

# Medical Evidence

2026-07-24

ChinaRxiv

---

View the original and related papers at  
<https://chinaxiv.org/items/chinaxiv-202607.00260>

Source: ChinaXiv. Machine translation. Verify with the original.

## Abstract

**Background:** The Grading of Recommendations Assessment, Development and Evaluation (GRADE) approach is a landmark methodological system in the field of evidence-based medicine and is undergoing continuous evolution. Therefore, timely review and synthesis of its latest methodological advances and practical applications are crucial for clarifying its developmental trajectory and guiding future directions. **Objective:** This article aims to systematically summarize these developments, with the goal of providing a comprehensive reference for domestic researchers and guideline developers to grasp cutting-edge trends and apply advanced methods. **Methods:** The official website of the GRADE Working Group and its related publications were systematically searched to review and summarize its organizational structure, series of articles and publications, and current use of GRADE. By summarizing the main contents of GRADE Guidance, Concept, and Notes from 2021 to 2025, as well as the Core GRADE series of articles, and the evolution from the GRADE Handbook to the GRADE Book, the latest advances in GRADE methodology were described. **Results:** (1) GRADE organizational structure: through the collaborative mechanisms of the steering group, centers, networks, and project groups, GRADE enables global knowledge co-creation and continuous updating; (2) GRADE series of articles: to date, 42 GRADE Guidance articles, 9 GRADE Concept papers, and 6 GRADE Notes have been published, covering methodological deepening and expansion of application scenarios, such as systematic guidance on the GRADE approach for diagnostic test accuracy studies, imprecision, inconsistency, network meta-analysis, and threshold setting; (3) GRADE Book and GRADE GPT: the GRADE Handbook has been upgraded into the modular and dynamically updated GRADE Book, and GRADE GPT, an intelligent assistant based on a large language model, has been launched; (4) Core GRADE: the latest GRADE series of articles integrates the core principles of GRADE through a more simplified pathway to improve usability and consistency; (5) practical application of GRADE: standards for assessing the use of GRADE already exist, but structured evaluation tools are still lacking, and current guidelines vary in their adherence to GRADE. **Conclusion:** This study systematically reviews the latest advances in the GRADE methodological system and reveals its core trend of evolving from a basic framework toward greater depth, refinement, and intelligence. Meanwhile, the launch of the Core GRADE series indicates that GRADE is attempting to address the tension between increasing methodological complexity and ease of practical application, thereby promoting its broader implementation. The current state of GRADE practice suggests that researchers need to pay attention to the effective implementation of GRADE methods in practice.

Full Text

Medical Evidence

## Recent Advances in the Evolution and Practical Application of the GRADE Methodology

WANG Liuyang<sup>1,2</sup>, LIU Changning<sup>3</sup>, SHEN Quan<sup>4</sup>, CAI Zican<sup>5</sup>, WANG Yongbo<sup>1</sup>, REN Xiangying<sup>1</sup>, HUANG Qiao<sup>1</sup>, JIN Yinghui<sup>1</sup>, YAN Siyu<sup>1\*</sup>

1. Center for Evidence-Based and Translational Medicine, Zhongnan Hospital of Wuhan University, Wuhan 430071, Hubei Province, China
2. The Second Clinical College, Wuhan University, Wuhan 430071, Hubei Province, China
3. Dongzhimen Hospital, Beijing University of Chinese Medicine / Key Laboratory of Internal Medicine of Traditional Chinese Medicine, Ministry of Education and Beijing Municipality, Beijing 100700, China
4. Department of Obstetrics and Gynecology, The Third Xiangya Hospital of Central South University, Changsha 410000, Hunan Province, China
5. Wuhan Huangpi District Hospital of Traditional Chinese Medicine, Wuhan 430071, Hubei Province, China

\*Corresponding author: YAN Siyu, Assistant Researcher; E-mail: [yan-siyu@whu.edu.cn](mailto:yan-siyu@whu.edu.cn)

WANG Liuyang and LIU Changning contributed equally as co-first authors.

**[Abstract] Background** The Grading of Recommendations Assessment, Development and Evaluation (GRADE) approach is a landmark methodological system in the field of evidence-based medicine and remains in continuous evolution. Therefore, timely synthesis and summarization of its latest methodological advances and practical applications are crucial for clarifying its developmental trajectory and guiding future directions. **Objective** This article aims to systematically summarize these advances, in order to provide comprehensive reference for domestic researchers and guideline developers in keeping abreast of frontier developments and applying advanced methods. **Methods** The official website of the GRADE Working Group and its related publications were systematically searched, and the organizational structure, series of articles and publications, and current use of GRADE were systematically reviewed and summarized. By compiling the main contents of GRADE Guidance, Concept papers, Notes, and Core GRADE series articles from 2021 to 2025, as well as the evolution from the GRADE Handbook to the GRADE Book, the latest progress in GRADE methodology was described. **Results** (1) **GRADE organizational structure:** through a collaborative mechanism involving the guidance group, cen-

ters, networks, and project groups, GRADE has enabled global co-creation of knowledge and continuous updating; (2) **GRADE series articles**: to date, 42 GRADE Guidance articles, 9 GRADE Concept papers, and 6 GRADE Notes have been published. Their content covers the deepening of methodology and the expansion of application scenarios, providing systematic guidance in areas such as GRADE methods for diagnostic test accuracy studies, imprecision, inconsistency, network meta-analysis, and threshold setting; (3) **GRADE Book and GRADE GPT**: the GRADE Handbook has been upgraded into the modular and dynamically updated GRADE Book, and GRADE GPT, an intelligent assistant based on large language models, has been launched; (4) **Core GRADE**: as the latest series of GRADE articles, Core GRADE integrates the core principles of GRADE through a more simplified pathway, thereby improving usability and consistency; (5) **Practical application of GRADE**: although relevant standards for judging the use of GRADE already exist, structured assessment tools remain lacking, and current guidelines vary in their adherence to GRADE. **Conclusion** This study systematically reviews the latest advances in the GRADE methodological system and reveals its core trend of evolving from a foundational framework toward greater depth, refinement, and intelligentization. Meanwhile, the launch of the Core GRADE series marks GRADE's formal attempt to resolve the tension between increasing methodological complexity and the convenience of practical application, thereby promoting broader implementation. The current status of GRADE practice suggests that researchers need to pay attention to the effective implementation of GRADE methods in practice.

**[Keywords]** GRADE methodology; certainty of evidence; guidelines; strength of recommendations; GRADE Guidance; Core GRADE; GRADE Book; GRADE GPT

**[Chinese Library Classification No.]** R 19      **[Document Code]** A  
DOI: 10.12114/j.issn.1007-9572.2025.0424

## Recent Advances in the Evolution and Practical Application of the GRADE Methodology

WANG Liuyang<sup>1,2</sup>, LIU Changning<sup>3</sup>, SHEN Quan<sup>4</sup>, CAI Zican<sup>5</sup>, WANG Yongbo<sup>1</sup>, REN Xiangying<sup>1</sup>, HUANG Qiao<sup>1</sup>, JIN Yinghui<sup>1</sup>, YAN Siyu<sup>1\*</sup>

1. Center for Evidence-based and Translational Medicine, Zhongnan Hospital of Wuhan University, Wuhan 430071, China

2. The Second Clinical College, Wuhan University, Wuhan 430071, China

3. Dongzhimen Hospital, Beijing University of Chinese Medicine / Key Laboratory of Internal Medicine of TCM, Ministry of Education

**Funding:** General Program of the National Natural Science Foundation of China (82174230); Young Scientists Fund of the National Natural Science Foundation of China (82174230)

dation of China (82505373)

**Cite this article:** WANG Liuyang, LIU Changning, SHEN Quan, et al. Recent advances in the evolution and practical application of the GRADE methodology[J]. Chinese General Practice, 2026. DOI: 10.12114/j.issn.1007-9572.2025.0424. [Epub ahead of print] [www.chinagp.net]

Wang L Y, Liu C N, Shen Q, et al. Recent advances in the evolution and practical application of the GRADE methodology[J]. Chinese General Practice, 2026. [Epub ahead of print]

© Editorial Office of Chinese General Practice. This is an open access article under the CC BY-NC-ND 4.0 license.

