the guideline are listed in this section.

Depending on the topic of the guideline, the clinical introduction contains disease definition and terminology, prevalence, variation in practice, provision of suboptimal care, personal and societal costs.

For the key question related part, the guideline development group determines a logical order for reporting the total collection of key questions with their recommendations. One such method is to follow the patients' pathway, starting with the clinical evaluation (e.g. symptoms, risk factors) followed by the diagnostics, treatment options, follow-up, complications, information provision. Per key question, the following items are reported:

- Key question
- Evidence: a descriptive summary of the selected clinical evidence, with GRADE summary of findings tables (or a reference to the tables in annex).
- Recommendations: one or more recommendations, in boxes with appropriate GRADE.
- Justification: a summary of the relevant evidence (with most important limitations) and the considerations taken into

account when determining the strength of the recommendations. In the case of non-consensus, practice statements about the different schools of thought should be recorded. Furthermore, the explanatory text gives room for considerations from ethical or legal perspectives.

- References.

If there are no recommendations for a certain section, a conclusion and justification can be written.

The annexes contain:

- List of abbreviations
- Glossary
- Summary of findings tables
- Evidence tables
- Details on the literature study: flowcharts, list of excluded studies
- Guideline development group, with list of declared conflicts of interest
- Research recommendations: describe gaps in scientific knowledge for future investigation
- Methodology (Appendix 5), that describes the guideline development in detail:
 - Guideline development process in 12 steps, according to the manual: funding, tailored information on the scoping procedure, details on the literature searches (searched databases, timeframe, inclusion and exclusion criteria), methodology of writing recommendations, and the used guidelines' manual version.
 - Information on the quality of evidence, grades of recommendations and phrasing, i.e. basic information needed to understand and interpret the recommendations.
 - Strategy for review of the guideline draft: invited reviewers, review deadlines, processing of comments.
 - Guideline implementation strategy, tailored to the guideline.
- Schedule for updating the guideline
- List of reviewers in stakeholder consultation
- Guideline versions and dissemination (e.g. existence of additional tools), if relevant.
- Relationship with other existing guidelines or ESHRE documents, if relevant.
- Key priorities for implementation, if relevant.

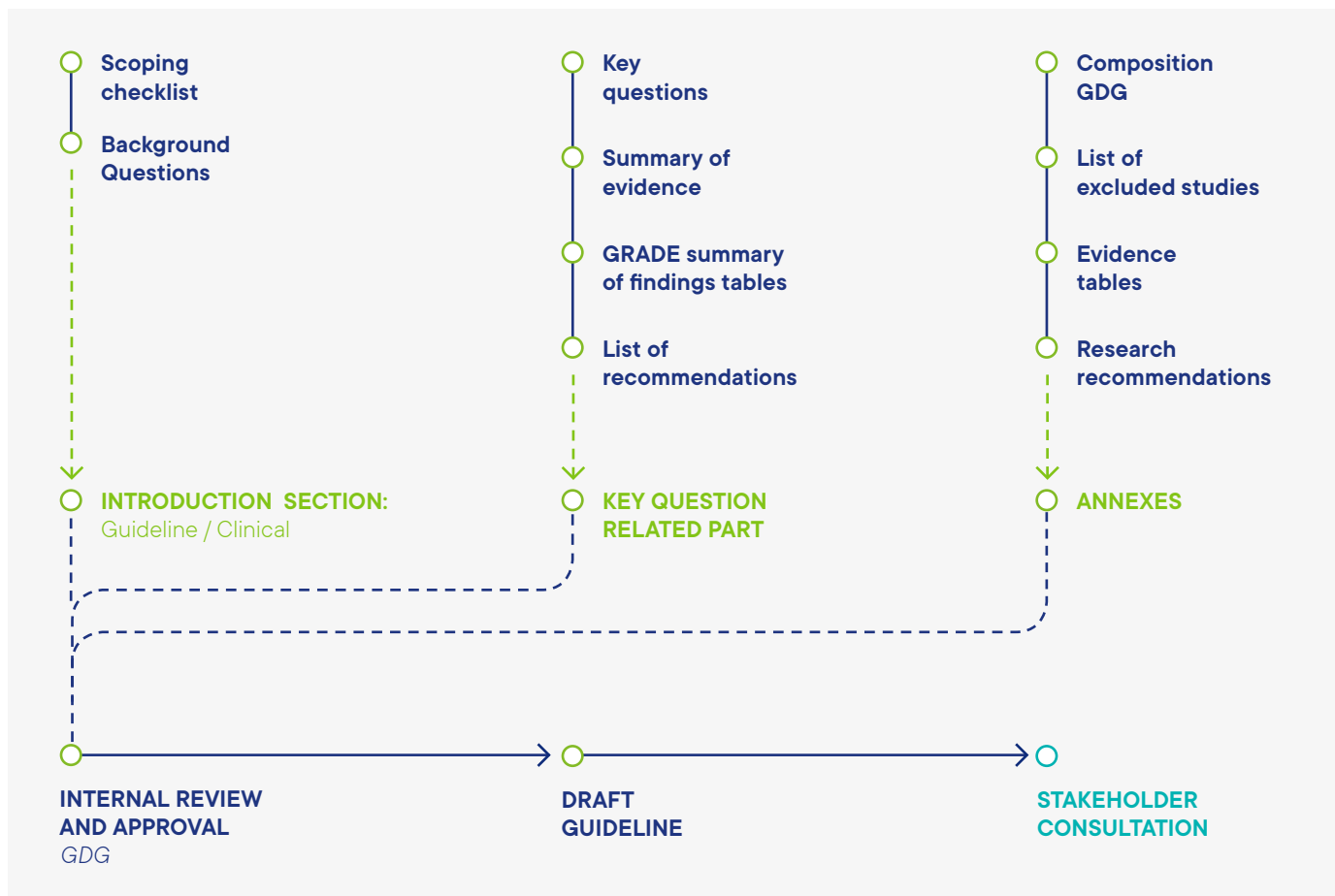
A legal disclaimer is also added at the back of the cover of the guidelines.

A summary of all recommendation (condensed version) will be published in one of the ESHRE journals. This version will contain the most important content of full guideline, at least all key questions and recommendations.

Tips

- Check if recommendations answer the key questions.
- Use the AGREE Instrument (www.agreetrust.org) as a checklist.
- Develop or adopt a standardized format for reporting the guideline, with specific structure, headings, and content.
- Decide on the format(s) to be prepared for the guideline product(s) (e.g. full guideline, full guideline with technical report/ systematic reviews, brief guideline for clinicians or policymakers, consumer version for patients) that will correspond to the dissemination plan.
- Seek approval from all members of the guideline development group for the final document(s).
- Link the recommendations explicitly to the supporting evidence.
- Present the different management options clearly.
- Present if appropriate expected exceptions for recommendation application.
- Facilitate recommendation identification (e.g. bullets, numbering, boxes).
- Discuss potential barriers in applying the recommendations.
- Consider potential cost implications of applying the recommendations.

Summary




09. Stakeholder consultation

The final stages of guideline development involve review by future users and approval by the parties involved. Within this phase the adequacy of the guideline document is evaluated, especially for its methodological quality, its clinical content and its applicability.

Review procedure

The review phase starts with a review of the guideline draft by several stakeholders. Their consultations concern in particular the guidelines' comprehensiveness, the accuracy of evidence interpretation and the acceptance of recommendations.

A sample of the target group, all members of the involved SIGs, patients' representatives (lay reviewers), and representatives of national organizations are invited by email to review the draft. At the same time, the draft is web posted with an invitation to review for all ESHRE members. Interested reviewers must sign a statement of confidentiality and submit their review comments within six weeks. Following this procedure results in an addition to the reviewers' list which will be mentioned as an appendix to the guideline (Reviewer comments form ).

For adapted (parts of) or updated clinical guidelines the ESHRE draft guideline should be also send for review to the original developers.

The comments received from reviewers are tabulated and discussed in the stakeholder review report. The GDG will respond to each of the comments but does not commit to altering the recommendations in all instances. If no change is made, the reasons for this will be recorded in the report.

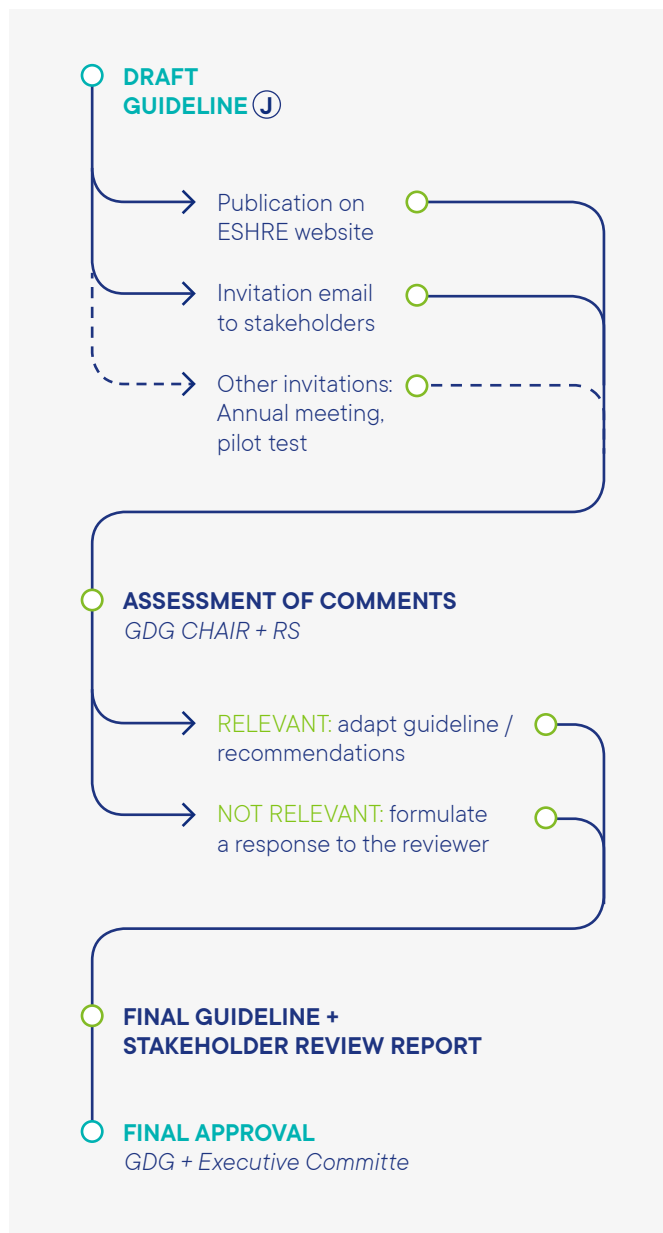
Any alterations to the recommendations must be made with the agreement of the whole GDG and noted in the report. This reviewer comments processing report is published on the ESHRE website alongside the guideline.

Additional options for stakeholder consultation

An additional open meeting at the ESHRE annual meeting is also an option for review. This provides the opportunity to present preliminary conclusions and draft recommendations to a wider audience and to hear valuable suggestions for additional evidence or alternative evidence interpretation. Because participation in such a meeting generates a sense of ownership across geographical and disciplinary boundaries, the organization of such a meeting might accelerate the internal consensus process, the review procedure and final implementation.

A draft guideline can also be pilot tested before a wider launch. This step can detect problems in formatting, usability and acceptance.

Summary



Tips

- Use the reviewing and piloting phase as an opportunity to advertise the existence of a new guideline.
- Set a policy and process for handling consumer and stakeholder feedback and dealing with different perspectives (e.g. ensure that diverse perspectives are taken into account in making decisions, provide transparent rationale for judgments made, provide an appeal process for stakeholders, publish consultation comments and the guideline development panel's responses).
- Record the stakeholder review reporting methodology, document the internal and external peer review process and, if applicable, publish consultation comments and the guideline development group's responses.
- Document the enrollment and selection of consumers and stakeholders for the guideline panel and the involvement and consultation with all other consumers and stakeholders to ensure explicit and transparent methods.

