Manual for creating a living guideline (Version 3.1, March 2023)

Editors:
Prof. Wolfgang Gaebel[a,b]
Prof. Peter Falkai[c]
Prof. Alkomiet Hasan[d]
Prof. Stefan Leucht[e]

Co-authors:
Marco Schneider[c]
Dr Eva Salveridou-Hof[a,b]
Charline Pielenz[a,b]
Dr Elias Wagner[c]
Marisa Flick[a,b]
Gabriele Gaigl[d]
Naiiri Khorikian-Ghazari[d]
Duygu Güler[d]
Theresa Halms[d]
Angelika Burschinski[e]
Carolin Lorenz[e]
Dr Astrid Röh[d]

[a]Department of Psychiatry and Psychotherapy, LVR Hospital Düsseldorf, Faculty of Medicine, Heinrich Heine University, Düsseldorf, Germany
[b]WHO Collaborating Centre on Quality Assurance and Empowerment in Mental Health DEU-131
[c]Department of Psychiatry and Psychotherapy, Hospital of the Ludwig Maximilian University (LMU) Munich, Germany
[d]Department of Psychiatry, Psychotherapy and Psychosomatics of the University of Augsburg, Faculty of Medicine, District Hospital Augsburg, Germany
[e]Department of Psychiatry and Psychotherapy, Technical University of Munich, TUM School of Medicine, Klinikum rechts der Isar, Munich, Germany

We thank the following organizations, professional associations, and guideline developers for their contributions in the form of a review process of the manual:

- Prof. Anil Batra, co-author of publications including the *S3 Guideline on Screening, Diagnostics and Treatment of Harmful and Addictive Tobacco Use*
- Prof. Achim Berthele, co-author of publications including the *S2k living guideline on diagnosis and treatment of multiple sclerosis, neuromyelitis optica spectrum disorder and MOG-IgG-associated disorders*
- Jurand Daszkowski, State Association of (ex-) Users of Psychiatry, Hamburg
- Prof. Thomas Fischer, co-author of publications including *the S2k living guideline on in-home care, social participation and quality of life of patients requiring care in the context of outpatient nursing care under the conditions of the COVID-19 pandemic*
- Dr Gabriel Gerlinger, German Association for Psychiatry, Psychotherapy and Psychosomatics (DGPPN)
- Prof. Euphrosyne Gouzoulis-Mayfrank, co-author of publications including the *S3 guideline on drug-induced psychosis* (in preparation)
- Prof. Julia Haberstroh, co-author of publications including the *S2k guideline on Capacity to consent of people with dementia in medical interventions. Interdisciplinary S2k guideline for medical practice.*
- Wulf-Peter Hansen, Klinikum rechts der Isar, BASTA – Alliance against discrimination, Munich, Germany
- Prof. Ursula Havemann-Reinecke, co-author of publications including the *S3 guideline on drug-related disorders*.
- Prof. Frank Jessen, German Association for Psychiatry, Psychotherapy and Psychosomatics (DGPPN)
- Prof. Ina Kopp, Association of the Scientific Medical Societies of Germany (*AWMF*)
- Prof. Wolfgang Maier, co-author of publications including the *S3 guideline on dementia*
- Prof. Thomas Messer, co-author of publications including the *S2K guideline on emergency psychiatry*
- Dr Cathleen Muche-Borowski, Association of the Scientific Medical Societies of Germany (*AWMF*), co-author of publications including the *S2e living guideline on protection against overuse and underuse of health care - making joint decisions*
- Prof. Steffi Riedel-Heller, co-author of publications including the *S3 guideline on psychosocial therapies in severe mental illness*
- Prof. Tilman Steinert, co-author of publications including the *S3 guideline on prevention of coercion: prevention and treatment of aggressive behaviour in adults*
- Prof. Kai Vogeley, co-author of publications including the *S3 guideline on childhood, youth and adult autism spectrum disorders*
- Prof. Uwe Wagner, co-author of publications including the *S3 living guideline on diagnostics, therapy and follow-up of malignant ovarian tumours*
- Prof. Joachim Weis, co-author of publications including the *S3 guideline on psycho-oncological diagnostics, counselling and treatment of adult cancer patients*
- Prof. Widder, co-author of publications including the *S2k guideline on the assessment of patients with psychological and psychosomatic disorders*
- Mr Kelvin Hill, co-author of publications including the *(Australian) Clinical Guidelines for Stroke Management* and *The Living Guidelines Handbooks* published by the Australian Living Evidence Consortium
- Prof. Elisabeth Kuipers, co-author of publications including the *NICE Guideline Psychosis and Schizophrenia in adults*
- Dr Ariel Izcovich, co-author of publications including a living systematic review on the treatment of covid-19 and member of the GRADE working group, Hospital Alemán de Buenos Aires, Internal medicine department

Guidance

Introduction

Developing conventional guidelines can take years. Given the exponential growth of scientific knowledge, recommendations may therefore already be outdated by the time they are published (Martinez García et al., 2014). Living guidelines aim to adapt the guideline development process, enabling individual obsolete recommendations to be updated as soon as relevant new findings become available (Akl et al., 2017). Central to a living guideline is the focus on individual recommendations and their timeliness, without necessarily revising the entire guideline (ALEC, 2022). The process of drawing up living guidelines should thus produce validated recommendations and, at the same time, streamline guideline development, as the short cycle allows for specific aspects to be updated (Akl et al., 2017). A recommendation is considered *living* once an optimised, timely update process has been established for it (for definitions of living recommendation and living guideline, see El Mikati et al., 2022). Consequently, decision-makers should have timely access to new recommendations (Akl et al., 2017). Living guidelines enable the rapid production of guidelines without affecting the strict gold standard methods of guideline development. The development process thus does not replace the prescribed standards of the AWMF regulations for the preparation of guidelines, but supplements them with living guideline-specific requirements. However, a dynamic process of this nature obviously entails its own challenges. For instance, the guideline development group must work closely and continuously with one another and the process requires a long-term funding strategy (Pieper et al., 2019). Furthermore, living guidelines can only provide timely treatment recommendations if they are published and disseminated in a timely manner (Akl et al., 2017). This manual offers guidance for overcoming these challenges and capitalizing on the overwhelming benefits provided by living guidelines.

This manual for creating living guidelines is based on the structure of the *AWMF Guidance Manual and Rules for Guideline Development* (2020, German and English version available) and thus also on the quality criteria for guidelines defined in AGREE II. Like the entire living process, the structure of the manual proposed here is also dynamic, so that it may be subject to changes itself.

This manual is geared towards the standards of S3 guidelines and is suitable for converting an already existing set of guidelines into a living format and for annually updating living guidelines. The manual is aimed at guideline developers. For creating conventional de novo guidelines, we recommend consulting the *AWMF Guidance Manual* (2020). For creating de novo living guidelines, the Living Guidelines Handbook published by the Australian Living Evidence Consortium may offer additional guidance (ALEC, 2022).

In four main sections, the manual provides recommendations to be considered when updating a guideline or when converting it into a living format. The chapters follow the structure of the *living cycle of guideline development* (figure 1) and the *flowchart* of the work process (figure 2). The four main sections are:

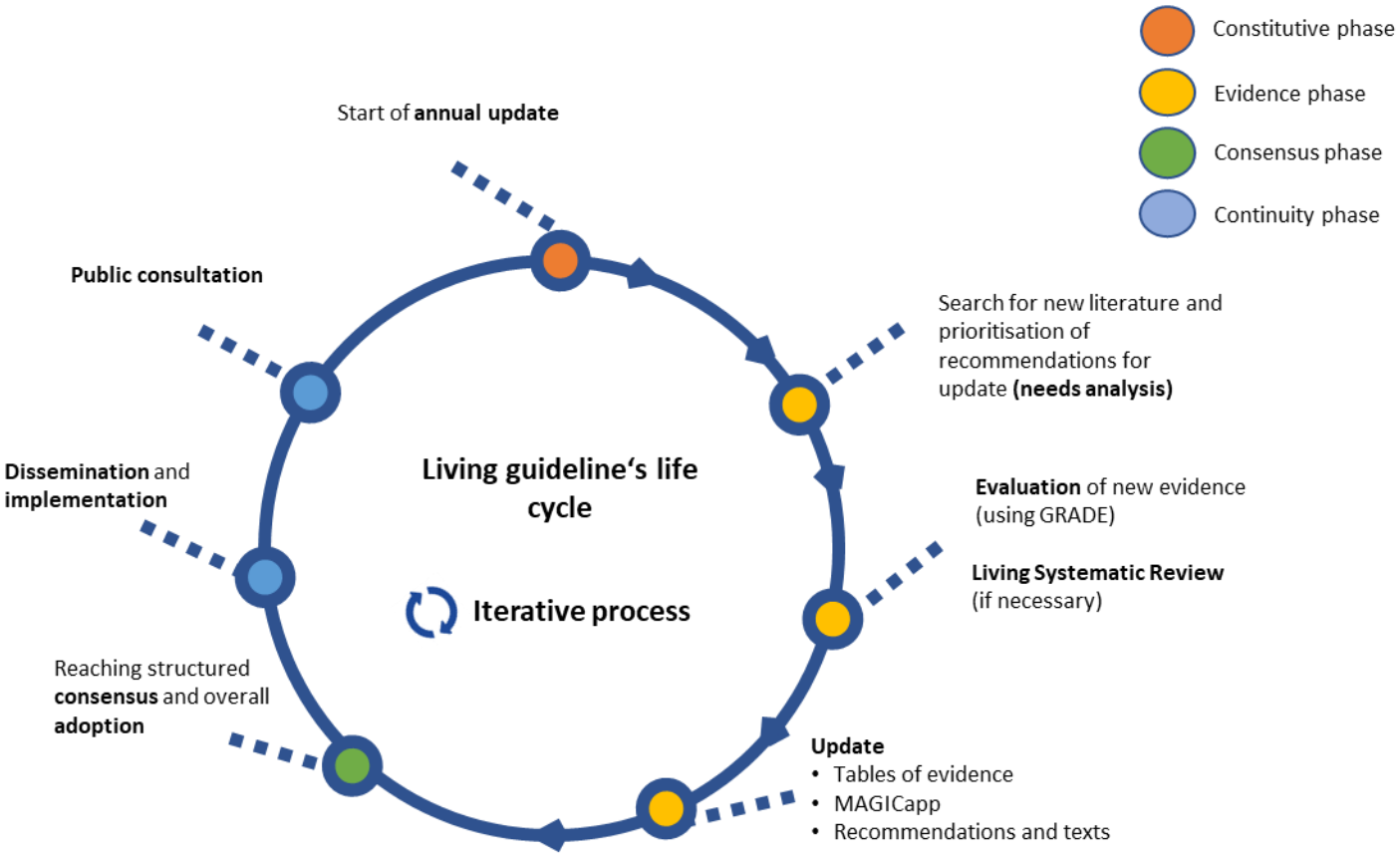
(1) Constitutive/planning phase (administrative processes, planning and constituent meeting).
(2) Evidence phase (research and methodological work)
(3) Consensus phase (consensus on revised recommendations)
(4) Continuity phase (editorial implementation of changes and continuous work processes to maintain the living guideline).

In some sections, alternative methods are mentioned in addition to the recommendation suggested first, with the first option constituting the version favoured by the editors of this manual.

The ideal process for recurring updates is illustrated in the living cycle schematic (Fig. 1) and will be described in more detail in later sections of the manual.

Figure 1

Living cycle of guideline development



Note. Schematic of the living guideline's life cycle. The figure depicts steps of the respective development phase in an abbreviated form. The update starts with a constitutive meeting of the guideline group, followed by an evidence phase of several months to assess needs and prepare proposals for adaptation. The annual update is concluded with the consensus finding and adoption. For a more detailed description of the work steps, see Figure 2.

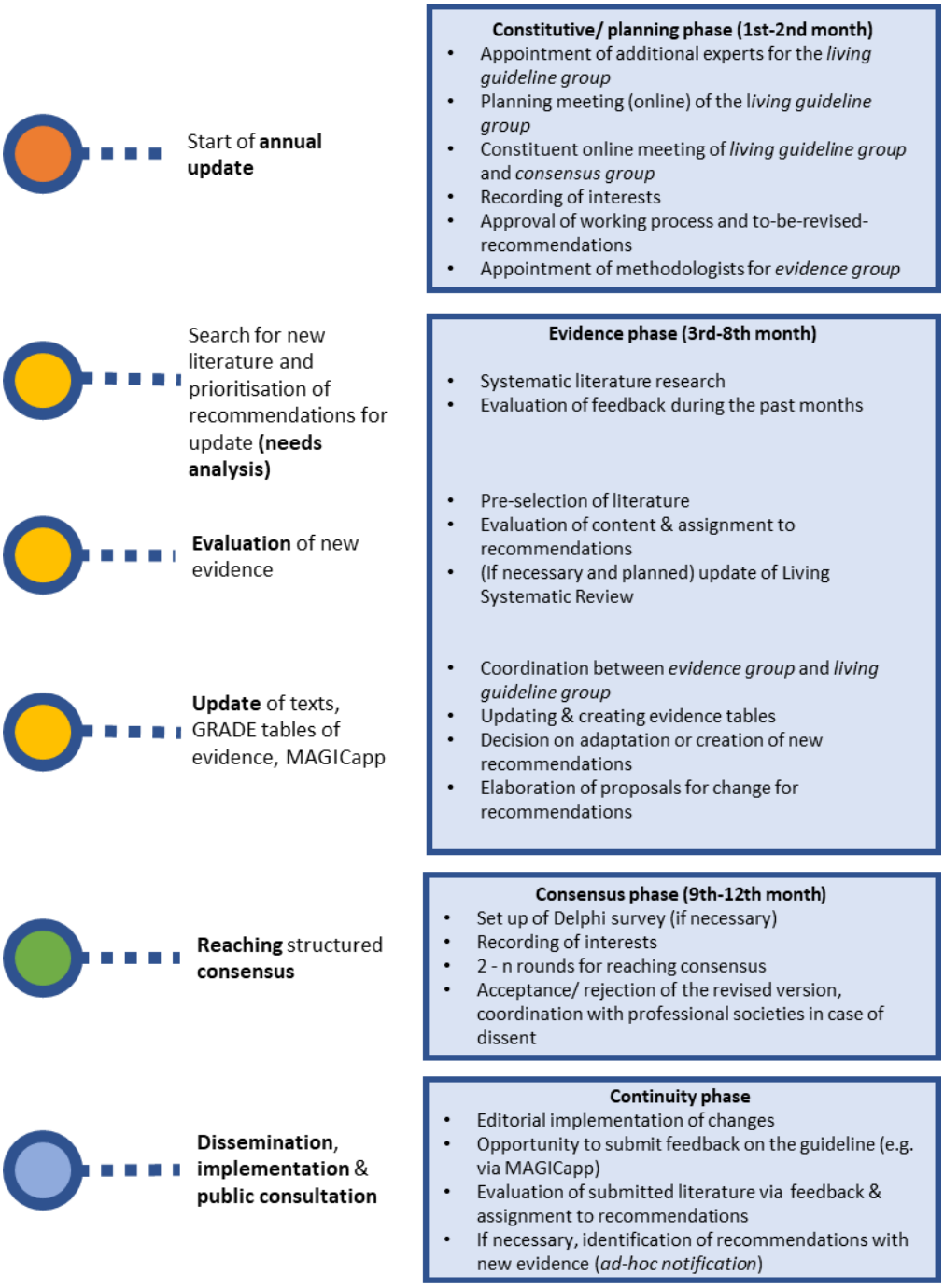
GRADE = Grading of Recommendations, Assessment, Development and Evaluation.