This version posted 2026-07-24.

and Beijing, Beijing 100700, China

4. Department of Gynecology and Obstetrics and Gynecology, The Third Xiangya Hospital, Central South University, Changsha 410000, China

5. Huangpi District Hospital of TCM, Wuhan 430071, China

\* Corresponding author: YAN Siyu, Assistant Researcher; E-mail: yan-siyu@whu.edu.cn

WANG Liuyang and LIU Changning are the co-first authors

**[Abstract] Background** The Grading of Recommendations, Assessment, Development and Evaluation (GRADE) approach is a seminal framework in the field of evidence-based medicine and continues to evolve. Therefore, timely synthesis of its latest methodological advances and practical applications is crucial for clarifying its developmental trajectory and guiding future directions. **Objective** This study aims to systematically summarize the GRADE approach to provide researchers and guideline developers with a comprehensive reference for keeping abreast of cutting-edge developments and applying advanced methods. **Methods** We conducted a systematic review of the GRADE Working Group's official website and related publications to summarize the organizational structure, series articles, publications, and current usage of GRADE. By compiling the main content of GRADE Guidance, Concept papers, Notes, and Core GRADE articles published between 2021 and 2025, as well as the evolution from the GRADE Handbook to the GRADE Book, we elucidated the latest methodological progress in GRADE. **Results** (1) GRADE organizational structure: Through the collaborative mechanisms of guidance groups, centers, networks, and project teams, GRADE achieves global knowledge co-creation and continuous updates. (2) GRADE series articles: To date, 42 GRADE Guidance articles, 9 GRADE Concept papers, and 6 GRADE Notes have been published, covering methodological deepening and expanded applications, such as systematic guidance for GRADE in diagnostic test accuracy studies, imprecision, inconsistency, network meta-analysis, and threshold setting. (3) GRADE Book and GRADE GPT: The GRADE Handbook has been upgraded to the modular

and dynamically updated GRADE Book, and an AI-based assistant, GRADE GPT, has been introduced. (4) Core GRADE: The latest series of GRADE articles streamline the integration of core GRADE principles to enhance usability and consistency. (5) Practical application of GRADE: Although standards exist for evaluating the use of GRADE, structured assessment tools are still lacking, and current guideline adherence to GRADE varies. **Conclusion** This study systematically reviews the latest developments in the GRADE methodological framework, highlighting the core trend of evolution from foundational frameworks toward depth, refinement, and intelligence. The launch of the Core GRADE series marks a strategic effort to reconcile the growing complexity of the methodology with the need for practical accessibility, thereby promoting broader implementation. Current practice also indicates the need to focus on the effective implementation of GRADE methods in real-world settings.

**[Key words]** GRADE approach; Certainty of evidence; Guidelines; Strength of recommendations; GRADE Guidance; Core GRADE; GRADE Book; GRADE GPT

Evidence is the cornerstone of evidence-based medicine. The development of evidence-based guidelines has three key elements: comprehensively collecting existing research evidence, systematically evaluating its quality, and grading recommendations accordingly. In this process, a scientific system for grading evidence and the strength of recommendations plays a vital role in the rigor, applicability, and ultimate implementation effectiveness of guidelines. The Grading of Recommendations Assessment, Development and Evaluation (GRADE) approach is currently the most widely used system in evidence-based medicine for grading the quality of evidence and determining the strength of recommendations.

GRADE began in 2000, with the initial aim of establishing a unified, transparent, and reasonable system for grading evidence quality and rating recommendation strength <sup>[1]</sup>. Benefiting from its open membership system, which has attracted scholars worldwide to jointly refine and deepen its methodology, the scope of GRADE has expanded from its initial focus on clinical interventions to various fields of evidence-based medicine, including public health, environmental health, diagnostic tests, prognostic research, and even qualitative research. At present, the GRADE approach has been endorsed and used by more than 120 international guideline-developing institutions and organizations from 19 countries worldwide, including the World Health Organization (WHO), the Centers for Disease Control and Prevention (CDC), the National Institute for Health and Care Excellence (NICE), the Cochrane Collaboration, the British Medical Journal (BMJ), and the UpToDate clinical advisory resource <sup>[2-3]</sup>.

GRADE is not merely a rating method; it provides a transparent and structured process for developing systematic reviews, clinical practice guidelines, and health technology assessment (HTA) in health care <sup>[2]</sup>. Compared with traditional evidence grading systems, the GRADE system has distinct advantages. First, it evaluates quality by “body of evidence” as the unit—that is, for a specific

clinical question, such as the treatment, diagnosis, or prognosis of a particular disease, it systematically and comprehensively collects, assesses, and integrates all available relevant and valid research evidence to form a summary of the scientific evidence on that question. It does not use a single study as the unit of assessment, but rather each important outcome under that clinical question, such as mortality, incidence of adverse reactions, and [[unclear: outcome term beginning with “生” continues beyond the visible page]].

(such as quality of life), and constructs and evaluates the body of evidence separately for each. Second, in the logic of rating evidence quality, GRADE further breaks through the single logic of traditional grading systems that take study-design type as the core criterion for judgment (for example, directly judging randomized controlled trials as high quality and non-randomized studies as low quality), and establishes an evaluation framework of “study design as the basis + upgrading/downgrading factor adjustment.” For example, even in randomized controlled trials, if there is serious bias or inconsistency of results, the quality of evidence still needs to be downgraded; whereas for non-randomized studies, when the study results show a clear dose-response association or a sufficiently large effect size, evidence from non-randomized studies that initially has lower quality can reasonably be upgraded, thereby breaking the inherent notion that “non-randomized studies must be low quality.” This rating mechanism not only retains the fundamental influence of study design on evidence quality, but also, through comprehensive consideration of multidimensional factors, reflects the true credibility of the body of evidence more accurately and comprehensively. It provides more flexible and rigorous bases for judging the reasonable application of evidence from different types of studies, and also makes the final evidence-grading results better able to guide the needs of real-world decision-making scenarios such as clinical practice and public health policy formulation.

The GRADE approach provides a complete and dynamic decision-making framework, clearly defining evidence quality and strength of recommendation

. This process systematically guides researchers in how to prepare evidence profile tables, how to weigh benefits and harms, how to incorporate patient values and preferences, and ultimately how to form clear guidance for strong or weak recommendations. This study aims to systematically summarize the latest advances and practical applications of the GRADE methodological system, comprehensively sort out its organizational structure, the development trajectory of its series of articles and publications, and its current applications in different research and guideline-development fields, grasp the development trends of its methodology, and provide references and lessons for domestic researchers and guideline developers to understand the frontiers of GRADE and improve the quality of evidence-informed decision-making.