Available forms/checklists:

Reviewer comments form 

10. Approval

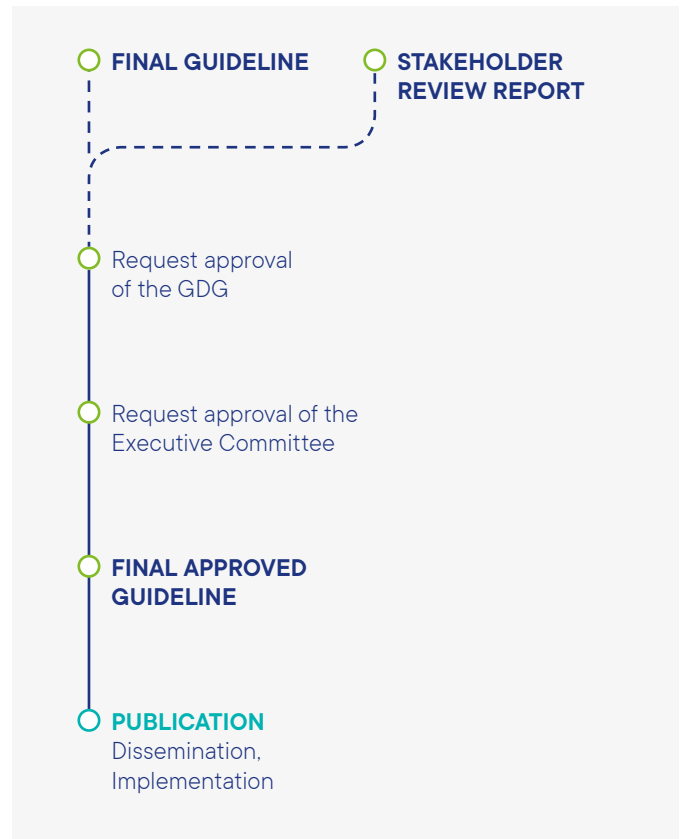
Final version & authorization

After stakeholder consultation, the entire GDG should formally approve the final version of the guideline.

After completion of all revisions, English language reviewers and proofreaders (and possibly lawyers) can be called upon when necessary.

The final step is to submit the guideline for formal approval by the ESHRE Executive Committee.

Summary



11. Publication, dissemination, translation and implementation

Dissemination of ESHRE guidelines is considered as a continuation of the work of the GDG and involves making guidelines accessible, advertising their availability, and distributing them widely. Guidelines are (most) effective if their dissemination and implementation are carefully considered and vigorously pursued. If not, the time, energy and costs devoted to their development are wasted and potential improvements in reproductive health care are passed.

A range of dissemination strategies can be effective, for instance the:

- use of short summaries
- promotion of guideline's development/existence
- publication in professional journal(s)
- publication on the internet and links on related websites

Currently there is too little evidence to support decisions about which guideline dissemination strategy is efficient under which circumstances. In general, the use of multi-faceted dissemination strategies is recommended.

The standard dissemination procedure for all ESHRE guidelines comprises publishing and announcement.

Publications

The document will be published at least in 2 formats:

- Publication of the full guideline on the ESHRE website (www.eshre.eu/guidelines)
- Publication of the summary guideline (including all recommendations) in one of the ESHRE journals. After publication, a link to the paper is added to the guideline page.

Additional options

Distribution of guidelines alone has been shown to be ineffective in achieving change in practice; guidelines are more likely to be effective if they are disseminated by a strategy based on barrier research, by an active educational intervention or by patient-specific reminders. However, the extent of potential clinical benefits and resources required to introduce guidelines - and the likely benefits and costs as a result of any provider's behavior change - need to be considered carefully before developing additional tools. The efficiency of a dissemination strategy is best evaluated in the presence of different barriers and effect modifiers.

Two more options are the development of algorithms and patient information:

- An algorithm is a flow chart of the clinical decision pathway described in the guideline, in which process steps and decision points are linked by arrows. Whenever relevant and useful,

flowcharts can be digitalized in decision-supports in a web-based or smartphone/tablet version.

- Patient information summarizes the recommendations in the ESHRE guideline in everyday language. It aims to help patients understand the guideline's recommendations and facilitates decision-making. Moreover, the patient information may be used by hospitals or patient organizations for developing their own information leaflets. Patient versions of guidelines will be developed in collaboration with the patient representative involved in the guideline and its accuracy and correctness will be checked by the chair, or a delegate guideline development group member. If possible, a review of the patient version will be organized by inviting all relevant patient organizations to send in comments. The final version will be distributed among all relevant patient organizations with an invitation to endorse it, and if necessary, translate it. The translated version should be checked by a guideline development group member of the specific country or the national representative of this country and should contain a disclaimer, provided by ESHRE, stating that the English version is the reference document.

Dissemination

All relevant ESHRE communication channels will be used to announce the release of a new guideline:

- A newflash on the ESHRE website's homepage
- A mailing to the members of the relevant SIGs, or all ESHRE members
- An announcement in "Focus on Reproduction"
- Promotion at the annual ESHRE meeting via different media. Optionally, participants will be informed about the development and release of new guidelines during a specific guideline session.
- A mailing to all related National Societies to inform them about the guideline release. They are asked to encourage local implementation by, for instance, translations or condensed versions, but they are also offered a website link to the original document.
- All appropriate remaining stakeholders - for instance, European policy makers, patient societies and industry representatives - will be separately informed.

Guideline translation and endorsement

An important factor facilitating guideline implementation is endorsement by professional groups. Endorsement of ESHRE guidelines is always sought from relevant National Societies by informing their presidents.

ESHRE and the guideline group members put significant effort into developing evidence-based guidelines. Furthermore, we try to involve different nationalities in the guideline groups, and to organize a broad stakeholder review. By doing so, we provide guidelines written to apply to a broad population which ideally should be endorsed by national societies, and if wished upon, translated verbatim.

National Societies and organizations can request permission to translate (verbatim) one of the guidelines, or any specific content related to the guidelines, in their language. For an official ESHRE approved translation, a straightforward 4-step procedure of approval, translation, validation and publication is outlined in a policy (see p38–39).

For reasons of consistency only one translation of a certain ESHRE guideline in any given language is accepted by ESHRE. ESHRE reserves the exclusive right to publish the first edition of all ESHRE guidelines and post its translation on the ESHRE website. National Societies must secure copyright protection in their own country.

When a verbatim translation of a guideline is insufficient for national uptake, ESHRE will allow for national societies to use the ESHRE guidelines as the basis of their national guidelines. To ensure transparency, the methodology should clearly refer to the ESHRE Guideline (including the weblink www.eshre.eu/guidelines) and state how the ESHRE guideline was used, including which text blocks / search strings have been used from the ESHRE guideline and for which topics the recommendations differ significantly between the documents. The resulting document will be considered a national guideline, not an ESHRE guideline.

ESHRE gives National Societies and organizations the optional right to publish the translated guideline in their own national journals. All costs of carrying out these rights and of translating the guideline are for the National Societies.

The above information applies only to documents to which ESHRE holds the copyright. For translation of the summary guideline published in one of the ESHRE journals, permission of Oxford University Press (OUP) should be requested.

Background information

Policy for the translation of ESHRE® Documents⁴

Please note that this policy sets out general rules with regard to the translation of ESHRE® Documents (as defined below). Depending on the type of ESHRE® Documents, specific provisions might also be applicable (as is for example the case for the ESHRE® guidelines). In case of a conflict between the provisions of this policy and specific provisions, the latter shall prevail.

Translation of ESHRE® Documents:

In summary, the following four steps must be followed in case of translation of an ESHRE® Document:

1. Request written permission of ESHRE® before endeavouring translation
2. Make an exact translation and ensure that the ESHRE® copyright statement and the ESHRE® disclaimer are foreseen on the document, as well as full reference to the ESHRE® Document
3. Request written validation of the translation from ESHRE®
4. Ensure that the translation is up-to-date and corresponds to the latest version of the ESHRE® Document

1. Prior permission to translate

A National Society shall have the right – at its own cost – to translate ESHRE® Documents and publish the translations thereof in its own country upon (i) prior written approval of ESHRE® and (ii) full endorsement of the corresponding parent ESHRE® Document.

For reasons of consistency, ESHRE® shall accept only one translation per ESHRE® Document in any given language.

At all times, ESHRE® retains full (copy)rights whatsoever on every ESHRE® Document and its translations.

2. Obligations for the translators and the National Society

General

All costs and expenses relating to the translation of an ESHRE® Document (including the cost of compensating translators) shall be borne by the National Society exclusively.

The National Society ensures that every translator transfers all rights whatsoever (which the latter might possibly possess with respect to the performed translation) to ESHRE®.

The National Society shall be responsible for the exact translation of the ESHRE® Document by the translator it appeals on. Each translation shall contain all textual, pictorial and diagrammatic material, as foreseen in the ESHRE® Document, without any alterations. Footnotes or annexes may be added to highlight national and/or regional practices. In no event, amendments to the original text shall be allowed.

Further, the National Society (and the translator it appeals on) undertake to:

1. give full credit to ESHRE® for the ESHRE® Document by including on the title page of the translated document:
 - the ESHRE® copyright statement (as mentioned below),
 - the ESHRE® logo,
 - full reference to the original publication of the ESHRE® Document on ESHRE's® website and in ESHRE's® official journals ('Human Reproduction');
2. foresee the appropriate ESHRE® disclaimer, as mentioned below, in the translated document;
3. mention in the title of the translated document the name of each ESHRE® working group member who is (co-) author of the ESHRE® Document; and
4. clarify in the (sub)title of the translated document that it entails a translation from an ESHRE® Document, whereby the full title of the parent ESHRE® Document needs to be mentioned.

Whenever possible, a back-to-back translation is recommended.

The National society that produces a translation of an ESHRE® Document may foresee the translated document of its own logo(s) and additional information about its society. The names of the translators, reviewers and/or other people involved in the translation of the ESHRE® Document, can also be foreseen on the translated document, provided that it has been made clear they were solely involved in the translation of the ESHRE® Document and thus took no part in the production and publication of the ESHRE® Document.

⁴ any document, produced and published by ESHRE®, to which ESHRE® exclusively possesses all rights of ownership. The English version always entails the original version of the document;

Background information

Translation sponsored by companies

In case a National Society obtains sponsoring from commercial organisations in order to finance the translations of ESHRE® Documents, it shall be strictly prohibited to foresee in any kind of product advertising on the translated document.

However, corporate logos of the sponsoring company(ies) in question can be displayed with the following statement: 'The translation of this ESHRE® document was made possible through an educational grant from [name sponsor]. [Name sponsor] acknowledges explicitly that it was not involved in the actual production and publication of the parent ESHRE® document, hence influenced in no way the content thereof.'

3. Validation of the translation

All documents translated in line with the above can only be published upon prior written validation of ESHRE®. Such validation shall:

1. be organised by the ESHRE® central office;
2. be performed by a native speaker from the ESHRE® working group or the committee of national representatives; and
3. only relate to the translation itself and in no case entail a review of the content, meaning that ESHRE® shall not verify if the scientific value of the parent ESHRE® Document has been preserved in the translated document.

ESHRE® strives to inform the National Society of the outcome of the performed validation within four weeks upon receipt of the translation by ESHRE®.

Validated translations of ESHRE® Documents will be published by ESHRE® on its website, upon prior written approval of the respective National Society.

4. Keep the translation up to date

It is the responsibility of a National Society to ensure that the translated document is kept up-to-date and corresponds to the latest version of the parent ESHRE® Document.

ESHRE® strives to inform the National Society of any updates on the parent ESHRE® Document, and this within due time.

ESHRE® copyright statement

"Copyright © European Society of Human Reproduction and Embryology ('ESHRE®') – All rights reserved"

ESHRE® disclaimer

"This publication entails a translation of an original ESHRE® document – as fully referred to on the title page of this document – whereby such translation was performed in line with the provisions of the 'Policy for the translation of ESHRE® Documents', as consultable on the ESHRE® website (www.eshre.eu).

The translation of the original ESHRE® document is made by and under supervision of [name of the National Society], which is solely responsible for the content of this translation. Prior validation of ESHRE® of this translation does not affect such responsibility.

If any questions arise related to the accuracy of the information contained in the translation and/or its scientific value, please refer to the original ESHRE® document. Any discrepancies or differences created in the translation are not binding to ESHRE® and shall have no legal effect for compliance or enforcement purposes. The English version, being the language in which the original ESHRE® document is published, shall always prevail."