Adapted from: Pielenz/Schneider et al., 2022.

The updating cycle can be subdivided into four phases (see Fig. 2). The flowchart serves as a rough roadmap for updating a living guideline and is described in more detail in the sections below.

Figure *2*

Flowchart – Living guideline's modular revision



Note. Individual work steps in the update process differentiated according to the respective development phase. Start of the update with the constituent meeting of the stakeholders involved (orange – constitutive/ planning phase). This is followed by the evidence phase (yellow) for recording the need for updating and drawing up proposals for changes to recommendations. These are agreed with the stakeholders involved in the consensus phase (green) and a revised version of the guideline is then produced and published (continuity phase) (blue). Dissemination via the selected platforms (e.g. *MAGICapp*) takes place on an ongoing basis (blue), as does the opportunity to provide feedback on the guideline and its recommendations (e.g. via a comment function).

1 Constitutive/ planning phase

Before actually starting the annual update a sound funding strategy should be established, the members of the guideline development group should be appointed, the tasks and roles should be clearly divided and all relevant interests should be recorded.

As a part of annual needs and status analysis, it should ideally be ascertained which changes have resulted from the availability of the guideline in care and which care problems still exist. Due to the short update interval, this can hardly be realised in annual intervals. Criteria must be defined according to which existing, previously adopted recommendations are to be revised and thus re-adjusted. An essential criterion is the availability of new robust scientific data. It is recommended that the in-year needs analysis be based on new, available evidence and should be collected through a systematic literature review (see 2.2 Systematic literature review). Accordingly, a comprehensive inventory analysis should not take place annually, but, following the update of a conventional guideline, for example every 5 years.

Within the first constitutive meeting of the *living guideline group* (see 1.2 for the constitution of the guideline development group) to-be-revised recommendations, clinical questions and outcomes have to be determined. Followed by a constitutive meeting of the entire guideline group when the working process and the recommendations for the guideline updates have to be approved. One possibility of defining the update standardized is by using the *Up-Priority-Tool* (Sanabria et al., 2020). The tool assists for the decision which clinical questions have to be updated with priority (see Appendix for additional information).

1.1 Funding strategy

It is imperative to develop a financing structure in advance that will ensure funding for the living guideline. The professional society publishing the guideline or funding from a governmental body (e.g. for Germany the German Federal Joint Committee (G-BA)) should ensure that the essential funding for the update is in place on an ongoing basis. Guidelines funded by third parties that have a direct influence on content are not permitted due to the ensuing conflicts of interest. A publication would not be accepted by the AWMF in such a case (AWMF Guidance Manual, 2020).

List of expenses/required resources (based on the *AWMF Guidance Manual*, 2020)

- Databases (if pay-for-use)
- Costs for staff
  - Guideline office
  - Commissioning external methodologists for the *evidence group*
- Costs for materials and equipment (office, materials, etc.)
- Publication, layout
- The preparation of a living systematic review (LSR), if needed, whereby ongoing updates may necessitate more manpower
- Digital tools:
  - Use of literature management programmes
  - Increased use of digital conference platforms (reduces travel expenses) by conducting consensus processes online

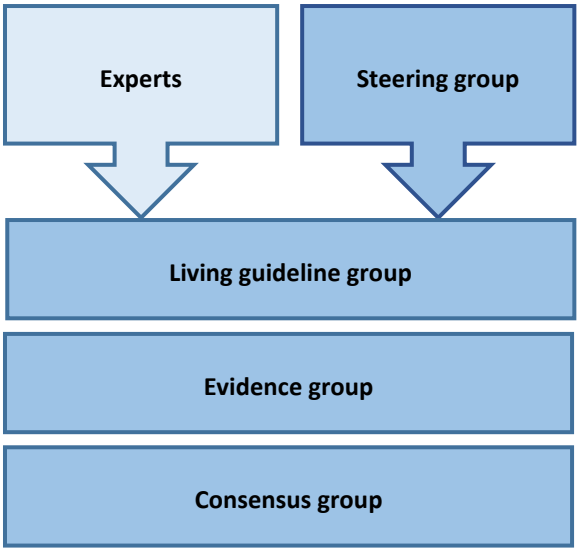
1.2 Constitution of the guideline development group

The composition of the guideline development group should be evenly balanced. This is indispensable with a view to professional legitimation and acceptance as well as to replicating the state of the art as impartially as possible (AWMF Guidance Manual, 2020). One way to ensure this is to involve as many organizations and professional societies dedicated to the subject area as possible (AWMF Guidance Manual, 2020) and to include all the professional groups involved in treatment. In addition, it is essential to include/consider the perspective of patients and their families addressed by the guideline (see special vote under 3.1, Voting process and structured consensus). If a professional society publishing living guidelines coordinates more than one such publication, parallel working structures will emerge that could be combined (for example, one *evidence group* for all living guidelines).

The guideline development group should consist of four sub-groups (see Fig. 3). While the size of the guideline development groups should be kept consistent, specialists and professional societies and organizations that are not yet involved should have the opportunity to request participation. After three updating cycles, for example, the composition of the guideline development group should be reviewed in order to involve new specialists in the update or offer current members an opportunity to leave the group.

Figure 3

Constitution of the guideline development group



Note. The four guideline subgroups (dark blue) in the update process consist of the *steering group* to lead the update process with expertise in guideline development, the *living guideline group*, which, in addition to the members of the *steering group*, gathers other experts to assess the content of the evidence and develop recommendations, the *evidence group* with methodologists to research and process the evidence, and the *consensus group* with stakeholders to agree on the revised version of the guideline. The roles and responsibilities are described below.

The *steering group*

The *steering group* coordinates the entire guideline development process. Members are not entitled to vote in the consensus phase. In order to simplify and accelerate informal coordination processes in the evidence phase, it is recommended that a member of the *steering group* is familiar with the work of the *evidence group* (especially with the GRADE process). This helps to link the cooperation between these groups.

Roles and responsibilities of the *steering group*

- Steering the entire guideline development process
- Establishing a funding strategy for the ongoing update
- Assigning responsibility for administrative, planning and editorial responsibilities (including organization of the consensus, the global adoption and the publication process)
- Inviting experts to join the *living guideline group*
- Commissioning the *evidence group* and appointing the methodologists
- The *steering group* is part of the *living guideline group* and coordinates the editing of the evidence tables presented on the recommendations to be revised (see roles and responsibilities of the *living guideline group*)

The *living guideline group*

The second sub-group, the *living guideline group*, is composed of members of the *steering group* and experts as needed. The annual update decreases the extent of updates but also reduces the period of time available for revision, meaning that the size of the *living guideline group* should be kept small in order to expedite consultations and agreement. At the same time, despite the smaller size, the group must ensure that it has the requisite expertise and work capacity. To this end, the *steering group* determines how many experts are required to work on/manage the content of the update cycle and invites the additional experts to join the group at the beginning of the constitutive phase accordingly. In addition, experts can register and be nominated for participation in the *living guideline group*. For this reason, these experts should have the specialist expertise in the (sub-)speciality of the guideline. Conflicts of interest in the process should be avoided. The *living guideline group*, so also the additional experts, is not eligible to vote in the consensus phase.

Roles and responsibilities of the *living guideline group*

- Constitutive/planning phase
  1st meeting for planning
  - Coordination of relevant clinical questions (PICOs) and outcomes (note: new PICOs should only be introduced if new evidence is available)
  - Coordination of to-be-revised recommendations of the guideline
- Evidence phase
  - Screening of the literature prepared by the *evidence group*, appraisal of the content of the new body of evidence
  - Informal coordination between *evidence group* and *living guideline group* to clarify methodological questions concerning the GRADE-process, *living guideline group* will carry out the *judgements* of evidence within the *GRADE-process*
  - Preparing the proposals for amending recommendations to be revised
- Consensus phase
    - Presentation of results in front of the *consensus group* (including a brief introduction about the *GRADE-process* and explanation how and why results have been down- or upgraded and which criteria have been used for the decision)
- Continuity phase – public consultation

- Content assessment of the comments from the (specialist) public prefiltered by the *evidence group*

The *evidence group*

The *evidence group* performs the methodological work in the evidence phase. The methodologists in this group are commissioned by the *steering group* or by the issuing professional society. They perform the systematic literature search for new studies and review articles (see 2.2, Systematic literature search) and assess and compile the evidence (see 2.2.2, Standards for rating the quality of evidence and following chapters). During the evidence phase close coordination with the *living guideline group* is necessary to clarify methodological questions and to delegate important steps in the evaluation process to experts when more expertise is required. The methodologists develop proposals regarding the *GRADE*-evaluations and strength of recommendations. They also filter the literature reported during the course of the year via the feedback pathways (see 4.3, Public consultation). An assessment of the content of the literature is in turn performed by the living guideline group (see roles and responsibilities of the *living guideline group*). Ad-hoc feedback on new evidence leads to a note on the respective guideline recommendation after consultation with the *steering group*. If the decision is made in favour of the *online Delphi technique* in the consensus phase (see 3.1, Voting process and structured consensus), the *evidence group* will take over the evaluation of the comments in the *Delphi rounds*.

In the event that a professional society publishing living guidelines coordinates more than one such publication, it may be possible to combine some steps, particularly in the evidence phase. An *evidence group* working on several guidelines can therefore coordinate search cycles and steps in the update process.

The *consensus group*

As the *consensus group* is formed from representatives of the professional societies and organizations concerned with the discipline in question, it is assumed that the appointed members have the requisite expertise. Any professional group involved in treatment and patient representatives should also be included in the group. It is advisable to give the professional society publishing the guideline responsibility for selecting and inviting the additional professional societies and organizations to be involved. The members of the *consensus group* are the only individuals eligible to vote in the consensus process. The respective professional societies and organizations are thus requested in a letter (see AWMF Guidance Manual, 2020, nomination of appointed representatives, https://www.awmf.org/regelwerk/zusammensetzung-der-leitliniengruppe-beteiligung-von-interessengruppen) to nominate one representative and one deputy. Each professional society or organization has one vote in the voting procedures. Double nominations are not possible, each person has only one vote. Transfer of votes is also not permitted. The representatives are the members of the *consensus group*.

Roles and responsibilities of the *consensus group*

- Constitutive/planning phase
  Constitutive meeting

- Adopting the update process (in accordance with the manual) and to-be-revised recommendations – if new evidence is available and sufficient new PICOs can be introduced
- Consensus phase
  - Voting on revised or new recommendations and global adoption of the updated living guideline

The guideline groups (*steering* and *consensus group*) can be reactivated in the subsequent update cycles. The experts additionally invited to join the *living guideline group* may be replaced by other experts every year.

1.3 Conflicts of interest

Whenever a living cycle is launched or restarted, it is advisable to use a standardized procedure to record the financial and non-financial interests of all guideline group members (members of the *steering group*, additionally invited experts in the *living guideline group* and members of the *consensus group*) and assess them with regard to existing conflicts of interest (before the constitutive meeting in the constitutive/planning phase). Maximum transparency is imperative. In addition, any changes to the members' interests should be queried prior to obtaining consent (in the consensus phase).

The *AWMF web portal* "*Declaration of Interest Online*" can be used to manage declarations of interest and their evaluation (see https://interessenerklaerung-online.awmf.org/, German and English version available), as the reuse of registered declarations of interest saves time with each subsequent registration. A declaration of interest can be submitted via the portal as well as evaluated by assigned persons. The *living guideline group* decide before the constituent meeting whether to use the portal and/or the classic form as a simple document. When the multistage *Delphi technique* is used (see 3.1, Voting process and structured consensus), this information can be elicited through a question in the questionnaire in use. If any new interests have emerged, the individual can submit a new declaration on the conflict of interests by using a link to the form in use or via the *AWMF web portal*.

The external assessment may be performed either by external individuals or by responsible parties ('Conflict of interest officer') selected from the guideline group; not, however, from the *steering group/living guideline group*. If the assessment is to be conducted by a small group of guideline group members, assessment criteria for low/moderate/high thematically related conflicts of interest should be jointly laid down (AWMF Guidance Manual, 2020); for example, voting rights are granted only if there are no conflicts of interest or the conflict of interest has been assessed as low. Certified guideline consultants can offer support in this regard.
The *AWMF Guidance Manual* mentions several principles which may provide help with dealing with existing conflicts of interest (see 02_20180117_AWMF-Regel_Interessenkonflikte_V2.4.pdf). The AWMF portal offers detailed instructions on the roles and rights (see Instructions — AWMF-Portal Declaration of Interests).

2 Evidence phase

At the beginning, the validity of all recommendations should be checked (AWMF Guidance Manual, 2020), whereby only those recommendations in need of change are subject to closer examination. If new clinical questions arise from the provision of care, these should be formulated by the *living guideline group* (for the constitution of the guideline development group, see point 1.2) and included in the literature search.