# 1 GRADE Organizational Structure

## 1.1 GRADE Working Group (GWG)

The establishment, promotion, and continuous evolution of the GRADE approach are driven at their core by the GWG and the global academic community on which it relies. The GWG was established in 2000. It is a fully open international academic collaborative community that broadly includes and welcomes participation by researchers, clinicians, methodologists, and policy makers from different countries and with multidisciplinary backgrounds. Becoming a member of this community means not only participating in methodological frontier exploration, but also committing to abide by the highest standards of academic integrity, including the reporting guidelines issued by the Enhancing the Quality and Transparency of Health Research (EQUATOR) Network, the author guidelines formulated by the International Committee of Medical Journal Editors (ICMJE), and the ethical principles advocated by the Committee on Publication Ethics (COPE). Users can learn specific information and the latest updates by visiting the GWG official website (<https://www.gradeworkinggroup.org/>).

## 1.2 GRADE Guidance Group (GGG)

To efficiently coordinate this large and dispersed global organization, the GWG established its executive committee—the GGG. This group consists of 12 core members, including the two co-chairs of the GWG. The GGG undertakes vitally important administrative and leadership functions: it provides guidance to the GWG chairs and ensures accountability to the GWG; assists in formulating GRADE policies and development directions; and supervises implementation of processes. To ensure organizational vitality and diversity of perspectives, the GGG implements a member-rotation system, replacing two members every two years. The most recent replacement was in 2024, and the next will be in 2026. The GGG maintains efficient operations through regular monthly meetings and special meetings before and after large working-group meetings.

## 1.3 GRADE Centers and GRADE Networks

The global dissemination and localized application of GRADE methodology cannot be separated from a large and solid support network. This network is mainly composed of GRADE centers and GRADE networks distributed around the world. GRADE centers are usually physical entities established within leading universities or research institutions, and they are devoted to the in-depth promotion, application, and study of the GRADE approach within specific geographic regions or disciplinary fields. These centers are not only bases for methodological training, but also play a role in the process of integrating the GRADE framework with local health systems, cultural backgrounds, and decision-making needs. There are 17 centers worldwide, four of which are in China: the GRADE Lanzhou University Center, the Nottingham-Ningbo GRADE Center, the GRADE Fudan University Center,



and the Beijing GRADE Center. At the same time, GRADE networks are more like collaborative alliances within a country or region, bringing together multiple institutions or individual researchers to jointly promote the use of GRADE locally. There are currently 12 networks; active networks include the US GRADE Network, the UK GRADE Network, and the Netherlands GRADE Network. This dual “center-network” structure enables the GRADE system both to have solid academic support and to possess broad and flexible coverage capacity, jointly underpinning its vision that “decisions in all departments worldwide should be based on the best evidence and science.”

### 1.4 GRADE Project Groups

The academic vitality of the GRADE system is reflected in its continuous methodological innovation and knowledge output, and this work is mainly accomplished through its project-group mechanism. GRADE project groups are the core within the GWG for promoting methodological advances and producing academic publications. Unlike the GWG, which is an academic community responsible for the development and maintenance of methodology on a global scale, GRADE project groups are working groups that conduct research on specific topics. Any GWG member may initiate the establishment of a project group, but must go through the formal process of submitting terms of reference (TOR) and obtaining approval from the GGG. Each project group is usually led by two or more persons in charge, and has one liaison who maintains close contact with the GGG and is responsible for coordinating the approval process. Within a project group, a model of “core team” plus “writing subgroup” is adopted to ensure efficient progress of the research. Project groups are open to interested members and organize project-group meetings regularly (once every 1-2 months). To date, 49 GRADE project groups have been established and are currently in...

there are 35 in operation, among which the GRADE and Artificial Intelligence group and the GRADE Planetary Health methodology project group are currently relatively active. Recently approved project groups include GRADE for Rare Diseases and GRADE for Primary Research ( “from GRADE to research” ), reflecting that GRADE methodology is expanding into more challenging and frontier areas. The above organizational structure of GRADE is shown in Figure 1.

In terms of academic exchange, the GWG brings together the global community, shares the latest achievements, and plans future directions through regularly held international conferences. For example, the GWG successfully held its annual meeting in Shanghai, China, on 2025-04-08–09. The theme of this annual meeting was to improve the quality and transparency of evidence used in healthcare decision-making. These meetings are not only platforms for presenting the latest advances in GRADE methodology, but also important occasions for global evidence-informed decision-making practitioners to exchange experience and establish collaboration.

**Text appearing in Figure 1**

- GRADE organization
  - 12 Centers
  - GGG
  - GWG
  - 17 Networks
  - Note: indicates a GRADE project group
- GRADE publications
  - GRADE Handbook / GRADE Book (GRADE [[unclear: small parenthetical text]])
  - Series articles:
    - \* GRADE Guidance Paper series
    - \* GRADE Concept Paper series
    - \* GRADE Notes
    - \* Other GRADE articles
    - \* Core GRADE series

Note: GRADE = Grading of Recommendations Assessment, Development and Evaluation; GWG = GRADE Working Group; GGG = GRADE Guidance Group.

**Figure 1** The organizational structure and series of publications of the GRADE Working Group

## 2 GRADE Article Types

In 2021, McGowan et al.[5], in an article published in the *Journal of Clinical Epidemiology*, classified GRADE-related articles into four types, which differ in their degree of formality and intended purposes. They are introduced in detail below:

### 2.1 GRADE Guidance Paper Series (formerly known as the GRADE Guidelines Paper Series)

This series of articles represents the official position and methodological standards of the GWG and provides operational methodological guidance. It is intended to offer global users specific guidance on “how to use GRADE.” Usually illustrated with examples, it describes in detail the steps and considerations for applying GRADE methods in systematic reviews, guideline development, and HTA to solve specific problems, such as assessing risk of bias and dealing with imprecision. Since the first guidance paper was published in 2011, 42 GRADE guidance papers have been published to date. The most recent, No. 42,

published by Wiercioch et al.[6], *GRADE Guidance: Using Thresholds for Judgments on Health Benefits and Harms in Decision Making* (using thresholds to judge health benefits and harms in health decision-making), focuses on how to use decision thresholds (DTs) to judge the magnitude of health effects, including both beneficial and harmful effects. (For the latest progress, see Part 3.)

## 2.2 GRADE Concept Paper Series

Next is the GRADE Concept Paper series, which has the same official status as the guidance papers. Although the two types of articles differ in purpose and background, their development and approval processes are the same. Before formal publication, both guidance papers and concept papers must undergo the GWG's rigorous formal approval process, including drafting, reporting to all working-group members at least twice to solicit comments, revision based on members' feedback, and final voting organized by the GGG. This rigorous process ensures the methodological rigor, transparency, and collective consensus of GRADE guidance papers and GRADE concept papers.