Implementation and evaluation

Guidelines do not implement themselves. Local ownership of the implementation process is crucial for changing practice. ESHRE is responsible for the development of European guidelines and their implementability, but not directly for their implementation into local practice. Nevertheless, the identification of barriers to guidelines' acceptance is one of the first steps of an implementation process and has ideally been part of the guideline developmental phase. Instruments like The Guideline Implementability Appraisal instrument can be helpful for identifying obstacles to guideline implementation [27] (http://nutmeg.med.yale.edu/glia/doc/GLIA_v2.pdf).

There are different types of barriers to guideline implementation:

- internal to the guideline itself
- factors related to the individual care providers (e.g. attitude and skills)
- factors related to the (social) setting (e.g. patients' and colleagues' characteristics)
- external factors related to the system (e.g. reimbursement).

After the determination of factors affecting guideline adoption, the currently recommended approach is to plan a targeted intervention. However, there is no specific guidance available for translating identified barriers into tailor-made implementation interventions. Each implementation strategy is effective under certain circumstances, and a multifaceted approach is more likely to succeed than a single approach. Evaluation of such complex interventions is therefore important and mostly undertaken by investigators with research funding.

Focusing on individual recommendations rather than on the guideline as a whole makes the implementation initiative more manageable. Criteria reflecting one or more of the six quality domains defined by the Institute of Medicine (safety, effectiveness, patient-centeredness, timely, efficiency and equitability) can help to prioritize guideline's recommendations for this purpose.

At an appropriate time after dissemination and implementation an evaluation is necessary for insight into the impact of the guideline. Such an evaluation consists of several components, namely an assessment of:

- guideline dissemination
- change in practice performance
- change in health outcomes
- change in consumer's knowledge and understanding
- economic consequences.

Practice performance is usually measured by a clinical audit and indicators. The frequently used definition for an indicator is "a measurable element of practice performance for which there is

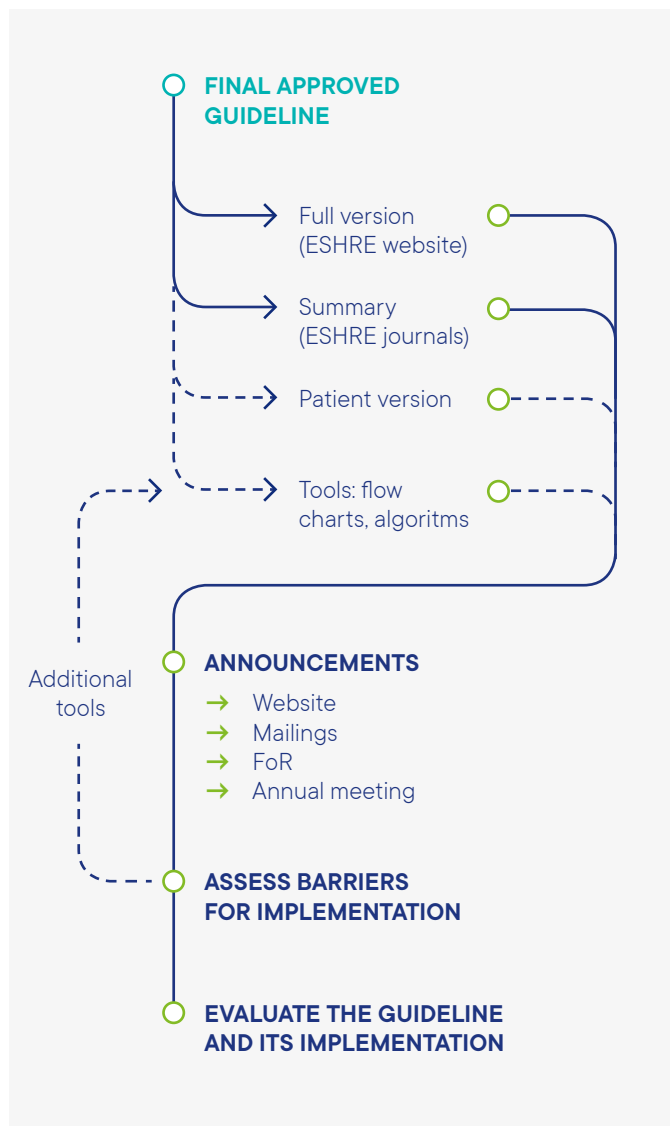
evidence or consensus that it can be used to assess the quality, and hence change in the quality of care provided". Based on the manual by the Agency for Healthcare Research and Quality (AHRQ) and additional literature on quality indicators in infertility, a set of quality indicators for each ESHRE guideline can be developed in a 3-step process:

- The GDG members rank the recommendations on priority for implementation to obtain key recommendations.
- The GDG members propose quality indicators for each key recommendation.
- The GDG members determine the importance and the preparedness to measure for each quality indicator, in a stepwise process.

For each step specific structured questionnaires should be developed.

The resulting set of key quality indicators will be used to evaluate the quality of care and the impact of the guideline on the quality of care within Europe.

Summary



Tips

- Develop or adapt tools, support, and derivative products to provide guidance on how the recommendations can be implemented into practice (e.g. mobile applications, integration with clinical decision support systems, make guideline adaptable as an educational resource for target audience for education outreach).
- Make considerations for adaptation of the guideline and provide specific instructions for how target end users who would like to adapt the guidelines to other contexts can do so in a systematic and transparent way (e.g. modifying a recommendation based on local resources and baseline risk, implications that deviate from the judgements made by the guideline panel).
- Conduct an internal evaluation (i.e. self-assessment) of the guideline development process, including the guideline panel meeting(s) held to formulate recommendations, by asking guideline group members for feedback.
- Consider pilot testing the guideline with the target end users (e.g. with members of target audience and stakeholders who participated in the guideline development group).
- Provide criteria and tools for target end users to monitor and audit the implementation and use of the guideline recommendations (e.g. identify outcomes that should change with implementation and suggest methods for measuring the outcomes).
- Provide support and tools for prospective evaluation of the guideline to determine its effectiveness after implementation (e.g. using randomized evaluations where possible, using before-after evaluations cautiously due to uncertainties regarding the effects of implementation).
- Consider the potential involvement of the guideline development group in prospective evaluation(s) of the guideline (e.g. partnering with organizations that implement the guideline to plan evaluation studies).
- Plan to collect feedback and evaluations from users to identify how to improve the intrinsic implementability of the recommendations in subsequent versions of the guideline.
- Support the guideline with application tools and record those within the guideline.

12. Updating the guideline

Updating of guidelines is an essential part of guideline development, to ensure guidelines remain current and their quality is maintained [28]. New evidence is emerging constantly, and recommendations may be compromised when they are no longer reflective of the current clinical evidence. An analysis of NICE guidelines showed that the median life span of the clinical guidelines was 60 months (95% CI 51 – 69) [29]. The authors also suggested a system of monitoring to detect guidelines that are quickly outdated. As new evidence is published at different rates in different fields, a full revision of guidelines after a fixed time period is not always appropriate. Recent publications on guideline updating propose systems of monitoring and assessment of the need for a review, before endeavoring on a formal update [28].

Guideline monitoring

Guideline monitoring includes guideline assessment, streamlined systematic review and formal update. Guideline assessment aims to identify documents that are no longer applicable to current practice and documents with recommendations that may have been, have the potential to be, or are expected to be invalidated by new evidence. Guideline assessment can be performed by the document assessment questionnaire (6 questions, see tool [\(K\)](#)).

The application of the document assessment questionnaire classifies each guideline into one of the following four groups: (1) endorse, (2) defer, (3) review, or (4) archive the guideline.

Only for guidelines categorized as review from the assessment are eligible for the next step. The research specialist performs a streamlined systematic review (without a full quality assessment) of new evidence using the original search strategy and study selection criteria and summarize the new evidence from studies and reviews.

The clinical expert (or the entire GDG) reviews the new evidence to determine:

1. if it supports or contradicts current recommendations;
2. if the current recommendations cover all relevant subjects addressed by the new evidence, and
3. if strong evidence that may change the current recommendations is expected to be published in the near future.

Based on the assessment for each PICO question, it is categorized as endorse, update or archive. Endorse means that the newly identified evidence supports the current recommendations with only minor changes or new qualifying statements; update means that the new evidence requires changes to the existing recommendations; and archive means that the document cannot be endorsed or deferred, and a full update is not either feasible or desired.

Complete or partial guideline update

If a need for review is identified for one or more PICO questions (partial review), or the full guideline (complete review), approval for the update must be requested from the ESHRE Executive Committee by completing the application form, and a report of the assessment. A complete review approved by the ESHRE Executive Committee will follow the usual process described in this manual. Updated guidelines are also subject to consultation and will follow the usual validation process.

For all ESHRE guidelines, an annex will be added documenting the outcomes of the annual assessment and details of the updating procedure.

Tips

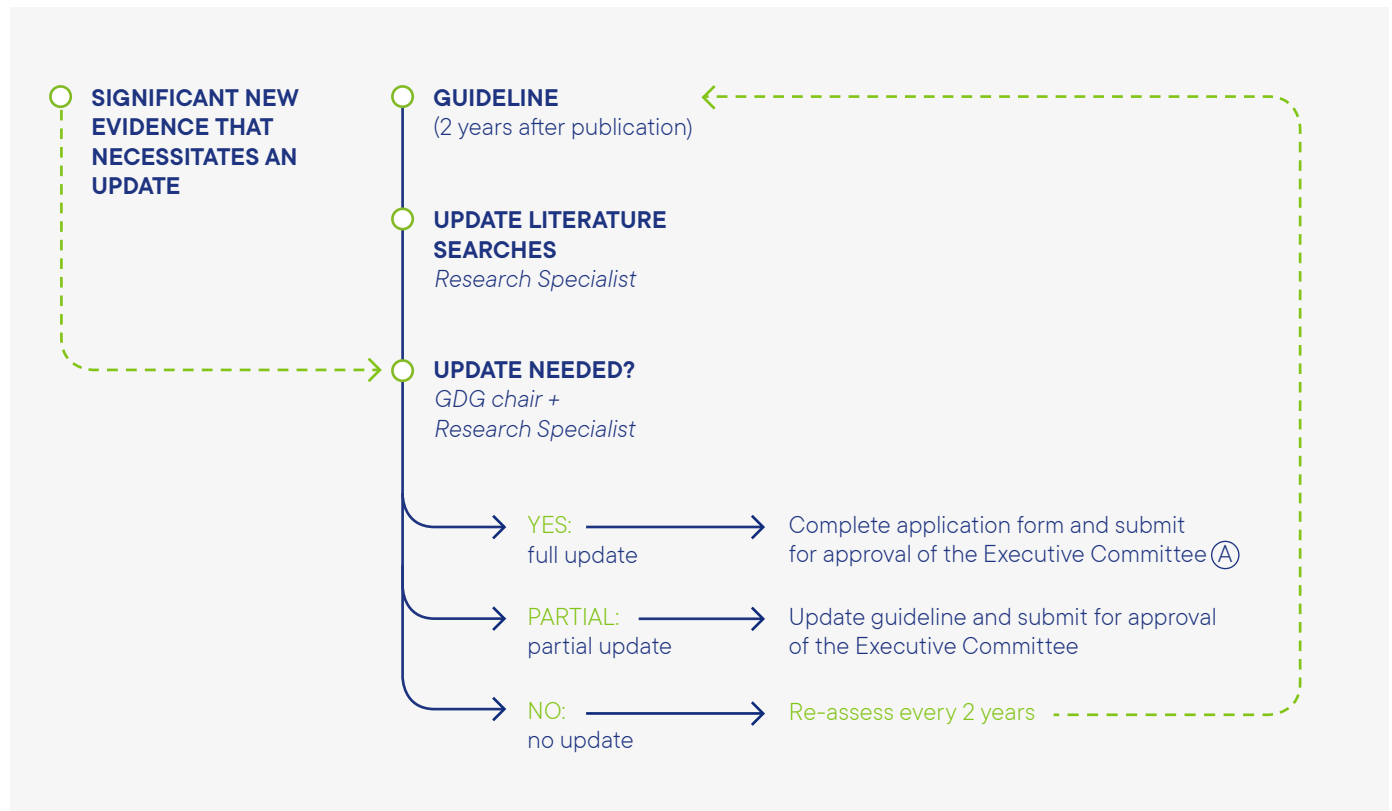
- Decide who will be responsible for routinely monitoring the literature and assessing whether new significant evidence is available (e.g. consider involvement of experts not previously involved in the guideline development group to periodically review the guideline).
- Make arrangements for guideline group membership and participation after completion of the guideline (e.g. rotating membership every 1–2 years, selection of a new group at time of updating, continuing participation by guideline panel chair).
- Plan the logistics for updating the guideline in the future.
- Refer to the procedure for guideline updating.