For the annual update of individual recommendations of a living guideline, it should be decided on the basis of the literature search or the type of publication which recommendations need to be revised. Established hierarchical levels of evidence structure the need for updating (see 2.2. Systematic literature search). Assessment tools or checklists can be used for this purpose (see 2.2.1 Critical appraisal of evidence). Another possibility to assess the extend for an update the standardised *Up-Priority-Tool* (Sanabria et al., 2020) can be used. With this tool clinical questions (PICOs) can be evaluated along six priority items (e.g. impacts of outdated recommendations on safety). This may help in the decision making and keep the decision transparent (for further information see Appendix). Another option is for the *living guidelines group* to determine which recommendations to revise/update based on the available literature. Besides criteria for the hierarchy of included studies (see 2.2 Systematic literature search) also further selection criteria can be definded, e.g. type of study, quality of evidence, effect size or sample size (ALEC, 2022). If, according to the majority opinion of the guideline development group, there are one or more recommendations in need of revision, this can also lead to an update after consultation with the *consensus group*. After eligible publications are identified and assessed regarding quality and potentials for content change of the guideline the current evidence will be presented in an aggregated form (evidence tables). This process will be described in the following sections (2.1 Screening and update frequency and 2.2 Systematic literature search).

2.1 Screening and update frequency of a living guideline

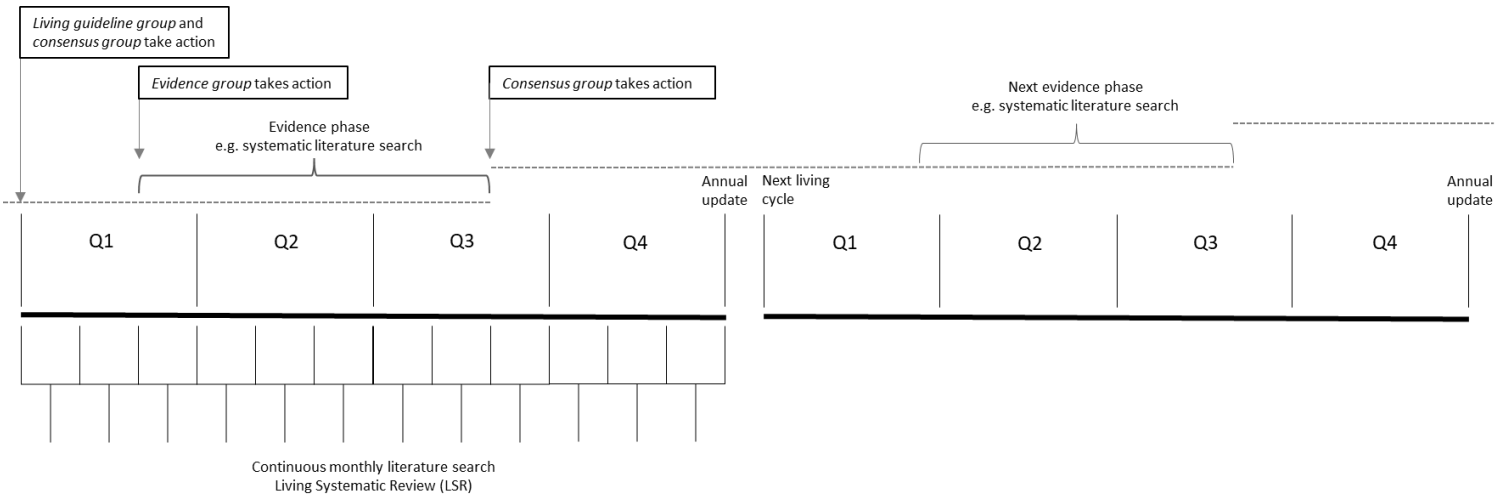
It is advisable to screen for new literature at 12-month intervals (see Fig. 4). The Australian Living Evidence Consortium recommends in its *Living Guidelines Handbook* (2022) a shorter screening frequency (latest every three months). Depending on the medical area differences in the expected dynamics of new evidence may arise. Somatic or highly researched topics require a more frequent screening. The screening is undertaken in the first step of the literature search. It is recommended publishing an annual update of the living guideline by revising individual predefined recommendations (AWMF Guidance Manual, 2020; Pielenz/Schneider et al., 2022).
In the subsequent update cycles, the systematic literature search should continue where the previous literature search left off in order to account for newly published literature (see Fig. 4). Publications that are not released until after the evidence phase of an annual revision cycle and thus cannot be considered in the update but have sufficient potential for prompting adaptation of the associated recommendation, should be marked in the guideline (for example, using the *MAGICapp* (see 4.1.2, *MAGICapp* presentation format)) accordingly (e.g. ad hoc report on the recommendation - 'New body of evidence').

Within an update cycle, several recommendations can be updated in the living format and the recommendations requiring a longer revision period may be marked accordingly and revised in the subsequent cycle (see 4.1.1, Clarity and presentation). If the recommendations to be revised exceed the capacity of the guideline development group, it is up to the *living guideline*

group to decide which recommendations to prioritise and revise in the current cycle. A systematic review of the need for revision yielding the result that no changes are needed is considered to be updated (AWMF Guidance Manual, 2020).

Figure 4

Screening frequency of living guideline's update cycles.



Note. Systematic literature search in annual time interval, annual publication of a living guideline. In subsequent update cycles, the systematic literature review should follow the previous one to include all new publications (dashed line). *The living guideline group and consensus groups* adopt the update work process (by manual) in the constitution phase (Q1), the *evidence group* becomes active in the evidence phase (Q1-Q3), and the *consensus group* consents to the revised recommendations (end of Q3) before editorial work begins (Q4). The guideline development sub-groups are not active on a continuous basis, but on an as-needed basis. Instead of a low-frequency systematic literature search, it would also be possible to prepare a living systematic review with a high-frequency literature search (lower range of the timeline).

2.2 Systematic literature search

It is advisable to conduct the literature search as a rule by the third month of the annual update cycle and to include all new publications that have appeared since the previous evidence phase (see Fig. 4, dashed line). The search period defined in the timeline may serve as guidance to ensure completion of all the subsequent steps of the update within the 12-month deadline. It is incumbent on the guideline group to define shorter search intervals. It is up to the guideline development group to set shorter search intervals (for example, every three months, see ALEC, 2022).

For particularly relevant questions in the guideline with greater implications for practical application (such as changes to treatment regimens), preparing a *living systematic review* (Living Evidence Network, 2019) as a body of evidence may be more worthwhile than conducting a low-frequency annual literature search. Delays caused by the publication process can thus be avoided and statements from the *living systematic review* can be used for the update in a timely fashion (AWMF Guidance Manual, 2020; Pielenz/Schneider et al., 2022.) This is a systematic review that is continuously updated and includes new relevant findings as soon as they become available (Elliott et al., 2017). The parallel preparation of the review requires greater effort and is mentioned here as an alternative high-frequency literature search that should be reserved for exceptional situations. For further information, please consult the appendix or the Cochrane guidance (2019).

Methodologists from the *evidence group* are responsible for performing the systematic literature search (see 1.2, Constitution of the guideline development group). When planning a systematic literature search, it is helpful to narrow down the clinically relevant questions in accordance with the **P**atient **I**ntervention **C**omparison **O**utcome (PICO) scheme (AWMF Guidance Manual, 2020). These should be worked on by the *living guideline group* before the start of the research full stops and adopted by the *consensus group* at the constitutive meeting (see 1 Constitutive/planning phase).Changes in the search strategy, the search terms and sources used, the period of the literature search and the number of hits should be described in detail and listed in the guideline report (AWMF Guidance Manual, 2020). The inclusion and exclusion criteria (target population, study design, comparisons, endpoints, language, context) should be presented. The results should be documented in the form of a PRISMA statement for the guideline report of the living guideline (Cochrane Deutschland Stiftung et al., 2020; Page et al., 2021).

Standardized search terms (general or specific keywords) can be used for the search. In the case of an already existing guideline, the keywords used originally can be applied and/or supplemented with new keywords. As the first step, the publications that are found are preselected at the title, abstract and full text level. The assessment of studies with aggregated evidence (systematic reviews, meta-analyses, HTA reports, IQWIG benefit assessment) often delivers more up-to-date results than existing guidelines, which is why they should be ranked highest in the hierarchy (AWMF Guidance Manual, 2020). The timeliness of data should be considered when using aggregated studies. Followed by related (living) guidelines and concluded with the assessment of primary studies (AWMF Guidance Manual, 2020). The decision as to whether studies should be considered at the next lower level or whether they must be processed in the expert consensus is taken by the *evidence group* based on the search results. It is recommended to consider studies at the next lower level if the research question cannot be answered adequately, if a high-quality study (e.g. primary study/RCT) is not yet included in the aggregated evidence, or if no aggregated evidence is (yet) available for the research question in general. In any case, the risk of bias assessment must be taken into account. The search in the study registries for ongoing or unpublished studies may prove to be a worthwhile complement to the systematic literature search (AWMF Guidance Manual, 2020). The 'Systematic Search for Evidence Syntheses and Guidelines' manual contains a detailed list of sources for systematic reviews and bibliographic databases that can be used to search for publications of clinical studies and other sources (Cochrane Deutschland Stiftung et al., 2020).

The publications that are found must meet the previously defined evidence criteria (see 2.2.1, Critical appraisal of the evidence) in order to undergo further assessment. To this end, the quality of evidence and strength of recommendation are assessed in a further step using one of the established instruments (see 2.2.2, Standards for quality of evidence). At this point in the evidence phase, there should be close cooperation and a coordination process with the *living guideline group* to clarify methodological queries. The *living guideline group* is then given the new assessed evidence for inclusion in the recommendations and for evaluating and agreeing on its potential for changing the recommendations in question. The recommendation is adapted only if there is a potential for changing the existing recommendation(s) (relevant outcomes). It is important to keep in mind that some recommendations are interdependent and thus must be updated simultaneously as needed (Akl et al., 2017).

The *ResearchRabbit* (https://www.researchrabbit.ai/) tool may offer good prospects for quicker literature searches in future. The tool filters reviews and studies on certain subjects, uses e-mail alerts to notify users about newly found reviews and enables clearly structured visualization of the literature. While the tool is not yet fully developed (for example, the tool accesses only two databases to date), it should be kept in mind for upcoming revision cycles.

2.2.1 Critical appraisal of evidence

Experts with experience in methodology, namely the *evidence group*, use the tools listed below to assess the evidence. The following checklists (AWMF Guidance Manual, 2020) are suitable for general methodological assessment:

- For systematic reviews with and without meta-analyses: AMSTAR (A Measurement Tool to Assess systematic Reviews) Checklist II (Shea et al., 2017)
- For guidelines: The *AGREE II* tool (AGREE Collaboration, 2014)
- For primary studies (dependent on study design): the *Cochrane Risk of Bias Tool II* for randomized studies (Sterne et al., 2019); the Scottish Intercollegiate Guidelines Network (SIGN) and other organizations provide editable checklist templates (SIGN, 2021)

2.2.2 Standards for quality of evidence

As described above, the *evidence group* should also be responsible for the ongoing/continued appraisal of the quality of evidence and strength of recommendation. We recommend using *GRADE* (**G**rading of **R**ecommendations, **A**ssessment, **D**evelopment and **E**valuation) as a standard instrument, as is already the case for a number of living guidelines (Pielenz/Schneider et al., 2022). However, before discussing the *GRADE* process in greater detail below, we would like to mention two other standards often used in developing living guidelines: the **O**xford **C**entre for **E**vidence-**B**ased **M**edicine; (OCEBM Levels of Evidence Working Group) and the **S**cottish **I**ntercollegiate **G**uidelines **N**etwork (SIGN). Table 1 presents a comparison of these instruments. The table does not claim to be exhaustive.

Table 1

Comparison of instruments for assessing the quality of evidence and strength of recommendations

Standard	**Pros**	**Cons**
GRADE (Grading of Recommendations, Assessment, Development and Evaluation)	▪ Available evidence is considered from the perspective of endpoint/outcome.	▪ Problematic in terms of simplicity and efficiency
	▪ Observational studies with strong effects can be 'upgraded' and studies can be 'downgraded' for quality-related and other reasons, enabling evidence from observational studies to be prioritized over evidence from RCTs, if necessary (Howick et al., 2011).	▪ Takes time to master and is primarily intended for appraising systematic reviews used for guideline development (Howick et al., 2011)
	▪ Key factors such as directness, precision, and consistency are considered when evaluating the evidence (Howick et al., 2011)	▪ Not fast enough for busy clinicians to find and use the likely best evidence to answer a clinical question (Howick et al., 2011)
	▪ Internationally recognized and most widely used according to a scoping review for living guidelines (Pielenz/Schneider et al., 2022).	
OCEBM (Oxford Centre for Evidence-Based Medicine)	▪ Levels of evidence for therapy/prevention/ etiology/harm, prognosis, diagnosis, differential diagnosis, and economic decision analyses (more diverse than other evidence schemes) (Howick et al., 2011).	▪ Not as detailed as *GRADE*; thus may not be as accurate in application
	Economical and efficient in use, thus saving time for clinicians (Howick et al., 2011).	▪ Rarely used for living guidelines according to a scoping review for living guidelines (Pielenz/Schneider et al., 2022).
	▪ Simple hierarchy	

	▪ After *GRADE,* the most frequently used standard for living guidelines	
SIGN (Scottish International Guidelines Network)	▪ Simple and clear to use, making it suitable for small or resource-constrained guideline development groups (Baker et al., 2010)	▪ Rarely used for living guidelines according to a scoping review for living guidelines (Pielenz/Schneider et al., 2022).
	▪ Highlights aspects of the study design that may lead to biased results and identifies the direction of these biases (Baker et al., 2010)	
	▪ Ensures that internal and external validity are assessed and incorporated into the evaluation of a recommendation (Baker et al., 2010)	

Note. Comparison of three instruments used for developing living guidelines. This list is not exhaustive.