Unlike the guidance papers described above, when a methodological issue still requires further validation or operationalization, GRADE usually first develops a concept paper[7]. After being tested in practice, a concept paper may develop into a formal guidance paper. This type of article focuses on exploring forward-looking concepts, such as proposing links between guideline recommendations and trustworthy essential medicines lists[8], and judging the certainty of estimates of the discriminatory performance of prognostic models across a series of validation studies[9]. The core value of concept papers lies in stimulating discussion and promoting methodological innovation; therefore, only some of them will become formal guidance papers in the future. Since the first concept paper was published in 2021, a total of 9 GRADE concept papers have been published to date. The latest, by Whaley et al.[10], *GRADE Concept Paper 9: Rationale and process for creating a GRADE Ontology* (GRADE Concept Paper 9: rationale and process for creating a GRADE ontology), focuses on the processes for developing and maintaining the GRADE ontology, and describes the specific procedures for building and maintaining a useful, machine-acceptable ontology to support the use of GRADE methods.

## 2.3 GRADE Notes

To address complex real-world problems encountered in the broad application of GRADE and to incorporate practical experience from a wider user community, the GRADE Notes series was launched. This series does not represent the official position of the GWG; its nature is exchange and discussion, and the articles are usually relatively short. It provides a platform for GRADE users worldwide to share case studies, practical experiences, or interim solutions for applying GRADE in specific or challenging situations that are not sufficiently covered by official guidance or concept papers. Some of these ideas may in the future develop, through the formal process, into concept papers or guidance papers.

For example, one note discusses how to use an “expert evidence” approach to formulate recommendations when published direct evidence is lacking[11]. Since the first GRADE Note was published in 2021, the GRADE Notes currently published

Together, there were 6 notes. The most recent, by Yao et al. [12], *GRADE Notes 6: three strategies to determine the clinically important thresholds for outcomes in evidence-based guideline development* (GRADE Note 6: three strategies for determining thresholds for clinically important outcomes in evidence-based guideline development), focuses on the basis for threshold setting in order to enhance the clinical applicability and credibility of guideline recommendations. The three types of GRADE articles published from 2021 to 2025 are shown in Table 1.

## 2.4 GRADE-Related Articles

In addition to the three specific series described above, the *Journal of Clinical Epidemiology* (JCE) and other journals also publish other research or review articles on GRADE. Such articles may be written by members of the GWG or by other researchers, and their content may involve evaluations of applications of the GRADE approach, empirical studies, or critical commentaries. However, because they have not undergone formal review and approval by the GWG, they represent only the individual academic views of the authors and do not constitute an official component of GRADE methodology.

## 3 Recent Advances in GRADE Guidance Articles

The continuous publication of GRADE guidance articles indicates that the system is evolving from providing a basic framework to addressing, in greater depth and with greater refinement, the challenges of assessing specific complex bodies of evidence. The core objective of this series of articles is to systematically expand, clarify, and optimize the application of the GRADE approach in multiple frontier and complex scenarios. This paper summarizes the latest advances from the following three aspects:

### 3.1 Expansion of Application Scenarios and Deepening of Methodological Fields

Several guidance articles have sought to apply GRADE principles to new fields beyond traditional evaluations of therapeutic efficacy, and have developed dedicated assessment criteria for these fields. For example, YANG, Mustafa et al. [13–14] and related parts systematically constructed a complete method for assessing the quality of evidence on diagnostic accuracy, covering domains such as study design, risk of bias, indirectness, inconsistency, imprecision, and publication bias. Foroutan et al. [15] applied the GRADE framework to prognostic-factor studies, clarifying the specific features of observational evidence with

respect to initial quality, downgrading, and upgrading criteria. In addition, Brozek et al. [16] provided the first systematic overview of the assessment of model evidence, emphasizing that the certainty of evidence is jointly determined by the nature of the model inputs and of the model itself, and proposed a hierarchical framework for selecting the best model evidence.

Beyond pairwise comparisons, GRADE has also begun to focus on methods for rating the certainty of evidence from network meta-analysis (NMA): Petersen et al. [16] established a standardized process for assessing imprecision [17], emphasizing the need to comprehensively consider the relationship between the boundaries of the 95% confidence interval and thresholds, the magnitude of the effect, and the optimal information size (OIS) calculated on the basis of the “effective sample size” ; within the evidence-downgrading framework

**Table 1 Summary of GRADE guidance papers, concept papers, and notes published from 2021 to 2025**



| Year                                       | GRADE Guidance papers                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                | GRADE Concept papers                                                                                                                                                                                                                                                                                                           | GRADE Notes                                                                                                                                                                                                                                          |
|--------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Year                                       | GRADE Guidance papers                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                | GRADE Concept papers                                                                                                                                                                                                                                                                                                           | GRADE Notes                                                                                                                                                                                                                                          |
| 2021                                       | GRADE Guidelines 29: Rating the certainty of evidence for time-to-event outcomes—study limitations caused by missing data accompanying censoring of participants in intervention studies; GRADE Guidelines 30: GRADE methods for assessing the certainty of modeled evidence—an overall overview in the context of health decision-making; GRADE Guidance 31: Assessing the certainty of a body of evidence from comparative studies of diagnostic test accuracy; GRADE Guidelines 32: GRADE guidance on selecting the target of certainty-of-evidence ratings; GRADE Guidelines 33: Addressing imprecision in network meta-analysis | GRADE Concept 1: Validating the effectiveness of the “F.A.C.E” tool in guideline implementation based on stakeholders’ perceptions of feasibility, acceptability, cost, and equity; GRADE Concept 2: A conceptual framework for assessing the certainty of prognostic model calibration in bodies of validation-study evidence | GRADE Notes: How to use GRADE when “evidence” does not exist?—a case study of an expert-evidence approach); GRADE Notes 2: Clarifying, at an early stage of guideline development, the criteria for searching for nonrandomized or indirect evidence |
| chinarxiv.org/items/chinarxiv-202607.00260 |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      |                                                                                                                                                                                                                                                                                                                                | Machine Translation                                                                                                                                                                                                                                  |

| Year | GRADE Guidance papers                                                                                                                                                                                                                                                                                                                                                        | GRADE Concept papers                                                                | GRADE Notes |
|------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------|-------------|
| 2022 | GRADE Guidance 24: Optimizing the integration of randomized and nonrandomized intervention studies in evidence synthesis and health guidelines; GRADE Guidance 34: Recent advances in rating imprecision using a minimally contextualized approach; GRADE Guidance 35: Updates on rating imprecision for assessing contextualized certainty of evidence and making decisions | GRADE Concept 3: Biological plausibility in environmental health systematic reviews | —           |



| Year | GRADE Guidance papers                                                                                                                                                                                                                                                                                                                                                                                                  | GRADE Concept papers                                                                                                                                                                                                                                                                                                                                                                                                                                                       | GRADE Notes                                                                                                                     |
|------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------|
| 2023 | GRADE Guidance 23: Incorporating consideration of cost-effectiveness evidence in the process from evidence to health-related recommendations; GRADE Guidance 36: Update of GRADE methods for addressing inconsistency; GRADE Guidance 36: Rating imprecision in bodies of evidence on diagnostic accuracy; GRADE Guidance 38: Latest guidance on increasing certainty of evidence based on dose-response relationships | GRADE Concept 4: Rating certainty of evidence when the intervention or comparator in a study is inconsistent with the PICO target; GRADE Concept 5: Addressing intransitivity in network meta-analysis; GRADE Concept 6: A new method for incorporating external indirect evidence into GRADE certainty-of-evidence ratings in network meta-analysis; GRADE Concept 7: Issues and insights in linking guideline recommendations with trustworthy essential medicines lists | GRADE Notes 3: Comparison of two methods used by two Dutch guideline organizations to assess bias due to commercial sponsorship |