Available forms/checklists:

Application form [\(A\)](#)

Document assessment tool [\(K\)](#)

Summary



Reference list

1. Committee on Clinical Practice Guidelines IoM (1992) Guidelines for clinical practice: from development to use. National Academy Press, Washington DC
2. Brouwers MC, Kerkvliet K, Spithoff K, Consortium ANS (2016) The AGREE Reporting Checklist: a tool to improve reporting of clinical practice guidelines. *BMJ* 352:i1152
3. AGREE Reporting Checklist (2016).
4. The GRADE Working Group (2013) GRADE handbook for grading quality of evidence and strength of recommendations. Updated October 2013. In: Schünemann H BJ, Guyatt G, Oxman A, (ed)
5. Scottish Intercollegiate Guidelines Network (SIGN) (2015) SIGN 50: A guideline developer's handbook. (SIGN publication no. 50). [November 2015]. Available from URL: <http://www.sign.ac.uk>, Edinburgh
6. Organization WH (2012) WHO handbook for guideline development.
7. Hilbink MA, Ouwens MM, Burgers JS, Kool RB (2014) A new impetus for guideline development and implementation: construction and evaluation of a toolbox. *Implement Sci* 9 (1):34. doi:10.1186/1748-5908-9-34
8. Schünemann HJ, Wiercioch W, Etzeandía I, Falavigna M, Santesso N, Mustafa R, Ventresca M, Brignardello-Petersen R, Laisaar K-T, Kowalski S, Baldeh T, Zhang Y, Rait U, Neumann I, Norris SL, Thornton J, Harbour R, Treweek S, Guyatt G, Alonso-Coello P, Reinap M, Brožek J, Oxman A, Akl EA (2014) Guidelines 2.0: systematic development of a comprehensive checklist for a successful guideline enterprise. *Canadian Medical Association Journal* 186 (3):E123–E142. doi:10.1503/cmaj.131237
9. GIN-McMaster Guideline Development Checklist.
10. Moses RE, Feld AD (2008) Legal risks of clinical practice guidelines. *Am J Gastroenterol* 103 (1):7–11. doi:10.1111/j.1572-0241.2007.01399.x
11. Fervers B, Burgers JS, Haugh MC, Latreille J, Mlika-Cabanne N, Paquet L, Coulombe M, Poirier M, Burnand B (2006) Adaptation of clinical guidelines: literature review and proposition for a framework and procedure. *Int J Qual Health Care* 18 (3):167–176. doi:10.1093/intqhc/mzi108
12. Darzi A, Abou-Jaoude EA, Agarwal A, Lakis C, Wiercioch W, Santesso N, Brax H, El-Jardali F, Schünemann HJ, Akl EA (2017) A methodological survey identified eight proposed frameworks for the adaptation of health related guidelines. *J Clin Epidemiol* 86:3–10. doi:10.1016/j.jclinepi.2017.01.016
13. Murphy MK, Black NA, Lamping DL, McKee CM, Sanderson CF, Askham J, Marteau T (1998) Consensus development methods, and their use in clinical guideline development. *Health Technol Assess* 2 (3):i-iv, 1–88
14. den Breejen EM, Hermens RP, Galama WH, Willemsen WN, Kremer JA, Nelen WL (2016) Added value of involving patients in the first step of multidisciplinary guideline development: a qualitative interview study among infertile patients. *Int J Qual Health Care* 28 (3):299–305. doi:10.1093/intqhc/mzw020
15. Richardson WS, Wilson MC, Nishikawa J, Hayward RS (1995) The well-built clinical question: a key to evidence-based decisions. *ACP J Club* 123 (3):A12–13
16. The Cochrane Collaboration (2011) Cochrane Handbook for Systematic Reviews of Interventions. In: Julian PT Higgins, Sally Green (eds)
17. Cook DJ, Greengold NL, Ellrodt AG, Weingarten SR (1997) The relation between systematic reviews and practice guidelines. *Ann Intern Med* 127 (3):210–216
18. Balshem H, Helfand M, Schünemann HJ, Oxman AD, Kunz R, Brozek J, Vist GE, Falck-Ytter Y, Meerpohl J, Norris S, Guyatt GH (2011) GRADE guidelines: 3. Rating the quality of evidence. *J Clin Epidemiol* 64 (4):401–406. doi:10.1016/j.jclinepi.2010.07.015
19. Guyatt GH, Oxman AD, Vist G, Kunz R, Brozek J, Alonso-Coello P, Montori V, Akl EA, Djulbegovic B, Falck-Ytter Y, Norris SL, Williams JW, Jr., Atkins D, Meerpohl J, Schünemann HJ (2011) GRADE guidelines: 4. Rating the quality of evidence—study limitations (risk of bias). *J Clin Epidemiol* 64 (4):407–415. doi:10.1016/j.jclinepi.2010.07.017
20. Guyatt GH, Oxman AD, Kunz R, Woodcock J, Brozek J, Helfand M, Alonso-Coello P, Glasziou P, Jaeschke R, Akl EA, Norris S, Vist G, Dahm P, Shukla VK, Higgins J, Falck-Ytter Y, Schünemann HJ, Group GW (2011) GRADE guidelines: 7. Rating the quality of evidence—inconsistency. *J Clin Epidemiol* 64 (12):1294–1302. doi:10.1016/j.jclinepi.2011.03.017
21. Guyatt GH, Oxman AD, Kunz R, Woodcock J, Brozek J, Helfand M, Alonso-Coello P, Falck-Ytter Y, Jaeschke R, Vist G, Akl EA, Post PN, Norris S, Meerpohl J, Shukla VK, Nasser M, Schünemann HJ, Group GW (2011) GRADE guidelines: 8. Rating the quality of evidence—indirectness. *J Clin Epidemiol* 64 (12):1303–1310. doi:10.1016/j.jclinepi.2011.04.014
22. Guyatt GH, Oxman AD, Kunz R, Brozek J, Alonso-Coello P, Rind D, Devereaux PJ, Montori VM, Freyschuss B, Vist G, Jaeschke R, Williams JW, Jr., Murad MH, Sinclair D, Falck-Ytter Y, Meerpohl J, Whittington C, Thorlund K, Andrews J, Schünemann HJ (2011) GRADE guidelines 6. Rating the quality of evidence—imprecision. *J Clin Epidemiol* 64 (12):1283–1293. doi:10.1016/j.jclinepi.2011.01.012
23. Guyatt GH, Oxman AD, Montori V, Vist G, Kunz R, Brozek J, Alonso-Coello P, Djulbegovic B, Atkins D, Falck-Ytter Y, Williams JW, Jr., Meerpohl J, Norris SL, Akl EA, Schünemann HJ (2011) GRADE guidelines: 5. Rating the quality of evidence—publi-

- cation bias. *J Clin Epidemiol* 64 (12):1277–1282. doi:10.1016/j.jclinepi.2011.01.011
24. Guyatt GH, Oxman AD, Sultan S, Glasziou P, Akl EA, Alonso-Coello P, Atkins D, Kunz R, Brozek J, Montori V, Jaeschke R, Rind D, Dahm P, Meerpohl J, Vist G, Berliner E, Norris S, Falck-Ytter Y, Murad MH, Schunemann HJ, Group GW (2011) GRADE guidelines: 9. Rating up the quality of evidence. *J Clin Epidemiol* 64 (12):1311–1316. doi:10.1016/j.jclinepi.2011.06.004
25. Andrews J, Guyatt G, Oxman AD, Alderson P, Dahm P, Falck-Ytter Y, Nasser M, Meerpohl J, Post PN, Kunz R, Brozek J, Vist G, Rind D, Akl EA, Schunemann HJ (2013) GRADE guidelines: 14. Going from evidence to recommendations: the significance and presentation of recommendations. *J Clin Epidemiol* 66 (7):719–725. doi:10.1016/j.jclinepi.2012.03.013
26. Andrews JC, Schunemann HJ, Oxman AD, Pottie K, Meerpohl JJ, Coello PA, Rind D, Montori VM, Brito JP, Norris S, Elbarbary M, Post P, Nasser M, Shukla V, Jaeschke R, Brozek J, Djulbegovic B, Guyatt G (2013) GRADE guidelines: 15. Going from evidence to recommendation—determinants of a recommendation's direction and strength. *J Clin Epidemiol* 66 (7):726–735. doi:10.1016/j.jclinepi.2013.02.003
27. Shiffman RN, Dixon J, Brandt C, Essaihi A, Hsiao A, Michel G, O'Connell R (2005) The GuideLine Implementability Appraisal (GLIA): development of an instrument to identify obstacles to guideline implementation. *BMC Med Inform Decis Mak* 5 (1):23. doi:10.1186/1472-6947-5-23
28. Agbassi C, Messersmith H, McNair S, Brouwers M (2014) Priority-based initiative for updating existing evidence-based clinical practice guidelines: the results of two iterations. *J Clin Epidemiol* 67 (12):1335–1342. doi:10.1016/j.jclinepi.2014.06.013
29. Alderson LJ, Alderson P, Tan T (2014) Median life span of a cohort of National Institute for Health and Care Excellence clinical guidelines was about 60 months. *J Clin Epidemiol* 67 (1):52–55. doi:10.1016/j.jclinepi.2013.07.012
30. Shea BJ, Grimshaw JM, Wells GA, Boers M, Andersson N, Hamel C, Porter AC, Tugwell P, Moher D, Bouter LM (2007) Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews. *BMC medical research methodology* 7 (1):10
31. Whiting P, Rutjes A, Dinnes J, Reitsma J, Bossuyt P, Kleijnen J (2004) Development and validation of methods for assessing the quality of diagnostic accuracy studies. *Gray Pub.*

Forms

- Ⓐ Application form
- Ⓑ Disclose form
- Ⓒ Confidentiality form
- Ⓓ Scooping checklist
- Ⓔ Pico checklist
- Ⓕ Relevance and quality check
- Ⓖ Quality assessment checklist
- Ⓗ Evidence tables
- Ⓘ Framework for recommendations
- ⓵ Reviewer comments form
- Ⓚ Document assessment tool for updating guidelines

A. Application form

Guideline / Good practice recommendations document

Applicants

CONTACT PERSON(S):

.....

ESHRE SPECIAL INTEREST GROUP(S):

.....

SUGGESTED MEMBERS OF THE WORKING GROUP (EXPERTS AND/ OR ESHRE SIG REPRESENTATIVES)⁵

.....

Topic

PROPOSED TITLE:

.....

GUIDELINE OR GOOD PRACTICE RECOMMENDATIONS:

.....

PROPOSED (CLINICAL) PROBLEM:

.....

THE RELEVANCE OF THE PROPOSED CLINICAL PROBLEM (E.G. VOLUME, COSTS AND PATIENT IMPACT):

.....

MAIN OUTCOME(S) TO BE ADDRESSED BY THE PROPOSED GUIDELINE/ GOOD PRACTICE RECOMMENDATIONS:

.....

INDICATION OF ACTUAL PRACTICE VARIATION:

.....

EXPECTED BENEFIT(S) FROM THE PROPOSED GUIDELINE/ GOOD PRACTICE RECOMMENDATIONS DEVELOPMENT AND IMPLEMENTATION:

.....

INDICATION OF THE SIZE AND STRENGTH OF THE EVIDENCE FOR THE PROPOSED TOPIC:

.....

OTHER COMMENTS: (IN CASE OF A GOOD PRACTICE RECOMMENDATIONS DOCUMENT, PLEASE CLARIFY METHODOLOGY, SCHEDULE AND COSTS FOR THE PROJECT)

.....