The use of *GRADE* is recommended. As *GRADE* is also compatible with the *MAGICapp* presentation format (see 4.1.2 *MAGICapp* presentation format), the two aforementioned instruments, *OCEBM* and *SIGN*, are not described in the sections below.
When using the *GRADE* approach, the available evidence is viewed from the perspective of the endpoint or outcome (assessment of the totality of the studies as a 'body of evidence' for each relevant endpoint, see Fig. 5 and Table 2) (AWMF Guidance Manual, 2020; *GRADE Handbook*, 2013). The *GRADE Handbook* (2013) limits the number of outcomes to seven. The endpoints are rated according to the criteria 'Critical for making a decision' (7-9), 'Important, but not critical for making a decision' (4-6) and 'Of limited importance for making a decision' (1-3).
The studies included are assessed not only in view of their risk of bias due to the study design (e.g., lack of blinding, selective outcome reporting), but also with regard to indirectness of evidence (examines whether the question to be answered has been answered by the evidence with regard to target population, intervention, control group and endpoints), heterogeneity (unexplained heterogeneity in the results across studies) and imprecision of results (defining a decision-making threshold, meeting criteria for optimal information size, number of events and sample sizes), as well as publication bias (GRADE Handbook, 2013).
Please refer to the *GRADE Handbook* (2013) for precise statements on the individual domains. It is important to refrain from rigidly deducting points for each applicable domain, because the individual domains are interconnected in some cases and viewing the individual domains in isolation could result in an assessment that is too strict. In case of high-quality meta-analyses and systematic reviews (e.g. Cochrane) who already used GRADE in their methodology the GRADE schemes can be adopted.
Figure 5 provides an overview of the *GRADE* process carried out for each individual endpoint.

Figure 5
Criteria for assessing the quality of evidence

<table>
<tr><th>Study design</th><th>Initial quality of a body of evidence</th><th>Lower if</th><th>Higher if</th><th>Quality of a body of evidence</th></tr>
<tr><td>Randomised trials</td><td>High</td><td>Risk of Bias
Inconsistency
Indirectness
Imprecision
Publication bias</td><td>Large effect
Dose response
All plausible residual confounding & bias
-Would reduce a demonstrated effect
-Would suggest a spurious effect if no effect was observed</td><td>A/High (four plus: ⊕⊕⊕⊕)
B/Moderate (three plus: ⊕⊕⊕◯)
C/Low (two plus: ⊕⊕◯◯)
D/Very low (one plus: ⊕◯◯◯)</td></tr>
<tr><td>Observational studies</td><td>Low</td><td></td><td></td><td></td></tr>
</table>

Note. Illustration of the *GRADE* process performed for each endpoint. Included studies are assessed for their risk of bias, inconsistency, indirectness, lack of precision, and publication bias. Final study quality is rated from very low to high. Adopted from "The *GRADE* approach: an introductory workshop" (2011) (https://slideplayer.com/slide/6254343/).

As part of a guideline process, the overall quality of the evidence should be determined on the basis of all critical endpoints. If the quality of evidence differs for the individual endpoints, the evidence with the lowest quality is always selected as the overall quality of evidence (see GRADE Handbook, 2013).

Table 2
Quality of evidence grades

Degree	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 the effect.

Note. In the *GRADE* approach, the available evidence is considered from the perspective of endpoints or outcomes (assessment of the totality of studies as a 'body of evidence' for each relevant endpoint). Final study quality is rated on a scale from 'very low' to 'high'. Taken from the *GRADE Handbook* (2013).

The *GRADE* process presented above applies to pairwise meta-analyses. *GRADE* provides for a different procedure for network meta-analyses. Please consult the relevant publications (Puhan et al., 2014; Brignardello-Petersen et al., 2018; Yepes-Nuñez et al., 2019) for guidance.

2.2.3 GRADE - from evidence to recommendations

The first step entails the preparation of a *Summary of Findings table* which clearly summarizes the information obtained from the appraisal of evidence and which can serve as an aid for the subsequent decision-making step (see Fig. 6).

Figure 6

Summary of Findings – table of evidence including evaluation per study

heparin compared to no heparin for patients with cancer who have no other therapeutic or prophylactic indication for anticoagulation

Bibliography: Akl EA, Gunukula SK, van Doormaal FF, Barba M, Kuipers S, Middeldorp S, Yosuico VE D, Dickinson HO, Schünemann H. Parenteral anticoagulation in patients with cancer who have no therapeutic or prophylactic indication for anticoagulation. Cochrane Database of Systematic Reviews [Year], Issue [Issue].

Outcomes	No of Participants (studies) Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with No heparin	Risk difference with Heparin (95% CI)
Mortality	2531 (8 studies) 12 months	⊕⊕⊕⊖ **MODERATE**[1,2,3] due to inconsistency	**RR 0.93** (0.85 to 1.02)	**Moderate**	
				649 per 1000	**45 fewer per 1000** (from 97 fewer to 13 more)
Symptomatic VTE	2264 (7 studies) 12 months	⊕⊕⊕⊕ **HIGH**[1]	**RR 0.55** (0.37 to 0.82)	**Moderate**	
				29 per 1000	**13 fewer per 1000** (from 5 fewer to 18 fewer)
Major bleeding	2843 (9 studies) 12 months	⊕⊕⊕⊖ **MODERATE**[1,4] due to imprecision	**RR 1.3** (0.59 to 2.88)	**Moderate**	
				7 per 1000	**2 more per 1000** (from 3 fewer to 13 more)
Minor bleeding	2345 (7 studies) 12 weeks	⊕⊕⊕⊖ **MODERATE**[1,4] due to imprecision	**RR 1.05** (0.75 to 1.46)	**Moderate**	
				27 per 1000	**1 more per 1000** (from 7 fewer to 12 more)
Health related quality of life the Uniscale and the Symptom Distress Scale (SDS); Better indicated by lower values	0 (1 study) 12 months	⊕⊕⊖⊖ **LOW**[6] due to risk of bias, imprecision	Not estimable[5]	See comment	-

*The basis for the **assumed risk** (e.g. the median control group risk across studies) is provided in footnotes. The **corresponding risk** (and its 95% confidence interval) is based on the assumed risk in the comparison group and the **relative effect** of the intervention (and its 95% CI).
CI: Confidence interval; **RR:** Risk ratio;

GRADE Working Group grades of evidence
High quality: Further research is very unlikely to change our confidence in the estimate of effect.
Moderate quality: Further research is likely to have an important impact on our confidence in the estimate of effect and may change the estimate.
Low quality: Further research is very likely to have an important impact on our confidence in the estimate of effect and is likely to change the estimate.
Very low quality: We are very uncertain about the estimate.

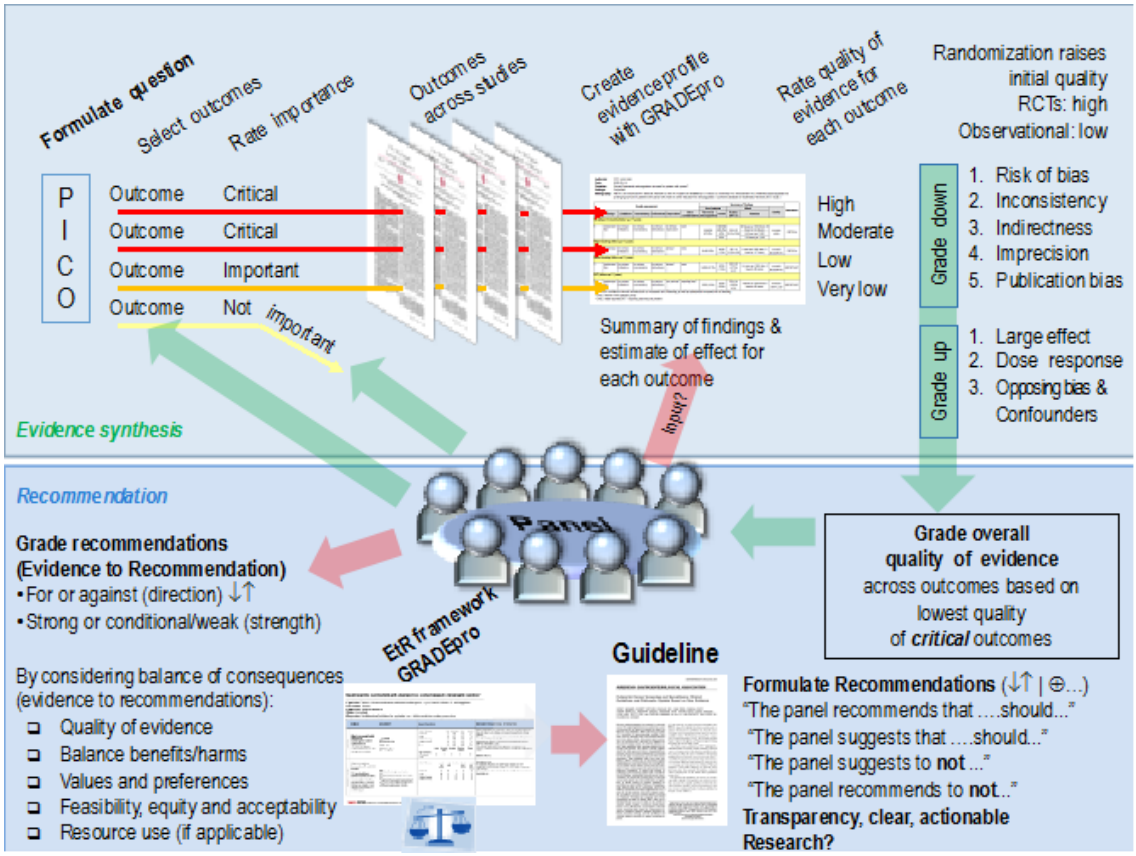
[1] Vast majority of studies had allocation concealment, and used blinded outcome and adjudication. We did not downgrade although there was some concern about lack of blinding in some studies; the overall risk of bias was felt to be very low.
[2] There is moderate heterogeneity among studies included in the analysis of death at 12 months (I2=35%). The subgroup analysis for mortality at 12 months was statistically significant and suggested survival benefit in patients with SCLC but not in patients with advanced cancer. Overall we decided to downgrade by one level when considering these issues along with imprecision.
[3] CI interval includes effects suggesting benefit as well as no benefit.
[4] CI includes possibility of both harms or benefits., [5] The scores for the 2 scales were similar for the 2 study groups, both at baseline and at follow-up, [6] High risk of bias and only 138 patients enrolled.

Note. A summary-of-findings table should be prepared using the *GRADE* approach. It clearly summarizes the information obtained from the assessment of the evidence and serves as an aid for subsequent decision making (shown here as an example). Adopted from *GRADE Handbook* (2013).

The next step entails the *evidence-to-decision framework*. The following figure illustrates how this framework is integrated in the guideline process (see Fig. 7).

Figure 7

Schematic representation of the GRADE process for developing recommendations

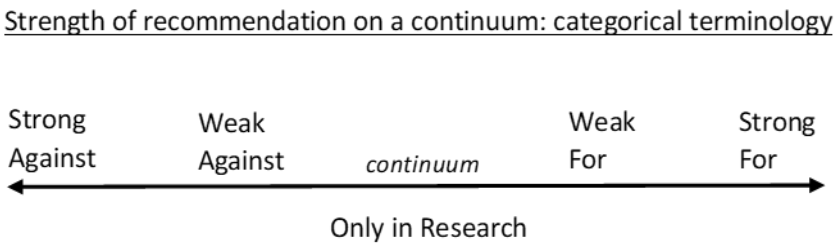


Note. Evidence-to-Decision- framework. Adapted from the *GRADE Handbook* (2013).

The *strength of recommendation* refers to the extent to which the *evidence group* is certain that desirable effects do not outweigh the undesirable effects.
There are two categories (with some exceptions): strong (for/against) and weak (for/against) (see Fig. 8). Optional recommendations ("can") are not suggested.

Figure 8

Strength of recommendation: a continuum divided into categories



Note. Assessment of the strength of the recommendation, to what extent the desirable effects outweigh the undesirable effects. The assessment of recommendation strength is based on the categories strong (for/against) and weak (for/against). Adopted from the *GRADE Handbook* (2013).

In this case, a strong recommendation means that the *evidence group* is certain that the desirable (strong recommendation for) effects outweigh the undesirable effects (strong recommendation against) or vice versa.

A weak recommendation is one for which uncertainties exist, meaning that the desirable effects probably outweigh the undesirable effects or the undesirable effects probably outweigh the desirable effects. A weak recommendation can also be referred to as a conditional recommendation, because a weak recommendation does not necessarily mean that the evidence is weak. A conditional recommendation can also be based on evidence of high quality but is weakened by other factors.

The evidence-to-decision framework

The *evidence-to-decision framework* is a continuation of the *GRADE* process. It contains various criteria included in the creation of a recommendation. These criteria may include the relevance of the question, the overall quality of the evidence (see above), the strength of the desirable and undesirable effects, the deliberation of these factors, the consideration of values and preferences, the use of resources, the implementability of the intervention and any inequality in the health care system as a possible consequence. This *evidence-to-decision framework* is not meant to be applied rigidly but is a flexible and transparent system that can be adapted to fit the question at hand.

Within this framework, the individual criteria are underpinned with evidence and the *evidence group* votes on the individual criteria (e.g. no, probably no, uncertain, probably yes, etc.).

Lastly, a summary of judgements is presented in a table which should conclude with a statement on the strength of recommendation: strong (for/against) or weak (for/against).

2.2.4 GRADE - preparing the evidence that has been assessed

GRADE proposes the following scheme for presenting the quality of evidence and strength of recommendation:

Table 3

Representations of quality of evidence and strength of the recommendations

Quality of Evidence	**Symbol**	**Letter** (varies)
High	⊕⊕⊕⊕	A
Moderate	⊕⊕⊕◯	B
Low	⊕⊕◯◯	C
Very low	⊕◯◯◯	D
Strength of Recommendation	Symbol	Number
Strong for an intervention	↑↑	1
Weak for an intervention	↑?	2
Weak against an intervention	↓?	2
Strong against an intervention	↓↓	1

Note. The *GRADE* approach is used for symbolic representation of evidence quality and strength of recommendation. Taken from the *GRADE Handbook* (2013).