| Year | GRADE Guidance papers                                                                                                                                                                                                                                         | GRADE Concept papers                                                                                                            | GRADE Notes                                                                                                                                                                                                                                            |
|------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 2024 | GRADE Guidance 39: Using the GRADE-ADOLOPMENT approach to adopt, adapt, or develop contextualized recommendations based on source guidelines and evidence syntheses;                                                                                          | GRADE Concept 8: Assessing the certainty of prognostic-model discrimination performance in bodies of validation-study evidence; | GRADE Notes 4: How to use GRADE when “evidence” does not exist?—a case study using unpublished registry data                                                                                                                                           |
| 2025 | GRADE Guidance 41: GRADE certainty-of-evidence rating: an assessment method based on thresholds rather than contextualized categories; GRADE Guidance 42: GRADE guidance for judging health benefits and risks in decision-making on the basis of thresholds; | GRADE Concept 9: Principles and processes for constructing the GRADE ontology;                                                  | GRADE Notes 5: Practical and methodological issues in comparing new interventions with existing treatments; GRADE Notes 6: Three strategies for determining thresholds for the clinical importance of outcomes in evidence-based guideline development |

Note: —indicates that there is no content for this item.

Among these, Petersen et al.[18] explicitly defined “intransitivity” as a specific form within the domain of “indirectness.” Its logic for downgrading is based on how differences in population, comparator, or outcome (PCO) characteristics across the direct evidence involved in indirect comparisons modify effect estimates. Assessment requires weighing the credibility, magnitude, and distribu-

tion of effect modifiers, and care must be taken to avoid duplicate downgrading for PICO indirectness or risk of bias. “Incoherence/Inconsistency,” by contrast, focuses on reflecting the statistical contradiction between direct and indirect evidence in network estimates. Although intransitivity is often the root cause of inconsistency, to avoid double-counting limitations in the evidence, the GRADE framework generally requires downgrading only once when the two are related. In addition, Hao et al.[19] introduced the application of external indirect evidence and transitive logic, proposing that when the logical inference “if A is superior to B and B is non-inferior to C, then A is superior to C” is satisfied and supported by strong external evidence, a lower certainty rating originally derived solely from internal NMA evidence may be revised and upgraded.

### 3.2 Refinement and Operational Updates of Assessment Domains in Evidence-Certainty Rating

The core assessment domains for rating certainty of evidence in the GRADE system, such as imprecision and inconsistency, have been interpreted in greater depth and updated operationally. With regard to the definition of “certainty of evidence,” before 2017 the wording used was that “the point estimate represents the true effect.” In 2017, the GWG proposed a correction, changing this to “the true effect lies on one side of a threshold or within a specified range” [20], and used “contextualization” for classification. However, to communicate the rating objective more clearly and resolve users’ confusion about difficulties with the previous classifications, Hultcrantz et al.[21] stated that the GWG had discontinued contextualization classifications—that is, terms such as “non-contextualized,” “partially contextualized,” and “minimally contextualized” are no longer used; instead, attention is directed to the choice of thresholds [such as the “null-effect threshold” and the “minimal important difference (MID) threshold”]. Nevertheless, determining thresholds is often a challenge in practice. Determining decision thresholds must follow a rigorous methodological pathway. Current research has explored multiple strategies, such as using evidence from the literature, surveying expert panels, deriving thresholds from generic coefficient models, and expert consensus[22–24]. First, the existing literature on threshold studies is one important route for determining thresholds: widely accepted thresholds can be obtained through searching and rigorous appraisal, or thresholds with general applicability can be determined through quantitative synthesis of threshold estimates from existing studies. Zeng et al.[23] proposed using the ping-pong method to survey guideline expert panels in order to obtain thresholds. GRADE Guidance 42 proposed a method for deriving large, moderate, and small thresholds by using generic coefficients obtained from empirical data[6] and integrating the utility values of outcome indicators. When threshold evidence is lacking or of low quality, or when major controversy exists, the expert consensus method uses a structured process to bring together domain experts’ clinical experience and judgments to reach agreement.

### 3.3 Standardization of Results Presentation and Linkage to Decision-Making

Another focus of this series of articles is optimizing the communication of evidence and its translation into recommendations. Santesso et al.[25] provided clear wording templates for how to write conclusive statements in systematic reviews in a standardized manner according to certainty of evidence. Skoetz et al.[26] and Goldkuhle et al.[27] focused on the special data type of time-to-event outcomes, respectively addressing methods for calculating absolute effects in the “summary of findings table, SoF table,” and issues in assessing risk of bias caused by loss-to-follow-up censoring data. In linking evidence with decision-making, Article 23[28] described in detail how cost-effectiveness evidence can be integrated into the Evidence to Decision framework. The most recently published Guidance Article 42, *GRADE Guidance: Using Thresholds for Judgments on Health Benefits and Harms in Decision Making*[6] (*Using thresholds to judge health benefits and harms in health decision-making*), emphasizes that when making decisions through a balance of benefits and harms during guideline development, decision thresholds should be set in advance to improve the transparency and structure of judgments in the “Evidence to Decision (EtD)” framework. These updates not only enhance the applicability and robustness of the GRADE approach across different study types and clinical questions, but also greatly promote the transparency, structure, and scientific rigor of the process from evidence to recommendation decision-making.

## 4 GRADE Handbook and GRADE Book

- (1) The *GRADE Handbook*[1] is written and maintained by the GWG. Its final update was completed in 2013. It contains 12 chapters and provides a relatively systematic and comprehensive introduction to the GRADE approach.
- (2) The *GRADE Book*[29] is a “next-generation” methodological book launched by the GWG. It is still being updated and is planned to fully replace the *GRADE Handbook* by the end of 2026. The book aims to systematically update and expand the content of the *Handbook*, covering emerging application areas such as diagnostic tests, public health, and policy decision-making, and to provide users with practical guidance to support scientific decision-making in the processes of certainty-of-evidence assessment and recommendation development.

This update reflects a trend in methodology toward dynamization, continuous updating, and systematized development. During the transition period, only some chapters have been updated and released publicly, while new content remains under continuous review and refinement by the GWG and GCG. For chapters that have not yet been updated, the original *GRADE Handbook* may still be consulted. Compared with the *Handbook*, which has not been updated since 2013, the *GRADE Book* adopts a modular structure, is more comprehen-

sive in content, and incorporates the latest guideline practices and examples, making it better able to meet the current complex environment for evidence application.

In addition to methodological updates, to improve the accessibility and interactivity of the *GRADE Book*, the GWG has launched GRADE GPT—an intelligent GRADE assistant based on an artificial-intelligence large language model. According to the official description, GRADE GPT is built on ChatGPT and is trained specifically on the content of the *GRADE Book*. GRADE GPT uses only officially approved content, including formally released GRADE guidance, concepts, and materials, thereby ensuring scientific rigor and consistency. The application is currently in the testing stage and is undergoing continuous development and optimization. Its purpose is to provide users with an intuitive and interactive way to use GRADE, enabling users to obtain guidance and resources from the *GRADE Book* in real time. However, large models still face challenges such as “hallucinations”; therefore, large models should be used only as auxiliary tools to help users understand the methodological principles of GRADE.

