



Education Article

GRADE Methods in traditional medicine

Jian-Ping Liu*

Centre for Evidence-Based Chinese Medicine, Beijing University of Chinese Medicine, Beijing, China



ARTICLE INFO

Article history:

Received 10 November 2021

Revised 7 January 2022

Accepted 11 January 2022

Available online 26 January 2022

Keywords:

Grade methods

Traditional medicine

Clinical practice guideline

Quality of evidence

Strength of recommendations

ABSTRACT

The GRADE (Grading of Recommendation, Assessment, Development and Evaluation) methods was developed to evaluate the quality of evidence and make recommendations, which has been widely adopted in clinical practice guidelines. The GRADE methods address the classification of outcomes, systematic collection, appraisal and synthesis of research evidence based on each outcome, evaluate the overall quality of the evidence, and making recommendations. This essay summarized the GRADE methods and its use in clinical practice guidelines of traditional, complementary, and integrative medicine as well as highlighting some of the challenges.

© 2022 Korea Institute of Oriental Medicine. Published by Elsevier B.V.

This is an open access article under the CC BY-NC-ND license

(<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

1. Introduction

Evidence is the core component in evidence-based medicine and practice. Since the introduction of clinical epidemiology in early 1980's, there has been an increasing amount of clinical evidence for healthcare. However, when using evidence for decision making, professionals have to systematically collect, critically appraise, and synthesize the evidence. As for the definition of evidence-based medicine by Dr David Sackett in his paper in 1996, he addressed "the conscientious, explicit, and judicious use of current best evidence in making decisions about the care of individual patients".¹ Furthermore, the best evidence should be updated, reliable, and accessible to professionals and decision makers. It suggests that we need to differentiate the best, reliable evidence from those unreliable evidence. Obviously, we need a validated tool or instrument to evaluate the evidence, and evidence should be graded according to their strength and weakness. There have been international organizations such as the World Health Organisation (WHO), the Cochrane Collaboration and the GRADE Working Group to develop tools for evidence grading in the necessity of development of clinical practice guidelines. The acronym of GRADE represents the Grading of Recommendation, Assessment, Development and Evaluation.

Since 2005, there has been an initiative on evidence-based approach from East Asian countries to develop clinical practice

guidelines in traditional medicine, supported by the WHO Western Pacific Regional Office, and this included traditional Chinese medicine, Korean medicine, and integrative medicine.² It has been a common sense that clinical practice guidelines should be based on high level of clinical evidence. Thus, different hierarchy models have been developed to grade the evidence and these include WHO, the Cochrane Collaboration, the Oxford evidence grading, and the GRADE systems. Based on the WHO Handbook for Guideline Development, the outline of the guideline development (www.who.int) can be presented at four stages (Fig. 1). It is also suggested that involvement of multidisciplinary development team, Delphi process, expert consensus, and consumer review are important methods for guideline development.

2. GRADE methods in clinical practice guideline

The GRADE working group (www.gradeworkinggroup.org) was formed in 2004,³ and it has developed series of methodology guidelines for development of clinical practice guidelines.⁴ Since then, the working group has developed into many centers and networks worldwide. For example, in China, four GRADE centers have been established, including GRADE Center Lanzhou University, GRADE Center The University of Nottingham Ningbo, GRADE Center Beijing University of Chinese Medicine, and GRADE Center Fudan University. These centers are hosted by universities which have demonstrated expertise and capacity in GRADE methodology, including capacity in systematic reviews and guideline development, and are able to provide the infrastructure to support and promote GRADE-related activities such as training on and support for methodology of guideline development.

* Corresponding author at: Centre for Evidence-Based Chinese Medicine, Beijing GRADE Center, Cochrane China Network Affiliate, Beijing University of Chinese Medicine, Beijing, 100029, China.

E-mail address: Liujp@bucm.edu.cn



Fig. 1. Flowchart of guideline development.

