

**SEMMELWEIS EGYETEM
DOKTORI ISKOLA**

Ph.D. értekezések

3534.

BLIN NAGAVCI

Klinikai hematológia és infektológia
című program

Programvezető: Dr. Masszi Tamás, egyetemi tanár
Témavezető: Dr. Lakatos Botond Tibor, egyetemi tanársegéd

THE SEMMELWEIS FRAMEWORK: STRUCTURING EVIDENCE FOUNDATIONS AND EXPERT OPINION IN CLINICAL PRACTICE GUIDELINES

PhD thesis

Blin Nagavci

Semmelweis University, Doctoral College
Károly Rácz Conservative Medicine Division



Supervisor: Botond Lakatos, MD, PhD

Official reviewers: Gergely Peskó, MD, PhD
István Klemencsics, MD, PhD

Head of the Complex Examination Committee: Tamás Masszi, MD, PhD, DSc

Members of the Complex Examination Committee: Gergely Peskó, MD, PhD
Marienn Réti, MD, PhD

Budapest
2026

Contents

1. INTRODUCTION	6
2. OBJECTIVES.....	8
2.1. WP 1: Mapping the use of expert opinion in infectious diseases clinical guidelines.....	8
2.2. WP 2: Defining expert opinion in clinical guidelines	8
2.3. WP 3: The Semmelweis Framework for structuring evidence foundations and expert opinion in clinical practice guidelines	8
3. METHODS.....	9
3.1. WP 1: Utilisation of Expert Opinion in Infectious Diseases Clinical Guidelines - A Meta-Epidemiological Study.....	9
3.2. WP 2: Defining Expert Opinion in Clinical Guidelines: Insights from 98 Scientific Societies – A Methodological Study.....	10
3.3. WP 3: The Semmelweis Framework for structuring evidence foundations and expert opinion in clinical practice guidelines	11
4. RESULTS.....	15
4.1. WP 1: Utilisation of Expert Opinion in Infectious Diseases Clinical Guidelines - A Meta-Epidemiological Study.....	15
4.2. WP 2: Defining Expert Opinion in Clinical Guidelines: Insights from 98 Scientific Societies – A Methodological Study.....	17
4.3. WP 3: The Semmelweis Framework for structuring evidence foundations and expert opinion in clinical practice guidelines	23
5. DISCUSSION.....	33
6. CONCLUSIONS	36
7. SUMMARY	37
8. REFERENCES	38
9. BIBLIOGRAPHY OF THE CANDIDATE'S PUBLICATIONS	41

10.	ACKNOWLEDGEMENTS	43
11.	APPENDIX	44

List of Abbreviations

CoE: Certainty of Evidence

DTA: Diagnostic Test Accuracy

EO: Expert opinion

FAQ: Frequently asked questions

GRADE: Grading of Recommendations Assessment, Development and Evaluation

IDSA-USPHS: Infectious Diseases Society of America–US Public Health Service System

OCEBM: Oxford Centre for Evidence-Based Medicine

PICO: Population, Intervention, Comparator, Outcome

PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses

RCT: Randomised controlled trial

RoB: Risk of bias

WP: Work package

1. INTRODUCTION

Clinical practice guidelines are systematically developed statements that include recommendations intended to support informed healthcare decisions (1). Recommendations are typically informed by systematic reviews of the best available evidence, including randomised and non-randomised studies, depending on the clinical question (2). However, when evidence is scarce, guideline developers might resort to expert opinion for providing clinical guidance (3).

Approximately one quarter of guidelines across different medical fields include recommendations based on expert opinion (4). In critical care, 10% of strong recommendations were based on expert opinion (5), while in cardiology, this proportion reaches 55% (6). In infectious diseases, 47% of strong recommendations were supported by low-quality evidence or expert opinion (7). These findings indicate a recurring practical role for expert opinion in addressing evidence gaps and enabling guideline panels to provide clinical recommendations.

However, despite its widespread use, the role of expert opinion in guideline development remains unstandardised. A meta-epidemiological study found that expert opinion was commonly used to support guideline recommendations, yet 91% of recommendations designated as expert opinion provided no explicit rationale for this classification (4).

There are two main issues encompassing the use of expert opinion in clinical guidelines. First, there is no widely accepted definition of what expert opinion is. Authors define expert opinion differently. For example, some authors describe expert opinion broadly as arising from clinicians' accumulated experiential knowledge and interpretation of multiple published and unpublished sources (8). In contrast, others distinguish expert opinion from expert evidence, defining the former as a judgement or conclusion and the latter as observations or experience supplied by experts (9). These differing interpretations not only illustrate the absence of a standardised definition but also influence how expert opinion is utilised in guideline development.

Second, different evidence-grading systems treat expert opinion in markedly different ways. More traditional systems categorise evidence largely according to study design and place expert opinion within their lowest evidentiary category, including the Oxford

Centre for Evidence-Based Medicine Levels of Evidence (OCEBM) (10), the American College of Cardiology Foundation/American Heart Association system (11), and the European Society of Cardiology system (ESC) (12). In contrast, Grading of Recommendations Assessment, Development and Evaluation (GRADE) does not recognise expert opinion as a distinct entity, but defines it as the process of interpreting evidence through the experience and knowledge of experts (13).

These two approaches are fundamentally different from each other, and both have their limitations. Although traditional grading systems recognise expert opinion as a distinct entity, allowing guideline developers to use it, they often fail to capture the full range of sources and roles on which expert opinion may be based. Conversely, GRADE provides a more advanced and systematic approach to assessing evidence and formulating recommendations, but it may not fully address the practical challenges faced by guideline developers when clinically important recommendations are required despite scarce or absent evidence.

This conceptual and methodological heterogeneity creates practical challenges for both guideline developers and clinicians. For guideline developers, the absence of a shared definition and consistent methodological approach may complicate decisions about when expert opinion is appropriate, how it should be elicited and documented, and how it should be distinguished from other forms of evidence. For clinicians, inconsistent terminology, classification, and grading may make the evidentiary basis of recommendations less clear and hinder assessment of their trustworthiness, potentially influencing how they are interpreted and applied in practice (4, 14).

This research aimed to address these challenges by mapping the use of expert opinion in clinical guidelines, synthesising an evidence-based definition of expert opinion, and developing a structured approach to support its transparent and consistent use in clinical guidelines, while incorporating methodological safeguards.

2. OBJECTIVES

This research project comprised three work packages (WPs), each resulting in a separate published study.

2.1. WP 1: Mapping the use of expert opinion in infectious diseases clinical guidelines

WP1 aimed to construct an analytical map of how expert opinion is used in international infectious diseases guidelines. This included describing its prevalence and application and examining relevant methodological characteristics of the included guidelines (15).

2.2. WP 2: Defining expert opinion in clinical guidelines

WP2 aimed to systematically review the methodological manuals of national and international guideline-developing organisations, examine the rationales and circumstances in which expert opinion is used, identify its underlying foundations, and develop an evidence-informed definition of expert opinion (16).

2.3. WP 3: The Semmelweis Framework for structuring evidence foundations and expert opinion in clinical practice guidelines

WP3 aimed to develop a structured framework to support the transparent and consistent use of expert opinion in clinical practice guidelines. The Framework was designed to classify the evidentiary foundations of recommendations and incorporate safeguards for situations in which direct evidence is limited or absent (17).

3. METHODS

3.1. WP 1: Utilisation of Expert Opinion in Infectious Diseases Clinical Guidelines - A Meta-Epidemiological Study

This section has been reproduced or rephrased from Nagavci et al. (15).

The first study of this research was a meta-epidemiological study, developed and reported in accordance with guidelines for reporting meta-epidemiological research (18). A protocol was developed for methodological consistency (15). Systematic literature searches were conducted in PubMed and Trip Medical Database in May 2023. The search aimed at identifying recent infectious diseases clinical guidelines, published in the last 5 years (search strategy is presented in the supplementary material of the published article) (15). The analysis was restricted to infectious diseases guidelines because the thesis originated from this clinical field, and because infectious diseases recommendations frequently require interpretation across heterogeneous pathogens, populations, interventions, and clinical contexts, making this area particularly relevant for examining the use of expert opinion in guideline development.

One reviewer screened the records in Rayyan, initially by title and abstract and subsequently in full text. Eligible documents were human clinical guidelines addressing infectious diseases, issued by international organisations involving at least two countries, based on primary research, written in English and published from 1 January 2018 onwards. The selection process and reasons for exclusion were documented using the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) flow diagram (19).

One reviewer collected the information using a piloted Excel form. Recorded variables covered bibliographic and organisational characteristics, methodological features such as literature searching, risk-of-bias assessment, evidence appraisal and screening diagrams, and details concerning expert-opinion use, terminology and the number of recommendations. When reporting within a guideline was insufficient, the organisation's website and methodological manual were consulted.

Frequencies were used to describe the occurrence of expert-opinion-based recommendations, while definitions and stated rationales were synthesised narratively.

Normality was examined with the Shapiro–Wilk test. Differences in methodological characteristics between groups were evaluated using the chi-square test and the Mann–Whitney U test, with alpha set at 0.05.

More detailed information on the methods can be found in the published study (15).

3.2. WP 2: Defining Expert Opinion in Clinical Guidelines: Insights from 98 Scientific Societies – A Methodological Study

This section has been reproduced or rephrased from Nagavci et al. (16).

The second study followed established reporting guidance for methodological and meta-epidemiological research (18, 20). An internal protocol was developed to ensure methodological consistency. A systematic process with four steps was used for this study:

Step 1) Systematic literature searches for clinical guidelines

Systematic searches were undertaken in PubMed for clinical guidelines published in the 2019-2024 period, in all fields of medicine (presented in the supplementary material of the article) (16). The retrieved references were screened in title/abstract and full-text by a single reviewer in Rayyan. Only clinical guidelines published in English by national or international organisations were considered for inclusion. Screening results with exclusion reasons were presented using the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) flow diagram (19).