## 5 Introduction to Core GRADE

Core GRADE is a series of articles published in *BMJ* beginning in April 2025 by Gordon and others, one of the co-founders of GRADE.[30–37] It aims to address problems that have emerged over more than 20 years in the development of the traditional GRADE approach, including scattered guidance articles, high operational complexity, and a high threshold for users to learn and apply it, by constructing a streamlined and systematic core guidance framework. It integrates and optimizes the core principles of GRADE, aiming to provide “efficient and precise” tools for evidence assessment and recommendation for systematic review authors, guideline developers, HTA practitioners, and educators in evidence-based medicine, and to avoid the dilemma in which users must search across multiple articles for guidance. At present, this series comprises eight papers. The first mainly explains why Core GRADE is needed; the remaining articles are briefly introduced as follows:

- (1) Core GRADE 1: overview of the Core GRADE approach[31] provides an overview of the overall framework of Core GRADE, explains the core principles of evidence certainty assessment, and describes how evidence is translated into recommendations or decisions.
- (2) Core GRADE 2: choosing the target of certainty rating and assessing imprecision[37] focuses on key issues in certainty assessment, including the target of certainty assessment, the positioning of different study designs, and how to downgrade for imprecision. It also emphasizes the “minimal important difference (MID)” as a key threshold in guideline development.
- (3) Core GRADE 3: rating certainty of evidence—assessing inconsistency[34] discusses inconsistency among study results. When the target of certainty rating is the null line, if the effect estimates of the great majority of studies

are in the same direction (i.e., all located on the same side of the null line), then even if there is statistical heterogeneity such as differences in point estimates or limited overlap of confidence intervals, downgrading for inconsistency is usually not recommended. Conversely, if the MID is used as the threshold, once the effect estimates of individual studies are distributed on both sides of the MID line (i.e., some studies suggest a clinically important effect whereas others do not), even if the overall pooled effect exceeds the MID, the inference of an “important benefit” will be weakened, and downgrading should therefore be considered. This article also provides new strategies for PICO construction and hypothesis setting for subgroup issues. It also introduces ICEMAN (Instrument to Assess the Credibility of Effect Modification Analyses)[38], a tool for assessing the credibility of subgroup analysis results, to help researchers judge the credibility of subgroup analyses.

- (4) Core GRADE 4: rating certainty of evidence—risk of bias, publication bias, and reasons for rating up certainty[30] concerns the assessment of risk of bias, the possibility of rating up evidence certainty because of a large effect size or dose-response relationship, and the impact of publication bias. The article notes that, because Cochrane RoB 2 is overly complex in design and algorithmically cumbersome, leading to poor agreement among assessors, ROBUST-RCT achieves the greatest possible simplification while maintaining methodological rigor and is more consistent with Core GRADE’s principle of ease of use; at the same time, it may be widely applied in the future. Therefore, when assessing risk of bias, Core GRADE recommends using ROBUST-RCT in randomized controlled trials to replace the traditional Cochrane RoB 2 tool; in non-randomized studies, it recommends using the revised CLARITY framework to replace ROBINS-I, in order to reduce operational complexity while maintaining scientific rigor.
- (5) Core GRADE 5: rating certainty of evidence—assessing indirectness[33] focuses on issues related to indirectness, including the applicability of research evidence to the target population, intervention, comparison, and outcomes, as well as the handling of surrogate outcomes.
- (6) Core GRADE 6: presenting the evidence in summary of findings tables[36] introduces the structure and creation methods of the Core GRADE SoF table. This table is the key single source for presenting the effects of interventions on outcomes important to patients; discussion of certainty and recommendations should be placed in the SoF table, with emphasis on attention to benefits, harms, and certainty (especially around the MID). The article also emphasizes distinguishing between “continuous” and “categorical” outcome indicators and provides guidance on how these two types of outcome indicators should be presented and interpreted.
- (7) Core GRADE 7: principles for moving from evidence to recommendations and decisions[35] discusses the major challenges and corresponding approaches faced by guideline developers and HTA practitioners in the

EtD process. The article emphasizes that, when formulating guideline recommendations, the primary factors should first be considered—namely the benefits and risks of interventions, certainty of evidence, and patients’ values and preferences—whereas secondary factors (such as cost, feasibility, acceptability, and equity) are mainly important for guidelines developed from a population perspective or at the public-health level. In addition, by determining the MID for each important outcome, it can be ensured that recommendations truly focus on outcomes that matter most to patients and have clinical significance.

The release of Core GRADE aims to enable the GRADE approach to fully realize its potential: to provide more reliable judgments of evidence certainty in systematic reviews, clinical practice guidelines, and HTA, and to provide more transparent and traceable bases for recommendations in guidelines, thereby optimizing clinical decision-making. However, it focuses on clinical decision-making scenarios involving a “single intervention and single comparator,” [32] and therefore, when complex clinical decisions involving non-one-to-one comparisons (such as complex interventions and NMA) are involved, researchers still need to refer to the complete GRADE methodological system for assessment.

## 6 Evaluation and Current Status of Use of the GRADE Method

Based on a cross-sectional study by Qureshi et al.[39] of 309 clinical practice guidelines recently issued by mainstream medical societies in the United States, the GRADE grading system has established a leading position in the development of clinical practice guidelines by mainstream U.S. medical societies and has been adopted by 61% of societies as the core evidence assessment tool. At the same time, studies have confirmed that use of the GRADE approach can not only statistically significantly improve the overall quality appraisal of guidelines, but can also specifically improve the transparency and standardization of guidelines in balancing clinical benefits and the costs of medical resources.[40] When used in combination with the Appraisal of Guidelines Research and Evaluation II (AGREE II) instrument, it can further optimize the methodological rigor and reporting quality of guidelines.[41] It is worth noting that, although the GRADE approach, as a systematic tool for assessing evidence quality and formulating clinical recommendations, has gained broad recognition and application worldwide, whether it is actually applied correctly and reasonably remains to be investigated.

The GRADE Handbook[1] first proposed evaluating the GRADE method criteria for adherence. In 2016, these criteria were updated on the GWG website<sup>[2]</sup>, and the GRADE Book<sup>[29]</sup> described the minimum requirements for the use of GRADE in statements published in *JCE* in 2024, including seven criteria (four for evaluating the certainty of evidence and three for developing recommendations). At the same time, the Australian GRADE Centre also developed

GRADE adherence assessment criteria for clinical practice guidelines, comprising seven categories and 18 questions<sup>[42]</sup>; all 18 questions can be answered “yes” or “no,” and all “yes” responses are summarized to calculate the percentage score for each guideline’s adherence to GRADE. Although the GWG has established explicit criteria, there is still a lack of authoritative tools for assessing whether guidelines have adopted the GRADE approach.

In 2020, after Barker et al.<sup>[42]</sup> surveyed Australian clinical guidelines, they found that the application of the GRADE approach in guidelines remained insufficient: even among guidelines claiming to use GRADE, adherence to the GRADE approach was still low in core steps such as ranking the importance of outcomes, using summary of findings tables (SoF tables), and applying EtD. In the same year, Dixon C et al.<sup>[43]</sup> investigated guidelines developed by American organizations based on the GRADE approach. The results showed that most guidelines clearly defined evidence quality and provided SoF tables, but only about one-tenth reported the basis for upgrading or downgrading evidence quality. Therefore, current guideline adherence to the GRADE approach still needs to be improved.