Other existing guidelines/consensus documents (to be completed by RS)

EXISTING GUIDELINES WITHIN THE FIELD OF THE PROPOSED TOPIC:

.....

OVERLAP WITH OTHER ESHRE DOCUMENTS:

.....

The completed application form should be sent to nathalie@eshre.eu

⁵ If feasible suggest a few names. A final list of WG members will have to be presented to and approved by ExCo before the working group can start.

B. Disclosure form

All ESHRE guideline development group members are expected to provide completed and signed disclosure statements about all financial, personal, or professional relationships with industry, individuals, or organizations to avoid the perception of a conflict of interest. Updates should be made if changes occur during the guideline development process.

ESHRE GUIDELINE

.....

Contact information of the guideline development group member

NAME

.....

INSTITUTION, ADDRESS:

.....

E-MAIL ADDRESS:

.....

Information on potential conflicts of interest from the last 3 years, or anticipated in the next 12 months

☐ I HAVE NO POTENTIAL CONFLICT OF INTEREST FROM THE LAST 3 YEARS TO REPORT

☐ I HAVE THE FOLLOWING POTENTIAL CONFLICT(S) OF INTEREST FROM THE LAST 3 YEARS TO REPORT:

☐ RESEARCH GRANT(S) FROM ONE OR MORE COMPANIES, FROM

.....

☐ CONSULTING FEE(S) FOR E.G. SERVICES ON AN ADVISORY BOARD OR LEGAL TESTIMONY, FROM

.....

☐ SPEAKER'S FEE(S) FOR INSTANCE AS COMPENSATION FOR LECTURING AND TRAVEL, FROM

.....

☐ SALARY OR POSITION FUNDING, FROM

.....

☐ OWNERSHIP INTEREST BY STOCK (OPTIONS) OR PARTNERSHIP OF A HEALTHCARE COMPANY, FROM

.....

☐ OTHER (FINANCIAL) BENEFIT E.G. BY INSTITUTIONAL CONFLICTS OF INTEREST IN THE TOPICS OR ISSUES ADDRESSED IN THE DOCUMENT

.....

SIGNATURE (OR STATE YOUR NAME):

.....

DATE:

.....



C. Confidentiality form

As a writer of an ESHRE guideline you have been or may be exposed to certain confidential and/or proprietary information, materials or data. It is important to the integrity of the writing process and final work that this information should be kept strictly confidential and not disclosed at any time under any circumstance.

ESHRE GUIDELINE

.....

Contact information of the guideline development group member

NAME

.....

INSTITUTION, ADDRESS:

.....

E-MAIL ADDRESS:

.....

Statement of confidentiality

I will not disclose any confidential and/or proprietary information, materials or data related to Guideline Development Group's work to any third party, but keep this information strictly confidential.

I will keep any confidential and/or proprietary information, materials or data in my possession in a safe and secure place to protect against inadvertent disclosure.

I will not use any confidential information and/or proprietary information, materials or data for any purpose other than participating in an ESHRE guideline development procedure.

SIGNATURE (OR STATE YOUR NAME)::

.....

DATE:

.....

D. Scoping checklist

1. WHAT IS/ARE THE OVERALL PURPOSE(S) OF THE PROPOSED GUIDELINE?

Specify health intents (i.e., prevention, diagnosis, treatment, etc.) and expected benefits or outcomes. E.g. preventing thromboembolic complications of patients undergoing elective orthopedic surgery.

.....

2. WHAT IS THE PROPOSED TARGET PATIENT POPULATION?

Specify subjects to whom those recommendations apply (i.e. patients, society, etc.). E.g. adults undergoing elective orthopedic surgery, all women 40 years of age or older, etc.

.....

3. WHAT IS THE PROPOSED HEALTH CARE SETTING?

Specify level of health care (i.e. primary, secondary, etc.) where these recommendations are supposed to be implemented.

.....

4. WHICH INTERVENTIONS SHOULD BE INCLUDED IN THE GUIDELINE?

Specify which preventive, therapeutic and diagnostic interventions will be covered and which will be not.

.....

5. WHICH OUTCOME(S) SHOULD BE ADDRESSED BY THE PROPOSED GUIDELINE?

Specify which outcome(s) would be preferred, which are commonly reported and which are preferred by patients.

.....

6. WHO ARE THE TARGET USERS OF THE PROPOSED GUIDELINE, AND WHO ARE THE KEY STAKEHOLDERS?

Specify all relevant professional groups, institutions, patients, public, etc. who are target users or beneficiaries of these guidelines and/or whose views should be sought.

.....

7. WHAT PREFERS THE PROPOSED PATIENT POPULATION?

Is this already included? Which methodology/methodologies will be used to include patients' preferences?

.....

8. WHAT ARE KEY RESOURCES TO CONSIDER?

Specify resources needed for the implementation of guidelines (i.e. need for additional human resources, equipment, infrastructure, system changes, etc.) and potential barriers to implementation.

.....

9. WHAT IS THE RELATION TO OTHER DOCUMENTS?

List all existing documents/guidelines on the same or similar topic that are likely to be currently used in practice (e.g. guidelines developed by other organizations).

.....

E. PICO checklist

For each PICO question, provide the following information:

| | |
|---|------------|
| KEY QUESTION: | |
| PATIENTS/POPULATION: | |
| INTERVENTION: | |
| COMPARISON: | |
| OUTCOMES: | CRITICAL: |
| | IMPORTANT: |
| | OTHER: |
| <i>Other databases to be searched for this question⁶</i> | |
| <i>Suggested key words/ search terms/synonyms:</i> | |

⁶ in addition to Medline/Pubmed and Cochrane Library

F. Relevance and quality check

| REFERENCE | TYPE | RELEVANT FULL-TEXT? | EXPLANATION (EXCLUSION CRITERION) | QUALITY FULL-TEXT | EXPLANATION (EXCLUSION CRITERION) |
|-----------|------|---|-----------------------------------|---|-----------------------------------|
| XX | SR | <input type="radio"/> YES <input type="radio"/> NO | ----- | <input type="radio"/> HIGH <input type="radio"/> MODERATE <input type="radio"/> LOW | ----- |

If a paper is found to be not-relevant, or of low quality, the GDG member should list the exclusion criteria.

(Remarks can be added in case of relevant good quality papers as well).

Possible exclusion criteria (non-exhaustive list):

→ Not relevant:

- ☐ Publication type : Case report - Expert opinion - Editorial
- ☐ Relevant patients are not included, or only as subgroup
- ☐ Relevant intervention is not included
- ☐ Relevant outcomes are not assessed or inappropriately assessed
- ☐ Study is included in meta-analysis / More recent meta-analysis available

→ Low Quality

- ☐ Selection bias
- ☐ Performance bias
- ☐ Attrition bias
- ☐ Detection bias
- ☐ Other sources of bias:
 - Study population: for instance too small
 - Methodology
 - Insufficient data

Instructions for the GDG members

- 1. You will receive all full text papers and a relevance and quality check table with all references.**
For some full text papers we do not have access to the full text.
- 2. Check whether the paper is relevant. If not, list an explanation/exclusion criterion**
- 3. If relevant, assess the quality of the paper (with the appropriate checklist for risk of bias assessment). List the most important criteria for your judgement on quality (especially in case of low quality)**
- 4. Send in the completed relevance and quality check table.**

G. Quality assessment checklist

AMSTAR – a measurement tool to assess the methodological quality of systematic reviews [30]⁷.

| | |
|--|--|
| <p>1. WAS AN 'A PRIORI' DESIGN PROVIDED?</p> <p>The research question and inclusion criteria should be established before the conduct of the review.</p> <p><i>Note: Need to refer to a protocol, ethics approval, or pre-determined/a priori published research objectives to score a "yes."</i></p> | <p><input type="radio"/> Yes</p> <p><input type="radio"/> No</p> <p><input type="radio"/> Can't answer</p> <p><input type="radio"/> Not applicable</p> |
| <p>2. WAS THERE DUPLICATE STUDY SELECTION AND DATA EXTRACTION?</p> <p>There should be at least two independent data extractors and a consensus procedure for disagreements should be in place.</p> <p><i>Note: 2 people do study selection, 2 people do data extraction, consensus process or one person checks the other's work.</i></p> | <p><input type="radio"/> Yes</p> <p><input type="radio"/> No</p> <p><input type="radio"/> Can't answer</p> <p><input type="radio"/> Not applicable</p> |
| <p>3. WAS A COMPREHENSIVE LITERATURE SEARCH PERFORMED?</p> <p>At least two electronic sources should be searched. The report must include years and databases used (e.g., Central, EMBASE, and MEDLINE). Key words and/or MESH terms must be stated and where feasible the search strategy should be provided. All searches should be supplemented by consulting current contents, reviews, textbooks, specialized registers, or experts in the particular field of study, and by reviewing the references in the studies found.</p> <p><i>Note: If at least 2 sources + one supplementary strategy used, select "yes" (Cochrane register/Central counts as 2 sources; a grey literature search counts as supplementary).</i></p> | <p><input type="radio"/> Yes</p> <p><input type="radio"/> No</p> <p><input type="radio"/> Can't answer</p> <p><input type="radio"/> Not applicable</p> |
| <p>4. WAS THE STATUS OF PUBLICATION (I.E. GREY LITERATURE) USED AS AN INCLUSION CRITERION?</p> <p>The authors should state that they searched for reports regardless of their publication type. The authors should state whether or not they excluded any reports (from the systematic review), based on their publication status, language etc.</p> <p><i>Note: If review indicates that there was a search for "grey literature" or "unpublished literature," indicate "yes." SIGLE database, dissertations, conference proceedings, and trial registries are all considered grey for this purpose. If searching a source that contains both grey and non-grey, must specify that they were searching for grey/unpublished lit.</i></p> | <p><input type="radio"/> Yes</p> <p><input type="radio"/> No</p> <p><input type="radio"/> Can't answer</p> <p><input type="radio"/> Not applicable</p> |

⁷ Additional notes (in italics) made by Michelle Weir, Julia Worswick, and Carolyn Wayne based on conversations with Bev Shea and/or Jeremy Grimshaw in June and October 2008 and July and September 2010.

(AMSTAR – a measurement tool to assess the methodological quality of systematic reviews [30]⁷.)

| | |
|---|--|
| <p>5. WAS A LIST OF STUDIES (INCLUDED AND EXCLUDED) PROVIDED?</p> <p>A list of included and excluded studies should be provided.</p> <p><i>Note: Acceptable if the excluded studies are referenced. If there is an electronic link to the list but the link is dead, select “no.”</i></p> | <p><input type="radio"/> Yes</p> <p><input type="radio"/> No</p> <p><input type="radio"/> Can’t answer</p> <p><input type="radio"/> Not applicable</p> |
| <p>6. WERE THE CHARACTERISTICS OF THE INCLUDED STUDIES PROVIDED?</p> <p>In an aggregated form such as a table, data from the original studies should be provided on the participants, interventions and outcomes. The ranges of characteristics in all the studies analyzed e.g., age, race, sex, relevant socioeconomic data, disease status, duration, severity, or other diseases should be reported.</p> <p><i>Note: Acceptable if not in table format as long as they are described as above.</i></p> | <p><input type="radio"/> Yes</p> <p><input type="radio"/> No</p> <p><input type="radio"/> Can’t answer</p> <p><input type="radio"/> Not applicable</p> |
| <p>7. WAS THE SCIENTIFIC QUALITY OF THE INCLUDED STUDIES ASSESSED AND DOCUMENTED?</p> <p>‘A priori’ methods of assessment should be provided (e.g., for effectiveness studies if the author(s) chose to include only randomized, double-blind, placebo controlled studies, or allocation concealment as inclusion criteria); for other types of studies alternative items will be relevant.</p> <p><i>Note: Can include use of a quality scoring tool or checklist, e.g., Jadad scale, risk of bias, sensitivity analysis, etc., or a description of quality items, with some kind of result for EACH study (“low” or “high” is fine, as long as it is clear which studies scored “low” and which scored “high”; a summary score/range for all studies is not acceptable).</i></p> | <p><input type="radio"/> Yes</p> <p><input type="radio"/> No</p> <p><input type="radio"/> Can’t answer</p> <p><input type="radio"/> Not applicable</p> |
| <p>8. WAS THE SCIENTIFIC QUALITY OF THE INCLUDED STUDIES USED APPROPRIATELY IN FORMULATING CONCLUSIONS?</p> <p>The results of the methodological rigor and scientific quality should be considered in the analysis and the conclusions of the review, and explicitly stated in formulating recommendations.</p> <p><i>Note: Might say something such as “the results should be interpreted with caution due to poor quality of included studies.” Cannot score “yes” for this question if scored “no” for question 7.</i></p> | <p><input type="radio"/> Yes</p> <p><input type="radio"/> No</p> <p><input type="radio"/> Can’t answer</p> <p><input type="radio"/> Not applicable</p> |