Step 2) Identification of guideline-developing organisations

The included clinical guidelines were checked in detail to identify organisations, mentioned either as developers of the guideline or as endorsing organisations. All mentioned organisations were listed in Excel. This list was further supplemented through additional hand searches in Google, and by adding organisations mentioned in other studies (15, 21, 22).

Step 3) Searching for methodological manuals

For each identified organisation, systematic searches were undertaken to retrieve methodological manuals for guideline development. Two reviewers independently examined official websites and supplemented these searches with a predefined Google

Advanced Search. Retrieved sources were assessed independently, with disagreements settled by discussion or consultation with a third reviewer. Eligible sources were English-language documents, articles or webpages published from 2010 onwards that described an organisation's guideline-development methods. When several editions were available, the latest was retained. The selection process was documented using PRISMA diagrams (19).

Step 4) Evidence synthesis and defining expert opinion

Information was entered into piloted Excel forms by one reviewer and verified by a second. Differences were resolved through discussion and, when necessary, referral to a third reviewer. The findings were integrated narratively and summarised using descriptive measures, including counts and means. All analyses were performed in Excel and independently checked.

For synthesising a definition of expert opinion, an inductive (bottom-up) approach was used, by analysing all included methodological manuals (16).

3.3. WP 3: The Semmelweis Framework for structuring evidence foundations and expert opinion in clinical practice guidelines

This section has been reproduced or rephrased from Nagavci et al. (17).

An international multidisciplinary team of clinicians and guideline methodologists was established to develop the Semmelweis Framework for structuring evidentiary foundations and expert opinion in clinical practice guidelines (17).

3.3.1. Conceptual basis of the Framework

The categories of expert opinion identified in WP2, and the corresponding definition, were used deductively to construct the initial Framework. These categories covered clinical experience, professional judgement, indirect and very-low-certainty evidence, and mechanism-based reasoning (16).

The initial recommendation classes were derived from these foundations. Indirect comparative evidence and non-comparative group-level evidence were treated as separate classes, while clinical experience, mechanism-based reasoning, case reports, case series,

and collective panel judgement were grouped under recommendations without relevant group-level evidence. Direct comparative evidence was also included as a separate class, because the Framework was intended to apply to all clinical recommendations, not only to those commonly labelled as expert opinion.

This resulted in a four-class system:

- Class I: direct comparative evidence;
- Class II: indirect comparative evidence;
- Class III: non-comparative group-level evidence; and
- Class IV: expert judgement in the absence of relevant group-level evidence.

The Framework therefore treats the relationship between evidence and expert opinion as a continuum. Expert input may contribute to all recommendation classes, but its relative importance increases as the supporting evidence becomes less direct, less comparative, or unavailable.

3.3.2. Iterative refinement using published recommendations

The preliminary Framework was refined through qualitative application to recommendations taken from published clinical guidelines. A structured extraction form was developed to capture the source guideline, the original recommendation, the grading system used, the rationale supplied by the guideline authors, and the study designs supporting the recommendation.

The extraction form was first piloted by three reviewers using two sample recommendations from two separate guidelines. The reviewers completed their assessments independently. Their results were then compared during a calibration meeting, after which unclear items and practical problems in the form were revised.

For the full refinement exercise, guideline-developing organisations were sampled from the list identified in WP2 (16). A computer-generated random sequence was used by the lead methodologist to allocate five organisations to each of three reviewers from the five reviewers participating in the project. Each reviewer selected one recent guideline, published during the preceding five years, from every assigned organisation.

Two recommendations were then selected from each guideline. One was chosen because it was explicitly described as expert opinion, consensus-based, or supported by very low-

certainty evidence. The second was selected because it was supported by higher-certainty evidence. This process yielded 30 recommendations from 15 guidelines and ensured representation of different specialties, organisations, clinical topics, grading systems, and evidentiary profiles.

The 30 recommendations were entered into an Excel database, reordered, and randomly allocated across the five reviewers. Each recommendation was assessed independently by three reviewers. Before the calibration meetings, reviewers could not view the assessments or class assignments made by the others.

For every recommendation, reviewers:

- Examined the rationale provided by the original guideline authors;
- Reviewed the cited evidence and identified the underlying evidence;
- Applied the current version of the Semmelweis Framework and assigned a recommendation class, and presented the reason;
- Reformulated the recommendation using the wording and notation proposed by the framework, without changing its intended direction;
- Documented uncertainties, practical difficulties, and proposals for improving the classification system.

The exercise was limited to the information reported in the original guidelines. No new systematic review, effect estimation, risk-of-bias assessment, or certainty-of-evidence reassessment was undertaken.

Assessments were discussed in batches of three to five recommendations. Nine online meetings were held. During each meeting, the recommendation, its evidentiary basis, the reviewers' class assignments, and the proposed wording were compared. Reviewers explained the reasons for divergent judgements and identified ambiguities in the classification criteria or flowchart.

These discussions informed successive revisions of the Framework. The flowchart, class descriptions, terminology, and implementation guidance were modified before the next group of recommendations was assessed. Several intermediate versions were therefore produced. Refinement continued until repeated application no longer identified major conceptual problems and only minor editorial changes remained.

Because the Framework was revised during the exercise, the same version was not applied to all recommendations. Formal agreement statistics, such as Cohen's or Fleiss' kappa, were therefore not calculated, as they would not have reflected the reliability of a fixed instrument. The exercise was developmental in purpose, and agreement, usability, and areas requiring clarification were assessed qualitatively during calibration discussions.

3.3.3. Development of the Preconditions Checklist

A Preconditions Checklist was developed in parallel with the classification system. Initial items were informed by the findings of WP2 and by the main methodological concerns associated with issuing recommendations when direct comparative evidence is limited or absent.

The Checklist was revised throughout the iterative refinement process. Problems encountered during recommendation assessment, such as unclear evidentiary foundations, insufficient justification for issuing guidance, unreported assumptions, or inadequate documentation of panel reasoning, were used to identify missing items or improve existing wording.

The final Checklist contains 11 questions grouped into four domains:

1. Evidence;
2. Necessity;
3. Patient and research impact; and
4. Panel integrity.

The Checklist was considered complete when further application did not reveal any major conceptual omissions and only minor wording adjustments were required.

Further details on the development process are provided in the WP 3 publication and its supplementary materials (17).

4. RESULTS

4.1. WP 1: Utilisation of Expert Opinion in Infectious Diseases Clinical Guidelines - A Meta-Epidemiological Study

Systematic literature searches identified 2,912 records, of which 178 were assessed in full text. Sixty-six clinical practice guidelines, containing a total of 2,296 recommendations, met the eligibility criteria and were included in the study (Figure 1). These guidelines were published or endorsed by 136 distinct societies or organisations (15).

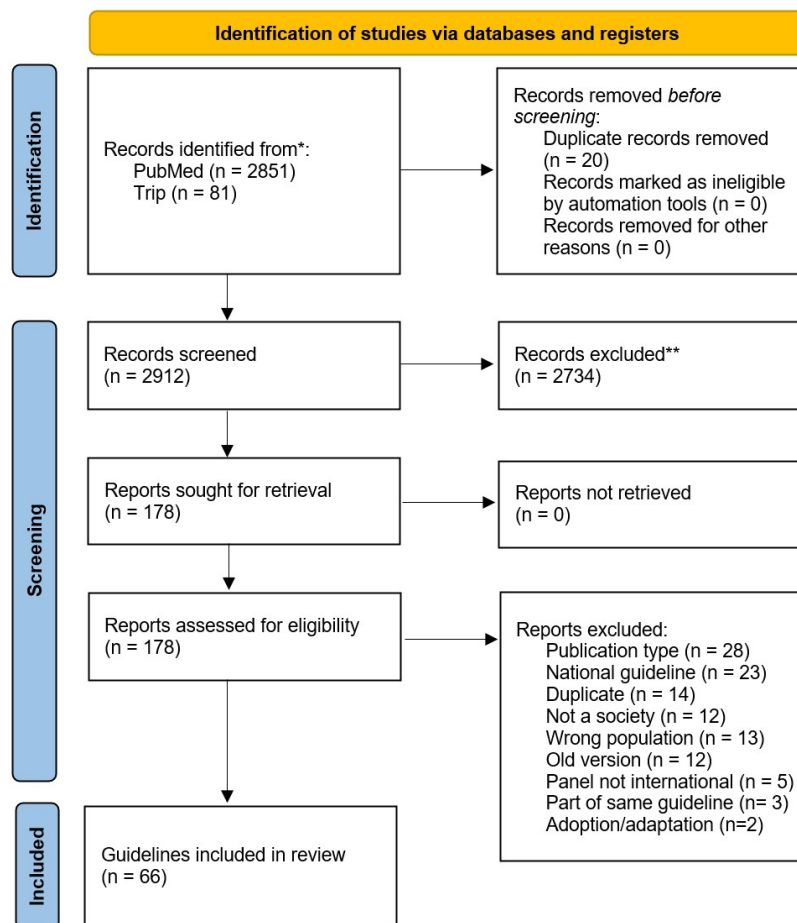


Figure 1. PRISMA flow diagram for screening of clinical guidelines

Source: reproduced from Nagavci B, Schwingshackl L, Martin-Loeches I, Lakatos B. Utilization of expert opinion in infectious diseases clinical guidelines-A meta-epidemiological study. *PLoS One*. 2024;19(6):e0306098. Used under the Creative Commons Attribution (4.0) license.

Of the 66 included guidelines, 32 (48.5%) reported allowing or using expert opinion in the formulation of recommendations. Conversely, 23 (34.8%) did not report allowing or using expert opinion, while the remaining 11 (16.7%) provided insufficient information to determine its status, mainly because of missing methodological information or the absence of evidence ratings for recommendations (15).