Since the establishment of the GRADE working group, GRADE guidance has been endorsed by many international organizations such as the WHO, the Cochrane Collaboration, the National Institute for Health and Care Excellence (NICE), Scottish Intercollegiate Guidelines Network (SIGN), US Center for Disease Control and Prevention (CDC), as well as many international peer review journals.

3. Quality of evidence by grade

GRADE system has two major components, rating the certainty of evidence and strength of recommendations.⁴ The systematic reviewers or guideline authors are developing clinical questions based on PICO model (in terms of patients, intervention, comparison, and outcome), classify the outcomes into critical, important, and less important outcomes. Thus, a systematic approach is adapted to collect, appraise, and synthesize the evidence to generate the estimate of effect from eligible studies for each outcome, and then, rating the quality of evidence body into four categories: high, moderate, low, or very low. There has been two-level of quality assessment. One is focusing on primary studies such as randomized clinical trials, and other is on overall evidence for each outcome. For the quality of evidence of randomized trials, the Cochrane Collaboration developed a tool called risk of bias, which covers six domains of bias: selection bias, performance bias, detection bias, attrition bias, reporting bias, and other bias.^{5,6} Within each domain, assessments are made for one or more items, which may cover different aspects of the domain, or different outcomes. The difference of the GRADE is to evaluate the overall quality of evidence body for each outcome based on systematic review or meta-analyses.^{7,8}

4. Factors that downgrade or upgrade the evidence

There are many determinants that influence the quality of evidence. These include study design and execution, inconsistency, indirectness, imprecision, and reporting bias. Existence of one or more above domains should downgrade the evidence to one or two levels. For study design and execution, randomized trials rank as high quality, while observational studies (cohort study and case-control study) rank as low. However, for individual randomized tri-

als, the Cochrane risk of bias tool can be used for the assessment of quality,^{5,6} which mainly include generation of allocation sequence, concealment of the allocation, blinding, and loss to follow up. Ideally, sample size estimation should be justified and baseline characteristics should be tested statistically for comparability between groups. Inconsistency (statistical term heterogeneity) represents the variations in results from meta-analysis. There are three ways to make judgement for consistency, that is, variations in effect size, the degree of overlap in confidence intervals, and statistical significance of heterogeneity through I square test. Imprecision represents a small sample size or lower event rate with wide confidence interval, which may bring uncertainty about magnitude of effect. Meta-analysis of two or more trials can reduce the imprecision by increasing the pooled sample and statistical power. If we can be sure reaching an optimal information size with narrow confidence interval and permit confident recommendation for or against, we don't downgrade the imprecision. In terms of directness of evidence, if we have head-to-head comparison, for example, drug A versus C, drug B versus C, we are confident about comparative effectiveness. But if we want to compare drug A with drug B, we don't have direct evidence. Then, an indirect comparison could be established between drug A and drug B. In addition, directness of evidence also relies on differences in patients (age, sex, ethnicity, condition, stage, severity), interventions (dose, class, duration), and outcomes (health related quality of life, functional capacity, laboratory biomarkers). For example, surrogate outcome is lower by one level for indirectness. Reporting bias comes from one of the two sources, including selective reporting of outcomes (especially those outcomes with positive results), and publication bias, where small sample size trials with positive findings are more likely to be published. Normally, we can make judgement on selective reporting of outcomes by checking the registered trial protocol against the full publication of trial. For detecting publication bias, reviewers can draw a funnel plot to check the asymmetry or performing Egger's test.

In what circumstances, we can upgrade the evidence? For large magnitude of effect from observational studies, dose-response gradient demonstrated, plausible confounders to be excluded for establishing association, these are the factors that can raise one or two level of evidence. Following quality assessment criteria are summarized in Table 1.

Table 1
Quality of evidence and downgrade/upgrade factors.