(AMSTAR – a measurement tool to assess the methodological quality of systematic reviews [30]⁷.)

| | |
|--|--|
| <p>9. WERE THE METHODS USED TO COMBINE THE FINDINGS OF STUDIES APPROPRIATE?</p> <p>For the pooled results, a test should be done to ensure the studies were combinable, to assess their homogeneity (i.e., Chi-squared test for homogeneity, I^2). If heterogeneity exists a random effects model should be used and/or the clinical appropriateness of combining should be taken into consideration (i.e., is it sensible to combine?).</p> <p><i>Note: Indicate “yes” if they mention or describe heterogeneity, i.e., if they explain that they cannot pool because of heterogeneity/variability between interventions.</i></p> | <p><input type="radio"/> Yes</p> <p><input type="radio"/> No</p> <p><input type="radio"/> Can't answer</p> <p><input type="radio"/> Not applicable</p> |
| <p>10. WAS THE LIKELIHOOD OF PUBLICATION BIAS ASSESSED?</p> <p>An assessment of publication bias should include a combination of graphical aids (e.g., funnel plot, other available tests) and/or statistical tests (e.g., Egger regression test, Hedges-Olken).</p> <p><i>Note: If no test values or funnel plot included, score “no”. Score “yes” if mentions that publication bias could not be assessed because there were fewer than 10 included studies.</i></p> | <p><input type="radio"/> Yes</p> <p><input type="radio"/> No</p> <p><input type="radio"/> Can't answer</p> <p><input type="radio"/> Not applicable</p> |
| <p>11. WAS THE CONFLICT OF INTEREST INCLUDED?</p> <p>Potential sources of support should be clearly acknowledged in both the systematic review and the included studies.</p> <p><i>Note: To get a “yes,” must indicate source of funding or support for the systematic review AND for each of the included studies.</i></p> | <p><input type="radio"/> Yes</p> <p><input type="radio"/> No</p> <p><input type="radio"/> Can't answer</p> <p><input type="radio"/> Not applicable</p> |

Note: the outcome of the quality assessment should be documented in the relevance and quality check document.

Checklist: Randomized controlled trials

| SELECTION BIAS (SYSTEMATIC DIFFERENCES BETWEEN THE COMPARISON GROUPS) | |
|---|---|
| An appropriate method of randomization was used to allocate participants to treatment groups (which would have balanced any confounding factors equally across groups). | <input type="radio"/> Yes <input type="radio"/> No <input type="radio"/> Can't answer <input type="radio"/> Not applicable |
| There was adequate concealment of allocation (such that investigators/participants cannot influence enrolment or treatment allocation) ⁸ | <input type="radio"/> Yes <input type="radio"/> No <input type="radio"/> Can't answer <input type="radio"/> Not applicable |
| The groups were comparable at baseline, including all major confounding and prognostic factors. | <input type="radio"/> Yes <input type="radio"/> No <input type="radio"/> Can't answer <input type="radio"/> Not applicable |
| Based on your answers to the above, in your opinion was selection bias present? If so, consider the likely direction of its effect? | <input type="radio"/> Yes <input type="radio"/> No <input type="radio"/> Can't answer <input type="radio"/> Not applicable |

⁸ Lack of allocation concealment: Those enrolling patients are aware of the group (or period in a crossover trial) to which the next enrolled patient will be allocated (a major problem in "pseudo" or "quasi" randomized trials with allocation by day of week, birth date, chart number, etc.).

(Checklist: *Randomized controlled trials*)

| PERFORMANCE BIAS (SYSTEMATIC DIFFERENCES BETWEEN GROUPS IN THE CARE PROVIDED, APART FROM THE INTERVENTION UNDER INVESTIGATION) | |
|---|---|
| The comparison groups received the same care apart from the intervention(s) studied | <input type="radio"/> Yes <input type="radio"/> No <input type="radio"/> Can't answer <input type="radio"/> Not applicable |
| Participants receiving care were kept 'blind' to treatment allocation | <input type="radio"/> Yes <input type="radio"/> No <input type="radio"/> Can't answer <input type="radio"/> Not applicable |
| Individuals administering care were kept 'blind' to treatment allocation ⁹ | <input type="radio"/> Yes <input type="radio"/> No <input type="radio"/> Can't answer <input type="radio"/> Not applicable |
| Based on your answers to the above, in your opinion was performance bias present? If so, consider the likely direction of its effect? | <input type="radio"/> Yes <input type="radio"/> No <input type="radio"/> Can't answer <input type="radio"/> Not applicable |

⁹ Lack of blinding : Patient, caregivers, those recording outcomes, those adjudicating outcomes, or data analysts are aware of the arm to which patients are allocated (or the medication currently being received in a crossover trial)

(Checklist: *Randomized controlled trials*)

| ATTRITION BIAS (DIFFERENCES BETWEEN THE COMPARISON GROUPS WITH RESPECT TO LOSS OF PARTICIPANTS) ¹⁰ | |
|---|---|
| All groups were followed up for an equal length of time (or analysis was adjusted to allow for differences in length of follow-up) | <input type="radio"/> Yes <input type="radio"/> No <input type="radio"/> Can't answer <input type="radio"/> Not applicable |
| The groups were comparable for treatment completion. (How many participants did not complete treatment in each group?) | <input type="radio"/> Yes <input type="radio"/> No <input type="radio"/> Can't answer <input type="radio"/> Not applicable |
| The groups were comparable with respect to the availability of outcome data (For how many participants in each group were no outcome data available?) ¹¹ | <input type="radio"/> Yes <input type="radio"/> No <input type="radio"/> Can't answer <input type="radio"/> Not applicable |
| Based on your answers to the above, in your opinion was attrition bias present? If so, consider the likely direction of its effect? | <input type="radio"/> Yes <input type="radio"/> No <input type="radio"/> Can't answer <input type="radio"/> Not applicable |

¹⁰ Incomplete accounting of patients and outcome events: Loss to follow-up and failure to adhere to the intention-to-treat principle in superiority trials; or in noninferiority trials, loss to follow-up, and failure to conduct both analyses considering only those who adhered to treatment, and all patients for whom outcome data are available.

The significance of particular rates of loss to follow-up, however, varies widely and is dependent on the relation between loss to follow-up and number of events. The higher the proportion lost to follow-up in relation to intervention and control group event rates, and differences between intervention and control groups, the greater the threat of bias.

¹¹ Incomplete or absent reporting of some outcomes and not others on the basis of the results.

(Checklist: Randomized controlled trials)

| DETECTION BIAS (BIAS IN HOW OUTCOMES ARE ASCERTAINED, DIAGNOSED OR VERIFIED) | |
|---|---|
| The study had an appropriate length of follow-up | <input type="radio"/> Yes <input type="radio"/> No <input type="radio"/> Can't answer <input type="radio"/> Not applicable |
| The study used a precise definition of outcome | <input type="radio"/> Yes <input type="radio"/> No <input type="radio"/> Can't answer <input type="radio"/> Not applicable |
| A valid and reliable method was used to determine the outcome | <input type="radio"/> Yes <input type="radio"/> No <input type="radio"/> Can't answer <input type="radio"/> Not applicable |
| Investigators were kept 'blind' to participants' exposure to the intervention and other important confounding and prognostic factors | <input type="radio"/> Yes <input type="radio"/> No <input type="radio"/> Can't answer <input type="radio"/> Not applicable |
| Based on your answers to the above, in your opinion was performance bias present? If so, consider the likely direction of its effect? | <input type="radio"/> Yes <input type="radio"/> No <input type="radio"/> Can't answer <input type="radio"/> Not applicable |
| OVERALL QUALITY OF THE RCTS | |
| Is the methodology used for the RCT sufficiently robust to permit a valid conclusion? | <input type="radio"/> Yes <input type="radio"/> No <input type="radio"/> Can't answer <input type="radio"/> Not applicable |
| What is your overall assessment of the methodological quality of this RCT | <input type="radio"/> Yes <input type="radio"/> No <input type="radio"/> Can't answer <input type="radio"/> Not applicable |

Note: the outcome of the quality assessment should be documented in the relevance and quality check document.

Checklist: Cohort studies

| SELECTION BIAS (SYSTEMATIC DIFFERENCES BETWEEN THE COMPARISON GROUPS) | |
|---|---|
| The method of allocation to treatment groups was unrelated to potential confounding factors (that is, the reason for participant allocation to treatment groups is not expected to affect the outcome[s] under study) ¹² | <input type="radio"/> Yes <input type="radio"/> No <input type="radio"/> Can't answer <input type="radio"/> Not applicable |
| Attempts were made within the design or analysis to balance the comparison for potential confounders ¹³ | <input type="radio"/> Yes <input type="radio"/> No <input type="radio"/> Can't answer <input type="radio"/> Not applicable |
| The groups were comparable at baseline, including all major confounding and prognostic factors | <input type="radio"/> Yes <input type="radio"/> No <input type="radio"/> Can't answer <input type="radio"/> Not applicable |
| Based on your answers to the above, in your opinion was selection bias present? If so, consider the likely direction of its effect? | <input type="radio"/> Yes <input type="radio"/> No <input type="radio"/> Can't answer <input type="radio"/> Not applicable |

¹² Failure to develop and apply appropriate eligibility criteria (inclusion of control population); Selection of exposed and unexposed in cohort studies from different populations

¹³ Failure to adequately control confounding; Failure of accurate measurement of all known prognostic factors – Failure to match for prognostic factors and/or adjustment in statistical analysis

Checklist: Cohort studies

| PERFORMANCE BIAS (SYSTEMATIC DIFFERENCES BETWEEN GROUPS IN THE CARE PROVIDED, APART FROM THE INTERVENTION UNDER INVESTIGATION) | |
|---|---|
| The comparison groups received the same care apart from the intervention(s) studied | <input type="radio"/> Yes <input type="radio"/> No <input type="radio"/> Can't answer <input type="radio"/> Not applicable |
| Participants receiving care were kept 'blind' to treatment allocation | <input type="radio"/> Yes <input type="radio"/> No <input type="radio"/> Can't answer <input type="radio"/> Not applicable |
| Individuals administering care were kept 'blind' to treatment allocation | <input type="radio"/> Yes <input type="radio"/> No <input type="radio"/> Can't answer <input type="radio"/> Not applicable |
| Based on your answers to the above, in your opinion was performance bias present? If so, consider the likely direction of its effect? | <input type="radio"/> Yes <input type="radio"/> No <input type="radio"/> Can't answer <input type="radio"/> Not applicable |

Checklist: Cohort studies

| ATTRITION BIAS (DIFFERENCES BETWEEN THE COMPARISON GROUPS WITH RESPECT TO LOSS OF PARTICIPANTS) | |
|---|---|
| All groups were followed up for an equal length of time (or analysis was adjusted to allow for differences in length of follow-up) | <input type="radio"/> Yes <input type="radio"/> No <input type="radio"/> Can't answer <input type="radio"/> Not applicable |
| The groups were comparable for treatment completion. (How many participants did not complete treatment in each group?) | <input type="radio"/> Yes <input type="radio"/> No <input type="radio"/> Can't answer <input type="radio"/> Not applicable |
| The groups were comparable with respect to the availability of outcome data (For how many participants in each group were no outcome data available?) | <input type="radio"/> Yes <input type="radio"/> No <input type="radio"/> Can't answer <input type="radio"/> Not applicable |
| Based on your answers to the above, in your opinion was attrition bias present? If so, consider the likely direction of its effect? | <input type="radio"/> Yes <input type="radio"/> No <input type="radio"/> Can't answer <input type="radio"/> Not applicable |

Checklist: Cohort studies

| DETECTION BIAS (BIAS IN HOW OUTCOMES ARE ASCERTAINED, DIAGNOSED OR VERIFIED) | |
|---|---|
| The study had an appropriate length of follow-up ¹⁴ | <input type="radio"/> Yes <input type="radio"/> No <input type="radio"/> Can't answer <input type="radio"/> Not applicable |
| The study used a precise definition of outcome | <input type="radio"/> Yes <input type="radio"/> No <input type="radio"/> Can't answer <input type="radio"/> Not applicable |
| A valid and reliable method was used to determine the outcome | <input type="radio"/> Yes <input type="radio"/> No <input type="radio"/> Can't answer <input type="radio"/> Not applicable |
| Investigators were kept 'blind' to participants' exposure to the intervention and other important confounding and prognostic factors | <input type="radio"/> Yes <input type="radio"/> No <input type="radio"/> Can't answer <input type="radio"/> Not applicable |
| Based on your answers to the above, in your opinion was performance bias present? If so, consider the likely direction of its effect? | <input type="radio"/> Yes <input type="radio"/> No <input type="radio"/> Can't answer <input type="radio"/> Not applicable |

¹⁴ Incomplete or inadequately short follow-up: Especially within prospective cohort studies, both groups should be followed for the same amount of time.