Sixteen of the 32 guidelines placed expert opinion in the same evidentiary category as low- or very low-quality evidence, whereas 14 either separated expert opinion from other forms of evidence or used systems in which it was not incorporated into the evidence hierarchy (Figure 2).

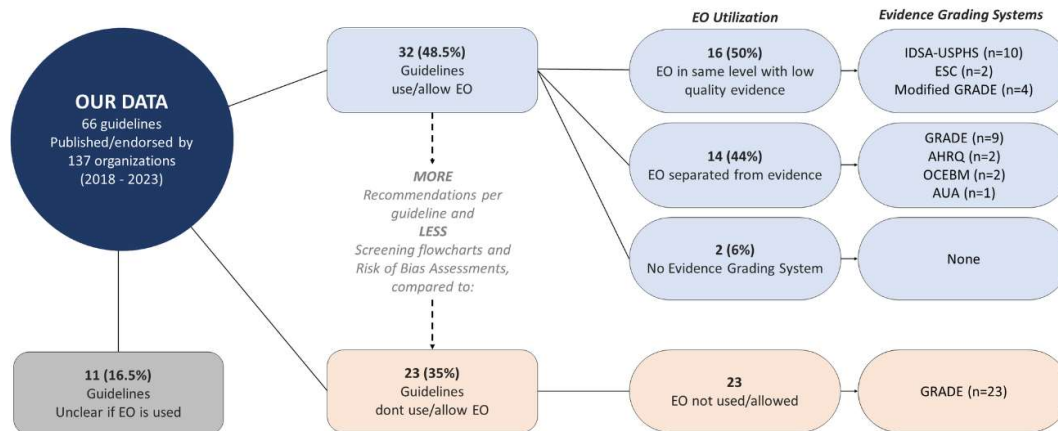


Figure 2. Visual description of expert opinion utilisation and evidence grading systems

EO: Expert Opinion, IDSA-USPHS: The Infectious Diseases Society of America-US Public Health Service, ESC: European Society of Cardiology, GRADE: Grading of Recommendations, Assessment, Development, and Evaluations, AHRQ: Agency for Healthcare Research and Quality, OCEBM: Oxford Centre for Evidence-Based Medicine, AUA: American Urological Association.

Source: reproduced unchanged from Nagavci B, Schwingshackl L, Martin-Loeches I, Lakatos B. Utilization of expert opinion in infectious diseases clinical guidelines-A meta-epidemiological study. *PLoS One*. 2024;19(6):e0306098. Used under the Creative Commons Attribution (4.0) license.

The 32 guidelines that allowed or used expert opinion contained 1,465 recommendations. Of these, 448 (30%) recommendations were classified either as being based on expert opinion or as being supported by an evidence category that included expert opinion among its foundations (e.g. “consensus of opinion of the experts and/or small studies, retrospective studies, registries”).

Regarding the terminology, in the 32 guidelines which allowed expert opinion, seven different terms for expert opinion were identified: "Opinions of respected authorities," "Expert opinion," "Expert judgment", and "Consensus of expert opinion", followed by "Consensus recommendation", "Good practice principle", and "In our practice statement" (15).

Several statistically significant methodological differences were identified between guidelines that allowed expert opinion and those that did not. Guidelines allowing expert opinion issued a higher mean number of recommendations than guidelines not allowing it (48.82 vs 19.13 recommendations per guideline; $p<0.001$). They were also less likely to present screening flow diagrams (25% vs 65%; $p=0.002$) or report conducting risk-of-bias assessments (19% vs 78%; $p<0.001$). No statistically significant differences were observed in the reporting of literature searches, electronic databases, search strategies, or grading systems (15).

The use of a grading system to assess the certainty of evidence was reported in 94% of the included guidelines. Among these, 47% used GRADE, followed by the Infectious Diseases Society of America-US Public Health Service System (IDSA-USPHS) with 16%, while 10.6% used different modifications of GRADE. Two guidelines did not use an evidence-grading system (15).

Further methodological details and results are available in the published study corresponding to WP1 (15).

4.2. WP 2: Defining Expert Opinion in Clinical Guidelines: Insights from 98 Scientific Societies – A Methodological Study

The systematic literature searches resulted in 692 guideline-developing organisations. In addition, 136 organisations were identified from previous work (15), 28 from hand searches, and five from other sources (21, 22). After deduplication, 473 unique

organisations were finally included and systematically searched for methodological manuals. From these searches, 98 methodological manuals were identified and included in the analysis. (Figure 3).

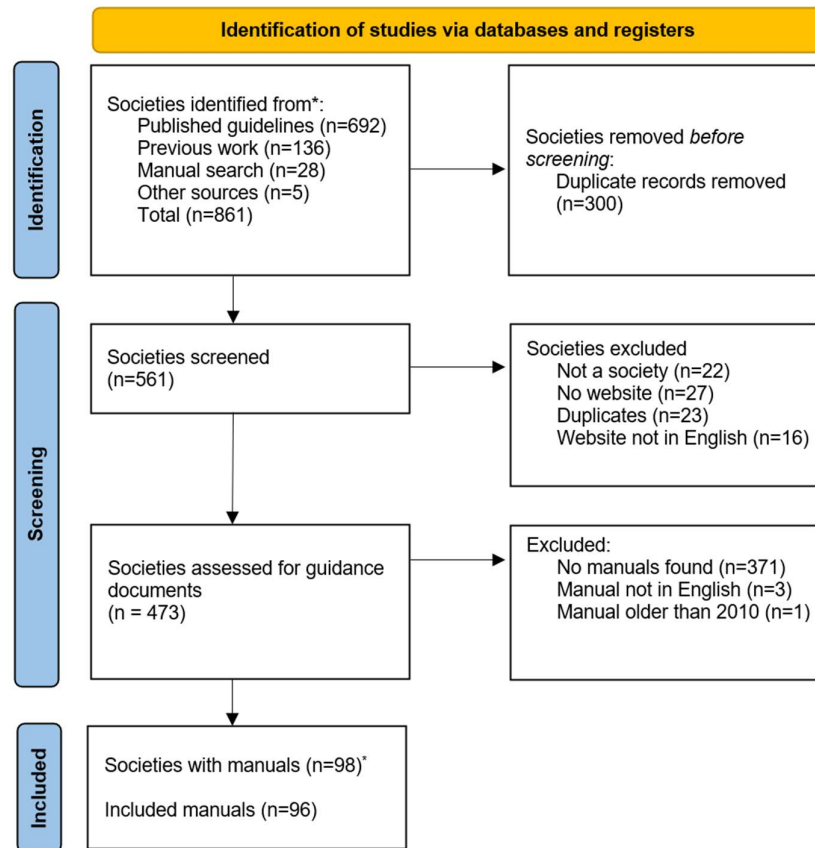


Figure 3. PRISMA flow diagram screening of methodological manuals

*Two guideline-developing organisations use the same manual.

Source: reproduced unchanged from Nagavci B, Gáspár Z, Lakatos B. Defining expert opinion in clinical guidelines: insights from 98 scientific societies - a methodological study. *BMC Med Res Methodol.* 2025;25(1):87. Used under the Creative Commons Attribution (4.0) license.

These manuals consisted of 61 handbooks, 29 journal articles, and 8 websites (16). Expert opinion was mentioned in 65 of them, with high variations in the level of detail regarding the utilisation, criteria and terminology. In 30 manuals expert opinion was not mentioned

at all, while in 3 it was mentioned but without sufficient information to determine whether it is utilised or not (16).

Among the 65 societies whose manuals mentioned expert opinion, two principal rationales for its use were identified:

1) Filling evidence gaps:

Expert opinion was most commonly used as a means to fill evidence gaps, with 47 (72%) organisations using it in this context. These organisations utilised expert opinion when reliable data are unavailable, ensuring that actionable guidance can still be provided. For example, (16):

- American Urological Association (AUA) uses expert opinion as "a statement achieved by panel consensus...for which there is no published evidence."
- The Scottish Intercollegiate Guidelines Network (SIGN) mentions that "if the group feels strongly that they want to make a recommendation even though there is no significant evidence, this should be done as a weak recommendation".
- The National Institute for Health and Care Excellence (NICE) also highlights the use of expert opinion when "there are gaps in the evidence base or subgroups are under-represented."
- The American College of Obstetricians and Gynecologists (ACOG) employs expert opinion in the form of Ungraded Good Practice Statements, which are incorporated "when a practice point is deemed necessary in the case of extremely limited or non-existent evidence".
- American Academy of Neurology (AAN) acknowledges that "compelling inferences alone can support practice recommendations without evidence".
- The Centers for Disease Control and Prevention (CDC) develops interim guidelines based on expert opinion stating that "...CDC developed these guidelines using either expert opinion or indirect or emerging evidence, and the recommendations might change when more and better evidence becomes available".
- European Society of Clinical Microbiology and Infectious Diseases (ESCMID) states: "We understand that there might be a clinical need for recommendations

even when published evidence is insufficient. In such cases recommendations should be provided with explicit acknowledgment that they are based on expert opinion”.

2) Interpretation of evidence

In this context expert opinion served to contextualise and synthesise the existing evidence. This use of expert opinion was observed in five organisations, with three main examples (16):

- The Infectious Diseases Society of America (IDSA) highlights that "Expert opinion is not categorised in any of the above classification (i.e., not a level of quality of evidence), but may be critical to interpret studies included in the systematic review".
- U.S. Preventive Services Task Force (USPSTF) mentions “The Task Force is particularly interested in receiving comments on the sufficiency of the systematic review process and interpretation of the body of evidence. However, expert opinion and clinical experience cannot substitute for the body of evidence that the Task Force reviews through a systematic process”.
- The European Reference Network on Rare Endocrine Conditions (Endo-ERN) states that "Expert opinion represents an interpretation of evidence in the context of experts' experiences and knowledge. An expert opinion may be based on the interpretation of studies ranging from uncontrolled case series to randomized controlled clinical trials, thus it is important to describe what type of evidence is being used as the basis for interpretation".