It is always advisable to use the symbols for the quality of evidence (+) and for the strength of recommendation (↑) rather than using letters or numbers, because the latter are used in guidelines in different ways and can more easily lead to misunderstandings. The *GRADE* scheme uses a two-level gradation (we recommend/we suggest), as presented in Table 4.

Table 4

Two-level scheme for gradation of recommendations

Symbol	Description	Wording
↑↑	Strong recommendation	We recommend / Clinicians should/should not
↑	Conditional recommendation	We suggest / We conditionally recommend

Note. According to the *GRADE* scheme, a two-level gradation (we recommend/we suggest) is used for strong and conditional recommendations respectively. Adopted from the *AWMF Guidance Manual* (2020).

3 Consensus phase

To legitimate the decisions made by the *living guideline group* and to agree on the updated recommendations a structured process of reaching consensus is needed. This process should not only legitimate but also ensure the guidelines acceptance and prevent potential biases.

3.1 Voting process and structured consensus

The representatives must vote on the life cycle of a living guideline on at least two different occasions. It is recommended holding a vote at the start of the process in a constitutive phase (see chapter 1) in order to legitimize the planned procedure for the living guideline process. In addition, interests should be recorded and assessed. A second round of voting takes place upon conclusion of the living guideline cycle as part of the consensus phase in order to adopt the amendments and confirm the validity of the recommendations. The participating professional societies, organisations and associations of those affected should be granted the right to a special vote in the consensus phase. This offers the opportunity to revise the supporting agreement on the recommendation(s) and submit an alternative text for explanation (see below, procedure in case of persistent dissent (adapted based on the AWMF Guidance Manual, 2020)). This must be clearly indicated in the publication.

To accelerate the voting processes and simplify their coordination, it is suggested using an online format for voting rather than face-to-face conferences that are otherwise conventionally used and often last for several days.
Consensus can be reached both in a digital nominal group process (via a direct, face-to-face presentation, discussion and voting during the online conference) and via the *Delphi technique* (time-delayed multi-stage questioning method by means of an online survey or questionnaire with predefined response deadlines). The *steering group* reserves the right to propose face-to-face meetings. For example, the change proposal can be sent to the *consensus group* for review in advance, and the new evidence situation can be presented and discussed in a digital or face-to-face meeting. In the case of a nominal vote, the involvement of a neutral moderator is recommended.

Formal voting techniques are described in the *AWMF Guidance Manual* (2020) and distinguish between the following: nominal group process (recommended for around 15 to 20 participants, direct, personal plenary discussion and voting, thus not allowing for anonymous participation), structured consensus conference (recommended for around 20 to 60 participants, discussion and preparation of the statements in small groups, subsequent presentation in the plenary session of the consensus group and direct voting) or *Delphi technique* (round-based, anonymous voting, discourse while viewing the anonymous comments). In the following the option of online and time-delayed voting using the *Delphi technique* is described in detail, because in the context of the shortened update period a time-delayed online voting procedure that does not require set deadlines is preferable to a (face-to-face) procedure in real time.

Delphi technique (*AWMF Guidance Manual*, 2020)

The *Delphi technique* is a multistage survey method conducted in writing to achieve consensus between experts from different disciplines. All participants are presented with statements and supplementary information and are then asked whether they agree or disagree. The participants may submit comments to justify their decision. All responses are evaluated and are reported in anonymous form to all members of the consensus groups in further rounds. In view of the time constraints, we recommend limiting the survey to two rounds. A third round is only necessary in the event of dissent (see below). All participants are surveyed again; the knowledge connected to the group opinion and additional information gleaned from the comments offer the opportunity to revise the judgement made (Hasson et al., 2000).

The term 'online Delphi' will be used below to refer to the *Delphi technique* used as a digital method.

Using the online Delphi in the living guideline update process

- Define responsibility for conducting and evaluating the *Delphi* survey
- Select the tool for conducting the survey, prepare the questionnaires
- Define binding reporting deadlines and quotas for a quorum for the online voting
- Define the required consensus strength and how to handle dissent

The individuals responsible for supporting the *online Delphi* must first be appointed. A suitable tool for conducting the *online Delphi* can be sourced from a commercial supplier of dedicated software. Alternatively, an application for web-based surveys can be used, as is common practice at many hospitals and other institutions. In the next step, the structure of the survey, the planned number of *online Delphi* rounds and the type of feedback options should be determined.

For the *online Delphi* procedure, its recommended sending an invitation ahead of time with the instructions and response deadlines (e.g. at least 30 days for a two-round Delphi, cf. McMillan et al., 2016). Data protection consent should be obtained from all members in advance and any conflicts of interest should be re-examined, which may affect the participants' voting rights (see 1.3, Conflicts of interest). Abstentions due to conflicts of interest should be subtracted from the population of voting mandate holders when determining the consensus rate. In the first round, the initial questionnaire is sent out. Quantitative response options can be defined nominally (e.g. abstention, agreement, disagreement) or ordinally (rating of agreement and disagreement on Likert scales). Prior to conducting the *online Delphi* survey, values for reaching consensus or dissent must be defined (for example, for a 9-level Likert scale, agreement is reached when the median score is greater than or equal to 7, disagreement if the median is less than or equal to 3) (McMillan et al., 2016). It is also advisable to provide a qualitative response option, such as a comment box for justifying the response. No new studies or publications should be introduced in the comments, because this is already covered by the comprehensive literature search and defined criteria in the evidence phase (see chapter 2) and could cause too great a delay in consensus in an annual update. Objections or clarifications can be made in writing to the *living guideline group*. The comments reported back should be documented.

If no response is received, one reminder is sent stating the final deadline and the consequences of exclusion of the professional society if it does not respond.

To ensure the quorum for voting, a minimum response rate should be defined prior to launching the *online Delphi*. In different studies on the use of the *Delphi technique*, return rates of between 72% and 92% are described (Bisson et al., 2010; McMillan et al., 2016; Niederberger

& Renn, 2019; Nurek et al., 2021). Reminders sent prior to the response deadline should ensure that the minimum number of responses is received. If this number is not reached, this must be documented in the guideline report. The participating professional societies will have declared their willingness to participate in advance. If a professional society fails to vote even after a reminder has been sent, the society should be excluded from participating in the guideline process so that the guideline group is still able to work and make decisions, also in view of further update cycles.

Consensus criteria

The first *online Delphi round* is followed by the evaluation of the answers. The criteria for reaching consensus and the procedure for failure to reach consensus must be defined. It is recommended using the percentages of agreement for achieving consensus suggested in the *AWMF Guidance Manual* (2020) (see Table 5).

Those recommendations where sufficient consensus could not be reached (≤75%) are presented again to the *consensus group*, with anonymised feedback from all members noted.

Recommendations on which consensus or strong consensus (>75%) is already achieved in the first round do not need to be presented again. The *steering group* is responsible for deciding whether to conduct additional rounds. For an annual living cycle, a total of two *Delphi rounds* can be expected. A third round should be carried out only in the event of dissent.

Table 5

Determining consensus strength

Strong consensus	Agreement by > 95 % of the participants
Consensus	Agreement by > 75 to 95 % of the participants
Majority approval	Agreement by > 50 to 75 % of the participants
No majority approval	Agreement by ≤ 50 % of the participants

Note. Consensus thresholds in a structured voting process stated in percentage terms. Adopted from the AWMF Guidance Manual (2020).

Procedure in case of persistent dissent

If there is a lack of consensus on a recommendation with a new body of evidence, it is recommended conducting a third *online Delphi round*. In this round, the new body of evidence is presented with the recommendation in the previous wording and votes are cast again. If consensus has been reached, the previous recommendation is retained and a note is made indicating the new body of evidence (e.g. in the background text). In the case of dissent, the previous recommendation should be deleted from the guideline. This must be documented in the guideline report.

Recommendations for presenting dissent (adapted based on the AWMF Guidance Manual, 2020):

1. Dissent with regard to recommendations: The professional society/organization requests inclusion of a statement in the guideline report that they were involved in the development process but they do not approve of the relevant recommendations of the guideline. In this case, the guideline text remains unchanged in the version agreed between the members of the guideline group and adopted by the other professional societies/organizations.
2. Dissent with regard to the entire guideline: The professional society/organization withdraws its participation altogether and is no longer mentioned as a participant. As in item 1, here too the guideline text remains unchanged.
3. Dissent by means of a special vote: At the time of the first Voting round, the patient association and the professional society/organisation can express its dissent by means of a special vote (to be submitted directly to the living guideline group) and is given the

opportunity to formulate alternative texts regarding recommendations that are not supported. The relevant recommendation is marked with the words 'Special vote in place' and the alternative text appears in the guideline supplement.

Alternatively, the *real-time Delphi* approach may be used, in which the participants' responses are provided in real- time and no iterative rounds are required (Gerhold, 2019). Participants are granted access to the portal for a limited period and can log in at any time during this period to review and edit their own assessment as needed and view the assessments of the other participants. The first *real-time Delphi* studies were conducted back in 2006 (Cuhls et al., 2007; Gordon & Pease, 2006; Zipfinger, 2007).

3.2 Global adoption

The annual global adoption of the living guideline confirms the continued validity of all recommendations. For global adoption of the updated living guideline, it is advisable to follow the guidance in the *AWMF Guidance Manual (2020)*, which suggests several alternatives. Due to the limited time available for the update, the following procedure is recommended: 'The next step is formal adoption by the boards of the participating medical societies/organizations. This ensures that all parties involved in developing the guideline and the co-editing medical societies bear mutual responsibility for the contents of the guidelines' (AWMF Guidance Manual, 2020). Feedback from the organizations, professional societies and patient associations on the living guideline and the modified/adapted recommendations must be obtained in writing by e-mail (AWMF Guidance Manual, 2020). 'The representatives are authorized by their professional society to vote on behalf of the board' (AWMF Guidance Manual, 2020). A deadline should be set for submission of feedback. To prevent delays, we recommend that agreement be assumed if no feedback is received by the deadline. The *steering group* can set a relevant deadline in view of the available time. The ongoing update process offers professional societies and organizations the opportunity to modify their positions as needed at shorter intervals. The *steering group* decides how to handle disagreement – for example, whether to mention in the living guideline that one or more professional societies or organizations have not agreed with a certain recommendation or the entire guideline. The procedure for handling changes must be defined in the guideline report (AWMF Guidance Manual, 2020).

4 Continuity phase

The content work is followed by the editorial implementation and ongoing maintenance of the living guideline. This contains the structure of guideline chapters, dissemination via digital platforms (e.g. MAGICapp or guideline homepages) but also handling feedback from the public. Finally strategies for enhancing the implementation of guidelines will be necessary especially when using new formats for guideline presentation (Lorenz et al., under review).

4.1 Editorial work

4.1.1 Clarity and presentation

The wording of the guideline title should clearly indicate that it is a living guideline. The rules should be formulated clearly, should be attractively presented, and be easily identifiable for the guideline's users (AWMF Guidance Manual, 2020). The *AWMF Guidance Manual* (2020) recommends that for a living guideline the most important updates should be presented at the beginning of the updated guideline ('What's new?') and the recommendations should be marked with 'Reviewed', 'Modified', 'New' and the year (Akl et al., 2017; GIN-McMaster Guideline Development Checklist, 2013). The methodological strategy should be included in the guideline report (Vernooij et al., 2017).

In general, the presentation of the living guideline should be of practical use to the target user and be understandable (Gaigl et al., 2021). The following structure is recommended (see also AWMF Guidance Manual, 2020, 'S3 template'):

Title including the words 'living guideline'

1. Brief summary/summary: clinical questions, recommendations, methodology, latest findings
2. Reading guide
3. Introduction
4. Methods
5. Clinical questions
6. Recommendations
7. Abbreviations and acronyms
8. References
9. Guideline report

4.1.2 Presentation formats

As described in a review (Pielenz/Schneider et al., 2022), current living guidelines generally use a static PDF document as the publication format and the second most frequent format is *MAGICapp*. In some cases, the organizations' own websites, in which the guidelines are embedded (e.g. https://immunisationhandbook.health.gov.au/), or mobile applications developed specifically for the guidelines (e.g. https://www.leitlinienprogramm-onkologie.de/app/) are used.