Checklist: Cohort studies

| OVERALL QUALITY OF THE COHORT STUDIES | |
|--|---|
| Is the methodology used for the cohort study sufficiently robust to permit a valid conclusion? | <div><input type="radio"/> Yes</div> <div><input type="radio"/> No</div> <div><input type="radio"/> Can't answer</div> <div><input type="radio"/> Not applicable</div> |
| What is your overall assessment of the methodological quality of this study | <div><input type="radio"/> HIGHT QUALITY</div> <div><input type="radio"/> MODERATE QUALITY</div> <div><input type="radio"/> LOW QUALITY</div> <div><input type="radio"/> VERY LOW QUALITY</div> |

Note: the outcome of the quality assessment should be documented in the relevance and quality check document.

Checklist: Case control studies

| SELECTION BIAS (SYSTEMATIC DIFFERENCES BETWEEN THE COMPARISON GROUPS) | |
|---|---|
| The cases and controls are taken from comparable populations | <input type="radio"/> Yes <input type="radio"/> No <input type="radio"/> Can't answer <input type="radio"/> Not applicable |
| The same exclusion criteria are used for cases and controls | <input type="radio"/> Yes <input type="radio"/> No <input type="radio"/> Can't answer <input type="radio"/> Not applicable |
| The participation rate was similar between cases and controls, and participants and non-participants are compared to establish their similarities and differences | <input type="radio"/> Yes <input type="radio"/> No <input type="radio"/> Can't answer <input type="radio"/> Not applicable |
| Cases are clearly defined and differentiated from controls | <input type="radio"/> Yes <input type="radio"/> No <input type="radio"/> Can't answer <input type="radio"/> Not applicable |
| PERFORMANCE BIAS (SYSTEMATIC DIFFERENCES BETWEEN GROUPS IN THE CARE PROVIDED, APART FROM THE INTERVENTION UNDER INVESTIGATION) | |
| Measures were taken to prevent knowledge of primary exposure from influencing case ascertainment | <input type="radio"/> Yes <input type="radio"/> No <input type="radio"/> Can't answer <input type="radio"/> Not applicable |
| Exposure status is measured in a standard, valid and reliable way | <input type="radio"/> Yes <input type="radio"/> No <input type="radio"/> Can't answer <input type="radio"/> Not applicable |
| The main potential confounders are identified and taken into account in design and analysis | <input type="radio"/> Yes <input type="radio"/> No <input type="radio"/> Can't answer <input type="radio"/> Not applicable |

Checklist: Case control studies

| DETECTION BIAS (BIAS IN HOW OUTCOMES ARE ASCERTAINED, DIAGNOSED OR VERIFIED) | |
|--|--|
| Have confidence intervals been provided | <input type="radio"/> Yes <input type="radio"/> No <input type="radio"/> Can't answer <input type="radio"/> Not applicable |
| OVERALL QUALITY OF CASE-CONTROL STUDIES | |
| Is the methodology used for the cohort study sufficiently robust to permit a valid conclusion? | <input type="radio"/> Yes <input type="radio"/> No <input type="radio"/> Can't answer <input type="radio"/> Not applicable |
| What is your overall assessment of the methodological quality of this study | <input type="radio"/> HIGHT QUALITY <input type="radio"/> MODERATE QUALITY <input type="radio"/> LOW QUALITY <input type="radio"/> VERY LOW QUALITY |

Note: the outcome of the quality assessment should be documented in the relevance and quality check document.

Checklist: Studies of diagnostic accuracy

Checklist based on QUADAS tool [31]

| | |
|--|---|
| Was the spectrum of patients representative of the patients who will receive the test in practice? | <input type="radio"/> Yes <input type="radio"/> No <input type="radio"/> Can't answer <input type="radio"/> Not applicable |
| Were selection criteria clearly described? | <input type="radio"/> Yes <input type="radio"/> No <input type="radio"/> Can't answer <input type="radio"/> Not applicable |
| Is the reference standard likely to classify the target condition correctly? | <input type="radio"/> Yes <input type="radio"/> No <input type="radio"/> Can't answer <input type="radio"/> Not applicable |
| Is the period between reference standard and index test short enough to be reasonably sure that the target condition did not change between the two tests? | <input type="radio"/> Yes <input type="radio"/> No <input type="radio"/> Can't answer <input type="radio"/> Not applicable |
| Did the whole sample or a random selection of the sample receive verification using a reference standard of diagnosis? | <input type="radio"/> Yes <input type="radio"/> No <input type="radio"/> Can't answer <input type="radio"/> Not applicable |
| Did patients receive the same reference standard regardless of the index test result? | <input type="radio"/> Yes <input type="radio"/> No <input type="radio"/> Can't answer <input type="radio"/> Not applicable |
| Was the reference standard independent of the index test (i.e. the index test was not part of the reference standard)? | <input type="radio"/> Yes <input type="radio"/> No <input type="radio"/> Can't answer <input type="radio"/> Not applicable |

Checklist: Studies of diagnostic accuracy

| | |
|---|---|
| Was the execution of the index test described in sufficient detail to permit replication of the test? | <input type="radio"/> Yes <input type="radio"/> No <input type="radio"/> Can't answer <input type="radio"/> Not applicable |
| Was the execution of the reference standard described in sufficient detail to permit its replication? | <input type="radio"/> Yes <input type="radio"/> No <input type="radio"/> Can't answer <input type="radio"/> Not applicable |
| Were the index test results interpreted without knowledge of the results of the reference standard? | <input type="radio"/> Yes <input type="radio"/> No <input type="radio"/> Can't answer <input type="radio"/> Not applicable |
| Were the reference standard results interpreted without knowledge of the results of the index test? | <input type="radio"/> Yes <input type="radio"/> No <input type="radio"/> Can't answer <input type="radio"/> Not applicable |
| Were the same clinical data available when test results were interpreted as would be available when the test is used in practice? | <input type="radio"/> Yes <input type="radio"/> No <input type="radio"/> Can't answer <input type="radio"/> Not applicable |
| Were uninterpretable/intermediate test results reported? | <input type="radio"/> Yes <input type="radio"/> No <input type="radio"/> Can't answer <input type="radio"/> Not applicable |
| Were withdrawals from the study explained? | <input type="radio"/> Yes <input type="radio"/> No <input type="radio"/> Can't answer <input type="radio"/> Not applicable |



Checklist: Studies of diagnostic accuracy

| OVERALL QUALITY OF THE DIAGNOSTIC STUDIES | |
|---|---|
| Is the methodology used for the diagnostic cohort study sufficiently robust to states the conclusion is reliable? | <div><input type="radio"/> Yes</div> <div><input type="radio"/> No</div> <div><input type="radio"/> Can't answer</div> <div><input type="radio"/> Not applicable</div> |
| What is your overall assessment of the methodological quality of this study | <div><input type="radio"/> HIGHT QUALITY</div> <div><input type="radio"/> MODERATE QUALITY</div> <div><input type="radio"/> LOW QUALITY</div> <div><input type="radio"/> VERY LOW QUALITY</div> |

Note: the outcome of the quality assessment should be documented in the relevance and quality check document.

H. Evidence tables

Intervention studies

Evidence table

| REFERENCE | STUDY TYPE | PATIENTS NO. OF PATIENTS PATIENT CHARACTERISTICS + GROUP COMPARABILITY | INTERVENTIONS (+COMPARISON) INCLUDE: STUDY DURATION / FOLLOW-UP | OUTCOME MEASURES INCLUDE: HARMS / ADVERSE EVENTS | EFFECT SIZE | AUTHORS CONCLUSIONS | COMMENTS |
|-----------|------------|--|---|--|-------------|---------------------|----------|
| | | | | | | | |

Details

| | |
|---|---|
| REFERENCE (+PMID) | Author, journal, year, volume and pages + PMID |
| STUDY TYPE | Meta-analysis/ systematic review (SR), randomized controlled trial (RCT), non-randomized cohort study (CS), case report (CASE), opinion documents (OPINION), other (specify)... |
| PATIENTS NO. PATIENTS CHARACTERISTICS + GROUP COMPARABILITY | Total number of patients, and the number of patients involved in each group Describe relevant baseline characteristics (age, disease status, inclusion / exclusion criteria,...), if stated per group and add comment for comparability of groups. |
| INTERVENTIONS (+COMPARISON) INCLUDE STUDY DURATION / FOLLOW-UP | Specify the interventions per group (treatment/procedure (dose, regimen, length...) / placebo/ alternative treatment / expectant management) duration patients participate the study from inclusion to a specified end-point (e.g. implantation) or the end of data collection, mention per outcome if relevant |
| OUTCOME MEASURES (PRIMARY/SECONDARY) INCLUDE HARMS / ADVERSE EVENTS | All outcome measures (positive and negative): e.g. OHSS occurrence rate, implantation rate, pregnancy, quality of life, satisfaction,... Divide between primary (1) and secondary (2) outcomes when this is specified by the author |
| EFFECT SIZE (INCLUDE HARMS) | Absolute risk reduction, relative risks, numbers needed to treat/harm, or odds ratios with confidence intervals; State clearly which outcome measure is used. Add p-values, if available. Define and describe observed harms per group as reported in the paper. Precise mean, percentages and p-values, if available |
| AUTHORS CONCLUSION | Report the Authors conclusion |
| COMMENTS | Additional characteristics/interpretations or flaws of the study, additional calculations made by the reviewer (NNT, RR, OR, CI,...) If relevant, mention the source of funding, or any competing interests. If the paper is to be excluded, mention the exclusion criterion here. |

When no element can be added, include:

- "Not applicable (NA)" when an item is not to be informed, or
- "Not described (ND)" when an item must be informed but no information is given in the publication)

Diagnostic studies

Evidence table

| REFERENCE | STUDY TYPE | PATIENTS NO. OF PATIENTS PATIENT CHARACTERISTICS + GROUP COMPARABILITY | DIAGNOSTIC TEST EVALUATED REFERENCE STANDARD TEST INCLUDE: TIME INTERVAL AND TREATMENT | OUTCOME MEASURES: PREVALENCE ACCURACY REPRODUCIBILITY | AUTHORS CONCLUSIONS | COMMENTS |
|-----------|------------|--|--|---|---------------------|----------|
| | | | | | | |

Details

| | |
|--|---|
| REFERENCE (+PMID) | Author, journal, year, volume and pages + PMID |
| STUDY TYPE | Meta-analysis/ systematic review (SR), randomized controlled trial (RCT), non-randomized cohort study (CS), case report (CASE), opinion documents (OPINION), other (specify)... |
| PATIENTS NO. PATIENTS CHARACTERISTICS + GROUP COMPARABILITY | <p>Total number of patients, and the number of patients involved in each group</p> <p>Describe relevant baseline characteristics (age, disease status, inclusion / exclusion criteria,...), if stated per group and add comment for comparability of groups.</p> <p>Setting: Multicentre, countries, healthcare setting..</p> |
| DIAGNOSTIC TEST EVALUATED REFERENCE STANDARD TEST INCLUDE: TIME INTERVAL AND TREATMENT | <p>Describe the evaluated test (what, by whom and how, when,...), cut offs, blinding to clinical information and/or index test results</p> <p>Describe the evaluated test (what, by whom and how, when,...), cut offs, blinding to clinical information and/or index test results</p> <p>Specify the time interval and treatments administered between the tests (if any)</p> |
| OUTCOME MEASURES : | Disease prevalence |
| → PREVALENCE | Accuracy: Give all available figures with confidence intervals (if available): |
| → ACCURACY | Sensitivity (Se), Specificity (Sp), positive predictive value (PPV), negative predictive value (NPV), likelihood ratios (LR, LR-), area under the ROC curve |
| → REPRODUCIBILITY | <p>Reproducibility: Give all available figures with confidence intervals (if available):</p> <p>→ Quantitative test (number of repetitions, extent of values, agreement, correlation)</p> <p>→ Qualitative test (reliability, correlation coefficient,...)</p> |
| AUTHORS CONCLUSION | Report the Authors conclusion |
| COMMENTS | <p>Additional characteristics/interpretations or flaws of the study, additional calculations made by the reviewer</p> <p>If relevant, mention the source of funding, or any competing interests.</p> <p>If the paper is to be excluded, mention the exclusion criterion here.</p> |

When no element can be added, include:

- "Not applicable (NA)" when an item is not to be informed, or
- "Not described (ND)" when an item must be informed but no information is given in the publication)

Background information

Instructions for the GDG members

1. You will receive all full text papers and a relevance and quality check table with all references.

For some full text papers we do not have access to the full text.