A similar approach is also followed by American Physical Therapy Association (APTA) and Society of American Gastrointestinal and Endoscopic Surgeons (SAGES).

For 13 other organisations it was not possible to identify how exactly expert opinion was utilised, as no details were provided (16).

4.2.1. Foundations of expert opinion

Most of the guideline-developing organisations used more than one distinct foundation for issuing expert opinion recommendations. While there was a significant variability

between organisations, there was also a considerable overlap between them. Based on the assessed data, five main foundations for expert opinion could be identified (16):

1) Expert opinion stemming from clinical experience

This foundation relies on clinical expertise, judgement, and practical experience of panel members developing the guideline. For example (16):

- The American Urological Association (AUA) defines expert opinion as statements based on the clinical training, experience, and judgement of panel members.
- The Heart Rhythm Society (HRS) relies on expert consensus derived from clinical experience and standard of care.
- American Academy of Orthopaedic Surgeons (AAOS) mentions clinical opinion for making statements in such cases.

2) Expert opinion stemming from indirect evidence

When direct (high) quality evidence is unavailable, indirect evidence is used as the basis for expert opinion. For example (16):

- The European Society of Clinical Microbiology and Infectious Diseases (ESCMID) acknowledges expert opinion when evidence is insufficient, extending empirical evidence to related interventions when appropriate.
- Centers for Disease Control and Prevention (CDC) uses either expert opinion or indirect or emerging evidence for developing interim guidelines.

3) Expert opinion stemming from low-quality evidence

This foundation relies on low-quality evidence such as small observational studies, case series or case reports. For example (16):

- The World Health Organization (WHO) utilises expert opinion drawn from case reports and individual or national experiences in the absence of rigorous research evidence.
- The American College of Cardiology (ACC) and American Heart Association (AHA) derive expert opinion from opinion of experts, case studies, and standard of care.

- The Society for Vascular Surgery (SVS) states that committees occasionally have provided guidance using their clinical experience, unsystematic observations, and best interpretation of the low-quality evidence available.
- The American Association of Respiratory Care (AARC) considers indirect evidence from case studies or low-quality data.

4) Expert opinion stemming from mechanism-based reasoning or theoretical rationale

This foundation uses principles from physiology, theoretical models, or established scientific rationale to support recommendations. Several organisations have explicitly presented this as a distinct foundation (16):

- The American Academy of Neurology (AAN) describes expert opinion as generally accepted principles of care and inferences drawn from known principles of the course of the disease and Bayes's theorem.
- The Society for Healthcare Epidemiology of America (SHEA) integrates expert opinion based on limited evidence, theoretical rationale, current practices and practical considerations.
- The Society of Interventional Radiology (SIR) includes expert opinion based on physiology, bench research, or first principles.
- The Academy of Nutrition and Dietetics uses expert opinion based on usual practice, expert consensus, clinical experience, opinion, or extrapolation from basic research.

5) Expert opinion stemming from expert evidence and witnesses

When evidence is absent, this foundation uses experience in a structured way, with formulated questions or forms, or by involving external experts who provide their experience in a form of a testimony. For example (16):

- The American Society of Hematology (ASH) uses systematically collected expert evidence in areas in which published evidence is insufficient.
- The National Institute for Health and Care Excellence (NICE) incorporates expert opinion from expert witnesses to address gaps in the evidence base.

4.2.2. The proposed definition of expert opinion

Based on the analysis of the assessed methodological manuals of guideline-developing organisations, and using an inductive approach, a new definition of the expert opinion concept was developed:

"The concept of expert opinion in clinical guidelines refers to the synthesis of guidance derived from clinical experience, expert judgment, indirect evidence, very low-quality evidence or mechanism-based reasoning, aimed at supporting clinical decision-making in the absence of evidence or when inconclusive evidence requires contextualization and interpretation" (16).

Further methodological details and results are available in the published study corresponding to WP2 (16).

4.3. WP 3: The Semmelweis Framework for structuring evidence foundations and expert opinion in clinical practice guidelines

4.3.1. Overview

The Framework was named after Semmelweis University, where the work originated. It was developed as an umbrella system for classifying individual recommendations according to the evidentiary foundation that principally supports them.

The Semmelweis Framework differs from systems that classify evidence mainly according to study design or certainty. Its focus is whether the available body of evidence directly and appropriately supports the recommendation being issued. It therefore adds an explicit description of the evidentiary foundation to the usual assessment of certainty and recommendation strength.

The Framework is intended to be used alongside established guideline methods, particularly GRADE. It preserves the usual systematic review, certainty assessment, and evidence-to-decision processes, but adds a further classification step to clarify whether the main foundation of a recommendation is direct evidence, indirect evidence, non-comparative evidence, or expert judgement (Figure 4).

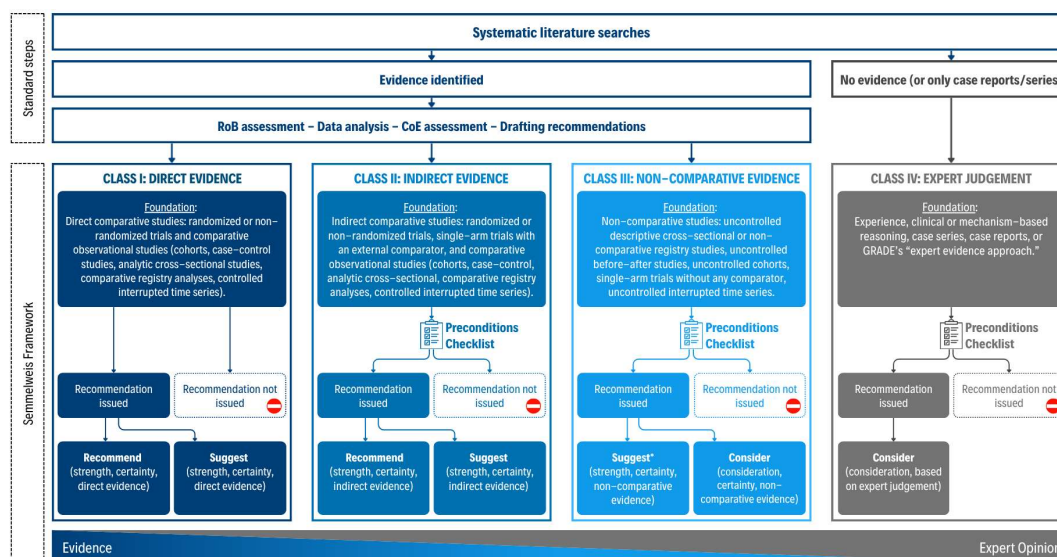


Figure 4. Flowchart of the Semmelweis Framework classification process

RoB: Risk of Bias, CoE: Certainty of Evidence.

Source: reproduced unchanged from Nagavci B, Andrews MR, Vikse J, Studenic P, Stamm T, Lakatos B. The Semmelweis Framework: structuring evidence foundations and expert opinion in clinical practice guidelines. *BMC Med Res Methodol.* 2026;26(1):199. Used under the Creative Commons Attribution (4.0) license.

Its intended scope is clinical practice guidance concerning treatment, diagnosis, and prevention. The Framework can be applied across the full range of evidence, but it is expected to provide the greatest added value when evidence is indirect, non-comparative, scarce, or absent.

Because expert opinion remains widely used in guideline manuals, the Framework retains the term “expert opinion” as a broad methodological concept. However, it avoids using it as a generic label for individual recommendations. Instead, the roles usually grouped under this term are assigned to four recommendation classes, each linked to a specific evidentiary foundation.

4.3.2. Applying the Framework

Application of the Framework begins only after the conventional evidence-synthesis process has been completed. The preceding work should include systematic searching,

study selection, data collection, risk-of-bias evaluation, synthesis of findings and assessment of evidence certainty.

When formulating each recommendation, the panel identifies the evidentiary foundation that had the greatest influence on the final judgement. The decision flowchart is then used to assign the corresponding Semmelweis class (Figure 4).

Classes are assigned separately to each recommendation rather than to the research question as a whole. Consequently, recommendations derived from the same question may receive different classifications when they rely on different evidentiary foundations. For instance, evidence may directly support the principal treatment choice but provide no comparable support for an associated statement concerning follow-up or management after treatment failure.

Once the class has been determined, the panel applies the corresponding wording, strength category, evidentiary notation, and safeguard requirements.

4.3.3. Description of recommendation classes

Class I: Recommendations based on direct evidence

Class I recommendations are based on comparative evidence answering the research question directly. Recommendations of this class are typically based on comparative studies, including randomized or non-randomized trials, and comparative observational studies (case-control studies, cohort studies, analytic cross-sectional studies, comparative registry analyses, controlled interrupted time series), or systematic reviews of these studies.

Recommendations in this Class can be either strong or conditional, as defined by the GRADE methodology (23). Strong recommendations should be labelled as “recommend” or “recommend against”, while conditional ones as “suggest” or “suggest against”. As an additional layer to the GRADE approach, the notation “direct” needs to be added in brackets for this class, i.e. “moderate certainty direct evidence” (Table 1).

In cases where evidence is of high or moderate certainty according to GRADE assessment, the evidence in most cases “speaks for itself”, and expert opinion has no role. However, when the certainty is low or very low, some interpretation of evidence might

be needed to reach conclusions and issue recommendations. In such cases expert opinion might help in contextualising the evidence, but the recommendation is firmly rooted in comparative direct evidence.