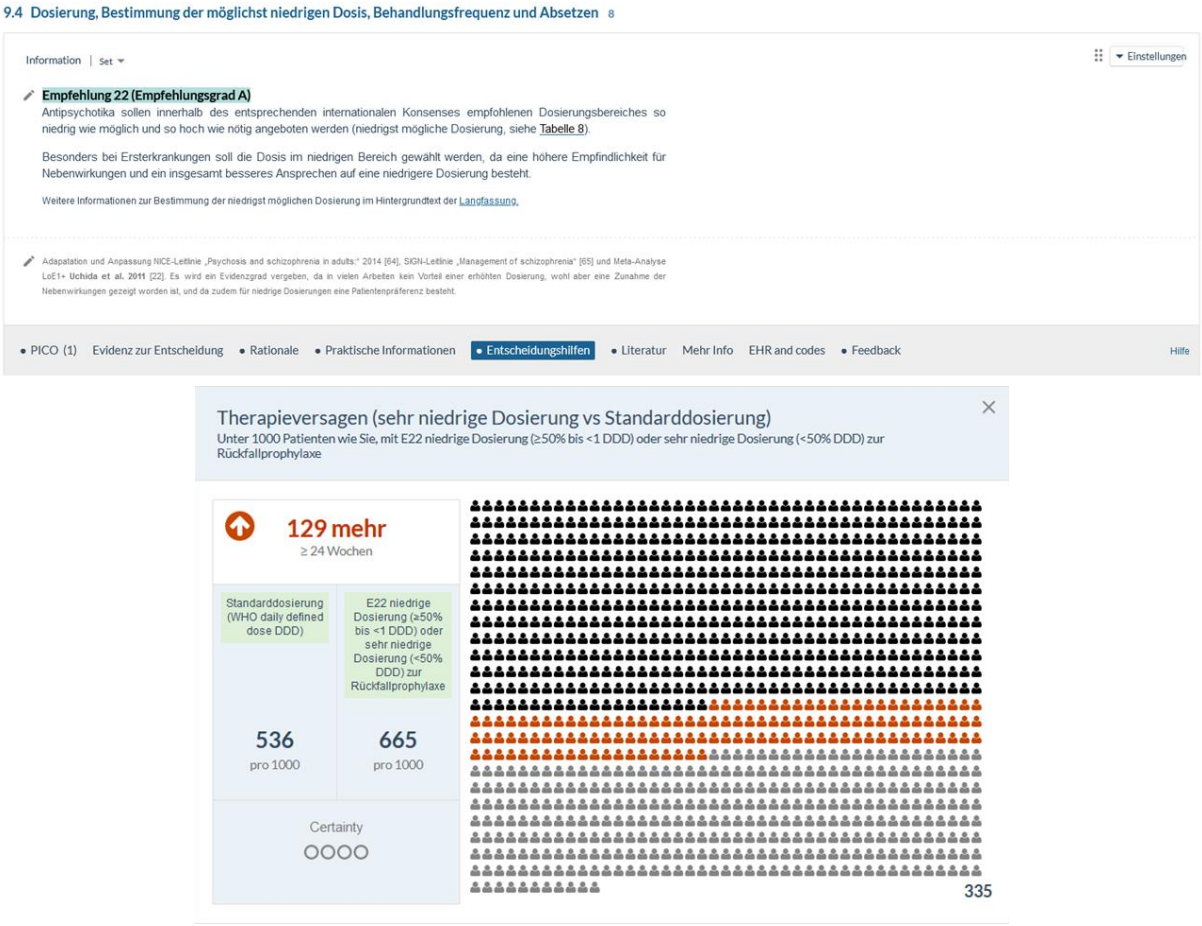
Presentation format *MAGICapp*

It is recommended using the *MAGICapp* digital platform *(*app.magicapp.org) for structured management and dissemination of the living guideline (*AWMF Guidance Manual*, 2020; Pielenz/Schneider et al., 2022). *MAGICapp* (**Ma**king **G**RADE the **I**rresistible **C**hoice) was founded by Per Olav Vandvik, Linn Brandt, Gordon Guyatt and Thomas Agoritsas. *MAGICapp* uses *GRADE* for developing a guideline, both for assessing evidence and for developing the recommendation (see 2.2.3 GRADE - from evidence to recommendations). It permits visualization of the evidence, which aims to facilitate shared decision-making (see Fig. 9). In addition, links to journals and *Pubmed* can be included in the references for the recommendation in question so that the relevant abstract or, in the case of open-access articles, the entire PDF, can be easily accessed by one click. Furthermore, practical tips on each recommendation and links to different guidelines can be stored on *MAGICapp*. *MAGICapp* offers the possibility to link recommendations from different guidelines. Especially in the case of independent update cycles, this offers the advantage of keeping linked recommendations up to date for living guidelines.

The status of a recommendation can be indicated in *MAGICapp* ('New'/'Updated'/'Under review'). In addition, a comment box can be inserted for public consultation. As publishing a full PDF version of the guideline is still standard (Pielenz/Schneider et al., 2022), it is also advisable to provide a PDF version of digital guidelines, which is possible with *MAGICapp*.

Figure 9

Screenshots of MAGICapp taken from the German schizophrenia guideline



Note. Recommendation 22 was taken from the following guideline and adapted for presentation in *MAGICApp*: German Society for Psychiatry and Psychotherapy, Psychosomatics and Neurology e.V. (DGPPN) (2019). S3 guideline for schizophrenia (037-009). AWMF online. https://www.awmf.org/leitlinien/detail/ll/038-009.html

4.2 Guideline report

The methodological procedure used to create a living guideline should be added to the guideline report (AWMF Guidance Manual, 2020). We refer here to the *AWMF Guidance Manual* (2020), which provides guidance on the guideline report (Appendix 10 "Guideline for the preparation of a guideline report for authors of S2k, S2e and S3 guidelines", https://www.awmf.org/regelwerk/formate-und-verbreitung-von-leitlinien). If *MAGICapp* is to be used as the presentation format, the guideline report becomes a sub-section of the living guideline and is not presented separately. Changes in search strategy, search terms used, sources and search period should be outlined.

4.3 Public consultation

It is recommended to have a public consultation phase also in the process of a living guideline (Pielenz/Schneider et al., 2022). This phase does not take place until after the first update has been disseminated, because the living guideline is not available until then (e.g. via *MAGICapp*). In the subsequent update cycles, comments from the (specialist) public may be collected on an ongoing basis.

Feedback can be submitted via a dedicated e-mail address or the comments feature of the *MAGICapp* platform under the 'Feedback' tab (see 4.1.2, *MAGICapp* presentation format)

(see also http://help.magicapp.org/knowledgebase/articles/1933261-comments-feature). Comments on the *AWMF-registered* guidelines can also be submitted via the *AWMF* website. Target users require their own account for commenting via *MAGICapp* so that the feedback can be matched to the submitter and cannot be submitted anonymously. Members of the *evidence group* can delete, edit or respond to all comments. The author of the comment is notified by e-mail when the comment is edited. For the sake of readability, the status of the comment may be modified ('unresolved', 'in progress', or 'resolved').

If feedback can be submitted on an ongoing basis, users should be informed that the comments will not be taken into account until the next annual update cycle, parallel to the evidence phase. It is suggested that the feedback process be formalized to stipulate that written feedback must contain references to new scientific studies. The feedback can then be processed as part of this public consultation with the standards employed for assessing evidence (see, 2.2.1 Critical appraisal of evidence and following). After consultation with the *steering group*, ad hoc feedback on a new body of evidence leads to a reference to the respective guideline recommendation. If the potential for changing a recommendation arises during the year, this should be indicated in *MAGICapp* (e.g. including the wording 'New body of evidence - recommendation under review').
Alternatively, the living guideline can be made available to the (specialist) public for comment for a specified, limited period (for example, one month) after the update is published (AWMF Guidance Manual, 2020).

The way feedback is handled, along with the substantive assessment of this feedback, should be coordinated in the *living guideline group* and documented in the guideline report (AWMF Guidance Manual, 2020).

5 Challenges and future developments

One opportunity for better networking between guidelines (groups) should be highlighted at this point. It is conceivable that a new body of evidence may be relevant for different recommendations and guidelines. The pooled coordination, standardized work resources via the professional society publishing the guideline and linking recommendations on *MAGICapp* enable changes to be made visible in all relevant guidelines simultaneously.
Implementing practical guideline recommendations in everyday clinical routine poses a huge challenge. It is a well-known problem that despite a number of existing clinical guidelines, implementation of the guidelines in clinical practice worldwide is still inadequate (Waldorff et al., 2003; Grol, 2001; Girlanda et al., 2017; Gaigl et al., 2021; Pereira et al., 2022).
The attitude towards both the implementation barriers and moderators of a living guideline for schizophrenia was studied for the first time in a survey of 439 participants (doctors, psychologists, psychotherapists, nursing staff, psychosocial therapists) (Khorikian-Ghazari et al., 2023).
Given the lack of a living guideline, two-thirds (64%) of the respondents said they would currently prefer a living guideline to a printed format and three-quarters (75%) stated that they were aware of the advantages of a living guideline compared to the printed version (Khorikian-Ghazari et al., 2023). Based on the images of the living guideline for schizophrenia that were presented (in *MAGICapp*), more than half of the participants felt that the living guideline was user friendly: 68% considered the layout to be attractive and the content clearly displayed, 77% could imagine handling the living format well, and 52% said that the living guideline was clearer than the previous printed version. With regard to practical clinical use/relevance, more than half the respondents considered the living guideline to be more practical than the printed

format (62%) and a valuable instrument for everyday clinical routine (80%) (Khorikian-Ghazari et al., 2023).

While the results indicate a beneficial effect, implementation entails various challenges.
More than two-thirds of the participants considered 'lack of experience' (80%) and 'lack of awareness' (64%) of living guidelines in general as a barrier to using the pending schizophrenia guideline in the living format (Khorikian-Ghazari et al., 2023). In addition, 64% of the respondents reported that they anticipate difficulty accessing the living guideline for schizophrenia once it has been published (Khorikian-Ghazari et al., 2023).

In order to completely harness the benefits of a living guideline, the barriers should be overcome by offering training on new dissemination formats, such as *MAGICapp*. National (AWMF, Kopp et al., 2020) and international guideline committees (*GIN*) prefer *MAGICapp* for digitization of a guideline. Time and resources can be saved thanks to the digital presentation of new and revised recommendations and e-mail alerts about new evidence. Knowledge can be disseminated in a timely manner with direct reference to the recommendation. In addition, professional development courses should be offered that deliver general information on living guidelines and promote practical use of living guidelines in everyday clinical routine, as have already been offered within the *SISYPHOS* project.

Methodological development of the manual

This manual was created as part of the G-BA-funded (German Joint Federal Committee) project on *Structured implementation of digital, systematically updated guideline recommendations for enhanced adherence in schizophrenia* (SISYPHOS) based on a comprehensive literature search and a review procedure with guideline development experts. In addition, the manual was piloted within the project team for revision of the S3 guideline for schizophrenia. The consensus group of the S3 guideline on schizophrenia was also invited to comment on the manual.

References

AGREE Collaboration. (2014). *Appraisal of Guidelines for Research & Evaluation II - AGREE II Instrument - Deutsche Version: AGREE NEXT STEPS Consortium*. https://www.agreetrust.org/wp-content/uploads/2014/03/AGREE_II_German-Version.pdf

Akl, E. A., Meerpohl, J. J., Elliott, J., Kahale, L. A., & Schünemann, H. J. (2017). Living systematic reviews: 4. Living guideline recommendations. *Journal of Clinical Epidemiology*, *91*, 47–53. https://doi.org/10.1016/j.jclinepi.2017.08.009

ALEC - Australian Living Evidence Constortium (2022). The Living Guidelines Handbook: Guidance for the production and publication of living clinical practice guidelines [Version 1.0]. https://livingevidence.org.au/

AWMF. (2020). *AWMF-Regelwerk Leitlinien.* AWMF Online. https://www.awmf.org/leitlinien/awmf-regelwerk.html [Accessed on 21.03.2022]

Baker, A., Young, K., Potter, J., & Madan, I. (2010). A review of grading systems for evidence-based guidelines produced by medical specialties. *Clinical Medicine*, *10*(4), 358-363. https://doi.org/10.7861/clinmedicine.10-4-358

Bisson, J. I., Tavakoly, B., Witteveen, A. B., Ajdukovic, D., Jehel, L., Johansen, V. J., Nordanger, D., Orengo Garcia, F., Punamaki, R.-L., Schnyder, U., Sezgin, A. U., Wittmann, L., & Olff, M. (2010). Tents guidelines: Development of post-disaster psychosocial care guidelines through a Delphi process. *The British Journal of Psychiatry: The Journal of Mental Science*, *196*(1), 69-74. https://doi.org/10.1192/bjp.bp.109.066266

Brignardello-Petersen, R., Bonner, A., Alexander, P. E., Siemieniuk, R. A., Furukawa, T. A., Rochwerg, B., Hazlewood, G. S., Alhazzani, W., Mustafa, R. A., Murad, M. H., Puhan, M. A., Schünemann, H. J., & Guyatt, G. H. (2018). Advances in the GRADE approach to rate the certainty in estimates from a network meta-analysis. *Journal of clinical epidemiology*, *93*, 36-44.

Cochrane Deutschland Stiftung, Institut für Evidenz in der Medizin, Institut für Medizinische Biometrie und Statistik, Freiburg, Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften - Institut für Medizinisches Wissensmanagement, Ärztliches Zentrum für Qualität in der Medizin. (2020). *Manual Systematische Recherche für Evidenzsynthesen und Leitlinien*. AWMF Online. https://www.awmf.org/fileadmin/user_upload/Leitlinien/Werkzeuge/20201214_Manual_Recherche_Evidenzsynthesen_Leitlinien_V2.1.pdf [Accessed on21.03.2022]

Cuhls, K., von Oertzen, J., & Kimpeler, S. (2007). Zukünftige Informationstechnologie für den Gesundheitsbereich. Ergebnisse einer Delphi-Befragung. *Schriftenreihe Des Projekts FAZIT, MFG, Karlsruhe/Stuttgart*. www.fazit-forschung.de

Deutsche Gesellschaft für Psychiatrie und Psychotherapie, Psychosomatik und Nervenheilkunde e.V. (DGPPN) (2019). *S3-Leitlinie Schizophrenie* (037-009). AWMF online. https://www.awmf.org/leitlinien/detail/ll/038-009.html

Elliott, J. H., Synnot, A., Turner, T., Simmonds, M., Akl, E. A., McDonald, S., Salanti, G., Meerpohl, J., MacLehose, H., Hilton, J., Tovey, D., Shemilt, I., & Thomas, J. (2017). Living systematic review: 1. Introduction-the why, what, when, and how. *Journal of Clinical Epidemiology*, *91*, 23-30. https://doi.org/10.1016/j.jclinepi.2017.08.010

El Mikati, I. K., Khabsa, J., Harb, T., Khamis, M., Agarwal, A., Pardo-Hernandez, H., Farran, S., Khamis, A. M., El Zein, O., El-Khoury, R., Schünemann, H. J., Akl, E. A., & The Living

Guidelines Group†. (2022). A framework for the development of living practice guidelines in health care. *Annals of Internal Medicine*, *175*(8), 1154-1160. https://doi:10.7326/M22-0514

Gaigl, G., Täumer, E., Merz, K., Zöscher, S., Wagner, S., Kösters, M., Falkai, P., Leucht, S., & Hasan, A. (2021). Multifactorial barriers in the implementation of schizophrenia and psychosocial therapies guidelines: a quantitative study across different professions. *Schizophrenia Research*, *228*, 425-434. https://doi.org/10.1016/j.schres.2021.01.010

GIN - Guidelines International Network & McMaster University. (2014). *GIN-McMaster Guideline Development Checklist* (retrieved on16.11.2021). https://cebgrade.mcmaster.ca/guidelinechecklistonline.html

Gerhold, L. (2019). Real-Time Delphi. In M. Niederberger & O. Renn (Eds.), *Delphi-Verfahren in den Sozial- und Gesundheitswissenschaften* (S. 101–123). Springer Fachmedien Wiesbaden. https://doi.org/10.1007/978-3-658-21657-3_5