2. For some papers, you do not have to complete the evidence table:

In the last column remarks have been added.

- “GRADE evidence profile”: this meta-analysis or RCT will be added to a GRADE Profile. You can add your comments to the meta-analysis or RCT, but you do not need to complete all details.
- “EXCLUDE”: Usually studies that are relevant but have been included in a meta-analysis. You can leave this blank. If at some point the meta-analysis is found to be of low quality, we may go back and complete the information in the evidence table.
- Any other comment on the paper – you can decide whether or not to complete the evidence table

3. Read the paper thoroughly

- Complete the evidence table: fill in as much information as possible that is relevant for answering the PICO question. (see detailed instructions below for intervention and diagnostic studies).
- Even though you have already judged the paper as “good quality” and “relevant”, you can exclude papers at this stage. If you would like to exclude a paper, you do not have to complete the evidence table, just add a remark in the last column: “excluded due to...”
- In the last column of the evidence table, formulate any concerns, comments, or questions you have with regard to the content of the paper.

4. Send in the completed evidence table

- The evidence table and GRADE Profiles are the basis to write a summary of evidence and recommendations.
- You can continue with writing the summary of evidence and draft recommendations.
- We will check and if needed complete the evidence table. Please remember that each paper in the evidence table should be mentioned in the summary of evidence or excluded with an appropriate exclusion criterion.

I. Framework for recommendations

The justification should comprise the following considerations:

| | CRITERIA | JUDGEMENTS | RESEARCH EVIDENCE | ADDITIONAL INFORMATION |
|----------------|---|---|--------------------|------------------------|
| BENEFITS/HARMS | Are the desirable effects large relative to the undesirable effects? | <input type="radio"/> Benefits outweigh harms/burden <input type="radio"/> Benefits slightly outweigh harms/burden <input type="radio"/> Benefits and harms/burden are balanced <input type="radio"/> Harms/burden slightly outweigh benefits <input type="radio"/> Harms/burden outweigh benefits | <i>Per outcome</i> | |
| EVIDENCE | What is the overall quality of evidence? | <input type="radio"/> High <input type="radio"/> Moderate <input type="radio"/> Low <input type="radio"/> Very low <input type="radio"/> No included studies | <i>Per outcome</i> | |
| VALUES | What are the patient values and what certainty do we have about them? | <input type="radio"/> Little uncertainty and similar values <input type="radio"/> Some uncertainty or some variation <input type="radio"/> Significant uncertainty or large variation | | |
| RESOURCES | Is the incremental cost (or resource use) small relative to the benefits? | <input type="radio"/> Cost is very small relative to the benefits <input type="radio"/> Cost is small relative to the benefits <input type="radio"/> Cost is borderline relative to the benefits <input type="radio"/> Cost is high relative to the benefits <input type="radio"/> Cost is very high relative to the benefits | | |

| | | | | |
|---------------------------|---|--|--|--|
| HEALTH SYSTEM PERSPECTIVE | What would the impact be on health equity? | <input type="radio"/> Reduced <input type="radio"/> Probably reduced <input type="radio"/> Probably no impact <input type="radio"/> Probably increased <input type="radio"/> Increased <input type="radio"/> Varies <input type="radio"/> Don't know | | |
| | Is the intervention acceptable to key stakeholders? | <input type="radio"/> No <input type="radio"/> Probably No <input type="radio"/> Probably yes <input type="radio"/> Yes <input type="radio"/> Varies <input type="radio"/> Don't know | | |
| | Is the intervention feasible to implement? | <input type="radio"/> No <input type="radio"/> Probably No <input type="radio"/> Probably yes <input type="radio"/> Yes <input type="radio"/> Varies <input type="radio"/> Don't know | | |
| OTHER | Subgroup considerations ¹⁵ | | | |
| | Implementation considerations ¹⁶ | | | |
| | Research priorities ¹⁷ | | | |

¹⁵ Are the recommendations applicable to subgroups, and is this mentioned in the recommendation?

¹⁶ Are there any barriers that could restrict the implementation of this recommendation? If so, how can we handle this?

¹⁷ Is there a need for further research on the topic, and is it a priority? If yes, please provide details on how new studies should ideally be performed (RCT, large multicenter) and what their focus should be (subgroup of patients, specific intervention). All recommendations for research will be added as an annex to the guideline.

J. Reviewer comments form

GUIDELINE:

.....

REVIEW PERIOD:

.....

Contact information of the reviewer

NAME

.....

COUNTRY:

.....

E-MAIL ADDRESS:

.....

I AM PARTICIPATING

☐ AS AN INDIVIDUAL

☐ ON BEHALF OF A (INTER)NATIONAL ORGANIZATION,
NAMESLY

.....

☐ ON BEHALF OF A COMPANY, NAMESLY

.....

Statement of confidentiality

As a reviewer of this ESHRE document you have been or may be exposed to certain confidential and/or proprietary information, materials or data. It is important to the integrity of the writing process and final work that this information should be kept strictly confidential and not disclosed at any time under any circumstance.

- I will not disclose any confidential and/or proprietary information, materials or data related to Working Group's work to any third party, but keep this information strictly confidential.
- I will keep any confidential and/or proprietary information, materials or data in my possession in a safe and secure place to protect against inadvertent disclosure.
- I will not use any confidential information and/or proprietary information, materials or data for any purpose other than participating in the review procedure.

SIGNATURE (OR STATE YOUR NAME):

.....

DATE:

.....

Comments to the document

| PAGE | LINE | COMMENT |
|------|------|---------|
| | | |
| | | |
| | | |
| | | |
| | | |
| | | |
| | | |
| | | |
| | | |

(Add more lines if you need to)

Please send completed forms (as word-document or pdf) to guidelines@eshre.eu before XX.

All comments will be revised by the working group and assessed. If the comment is accepted by the working group, it will result in a modification of the document. If not, the working group will formulate a reply to the reviewer. The details of the review procedure, the comments, modifications and replies will be summarized in a review report which will be available online.

By submitting this form, you will be listed as an expert reviewer of the guideline. The list of reviewers will be published in the review report and in an annex of the main guideline document.

For more information on the review, you can [contact guidelines@eshre.eu](mailto:guidelines@eshre.eu).

K. Document assessment tool for updating guidelines

Adapted from [28]

| | |
|---|--|
| NUMBER AND TITLE OF THE GUIDELINE | |
| CURRENT REPORT DATE | |
| LAST LITERATURE SEARCH DATE | |
| DATE ASSESSED | |
| RESEARCH SPECIALIST | |
| OUTCOME (FOR INTERNAL USE) | |
| ASSESSMENT: For each document, please respond YES or NO to all the questions below. Provide an explanation of each answer as necessary. | |
| 1. IS THE DOCUMENT STILL RELEVANT (CLINICALLY OR TO THE CARE SYSTEM AS A WHOLE IN SOME WAY)? | |
| 2. SHOULD FULL ASSESSMENT AND REVIEW OF THIS DOCUMENT BE DEFERRED UNTIL NEXT YEAR? Consider YES if: <ul style="list-style-type: none"> → The document is less than three years old, and there is no reason to doubt the recommendations → The document is between three and five years old, and a justification can be provided as to why the recommendations can be considered trustworthy for another year | |
| 3. DO THE QUESTIONS AND SEARCH CRITERIA AS THEY ARE IN THE DOCUMENT ADDRESS CURRENT NEEDS, SUCH THAT AN UPDATED LITERATURE SEARCH WOULD BE USEFUL AND IDENTIFY RELEVANT EVIDENCE? Consider NO if: <ul style="list-style-type: none"> → The standard of care has shifted significantly since the last version of the document, such that the questions only address the topic in part → There are new, significant options (for treatment, diagnosis, etc.) available that are not covered by the current questions, such that new questions would need to be added to the document → In general, if you believe that for the document to still be useful it will have to substantially be rewritten → The document has been repeatedly deferred, and is now older than five years | |

| | |
|---|--|
| 4. DOES THE DOCUMENT HAVE AN IMPACT ON ACCESS TO CARE (THAT IS, ARE DECISIONS ABOUT ACCESS OR PAYMENT FOR CARE MADE BY THE MINISTRY, CCO, OR OTHER ORGANIZATIONS BASED ON THE RECOMMENDATIONS IN THIS DOCUMENT)? <i>Consider YES if:</i> <ul style="list-style-type: none"> → Ministry funding decisions have been, are, or will be made on the basis of this document → An indication for a chemotherapy regimen was funded, or rejected, based on the document → Case by case review or out of country requests are known to be decided based on the document → Funding for some screening, diagnostic, staging or treatment procedure was or is determined | |
| 5. IS THERE KNOWN EVIDENCE THAT HAS BEEN PUBLISHED SINCE THIS DOCUMENT'S LAST LITERATURE SEARCH (SEE ABOVE) THAT WOULD RESULT IN SIGNIFICANT CHANGES TO THE RECOMMENDATIONS? | |
| 6. SHOULD THIS DOCUMENT BE TAKEN OFF THE WEBSITE WHILE IT AWAITS FULL REVIEW, OR CAN IT BE LEFT THERE WITH AN "IN REVIEW" WATERMARK? <i>Consider YES if:</i> <ul style="list-style-type: none"> → If followed, even in error, the recommendations have the potential to cause harm to patients. | |

PLEASE LIST ANY ADDITIONAL FACTORS THAT SHOULD BE CONSIDERED IN PRIORITIZING THIS DOCUMENT FOR REVIEW:

.....

| | | |
|-----------------|-------------------------------|---|
| OUTCOME: | <input type="radio"/> ENDORSE | The recommendations are still current and relevant for decision making. This can happen when there is a very strong justification to conclude that without a search for new evidence, the recommendations are still valid ¹⁸ . |
| | <input type="radio"/> DEFER | The document remains current and credible enough to wait until the next assessment. |
| | <input type="radio"/> REVIEW | The document will undergo a review for currency and relevance. |
| | <input type="radio"/> ARCHIVE | The document cannot be endorsed or deferred, and the recommendations will no longer be maintained. ¹⁹ |

¹⁸ For example, in cases in which added evidence will not change the recommendations because the existing evidence is so definitive, of high quality, and adequate quantity or in cases in which no additional evidence in that topic will be forthcoming because it is no longer an area of inquiry.

¹⁹ This may happen because the recommendations are no longer clinically relevant and applicable to current practice. Or, it may be because the developing group has little or no interest in maintaining them; for example, the topic areas may have changed so much that developing a new document is a more practical option than updating the existing one. Archived documents may still be useful for academic or other information purposes.