Class II: Recommendations based on indirect evidence

Class II recommendations are based on indirect comparative evidence, due to lack of direct evidence. Indirectness of evidence is defined by the differences between the retrieved evidence and some of the PICO (population, intervention, comparator, outcome, setting) elements of the research question, based on the GRADE guidance (24).

For this class indirect evidence might come from comparative studies, such as randomized or non-randomized trials, single-arm trials with a historical or external comparator, and comparative observational studies (case-control studies, cohorts, analytic cross-sectional studies, comparative registry analyses, controlled interrupted time series), or systematic reviews of these studies.

This class of recommendations needs to be chosen for recommendations when the dominant foundation is indirect evidence. To determine whether indirectness is substantial enough to constitute the main foundation of the recommendation, guideline panels need to consider two key aspects:

1. The Nature of Indirectness: If there are clinically important differences in population, intervention, comparator, setting or outcomes between the available evidence and the guideline question.
2. The Balance of Evidence: If the body of evidence is dominated in weight by indirect studies (e.g. number of included patients, or number of studies included), compared to direct studies.

The Semmelweis Framework does not define specific thresholds for such decisions, due to specificities of each guideline project and recommendation. Therefore, this decision should be made by the guideline panels on a case-by-case basis, through discussion and presented transparently in the guideline document.

Recommendations of this class usually need to be conditional, except in cases where strong ones are deemed necessary, according to the GRADE guidance (23). The verbs to be used should be “suggest/suggest against”, or “recommend/recommend against”. The

notation “indirect” needs to be added to the standard GRADE notation, e.g.: “low certainty indirect evidence” (examples in Table 1).

Expert opinion has a larger role than in Class I. The panel must decide whether the indirect evidence is sufficiently applicable and whether extrapolation to the target population and clinical context is justified.

Class III: Recommendations based on non-comparative evidence

Class III recommendations are based on studies that include only one group of subjects or patients, without a control arm. Such studies include single-arm trials without any kind of control group, uncontrolled descriptive cross-sectional studies, non-comparative registry studies, uncontrolled before-after studies, uncontrolled cohorts, uncontrolled interrupted time series.

This body of evidence has serious limitations in answering the research questions, due to the lack of a comparative study design. Therefore, recommendations in this class are labelled as “considerations” using the verb “consider/consider avoiding”. In rare situations where the effect estimates are considered large, or if there is an imminent risk of severe harm (e.g., clear safety signal), a conditional recommendation may be issued using the verb “suggest/suggest against”. Nevertheless, in both cases, the notation “non-comparative evidence” must be used (examples in Table 1). Strong recommendations are not warranted in this class.

Expert opinion has a substantial role in this class, because the panel must interpret findings without a concurrent comparator and determine whether they are sufficiently applicable to justify guidance.

Class IV: Recommendations based on expert judgement

The main foundations of recommendations in this class are clinical experience of the panel, mechanism-based reasoning and very weak evidence such as case series and case reports. The experience from the panel can be gathered through discussion, or formal methods such as Delphi, or through GRADE’s “expert evidence” approach (9, 25).

Since empirical evidence is not available, risk-of-bias assessment and certainty of evidence assessment are not applicable for evidence of this class.

Recommendations in this class should be phrased with verbs “consider” or “consider avoiding” and must include the notation: “based on expert judgement”, to indicate the reliance on experts’ judgement and experience, rather than empirical studies (examples in Table 1). Conditional and strong recommendations are not warranted in this class.

Since empirical evidence is absent in this class, except for case series and reports, the guideline panel must rely entirely on expert opinion in its traditional sense to formulate clinical recommendations.

Table 1. Classification of recommendation classes and accompanying statements with examples

Source: reproduced unchanged from Nagavci B, Andrews MR, Vikse J, Studenic P, Stamm T, Lakatos B. The Semmelweis Framework: structuring evidence foundations and expert opinion in clinical practice guidelines. BMC Med Res Methodol. 2026;26(1):199. Used under the Creative Commons Attribution (4.0) license.

Classes	Strength	Verb	Notation	Examples with verbs and notations
Class I: Direct evidence from comparative studies	Strong	Recommend	Certainty direct evidence	We recommend ... / We recommend against... (strong recommendation, high certainty direct evidence).
	Conditional	Suggest	Certainty direct evidence	We suggest ... / We suggest against ... (conditional recommendation, low certainty direct evidence).
Class II: Indirect evidence	Strong	Recommend	Certainty indirect evidence	We recommend ... / We recommend against...

from comparati ve studies				(strong recommendation, moderate certainty indirect evidence).
	Condition al	Suggest	Certainty indirect evidence	We suggest ... / We suggest against ... (conditional recommendation, low certainty indirect evidence).
Class III: Non- comparati ve evidence	Condition al*	Suggest	Certainty non- comparative evidence	We suggest ... / We suggest against ... (conditional recommendation, low certainty non-comparative evidence).
	Considerat ion	Consider	Certainty non- comparative evidence	Consider ... / Consider avoiding ... (consideration, low certainty non-comparative evidence).
Class IV: Expert judgement	Considerat ion	Consider	Based on expert judgement	Consider ... / Consider avoiding ... (consideration, based on expert judgement).
No recomme ndation	No recommendation could be made for or against ... due to lack of evidence (no recommendation, no evidence). No recommendation could be made for or against ... due to inconclusive evidence (no recommendation, inconclusive evidence). No recommendation could be made for or against ... due to lack of consensus (no recommendation, lack of consensus).			

**Conditional recommendations in Class III should be reserved only for exceptional cases (e.g. large benefits or certain harm). The default strength in this class is “Consideration”.*

4.3.4. Strength, wording, and evidentiary notation

The Semmelweis Framework adopted the GRADE system for labelling the strength of recommendations in classes I and II. Strong recommendations use the verb “recommend”, while conditional ones use the term “suggest” (23). For Classes III and IV, the Framework introduces a distinct category of recommendations, called “considerations”. This is the third category of strength, after strong and conditional recommendations, reflecting the lower certainty and interpretive nature of recommendations in these two classes.

Regarding the certainty of evidence, the Framework uses the standard GRADE terminology for classes I and II (“very low”, “low”, “moderate” and “high”). However, as an additional layer of rigour, recommendations must also contain the evidentiary foundations:

- Class I: direct evidence;
- Class II: indirect evidence;
- Class III: non-comparative evidence; and
- Class IV: expert judgement.

Where a recommendation cannot be issued, the reason should also be stated. Examples include insufficient evidence, inconclusive evidence, or lack of panel consensus (Table 1).

4.3.5. The Preconditions Checklist

The Preconditions Checklist is mandatory before issuing recommendations in Classes II, III, or IV. These classes involve foundations that are more dependent on extrapolation, interpretation, or expert judgement than Class I.

The Checklist is intended to ensure that the decision to issue guidance is explicit, justified, and documented. It asks whether a systematic search was conducted, why sufficient direct evidence is unavailable, whether the recommendation is clinically necessary, and whether its likely benefits and harms have been considered.

It also requires panels to document important uncertainties and assumptions, consider patient values and preferences, assess possible subgroup effects, and identify remaining evidence gaps and research priorities. Additional questions address whether the panel had the necessary clinical and methodological expertise, whether participation was balanced, and whether conflicts of interest were appropriately managed.

All items should be answered positively before a Class II–IV recommendation is issued. If one or more preconditions are not fulfilled, the panel should avoid issuing a recommendation on an insufficiently justified basis. Instead, it should summarise the available evidence, explain why a recommendation could not be made, and formulate a research recommendation where appropriate (Table 2).

Table 2. Preconditions checklist for recommendation classes II, III and IV

Source: reproduced from Nagavci B, Andrews MR, Vikse J, Studenic P, Stamm T, Lakatos B. The Semmelweis Framework: structuring evidence foundations and expert opinion in clinical practice guidelines. BMC Med Res Methodol. 2026;26(1):199. Used under the Creative Commons Attribution (4.0) license.

No.	Preconditions	Answer (Yes/No)	Justification (Explain briefly your answer)
Evidence			
1	Were systematic searches conducted to search for relevant evidence?		
2	Did the panel discuss the reasons why there is no (sufficient) direct evidence, including practical, ethical, or financial barriers to conducting relevant research?		
Necessity			
3	Is issuing this recommendation necessary and justified to improve clinical decision-making despite the evidence gap?		
4	Did the panel consider the important potential (direct and indirect) benefits and harms deriving from this recommendation?		

5	Is the rationale for issuing this recommendation (including key uncertainties and assumptions) documented in a way that is understandable and traceable for guideline users?		
Patient & Research Impact			
6	Did the panel discuss potential acceptability and likely values/preferences of patients and other stakeholders based on available information?		
7	Did the panel consider if this recommendation could have different effects on specific subgroups (e.g., with certain comorbidities, socioeconomic status, etc.)?		
8	Did the panel explicitly document key evidence gaps and research priorities to ensure that issuing this recommendation does not substitute the need for further research?		
Panel Integrity			
9	Did the panel include sufficient relevant clinical and methodological expertise to issue this recommendation despite the evidence gap?		
10	Was the decision-making process conducted in a way that minimised undue influence by any individual and ensured balanced contribution across panel members?		
11	Were potential conflicts of interest of all panel members disclosed, assessed, managed and documented, in a way that makes the risk of influencing this recommendation unlikely?		

4.3.6. Implementation resources

Four accompanying resources support practical use of the Framework. The flowchart guides class assignment, the classification table specifies the requirements and wording for each class, the Preconditions Checklist documents the methodological safeguards, and the FAQ addresses recurring questions concerning mixed evidence bases and different study designs (Appendix).

Further details and results are available in the published study corresponding to WP3 (17).