Girlanda, F., Fiedler, I., Becker, T., Barbui, C., & Koesters, M. (2017). The evidence–practice gap in specialist mental healthcare: systematic review and meta-analysis of guideline implementation studies. *The British Journal of Psychiatry*, *210*(1), 24-30. 10.1192/bjp.bp.115.179093

Gordon, T., & Pease, A. (2006). RT Delphi: An efficient, "round-less" almost real time Delphi method. *Technological Forecasting and Social Change*, *73*(4), 321-333. https://doi.org/10.1016/j.techfore.2005.09.005

Grol, R. (2001). Successes and failures in the implementation of evidence-based guidelines for clinical practice. *Medical care*, *39*, 46-54

Hasson, F., Keeney, S., & McKenna, H. (2000). Research guidelines for the Delphi survey technique. *Journal of Advanced Nursing*, *32*(4), 1008-1015. https://doi.org/10.1046/j.1365-2648.2000.t01-1-01567.x

Howick, J., Chalmers, I., Glasziou, P., Greenhalgh, T., Heneghan, C., Liberati, A., Moschetti, I., Phillips, B., & Thornton, H. (2011). *Explanation of the 2011 Oxford Centre for Evidence-Based Medicine (OCEBM) Levels of Evidence (Background Document)*. CEBM. https://www.cebm.ox.ac.uk/resources/levels-of-evidence/explanation-of-the-2011-ocebm-levels-of-evidence/

Khorikian-Ghazari, N., Lorenz, C., Güler, D., Halms, T., Röh, A., Flick, M., Kapfhammer, A., Pielenz, C., Salveridou-Hof, E., Schneider-Axmann, T., Schneider, M., Wagner, E., Falkai, P., Gaebel, W., Leucht, S., Hasan, A., Gaigl, G. (2023). Guideline for schizophrenia: implementation status and attitude towards an upcoming living guideline. *European Archives of Psychiatry and Clinical Neuroscience,* 1-12. https://doi.org/10.1007/s00406-023-01568-z

Kopp, I., Nothacker, M., Spies, C., Müller, A., Gamstätter, T., Langer, T., Wenzel, G., Starlinger, J., Karge, T., Vandvik, P. & Brandt, L. (2020). Digitalisierung von Leitlinienwissen zur Überwindung von Grenzen des Medizinischen Wissensmanagements: Modellprojekt für die Entwicklung hochwertiger Leitlinien und deren Verbreitung über Apps (DissoLVe).

Langer, G., Meerpohl, J. J., Perleth, M., Gartlehner, G., Kaminski-Hartenthaler, A., & Schünemann, H. (2012). GRADE-Leitlinien: 1. Einführung - GRADE-Evidenzprofile und Summary-of-Findings-Tabellen [GRADE guidelines: 1. Introduction - GRADE evidence profiles and summary of findings tables]. *Zeitschrift für Evidenz, Fortbildung und*

Qualitat im Gesundheitswesen, *106*(5), 357-368. https://doi.org/10.1016/j.zefq.2012.05.017

Living Evidence Network. (2019). *Guidance for the production and publication of Guidance for the production and publication of Cochrane living systematic reviews: Cochrane Reviews in living mode*. 201912_LSR_Revised_Guidance.pdf (cochrane.org)

Lorenz, C., Güler, D., Halms, T., Khorikian-Ghazari, N., Röh, A., Flick, M., Burschinski, A., Pielenz, C., Salveridou-Hof, E., Schneider-Axmann, T., Schneider, M., Wagner, E., Falkai, P., Gaebel, W., Leucht, S., Hasan, A., & Gaigl, G. (under review). Conventional and Living Guideline for schizophrenia - barriers and facilitating factors in guideline implementation.

Martinez García, L. M., Sanabria, A. J., Álvarez, E. G., Trujillo-Martín, M. M., Etxeandia-Ikobaltzeta, I., Kotzeva, A., Rigau, D., Louro-González, A., Barajas-Nava, L., Díaz Del Campo, P., Estrada, M.-D., Solà, I., Gracia, J., Salcedo-Fernandez, F., Lawson, J., Haynes, R. B., & Alonso-Coello, P. (2014). The validity of recommendations from clinical guidelines: a survival analysis. *Cmaj, 186*(16), 1211-1219.

McMillan, S. S., King, M., & Tully, M. P. (2016). How to use the nominal group and Delphi techniques. *International Journal of Clinical Pharmacy*, *38*(3), 655-662. https://doi.org/10.1007/s11096-016-0257-x

Niederberger, M., & Renn, O. (Eds.). (2019). *Delphi-Verfahren in den Sozial- und Gesundheitswissenschaften*. Springer Fachmedien Wiesbaden. https://doi.org/10.1007/978-3-658-21657-3

Nurek, M., Rayner, C., Freyer, A., Taylor, S., Järte, L., MacDermott, N., & Delaney, B. C. (2021). Recommendations for the recognition, diagnosis, and management of long COVID: A Delphi study. *The British Journal of General Practice: The Journal of the Royal College of General Practitioners*, *71*(712), e815-e825. https://doi.org/10.3399/BJGP.2021.0265

OCEBM Levels of Evidence Working Group. *The Oxford Levels of Evidence 2*. CEBM. https://www.cebm.ox.ac.uk/resources/levels-of-evidence/ocebm-levels-of-evidence

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *Systematic Reviews*, *10*(1), 89. https://doi.org/10.1186/s13643-021-01626-4

Pielenz, C./Schneider, M., Salveridou-Hof, E., Flick, M., Gaigl, G., Ghazari, N., Güler, D., Halms, T., Kapfhammer, A., Lorenz, C., Röh, A., Wagner, E., Falkai, P., Leucht, S., Gaebel, W., & Hasan, A. (2022). From conventional to living guidelines–faster updates for better informed guidance? A scoping review. *Zeitschrift für Evidenz, Fortbildung und Qualität im Gesundheitswesen*.

Pereira, V. C., Silva, S. N., Carvalho, V. K., Zanghelini, F., & Barreto, J. O. (2022). Strategies for the implementation of clinical practice guidelines in public health: an overview of systematic reviews. *Health research policy and systems*, *20*(1), 1-21.

Pieper, D., Ober, P., Dressler, C., Schmidt, S., Mathes, T., & Becker, M. (2019). Effizientere Leitlinienerstellung–eine narrative Übersichtsarbeit. *Zeitschrift für Evidenz, Fortbildung und Qualität im Gesundheitswesen*, *146*, 1-6.

Puhan, M. A., Schünemann, H. J., Murad, M. H., Li, T., Brignardello-Petersen, R., Singh, J. A., Kessels, A. G., & Guyatt, G. H. (2014). A GRADE Working Group approach for rating the quality of treatment effect estimates from network meta-analysis. *Bmj*, *349*.

Sanabria, A. J., Pardo-Hernandez, H., Ballesteros, M., Canelo-Aybar, C., McFarlane, E., Niño de Guzman, E., Penman, K., Posso, M., Roqué I Figuls, M., Selva, A., Vernooij, R. W. M., Alonso-Coello, P., & Martínez García, L. (2020). The UpPriority tool was developed to guide the prioritization of clinical guideline questions for updating. *Journal of Clinical Epidemiology*, *126*, 80-92. https://doi.org/10.1016/j.jclinepi.2020.06.018

Schünemann, H., Brozek, J., Guyatt, G., & Oxman, A. (2013). *GRADE Handbook*. https://gdt.gradepro.org/app/handbook/handbook.html#h.dce0ghnajwsm [Zugriff 28.03.2022]

Scottish Intercollegiate Guideline Network. SIGN. https://www.sign.ac.uk/what-we-do/methodology/checklists/ [Accessed on 28.03.2022]

Shea, B. J., Reeves, B. C., Wells, G., Thuku, M., Hamel, C., Moran, J., Moher, D., Tugwell, P., Welch, V., Kristjansson, E., & Henry, D. A. (2017). Amstar 2: A critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. *BMJ (Clinical Research Ed.)*, *358*, j4008. https://doi.org/10.1136/bmj.j4008

Sterne, J. A. C., Savović, J., Page, M. J., Elbers, R. G., Blencowe, N. S., Boutron, I., Cates, C. J., Cheng, H.-Y., Corbett, M. S., Eldridge, S. M., Emberson, J. R., Hernán, M. A., Hopewell, S., Hróbjartsson, A., Junqueira, D. R., Jüni, P., Kirkham, J. J., Lasserson, T., Li, T., & Higgins, J. P. T. (2019). Rob 2: A revised tool for assessing risk of bias in randomised trials. *BMJ (Clinical Research Ed.)*, *366*, l4898. https://doi.org/10.1136/bmj.l4898

Vernooij, R. W. M., Alonso-Coello, P., Brouwers, M., & Martínez García, L. (2017). Reporting Items for Updated Clinical Guidelines: Checklist for the Reporting of Updated Guidelines (CheckUp). *PLoS Medicine*, *14*(1), e1002207. https://doi.org/10.1371/journal.pmed.1002207

Waldorff, F. B., Almind, G., Mäkelä, M., Møller, S., & Waldemar, G. (2003). Implementation of a clinical dementia guideline. A controlled study of the effect of a multifaceted strategy. *Scandinavian journal of primary health care*, *21*(3), 142-147. 10.1080/02813430310005136

Yepes-Nuñez, J. J., Li, S. A., Guyatt, G., Jack, S. M., Brozek, J. L., Beyene, J., Murad, M. H., Rochwerg, B., Mbuagbaw, L., Zhang, Y., Flórez, I. D., Siemieniuk, R. A., Sadeghirad, B., Mustafa, R., Santesso, N., & Schünemann, H. J. (2019). Development of the summary of findings table for network meta-analysis. *Journal of clinical epidemiology*, *115*, 1-13.

Zipfinger, S. (2007). Computer-aided Delphi – An experimental study of comparing round-based with real-time implementation of the method (Linz: Austria).

Appendix

Glossary

Term	Definition
AGREE II	**A**ppraisal of **G**uidelines for **Re**search and **E**valuation; tool for evaluating the methodological quality of guidelines
AMSTAR	**A** **M**ea**s**urement **T**ool to **A**ssess Systematic **R**eviews – Checklist II; checklist for methodological assessment for systematic reviews with and without meta-analyses
Body of evidence	GRADE approach: Assessment of the totality of the studies as a 'body of evidence' for each relevant endpoint
Consensus group	Made up of representatives from the participating professional societies and organizations – the only individuals who are eligible to vote in the voting process and consent on the revised recommendations in the Delphi rounds
Consensus phase	The third and final phase in the living guideline preparation process in which the Delphi process for achieving consensus and the writing up of the amendments take place
Constitutive phase	The first phase in the living guideline preparation process in which, among others, the working process is adopted and the interests are elicited and assessed
Delphi technique	Multistage survey method for achieving consensus carried out in writing between experts from different disciplines
Evidence group	Consists of methodologists commissioned by the *steering group* or the professional society publishing the guideline to perform the methodological work in the guideline process (including systematic literature search)
Evidence phase	Second phase in the living guideline preparation process (months 3 to 8) in which the systematic literature search and appraisal of evidence take place
Evidence-to-decision framework	Continuation of the GRADE process in which individual key criteria are underpinned with evidence for preparation of a guideline recommendation. The evidence group subsequently votes on these criteria
GRADE	**G**rading of **R**ecommendations, **A**ssessment, **D**evelopment and **E**valuation; tool for assessing the quality of evidence and strength of recommendation
Levels of evidence	The overall quality of the evidence is determined on the basis of all critical endpoints. Classification of levels of evidence: High, moderate, low, very low
Living guideline	Ongoing guideline project in which recommendations are updated at short intervals (at least once a year)
Living guideline group	Composed of members of the *steering group* and experts: tasks include inspecting the literature screened by the *evidence group* and preparation of proposed changes
Living systematic review (LSR)	A systematic review which is updated on an ongoing basis and includes relevant new findings as soon as they become available

MAGICapp	**Ma**king **G**RADE the **I**rresistible **C**hoice; digital online platform for structured management and dissemination of a (living) guideline
OCEBM	**O**xford **C**entre for **E**vidence-**B**ased **M**edicine; standardized tool for assessing the quality of evidence and strength of recommendation
Online Delphi	Multistage online survey for achieving consensus
PICO	**P**atient/**P**opulation, **I**ntervention, **C**omparison, **O**utcomes; scheme used to narrow the scope of clinically relevant questions (for recommendations)
Real-time Delphi	Online Delphi consensus method in which the responses of the participants are provided in real time and no iterated rounds are required
ResearchRabbit	'Citation-based literature mapping tool' available online that aims to enable an optimized literature search using e-mail alerts and visualization of the literature review
Screening frequency	Frequency with which systematic searches for new literature are performed
SIGN	**S**cottish **I**ntercollegiate **G**uidelines **N**etwork; standardized tool for assessing quality of evidence and strength of recommendation
Steering group	Steering and coordination of the entire guideline process; members are also part of the *living guideline group*
Summary-of-findings table	Table summarizing the information obtained from the assessment of evidence in a clearly laid out format that can aid in the subsequent decision-making process
Update frequency	Frequency with which a living guideline is updated

Guidance for producing a living systematic review (LSR)

Elliott et al. (2017) have defined a *living systematic review* (*LSR*) as follows: 'A systematic review that is continually updated, incorporating relevant new evidence as it becomes available.'
The section below presents brief guidelines for creating a *living systematic review* (Living Evidence Network, 2019).