In terms of education and training, GRADE has become an important component of the evidence-based medicine teaching system<sup>[44]</sup>. Studies have compared the effects of online teaching and face-to-face teaching in training on the GRADE approach, indicating that its teaching models are gradually becoming more diverse<sup>[44]</sup>. From the perspective of development trends, the scope of application of the GRADE approach is continuously expanding: it is not limited to the development of traditional clinical guidelines, but is also gradually extending to health-resource allocation and cost-effectiveness analysis<sup>[28]</sup>, and has even been explored in environmental fields such as the assessment of water-resource carrying capacity<sup>[45]</sup>, demonstrating strong cross-disciplinary adaptability and potential for expansion.

## 7 Outlook

In recent years, in addition to updates and deepening in the methodological fields described above and the expansion of application scenarios, the GRADE approach has also attempted to integrate with AI and large language models, for example in intelligent and automated approaches to assessing the certainty of evidence. At present, the GRADE AI project team has begun exploring automated batch assessment and has developed the AI assistant GRADE GPT, which can provide real-time question answering and intelligent support based on the GRADE Book, guidance articles, and related materials. Meanwhile, threshold-based assessment methods provide a theoretical foundation for AI-driven automated evidence judgments. In the future, GRADE will further integrate AI, computable evidence systems, and global health issues—such as planetary health and essential medicines—to build a more open and standardized evidence-assessment ecosystem and promote the efficient and precise development of clinical decision-making and health-policy formulation.



Author contributions: Wang Liuyang was responsible for data collection and organization, figure and table preparation, and manuscript writing and revision; Liu Changning was responsible for data collection and manuscript review and revision; Shen Quan and Cai Zixuan were responsible for data collection and manuscript review; Ren Xiangying, Wang Yongbo, Huang Qiao, and Jin Yinghui were responsible for the conception and design of the article and manuscript review; Yan Siyu was responsible for the conception and design of the article, data collection, figure and table design, and manuscript review, and took overall responsibility for the article and supervision and management.

The authors declare no conflicts of interest.

## References

- [1] The Grade Working Group. GRADE Handbook[EB/OL]. [2025-09-11]. <https://gdt.gradepro.org/app/handbook/handbook.html>
- [2] The Grade Working Group. GRADE working group[EB/OL]. [2025-09-11]. <http://www.gradeworkinggroup.org>.
- [3] Werner S S, Binder N, Toews I, et al. Use of GRADE in evidence syntheses published in high-impact-factor nutrition journals: a methodological survey[J]. *J Clin Epidemiol*, 2021, 135: 54-69. DOI:10.1016/j.jclinepi.2021.02.010.
- [4] Deng Tong, Wang Yang, Huang Di, et al. Methods for developing clinical practice guidelines—Theoretical article on the GRADE approach[J]. *Chinese Journal of Evidence-Based Cardiovascular Medicine*, 2018, 10(12): 1441-5+9.
- [5] McGowan J, Akl E A, Coello P A, et al. Update on the JCE GRADE series and other GRADE article types[J]. *J Clin Epidemiol*, 2021, 140: 163-164. DOI:10.1016/j.jclinepi.2021.05.023.
- [6] Wiercioch W, Morgano G P, Piggott T, et al. GRADE guidance: using thresholds for judgments on health benefits and harms in decision making (GRADE guidance 42)[J]. *Ann Intern Med*, 2025, 178(11): 1644-1652. DOI:10.7326/annals-24-02013.
- [7] Schünemann H J, Brennan S, Akl E A, et al. The development methods of official GRADE articles and requirements for claiming the use of GRADE - A statement by the GRADE guidance group[J]. *J Clin Epidemiol*, 2023, 159: 79-84. DOI:10.1016/j.jclinepi.2023.05.010.
- [8] Piggott T, Moja L, Jenei K, et al. GRADE concept 7: issues and insights linking guideline recommendations to trustworthy essential medicine lists[J]. *J Clin Epidemiol*, 2024, 166: 111241. DOI:10.1016/j.jclinepi.2023.111241.
- [9] Foroutan F, Mayer M, Guyatt G, et al. GRADE concept paper 8: judging the certainty of discrimination performance estimates of prognostic models in a body of validation studies[J]. *J Clin Epidemiol*, 2024, 170: 111344. DOI:10.1016/j.jclinepi.2024.111344.

- [10] Whaley P, Alper B, Dehnbostel J, et al. GRADE concept paper 9: rationale and process for creating a GRADE Ontology[J]. *J Clin Epidemiol*, 2025, 187: 111921. DOI:10.1016/j.jclinepi.2025.111921.
- [11] Mustafa R A, Garcia C A C, Bhatt M, et al. GRADE notes: How to use GRADE when there is “no” evidence A case study of the expert evidence approach[J]. *J Clin Epidemiol*, 2021, 137: 231-235. DOI:10.1016/j.jclinepi.2021.02.026.
- [12] Yao X M, Xia J, Deodat M, et al. GRADE notes 6: Three strategies to determine the clinically important thresholds for outcomes in evidence-based guideline development[J]. *J Clin Epidemiol*, 2025, 188: 111976. DOI:10.1016/j.jclinepi.2025.111976.
- [13] Mustafa R A, El Mikati I K, Murad M H, et al. GRADE guidance 37: rating imprecision in a body of evidence on test accuracy[J]. *J Clin Epidemiol*, 2024, 165: 111189. DOI:10.1016/j.jclinepi.2023.10.005.
- [14] Yang B D, Mustafa R A, Bossuyt P M, et al. GRADE Guidance: 31. Assessing the certainty across a body of evidence for comparative test accuracy[J]. *J Clin Epidemiol*, 2021, 136: 146-156. DOI:10.1016/j.jclinepi.2021.04.001.
- [15] Foroutan F, Guyatt G, Zuk V, et al. GRADE Guidelines 28: Use of GRADE for the assessment of evidence about prognostic factors: rating certainty in identification of groups of patients with different absolute risks[J]. *J Clin Epidemiol*, 2020, 121: 62-70. DOI:10.1016/j.jclinepi.2019.12.023.
- [16] Brozek J L, Canelo-Aybar C, Akl E A, et al. GRADE Guidelines 30: the GRADE approach to assessing the certainty of modeled evidence—An overview in the context of health decision-making[J]. *J Clin Epidemiol*, 2021, 129: 138-150. DOI:10.1016/j.jclinepi.2020.09.018.
- [17] Brignardello-Petersen R, Guyatt G H, Mustafa R A, et al. GRADE guidelines 33: Addressing imprecision in a network meta-analysis[J]. *J Clin Epidemiol*, 2021, 139: 49-56. DOI:10.1016/j.jclinepi.2021.07.011.
- [18] Brignardello-Petersen R, Tomlinson G, Florez I, et al. Grading of recommendations assessment, development, and evaluation concept article 5: addressing intransitivity in a network meta-analysis[J]. *J Clin Epidemiol*, 2023, 160: 151-159. DOI:10.1016/j.jclinepi.2023.06.010.
- [19] Hao Q K, Gao Y, Zhao Y L, et al. GRADE concept 6: a novel application of external indirect evidence into GRADE ratings of evidence certainty in network meta-analysis[J]. *J Clin Epidemiol*, 2023, 163: 95-101. DOI:10.1016/j.jclinepi.2023.09.006.
- [20] Hultcrantz M, Rind D, Akl E A, et al. The GRADE Working Group clarifies the construct of certainty of evidence[J]. *J Clin Epidemiol*, 2017, 87: 4-13. DOI:10.1016/j.jclinepi.2017.05.006.