5. DISCUSSION

This research addressed a persistent methodological problem in clinical practice guideline development. The first WP showed that expert opinion is frequently used in infectious diseases guidelines, but often with variable terminology, limited methodological transparency, and unclear justification (15). The second WP expanded this analysis across guideline-developing organisations in all topics and demonstrated that expert opinion is mainly used in two situations: to fill evidence gaps, and to interpret weak, indirect, or inconclusive evidence. Based on these findings, a new definition of expert opinion was proposed, capturing its main foundations: clinical experience, expert judgement, indirect evidence, very low-quality evidence, and mechanism-based reasoning (16). The third and main output of this project was the development of the Semmelweis Framework, which translates this work into a practical methodological approach for guideline development (17).

The main conceptual change introduced by the Semmelweis Framework is the transparent and structured integration of expert opinion into clinical recommendations while preserving the principles of evidence-based guideline development. Instead of treating expert opinion as a vague or separate category, the Framework places it on a continuum with empirical evidence. As evidence becomes less direct or less robust, the contribution of expert opinion increases. This approach allows guideline panels to make explicit what supports each recommendation and how much interpretation is required.

A key novelty of the Framework is that it classifies recommendations according to their evidentiary foundation rather than only by certainty of evidence or study design. It separates recommendations based on direct evidence, indirect evidence, non-comparative evidence, and expert judgement. Each class has its own label, wording, and expected role of expert opinion. In this way, the Framework makes the use of the general term “expert opinion” obsolete, because the actual foundation of the recommendation is stated precisely (17).

The Framework also introduces “considerations” as a distinct category for recommendations based on weaker foundations, particularly non-comparative evidence or expert judgement. This category signals that such statements should be interpreted more cautiously than strong or conditional recommendations (17). This may reduce the

common problem of issuing strong recommendations based on low or very low certainty of evidence, an issue which has been reported to affect up to half of all strong recommendations in GRADE-based guidelines (5, 7, 26).

Another important implication concerns the utilisation of indirect evidence. In standard GRADE methodology, indirectness is one domain that can lead to downgrading certainty of evidence (24). However, recommendations based mainly on indirect evidence may still receive the same certainty rating as recommendations based on direct evidence. The Semmelweis Framework adds a further layer of transparency by explicitly labelling such recommendations as based on indirect evidence. This allows users to understand not only the certainty rating, but also the nature of the evidence foundation.

The proposed classification may also clarify the appropriate role of Good Practice Statements within the GRADE approach, which are ungraded statements that guideline panels issue when a recommendation is obviously beneficial (or harmful) and a full GRADE assessment would add little value (27). The Semmelweis Framework does not aim to redefine the original purpose of this approach but offers transparent alternatives for most cases where Good Practice Statements have been misused as expert opinion recommendations in the past (3). This reduces the need to rely on Good Practice Statements as a catch-all solution. Similarly, recommendations developed using GRADE's "expert evidence" approach can be labelled within Class IV as based on expert judgement, avoiding their misrepresentation as conventional evidence-based recommendations (25).

The main strength of this work is that it provides a structured system for using expert opinion transparently, supported by explicit recommendation classes, standardised labels, appropriate action verbs, and a mandatory Preconditions Checklist. To the best of our knowledge, this is the first framework to provide a dedicated checklist specifically for the use of expert opinion in guideline development.

Another major strength is its compatibility with GRADE. The Framework does not replace GRADE but complements it by addressing an area where GRADE offers limited practical guidance. This may be particularly useful for guideline-developing organisations that currently rely on modified GRADE systems, Good Practice Statements,

or poorly defined consensus-based recommendations when evidence is absent or insufficient.

This work also has its limitations. First, the Framework was developed through a qualitative, iterative refinement process. As the Framework was refined throughout this process, no single stable version was applied to all assessed recommendations, precluding formal inter-rater reliability testing. Second, implementation of the Semmelweis Framework will require orientation and some cultural change within guideline-developing organisations. Guideline panels will need to become familiar with the classification system, terminology, and Preconditions Checklist. However, this challenge is partly mitigated by the practical implementation tools provided with the Framework, including the decision flowchart, classification table, checklist, and FAQ guidance (17).

6. CONCLUSIONS

This thesis introduced a new evidence-informed definition of expert opinion. Building on this definition, the Semmelweis Framework provides a structured approach for integrating empirical evidence and expert opinion into clinical practice guidelines. It classifies recommendations according to their evidentiary foundations and distributes the roles traditionally grouped under “expert opinion” across clearly defined recommendation classes. Supported by a dedicated Preconditions Checklist, the Framework avoids using expert opinion as a generic recommendation label and instead promotes transparent and actionable guidance. In this way, it offers guideline developers a practical tool for supporting clinical decision-making when evidence is limited, indirect, or absent.

7. SUMMARY

Clinical practice guidelines are usually developed using the best available evidence from systematic reviews and formal evidence appraisal. However, in many areas of medicine, evidence is limited, indirect, inconclusive, or completely absent. In such situations, guideline panels often rely on expert opinion to issue clinically useful recommendations. Despite its frequent use, expert opinion remains poorly defined and inconsistently presented in guideline methodology. This thesis addressed this methodological gap by examining how expert opinion is used, how it is defined, and how it can be integrated more transparently into clinical guideline development.

The thesis consisted of three linked work packages. The first work package analysed the use of expert opinion in international infectious diseases guidelines. It showed that expert opinion is commonly used, but with substantial variation in terminology, reporting, and methodological transparency. These findings highlighted the need for a more structured and transparent approach.

The second work package examined methodological manuals from guideline-developing organisations across different medical fields. This analysis showed that expert opinion is mainly used for two purposes: to fill evidence gaps when direct evidence is unavailable, and to interpret evidence when it is weak, indirect, or inconclusive. Based on this work, a new definition of expert opinion was proposed. The concept was defined as guidance derived from clinical experience, expert judgement, indirect evidence, very low-quality evidence, or mechanism-based reasoning, aimed at supporting clinical decision-making when evidence is absent or requires contextual interpretation.

The main output of the thesis is the Semmelweis Framework, developed in the third work package. The Framework provides a structured method for classifying clinical recommendations according to their evidentiary foundation. It defines four classes of recommendations: recommendations based on direct evidence, indirect evidence, non-comparative evidence, and expert judgement. As the evidence becomes weaker, less direct, or absent, the role of expert opinion progressively increases. The Framework also introduces standardised labels, clear recommendation wording, and a mandatory Preconditions Checklist for recommendations relying on weaker foundations.

8. REFERENCES

1. Graham R, Mancher M, Miller Wolman D, Greenfield S, Steinberg E, editors. Clinical Practice Guidelines We Can Trust. Washington (DC): National Academies Press (US); 2011. Available from: <https://www.ncbi.nlm.nih.gov/books/NBK209539/>.
2. Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA, editors. Cochrane Handbook for Systematic Reviews of Interventions version 6.5 (updated August 2024): Cochrane; 2024. Available from: www.cochrane.org/handbook.
3. McArthur A, Klugárová J, Yan H, Florescu S. Innovations in the systematic review of text and opinion. *Int J Evid Based Healthc*. 2015;13(3):188-95.
4. Ponce OJ, Alvarez-Villalobos N, Shah R, Mohammed K, Morgan RL, Sultan S, Falck-Ytter Y, Prokop LJ, Dahm P, Mustafa RA, Murad MH. What does expert opinion in guidelines mean? a meta-epidemiological study. *Evid Based Med*. 2017;22(5):164-9.
5. Sims CR, Warner MA, Stelfox HT, Hyder JA. Above the GRADE: Evaluation of Guidelines in Critical Care Medicine. *Crit Care Med*. 2019;47(1):109-13.
6. Fanaroff AC, Califf RM, Windecker S, Smith SC, Jr, Lopes RD. Levels of Evidence Supporting American College of Cardiology/American Heart Association and European Society of Cardiology Guidelines, 2008-2018. *JAMA*. 2019;321(11):1069-80.
7. Miles KE, Rodriguez R, Gross AE, Kalil AC. Strength of Recommendation and Quality of Evidence for Recommendations in Current Infectious Diseases Society of America Guidelines. *Open Forum Infect Dis*. 2021;8(2):ofab033.
8. Eibling D, Fried M, Blitzer A, Postma G. Commentary on the role of expert opinion in developing evidence-based guidelines. *Laryngoscope*. 2014;124(2):355-7.
9. Schünemann HJ, Zhang Y, Oxman AD. Distinguishing opinion from evidence in guidelines. *BMJ*. 2019;366:l4606.
10. Howick J, Chalmers I, Glasziou P, Greenhalgh T, Heneghan C, Liberati A, Moschetti I, Phillips B, Thornton H. The 2011 Oxford CEBM Levels of Evidence (Introductory Document). Available from: <https://www.cebm.ox.ac.uk/resources/levels-of-evidence/ocebmllevels-of-evidence>.
11. Methodology Manual and Policies From the ACCF/AHA Task Force on Practice Guidelines: American College of Cardiology Foundation and American Heart Association; 2010. Available from: <https://www.acc.org/Guidelines/About-Guidelines-and-Clinical-Documents/Methodology>.