Please consult the following publication for a more detailed description:

> Living Evidence Network (2019). Guidance for the production and publication of Cochrane living systematic reviews: Cochrane Reviews in living mode. Available: 201912_LSR_Revised_Guidance.pdf (cochrane.org) [Retrieved on 23.03.2022]

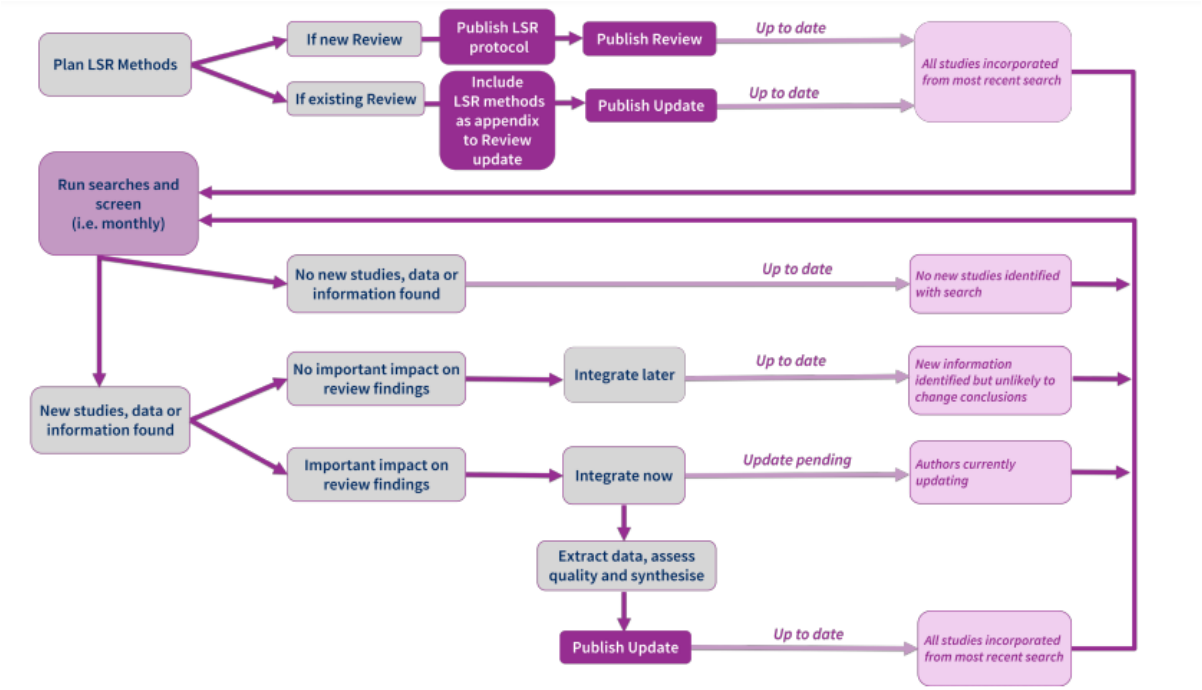


FIGURE 1. COCHRANE LSR WORKFLOW, WITH PUBLICATION OUTPUTS (taken from *Guidance for the production and publication of Cochrane living systematic reviews: Cochrane Reviews in living mode,* 2019, p. 14)

1. Publish an LSR protocol (should contain information such as search methods and frequency). The guidance recommends determining a pool of peer reviewers at the protocol planning stage who must agree on the LSR process and the protocol.
2. Publication of a baseline review (recommended: in the Cochrane Database of Systematic Reviews)
3. Conversion of the baseline review into living mode and production of the LSR
4. Ongoing screening and literature search (recommended: monthly screening of 'key databases'; less frequent screening of other sources; inclusion of an 'information specialist')
   - **Information specialist:**
     - Protocol: estimate the likely scope of monthly search results and advise on the suitability of technological enablers
     - Set up monthly searches and establish comprehensive workflows for processing search results.

- Manage monthly searches, identify and remove duplicates, apply any technological enablers, prepare search results for the author team, maintain accurate search records, update PRISMA flow diagram
- Annual review: reassess the appropriateness of the search methods and ensure that any amendments to the scope of the review are reflected in the search.

5. Continual updating of the LSR status (recommended: readers are informed about the status on a monthly basis; use of 'What's New table' so that the LSR does not need to be republished every month)
6. Peer review: a full review is required for the very first publication. For subsequent updates the guidance recommends: selective peer review of only those parts of the LSR that have changed from the previous version (saves time and helps accelerate updates)
7. Recommended: LSR 'screening': pre-publication quality assurance programme for checking the quality of the protocol and review against key criteria. The key criteria for ensuring the quality of the protocol and the review should be determined in advance. The protocol and review should then be regularly checked against these key criteria to ensure compliance.
8. Recommended: after an LSR has been in living mode for 12 months, a review of the following aspects is recommended: appropriateness of leaving the review in living mode, methods, author team, peer reviewers
9. Recommended: use of tools for quicker literature searches (such as ResearchRabbit, https://www.researchrabbit.ai/)

The *UpPriority Tool*

(Sanabria et al., 2020)

The following sections are adapted by the original version and in consent with the authors. For the original version please see https://doi.org/10.1016/j.jclinepi.2020.06.018 or contact ajsanabria@cochrane.es

Four recommendations for the use of the UpPriority Tool:

1. Organisations should determine the frequency of clinical question assessment based on available resources. In terms of living guidelines this must be done at least once a year.

2. The suggested process for applying the tool is as follows:

a) **Mapping of the clinical guideline (CG)**: The CG should be mapped before applying the Up Priority Tool. This process starts with identifying the clinical questions developed in the original CG. Each clinical question should then be linked to its respective recommendations and the supporting references. The process should also includes the compilation of the original literature search strategies, evidence syntheses, Summary of Findings tables, and Evidence to Decision frameworks, if available.
b) **Developing of the priority survey**: Online software can be used to design the survey and collect responses. The survey should include clinical questions, recommendations, references, and priority items.
c) **Assessing clinical questions according to six priority items**: We suggest assessing clinical questions according to six priority items. The "Priority items" section provides a description of each item, guidance on "where to look" and "how to rate" each item, as well as examples to facilitate the users' assessment.
d) **Calculating and ranking the priority scores**: We suggest different priority scores to support decision-making for updating clinical questions within a CG. The "Priority scores" section provides guidance for calculating and ranking scores.
e) **Deciding on prioritised clinical questions for updating**: Based on ranking priority scores, we suggest a consensual, contextualised, and justified decision of which clinical questions should be prioritised for updating.
f) **Developing a priority report**: We suggest a presentation format to communicate results of the prioritisation process. The "Priority report" section provides a description of what information should be considered in this report.

3. The authors of the UpPriority Tool suggest that the original guideline development group assess the clinical questions within a CG. If not possible, we suggest that appraisers are involved in the CG topic. We suggest that at least four appraisers assess each CG. All appraisers should update their conflict of interest disclosure form before and, if needed, during the prioritisation process. The management of conflicts of interests have to follow the same policy as in the CG developing process.

4. The UpPriority Tool was developed to assess any clinical question within a CG. The authors suggest assessing each CG as a whole, including all clinical questions.

The UpPriority Tool can be adapted according to users' needs and resources (for further information see table 2 of the original version of the UpPriority Tool).

Item	Explanation	Strongly disagree (lower priority for updating)	Uncertainty	Strongly agree (higher priority for updating)
Item 01 Impact of out-dated recommendations on safety	Evaluate whether potentially outdated recommendations have any implications on safety in the current CG healthcare context (e.g. safety alerts). For example, if followed, even in error, the recommendations have the potential to cause harm to patients. Medicines and healthcare products alerts could be identified with a literature search or without it (e.g. clinical expert input).	Following a potentially outdated recommendation is unlikely to result in harm to patients. The clinical question will be annotated with information about the review process (e.g. "under review").	There is uncertainty about whether outdated recommendations have any implications on safety of health care.	Following a potentially outdated recommendation is likely to result in harm to patients. The clinical question and its recommendations must be withdrawn until they will be reviewed and, if necessary, modified. For example, if followed, even in error, the recommendations have the potential to cause harm to patients.
Item 02 Availability of new relevant evidence	Assess the availability of new relevant evidence related to the clinical question and recommendations. New relevant evidence could be identified: 1) with a literature search, 2) without a literature search (e.g. clinical expert input), or 3) with a literature search and clinical expert input. New relevant evidence should be related to new studies that adequately address the clinical question (e.g.	There is no new evidence related to the clinical question and/or recommendations, or there is new evidence but it does not have an impact on current recommendations. The clinical question does not need to be updated.	There is uncertainty about the availability of new evidence related to the clinical question and/or recommendations, or there is uncertainty about its impact on current recommendations.	There is new evidence that may modify the clinical question and/or recommendations.

	systematic reviews, randomized controlled trials, or observational studies). Potential changes may be related to clinical questions (patients, intervention, comparisons, or outcomes), factors that influence the formulation of recommendations (e.g. quality of the evidence, balance between benefits and harms, values and preferences, use of resources and costs), or recommendations			
Item 03 Context relevance of the clinical question	Review if the clinical question is still supported by factors of interest (e.g. prevalence and burden of disease or variation in clinical practice in the current CG healthcare context.	The clinical question is not relevant to current practice. The clinical question does not need to be updated and/or should be removed.	There is uncertainty about the relevance of the clinical question to current practice.	The clinical question is still relevant to current practice.
Item 04 Methodological applicability of the clinical question	Review if the clinical question still addresses components of interest (population, intervention, comparison, and outcomes) in the current CG healthcare context. Consider whether: • The original literature search will be useful to identify new evidence. • The availability of new relevant evidence (Item 02) modifies one or more components of the clinical question.	There are new populations, interventions, comparisons, or outcomes that are not covered by the current clinical question. The clinical question needs to be developed de novo.	There is uncertainty about whether the clinical question still addresses components of interest.	The clinical question still addresses the components of interest (population, intervention, comparison, and outcomes). The original literature search will be useful to identify new evidence.

	• The clinical question needs to include a subgroup of patients. • The clinical question needs to include patient-reported outcomes. • The clinical question needs to include implementation considerations.			
Item 05 Users' interest	Estimate the current interest (e.g. citations, downloads, news, debate, or website visits) of patients, health care providers, healthcare system, or other stakeholders related to the clinical question and recommendations.	The clinical question and recommendations are not considered an influential topic to current practice.	There is uncertainty about the interest on behalf of patients, health care providers, healthcare system, or other stakeholders regarding the clinical question and recommendations.	There is a growing interest on behalf of patients, health care providers, or other stakeholders regarding the clinical question and recommendations.
Item 06 Impact on access to health care	Evaluate whether the recommendations have any implications on access and coverage in the current CG healthcare context. For example, inclusion of health treatments in a National Health Service, out of country requests, or access to orphan drugs.	The recommendations are not legally binding to funding decision and do not have an impact on access and coverage to health care.	There is uncertainty about whether the recommendations have any implications on access and coverage to health care.	The recommendations are legally binding to funding decision and may have an impact on access and coverage to health care.

Note. Items of the UpPriority Tool. Ratings on a 7-point-likert scale. Focus on the three anchor points for clarity. Please see original version for more information.

Procedure:
1) Select priority items to be used.
2) select clinical questions to be assessed
3) Create table (original table from UpPriority Tool):

	Item 1	Item 2	Item 3	Item 4	Item 5	Item 6	Overall
Appraiser 1	7	4	4	4	7	7	33
Appraiser 2	6	6	6	6	6	6	36
Appraiser 3	7	7	7	7	7	7	42
Appraiser 4	4	7	7	7	6	7	38
Item scores	6	6	6	6	6.5	6.75	
Priority score							**37.25**
Standard deviation	1.41	1.41	1.41	1.41	0.58	0.50	**3.77**
Range	1-7	1-7	1-7	1-7	1-7	1-7	6-42

4) In this table, the appraisers' scores are given for each clinical question and priority item.
5) See Assessment & Evaluation
6) Graphical representation of the results, if applicable.
7) Priority report
Notes on the presentation of the priority results (for more detailed information, see the UpPriority Tool Guideline, p. 21 onwards):

The final priority report should include the following information:

a. Executive summary;
b. Objectives;
c. Methods;
d. List of clinical questions;
e. ranking of priority scores;
f. Priority decision; and
g. Explanation for the priority decision.

Scoring and evaluation:
Priority items are scored on a 7-point Likert scale.
1 = Strongly disagree (low priority for updating),
4 = Unclear/Uncertain,
7 = Strongly agree (high priority for updating).

For each clinical question, an average score should be calculated per item (item score) and overall (priority score).
We recommend imputing a score of “four” for items where data were missing.

Score	Description	Calculation
Item score	Calculated by adding the values of each item given by each appraiser and dividing these by the number of appraisers.	$\text{Item score} = \frac{IsA1+IsA2+\cdots+IsAn}{An}$
Standard deviation for item score	Square root of the average of the squared deviations of the values of each item for their average value.	$s=\sqrt{\frac{\Sigma(x-\bar{x})^2}{n-1}}$
Range of item score	Minimum and maximum possible score.	Minimum = 1 (strongly disagree) Maximum = 7 (strongly agree)
Priority score	Calculated by adding the values of all items given by each appraiser and dividing these by the number of appraisers.	$\text{Priority score} = \frac{PsA1+PsA2+\cdots+PsAn}{An}$
Standard deviation for priority score	Square root of the average of the squared deviations of the values of each score for their average value.	$s=\sqrt{\frac{\Sigma(x-\bar{x})^2}{n-1}}$
Range of priority score	Minimum and maximum possible score.	Minimum = 1 (strongly disagree) * n items Maximum = 7 (strongly agree) * n items

Abbreviations: A: Appraisers; Is: Item score; n: Number of values in data set; Ps: Priority score; s: Standard deviation; x: Each value in the data set; $\bar{x}$: Mean of all values in the data set.

Ranking priority scores:

The clinical questions should be sorted by priority scores (higher to lower) followed by standard deviation (lower to higher).

We suggest presenting the results in tabulated and graphical formats.

Priority scores can be useful for comparing clinical questions and can inform whether a clinical question should be prioritised for updating. Item scores can be useful to adjust decision-making for clinical questions with similar priority scores.

Particular attention must be paid to clinical questions with a high score in “Item 03 - Context relevance” and low score in “Item 04 - Methodological applicability”. These clinical questions will be still context relevant but not methodologically applicable for updating. We suggest assessing these clinical questions as new clinical questions.

Priority decision:

The priority decision should be consensual, contextualised, and justified. We suggest considering the following factors:

- Item scores and priority scores
- Available resources
- Expected volume of new evidence
- Development of new clinical questions

Currently, the Up Priority Tool does not include thresholds to classify clinical questions according to their priority for updating (e.g. high, medium or low relevance for updating).