- [21] Hultcrantz M, Schünemann H J, Mustafa R A, et al. GRADE certainty ratings: thresholds rather than categories of contextualization (GRADE guidance 41)[J]. *Ann Intern Med*, 2025, 178(8): 1183-1186. DOI:10.7326/annals-25-00548.
- [22] Morgano G P, Wiercioch W, Piovani D, et al. Defining decision thresholds for judgments on health benefits and harms using the grading of recommendations assessment, development, and evaluation (GRADE) evidence to decision (EtD) frameworks: a randomized methodological study (GRADE-THRESHOLD)[J]. *J Clin Epidemiol*, 2025, 179: 111639. DOI:10.1016/j.jclinepi.2024.111639.
- [23] Zeng L N, Helsing L M, Bretthauer M, et al. A novel framework for incorporating patient values and preferences in making guideline recommendations: guideline panel surveys[J]. *J Clin Epidemiol*, 2023, 161: 164-172. DOI:10.1016/j.jclinepi.2023.07.003.
- [24] Zhan Yinghui, Zhou Pei, Zhao Dan, et al. Overview of the theory and methods for determining decision thresholds in clinical practice guidelines[J]. *Journal of Nurses Training*, 2025, 40(24): 2604-10.
- [25] Santesso N, Glenton C, Dahm P, et al. GRADE guidelines 26: informative statements to communicate the findings of systematic reviews of interventions[J]. *J Clin Epidemiol*, 2020, 119: 126-135. DOI:10.1016/j.jclinepi.2019.10.014.
- [26] Skoetz N, Goldkuhle M, Van Dalen E C, et al. GRADE guidelines 27: how to calculate absolute effects for time-to-event outcomes in summary of findings tables and Evidence Profiles[J]. *J Clin Epidemiol*, 2020, 118: 124-131. DOI:10.1016/j.jclinepi.2019.10.015.
- [27] Goldkuhle M, Bender R, Akl E A, et al. GRADE Guidelines: 29. Rating the certainty in time-to-event outcomes—Study limitations due to censoring of participants with missing data in intervention studies[J]. *J Clin Epidemiol*, 2021, 129: 126-137. DOI:10.1016/j.jclinepi.2020.09.017.
- [28] Xie F, Shemilt I, Vale L, et al. GRADE guidance 23: considering cost-effectiveness evidence in moving from evidence to health-related recommendations[J]. *J Clin Epidemiol*, 2023, 162: 135-144. DOI:10.1016/j.jclinepi.2023.08.001.
- [29] The Grade Working Group. The GRADE Bookversion 1.0[EB/OL]. [2025-09-11]. <https://book.grade.pro.org>.
- [30] Core Grade 4: Rating Certainty Of Evidence—Risk Of Bias, Publication Bias, And Reasons For Rating Up certainty[J]. *Bmj*, 2025, 390: r1468.
- [31] Guyatt G, Agoritsas T, Brignardello-Petersen R, et al. Core GRADE 1: overview of the core GRADE approach[J]. *Bmj*, 2025, 389: e081903. DOI:10.1136/bmj-2024-081903.
- [32] Guyatt G, Hultcrantz M, Agoritsas T, et al. Why Core GRADE is

needed: introduction to a new series in The BMJ[J]. *Bmj*, 2025, 389: e081902. DOI:10.1136/bmj-2024-081902.

[33] Guyatt G, Iorio A, De Beer H, et al. Core GRADE 5: rating certainty of evidence—assessing indirectness[J]. *Bmj*, 2025, 389: e083865. DOI:10.1136/bmj-2024-083865.

[34] Guyatt G, Schandelmaier S, Brignardello-Petersen R, et al. Core GRADE 3: rating certainty of evidence—assessing inconsistency[J]. *Bmj*, 2025, 389: e081905. DOI:10.1136/bmj-2024-081905.

[35] Guyatt G, Vandvik P O, Iorio A, et al. Core GRADE 7: principles for moving from evidence to recommendations and decisions[J]. *Bmj*, 2025, 389: e083867. DOI:10.1136/bmj-2024-083867.

[36] Guyatt G, Yao L, Murad M H, et al. Core GRADE 6: presenting the evidence in summary of findings tables[J]. *Bmj*, 2025, 389: e083866. DOI:10.1136/bmj-2024-083866.

[37] Guyatt G, Zeng L N, Brignardello-Petersen R, et al. Core GRADE 2: choosing the target of certainty rating and assessing imprecision[J]. *Bmj*, 2025, 389: e081904. DOI:10.1136/bmj-2024-081904.

[38] Schandelmaier S, Briel M, Varadhan R, et al. Development of the Instrument to assess the Credibility of Effect Modification Analyses (ICEMAN) in randomized controlled trials and meta-analyses[J]. *Can Med Assoc J*, 2020, 192(32): E901–E906. DOI:10.1503/cmaj.200077.

[39] Qureshi O, Ross J S, Ramachandran R, et al. Level of evidence and strength of recommendations in US medical society clinical practice guidelines, 2019–2023: a cross-sectional analysis[J]. *J Gen Intern Med*, 2025. DOI:10.1007/s11606-025-10088-6.

[40] Fotteler M L, Kocar T D, Willems J, et al. “What makes a good guideline – A systematic review and analysis of 120 clinical practice guidelines in orthogeriatric care and related fields using the AGREE II tool” [J]. *J Clin Epidemiol*, 2025, 184: 111830. DOI:10.1016/j.jclinepi.2025.111830.

[41] Ramakrishnan V M, Yim K, Wollin D A. Is the AGREE II framework the appropriate tool to evaluate global urolithiasis guidelines [J]. *Transl Androl Urol*, 2023, 12(9): 1366–1367. DOI:10.21037/tau-23-411.

[42] Barker T H, Dias M, Stern C, et al. Guidelines rarely used GRADE and applied methods inconsistently: a methodological study of Australian guidelines[J]. *J Clin Epidemiol*, 2021, 130: 125–134. DOI:10.1016/j.jclinepi.2020.10.017.

[43] Dixon C, Dixon P E, Sultan S, et al. Guideline developers in the United States were inconsistent in applying criteria for appropriate Grading of Recommendations, Assessment, Development and Evaluation use[J]. *J Clin Epidemiol*, 2020, 124: 193–199. DOI:10.1016/j.jclinepi.2020.01.026.

[44] Tokalić R, Poklepović Peričić T, Marušić A. Similar outcomes of web-based and face-to-face training of the GRADE approach for the certainty of evidence: randomized controlled trial[J]. J Med Internet Res, 2023, 25: e43928. DOI:10.2196/43928.

[45] Yang J L, Yang P, Zhang S Q, et al. Evaluation of water resource carrying capacity in the middle reaches of the Yangtze River Basin using the variable fuzzy-based method[J]. Environ Sci Pollut Res, 2023, 30(11): 30572-30587. DOI:10.1007/s11356-022-24103-4.

(Received: 2025-12-21; revised: 2026-01-30)

(Editor of this article: Wang Shiyue)

*Note: Figure translations are in progress. See original paper for figures.*

*Source: ChinaXiv. Machine translation. Verify with the original.*