12. ESC Clinical Practice Guidelines: Policies and Procedures: European Society of Cardiology; 2022. Available from: <https://www.escardio.org/Guidelines/Clinical-Practice-Guidelines/Guidelines-development/Writing-ESC-Guidelines>.
13. Schünemann H, Brożek J, Guyatt G, Oxman A, editors. GRADE handbook for grading quality of evidence and strength of recommendations: The GRADE Working Group; 2013. Available from: <https://gdt.gradepro.org/app/handbook/handbook.html>.
14. Venus C, Jamrozik E. Transparency in clinical practice guidelines: the problem of consensus-based recommendations and practice points. *Intern Med J*. 2021;51(2):291-4.
15. Nagavci B, Schwingshackl L, Martin-Loeches I, Lakatos B. Utilization of expert opinion in infectious diseases clinical guidelines-A meta-epidemiological study. *PLoS One*. 2024;19(6):e0306098.
16. Nagavci B, Gáspár Z, Lakatos B. Defining expert opinion in clinical guidelines: insights from 98 scientific societies - a methodological study. *BMC Med Res Methodol*. 2025;25(1):87.
17. Nagavci B, Andrews MR, Vikse J, Studenic P, Stamm T, Lakatos B. The Semmelweis Framework: structuring evidence foundations and expert opinion in clinical practice guidelines. *BMC Med Res Methodol*. 2026;26(1):199.
18. Murad MH, Wang Z. Guidelines for reporting meta-epidemiological methodology research. *Evid Based Med*. 2017;22(4):139-42.
19. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, Shamseer L, Tetzlaff JM, Akl EA, Brennan SE, Chou R, Glanville J, Grimshaw JM, Hróbjartsson A, Lalu MM, Li T, Loder EW, Mayo-Wilson E, McDonald S, McGuinness LA, Stewart LA, Thomas J, Tricco AC, Welch VA, Whiting P, Moher D. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*. 2021;372:n71.
20. Mbuagbaw L, Lawson DO, Puljak L, Allison DB, Thabane L. A tutorial on methodological studies: the what, when, how and why. *BMC Med Res Methodol*. 2020;20(1):226.
21. Schünemann HJ, Wiercioch W, Etzeandía I, Falavigna M, Santesso N, Mustafa R, Ventresca M, Brignardello-Petersen R, Laisaar KT, Kowalski S, Baldeh T, Zhang Y, Raid U, Neumann I, Norris SL, Thornton J, Harbour R, Treweek S, Guyatt G, Alonso-

- Coello P, Reinap M, Brozek J, Oxman A, Akl EA. Guidelines 2.0: systematic development of a comprehensive checklist for a successful guideline enterprise. *CMAJ*. 2014;186(3):E123-42.
22. Vernooij RW, Sanabria AJ, Solà I, Alonso-Coello P, Martínez García L. Guidance for updating clinical practice guidelines: a systematic review of methodological handbooks. *Implement Sci*. 2014;9:3.
23. Andrews J, Guyatt G, Oxman AD, Alderson P, Dahm P, Falck-Ytter Y, Nasser M, Meerpohl J, Post PN, Kunz R, Brozek J, Vist G, Rind D, Akl EA, Schünemann HJ. GRADE guidelines: 14. Going from evidence to recommendations: the significance and presentation of recommendations. *J Clin Epidemiol*. 2013;66(7):719-25.
24. Guyatt GH, Oxman AD, Kunz R, Woodcock J, Brozek J, Helfand M, Alonso-Coello P, Falck-Ytter Y, Jaeschke R, Vist G, Akl EA, Post PN, Norris S, Meerpohl J, Shukla VK, Nasser M, Schünemann HJ. GRADE guidelines: 8. Rating the quality of evidence--indirectness. *J Clin Epidemiol*. 2011;64(12):1303-10.
25. Mustafa RA, Garcia CAC, Bhatt M, Riva JJ, Vesely S, Wiercioch W, Nieuwlaat R, Patel P, Hanson S, Newall F, Wiernikowski J, Monagle P, Schünemann HJ. GRADE notes: How to use GRADE when there is "no" evidence? A case study of the expert evidence approach. *J Clin Epidemiol*. 2021;137:231-5.
26. Brito JP, Domecq JP, Murad MH, Guyatt GH, Montori VM. The Endocrine Society guidelines: when the confidence cart goes before the evidence horse. *J Clin Endocrinol Metab*. 2013;98(8):3246-52.
27. Guyatt GH, Alonso-Coello P, Schünemann HJ, Djulbegovic B, Nothacker M, Lange S, Murad MH, Akl EA. Guideline panels should seldom make good practice statements: guidance from the GRADE Working Group. *J Clin Epidemiol*. 2016;80:3-7.

9. BIBLIOGRAPHY OF THE CANDIDATE'S PUBLICATIONS

Publications related to the thesis

1. **Nagavci B**, Schwingshackl L, Martin-Loeches I, Lakatos B. Utilization of expert opinion in infectious diseases clinical guidelines-A meta-epidemiological study. *PLoS One*. 2024;19(6):e0306098.
2. **Nagavci B**, Gáspár Z, Lakatos B. Defining expert opinion in clinical guidelines: insights from 98 scientific societies - a methodological study. *BMC Med Res Methodol*. 2025;25(1):87.
3. **Nagavci B**, Andrews MR, Vikse J, Studenic P, Stamm T, Lakatos B. The Semmelweis Framework: structuring evidence foundations and expert opinion in clinical practice guidelines. *BMC Med Res Methodol*. 2026;26(1):199.

Publications not related to the thesis:

4. Gáspár Z, **Nagavci B**, Szabó BG, Lakatos B. Gut Microbiome Alteration in HIV/AIDS and the Role of Antiretroviral Therapy—A Scoping Review. *Microorganisms*. 2024;12(11):2221.
5. Bartoletti M, Azap O, Barac A, Ben Selma M, Ergonul O, Gkrania-Klotsas E, Grossi PA, Krause R, **Nagavci B**, Paño-Pardo JR, Pierrotti LC, Power N, Rodríguez-Baño J, Sibani M, Slavin MA, Szabo BG, Tazza B, Yung Tsang NN, Tsiodras S, Zollner-Schwetz I, Chemaly RF. Management of COVID-19 in immunocompromised patients: an European Society of Clinical Microbiology and Infectious Diseases consensus document. *Clin Microbiol Infect*. 2025;31(10):1655-66.
6. Righi E, Yahav D, Nasim A, Akova M, Barac A, Scudeller L, **Nagavci B**, Armellini M, Zanchi C, Shirin N, Tunali V, Paño-Pardo JR, Rodríguez-Baño J, Jamrozik E. Incorporating ethics into infectious disease clinical practice guidelines. *Clin Microbiol Infect*. 2025;31(4):560-7.
7. Yahav D, Nasim A, Shirin N, Armellini M, Zanchi C, Tunali V, Mirandola M, Barac A, Scudeller L, **Nagavci B**, Paño-Pardo JR, Rodríguez-Baño J, Jamrozik E, Akova M, Righi E. Incorporating Ethics in Clinical Guidelines in Infectious Diseases: A Scoping Review. *JAMA Netw Open*. 2025;8(7):e2519826.
8. Albers B, **Nagavci B**, Joean O, Tschudin Sutter S, Tacconelli E, Paul M, Langendam M, Kostyanov T, Molina J, Stahl JP, Pano Pardo JR, Gkrania-Klotsas E, Hassoun-Kheir N, Clack L, Scudeller L, Zervos M. Guiding principles for implementing ESCMID clinical practice guidelines. *Clin Microbiol Infect*. 2026;32(7):1107-14.
9. Joean O, Sermet K, Ashkenazi-Hoffnung L, Kiymaz YC, Blumenthal K, Bonazzetti C, Chiriac AM, Gomez-Zorrilla S, Karakike E, Khatamzas E, Powell N, Trubiano JA, Wijnakker R, Sandoe J, **Nagavci B**, De Boer MGJ. ESCMID clinical guidelines on the evaluation and management of a reported antibiotic allergy. *Clin Microbiol Infect*. 2026;32(5):767-87.

10. **Nagavci B**, Joean O, Tschudin-Sutter S, Tacconelli E, Paul M, Langendam M, Kostyanev T, Molina J, Stahl JP, Zervos M, Bodilsen J, Paño-Pardo JR, Scudeller L. The ESCMID framework for reaching consensus in clinical guidelines. Clin Microbiol Infect. 2026.
11. Tunalı V, **Nagavci B**, Carmena D, Robertson LJ, Doğruman Al F, Gentekaki E, Tsoulos AD. Relationship between Blastocystis infection and clinical outcomes: A scoping review protocol. Open Res Eur. 2025;5:15.

<i>Cumulative impact factor of publications related to the thesis:</i>	<i>10.0</i>
<i>Cumulative impact factor of publications unrelated to the thesis:</i>	<i>59.4</i>
<i>Total cumulative impact factor of the candidate's publications:</i>	<i>69.4</i>

10. ACKNOWLEDGEMENTS

I am very thankful to my supervisor, Botond Lakatos, who not only made this collaboration possible and provided immense support throughout my PhD, but whose friendship I also deeply value.

I would also like to thank the Doctoral Office of the University for making the entire process run so smoothly. I am grateful to my co-authors, with whom I had the great pleasure of working throughout this research.

My deepest thanks go to my wife and my children for their support and understanding whenever I had to spend weekends working or could not be there with them. Special thanks also go to my parents and to my friends, who have always been there for me.

11. APPENDIX

Supplementary Table 1. Frequently Asked Questions (FAQ) for adoption and implementation of the Semmelweis Framework

Reproduced unchanged from Nagavci B, Andrews MR, Vikse J, Studenic P, Stamm T, Lakatos B. The Semmelweis Framework: structuring evidence foundations and expert opinion in clinical practice guidelines. BMC Med Res Methodol. 2026;26(1):199. Used under the Creative Commons Attribution (4.0) license.

No.	Question	Answer
1	Why does the Framework use “expert opinion” in the title if it advises against using the term?	The term “expert opinion” is used in the title because it is widely recognised in guideline methodology and helps readers immediately understand the topic the Framework addresses. At the same time, the Framework discourages using “expert opinion” as a vague, catch-all justification within recommendations. Instead, it operationalises what is often labelled “expert opinion” by distributing such situations across structured classes (II–IV), each with a defined evidentiary foundation and explicit safeguards and conditions for use. For the traditional scenario in which guidance is based primarily on clinical expertise in the absence of meaningful evidence, the Framework uses Class IV (“expert judgement”), which explicitly signals the nature of this foundation. This makes the role of expertise traceable, reduces ambiguity, and improves transparency for guideline users.
2	Is expert opinion the same as consensus?	No, they are different concepts. Expert opinion is the substance (clinical experience, mechanistic reasoning, interpretation of evidence etc.) that fills an evidence gap. Consensus is the decision-making approach (e.g., discussion, Delphi, majority vote etc.) used to agree on a recommendation. These two terms should not be used interchangeably.
3	How do the Semmelweis classes and their terminology relate to GRADE certainty of evidence and strength of recommendations?	They are separate dimensions. While the Semmelweis class describes the type and structure of the evidentiary foundation (direct vs indirect, comparative vs non-comparative, versus expert judgement), GRADE certainty, in contrast, reflects the confidence in the effect estimate for each outcome, based on risk of bias, inconsistency, indirectness, imprecision, and publication bias. Strength of recommendation (strong vs conditional) is then determined by the full GRADE process (including benefits and harms, values, feasibility, equity, etc.). As a result, it is possible to have, for example, a Class I recommendation supported by randomized trials but with low certainty for a given outcome, or a Class II recommendation with moderate certainty for a specific outcome. Panels should therefore always report both the Semmelweis class and the GRADE certainty (see Table 1 for examples).

4	Which risk-of-bias (RoB) assessment tool should we use to assess the quality of studies?	The Semmelweis Framework does not suggest a specific RoB tool. Guideline-developing organisations should select tools according to their own methods manual and the study designs included (e.g. different tools for randomised trials, non-randomised comparative studies, diagnostic studies etc.). The key requirement is that validated, design appropriate RoB tools are applied consistently across all recommendations. The Framework is compatible with any such tools.
5	What if the recommendation seems to be supported by evidence which fits in more than one class?	When evidence spans more than one class, panels should assign the class that reflects the dominant foundation that had the largest influence on the final judgement. For example, if direct comparative evidence exists but the recommendation was primarily driven by indirect evidence because the direct evidence was not applicable or was too uncertain, the recommendation should be classified according to that dominant driver. If two classes contributed equally and the panel cannot reasonably distinguish which had the greater influence, the higher class may be chosen as the label for the recommendation.
6	How should we distinguish between an uncontrolled cohort and a case series when assigning classes?	In the Semmelweis Framework, uncontrolled cohorts are classified as Class III, whereas case series are classified as Class IV. The key distinction is the sampling frame and the presence of a denominator. In an uncontrolled cohort study, patients are sampled based on an exposure or membership in a clearly defined group (for example, all patients receiving a particular intervention over a specified period), and outcomes are assessed during follow-up. This design, at least in principle, allows calculation of the absolute risk for that exposed group. By contrast, a case series samples patients based on having the outcome (or disease), but without a well-defined source population of all exposed individuals. In such studies, the risk of outcome given exposure cannot be calculated. <i>More info: Dekkers OM, Egger M, Altman DG, Vandenbroucke JP. Distinguishing case series from cohort studies. Ann Intern Med. 2012;156(1 Pt 1):37-40</i>
7	How do we classify evidence from a systematic review?	Systematic reviews do not automatically fall into a single class. A systematic review is a method of synthesis, not a study design itself. Therefore, the classification depends entirely on the underlying evidence included within the review. Guideline panels must assess the primary studies synthesised in the review. The final class is determined by the highest-class evidence used to underpin the recommendation. For example, a systematic review of randomized trials would support a Class I or II recommendation, whereas a systematic review of case series would support a Class IV recommendation.
8	How does the Framework apply to Diagnostic Test Accuracy (DTA)?	The same principles apply. The classification depends on the study type and its directness to the patient-important outcomes:

		Class I: Based on a direct comparative study that measures patient outcomes (e.g., RCTs or prospective comparative cohorts comparing test-treat strategies). Class II: Based on a Diagnostic Test Accuracy (DTA) study. A DTA study is comparative (test vs. standard), but "accuracy" is a surrogate outcome. Using it to infer patient outcomes is a form of outcome indirectness, placing it in Class II. Class III: Evidence from non-comparative but systematically collected datasets, e.g., single-arm prospective cohorts, registries without a concurrent reference standard. Class IV: Case series, case reports, mechanistic or expert-opinion reasoning about tests (e.g., pathophysiological plausibility without data).
9	How does the Framework classify evidence based on modelling?	Modelling is an analytical approach, not a study design per se. When a model is used only to translate or project an existing effect estimate from a comparative study (e.g. RCT or observational study), the recommendation retains the classification of the underlying evidence—Class I if direct, Class II if indirect. However, if the underlying evidence is non-comparative and the model is used to construct a counterfactual or effect estimate, the model becomes the primary foundation, and the recommendation should be classified as Class III. The same applies when a model is the sole source of the effect estimate because adequate empirical studies are absent. In such cases, it functions as a form of mechanism-based reasoning, and the recommendation should be classified as Class IV.
10	How should we decide for or against a recommendation?	The Semmelweis Framework provides the structure for classifying and wording recommendations but does not determine the criteria or the system for the decision-making process. That decision is made by each guideline panel using the organisation's own established process for weighing benefits and harms. The Semmelweis Framework is designed to be compatible with these approaches and does not replace them.
11	What is a "consideration" and how should clinicians interpret it compared with a conditional recommendation?	In the Semmelweis Framework, considerations are guidance statements used when recommendations rely mainly on non-comparative evidence (Class III) or expert judgement (Class IV). Compared with a conditional recommendation, a consideration carries less prescriptive force. It does not define a default course of action for most patients. Instead, it flags an option or concern that clinicians should weigh carefully with patients, taking account of context, values, and local resources. Considerations should be interpreted as cautious, judgement-dependent guidance rather than as a standard of care.
12	Can we still issue strong recommendations when	The Framework prevents strong recommendations when the evidentiary foundation is non-comparative (Class III) or based on expert judgement (Class IV).

	certainty of evidence is low or very low?	For Classes I–II, the usual GRADE rules remain in place regarding the strength of recommendations (see GRADE guidance for more).
13	What if we do not issue a recommendation for a research question?	If no recommendation can be made, the authors must explicitly state the reason below. For example, “No recommendation could be made due to inconclusive evidence or very-low certainty or lack of consensus” etc. This helps guideline users understand right away why a recommendation was not issued. Also, suggestions for further research might be issued.
14	What if we believe an expert opinion-based recommendation is necessary, but does not meet all preconditions?	In such cases a recommendation should not be issued until further evidence becomes available. Instead, a statement identifying the gap, flagging it as a research priority, may be used.
15	Can remarks be used to introduce additional guidance that is not part of the recommendation?	Remarks are used to supplement recommendations by adding contextual or practical guidance (e.g., scope, implementation considerations, monitoring, exceptions). However, remarks should not be used to introduce “hidden” recommendations. If a remark contains a clear action statement, it should be written as a separate recommendation and classified accordingly. Remarks must also be transparent: cite supporting evidence when evidence-based; if they reflect clinical expertise, this must be labelled explicitly in brackets (see Table 1 for examples).
16	Why should we label the reason for issuing a “no recommendation”?	Because “no recommendation” can mean fundamentally different things. Labelling the reason makes the decision transparent, helps users interpret uncertainty correctly, and prevents misinterpretation of “no recommendation” as evidence of no effect or no importance. It also documents whether the reason for not issuing a recommendation was evidence-related (e.g., very low certainty, conflicting results, or scarcity of evidence), process-related (e.g., no consensus in the panel), or contextual (e.g., substantial variability in values/preferences, feasibility constraints). And, stating the reason highlights the key evidence gaps or decision barriers, helping to prioritise and design future research.
17	Can the Semmelweis Framework be used for public-health or health-system recommendations?	The current version of the Semmelweis Framework was primarily developed for clinical practice guidelines focused on individual patient care in treatment, diagnosis and prevention. Public-health or system-level guidance often involves additional considerations (population-level effects, implementation logistics, policy constraints) that are outside the current scope. However, until further specific guidance is published, organisations may freely adapt the current principles for public-health or health-system guidelines.

18	Does this Framework increase the workload for guideline developers?	As the Framework introduces a new method and more rigour, initially it might be more time consuming due to need for filling the preconditions, additional assessments or panel discussions. However, this is expected to be minimal. Such steps are required for a reproducible and transparent method for utilising expert opinion.
19	Can we modify the Framework to fit our needs?	Since one of the aims of this Framework is also to provide consistency between organisations, it is advised that modifications are kept to a minimum. Organisations can adapt certain elements of the Framework while maintaining its principles and provided the four classes, expert opinion preconditions, and phrasing rules remain intact.
20	How should we communicate Semmelweis-based recommendations to clinicians who are unfamiliar with the Framework?	The Semmelweis Framework uses a practical system and terminology derived from the needs of guideline-developing organisations. The terminology is easy to be understood by clinicians and lay persons (e.g. based on direct, indirect evidence, etc.).
21	Should we publish Semmelweis classes and checklists in the main manuscript, or should they be supplementary?	For transparency, it is advised that each recommendation's class is clearly displayed in the main guideline document, and the completed preconditions checklists (for Classes II–IV) published in supplementary material.
22	How should we apply the Semmelweis Framework when updating an existing guideline? Does the recommendation class change over time?	Yes. The Semmelweis class should be re-assessed whenever a recommendation is updated. During an update, panels should identify the new evidence and reconsider the overall evidentiary foundation of the recommendation. If higher-class evidence becomes available (for example, moving from case series to an uncontrolled cohort, or from observational studies to randomized trials), the recommendation may shift from Class IV to III, or from Class III/II to Class I.