Quality of Evidence	Study Design	Lower if*	Higher if*
⊕⊕⊕⊕ High	Randomised trial	Study limitations - <i>Serious</i>	Large effect - <i>Large</i>
⊕⊕⊕O Moderate	Observational study	- <i>Very serious</i>	- <i>Very large</i>
⊕⊕OO Low		Inconsistency - <i>Serious</i>	Dose response - <i>Evidence of a gradient</i>
⊕OOO Very low		- <i>Very serious</i>	All plausible confounding - <i>Would reduce a demonstrated effect</i>
		Indirectness - <i>Serious</i>	- <i>Would suggest a spurious effect when results show no effect</i>
		Imprecision - <i>Serious</i>	
		- <i>Very serious</i>	
		Publication bias - <i>Likely</i>	
		- <i>Very likely</i>	

5. Strength of recommendations and evidence to decision (EtD) framework

The main compositions in guidelines are the recommendations. In GRADE methods, the strength of recommendations is defined as degree of confidence that desirable effects of adhering to recommendation outweigh undesirable effects. The recommendation can be strong or weak depending on the confidence that benefit outweigh risk, and the direction of recommendation can be favouring or against the recommendations. In the perspectives of patient value and preference, a strong recommendation suggest lower variability (for example, >90% patients would make the same choice), less interaction with patients, and no need for decision aid. Weak recommendations can also be named as conditional, or discretionary recommendations. For this category, there is a large variability in patient preference, ensuring interaction with patients, need for decision aid, and don't consider quality of care criterion.

In 2016, the DECIDE (Developing and Evaluating Communication Strategies to Support Informed Decisions and Practice Based on Evidence) Project, led by a GRADE Working Group and funded by the European Union, developed GRADE Evidence to Decision (EtD) framework for different types of decisions.^{9,10} The GRADE EtD framework takes an explicit and transparent system to help decision makers informed by the best available research evidence to make clinical recommendations, coverage decisions, and health system or public health recommendations and decisions. The framework includes formulation of the question, an assessment of the evidence, and drawing conclusions in general. For health care interventions, following criteria are considered: priority of the problems, benefits and harms, certainty of the evidence, outcome importance, balance between desirable and undesirable effects, resource use, equity, acceptability, and feasibility.⁹

6. Application of GRADE in guidelines of traditional and integrative medicine

In some countries such as China and South Korea, traditional medicine is widely used in clinical practice. One survey in China was done in 2015 in 604 conventional medicine guidelines which demonstrated about 12% of the guidelines recommending traditional Chinese medicine.¹⁰ Among the 74 guidelines that recommended Chinese medicine, only five guidelines used evidence grading. One of the clinical practice guidelines using GRADE is the guideline of traditional medicine for primary osteoporosis published in 2011.¹¹ In 2020, a critical review investigated the proportion of traditional Chinese medicine (TCM) guidelines adopting grading systems and the level of evidence used to support TCM recommendations.¹² Till 2018, 142 TCM guidelines were identified,

among which, 68 (47.9%) adopted a total of eight grading systems. A total of 1284 recommendations were included in the TCM guidelines. More than 60% recommendations were based on a low and very low level of evidence (33.4% and 30.2%) based on GRADE systems. Only 7.8% recommendations were rated as strong recommendation, while 76.2% recommendations were rated as conditional recommendations. The GRADE methods were also recommended for development of clinical guidelines on acupuncture and moxibustion.¹³ Eighteen evidence-based guidelines on acupuncture used GRADE methods in their development in China, and the guideline developers summarized advantages and limitations of the GRADE approach in the guideline development.¹³ The advantages were rating the quality of evidence, outcome-centric direction, and transparent process of recommendation development. However, they also addressed some limitations of the GRADE approach in acupuncture guidelines such as lack of evidence grading for ancient literature and literature on prestigious Chinese medicine experts' experience, and specific guidance concerning the characteristics of acupuncture therapy for formulating recommendations.¹³ Thus, the authors suggest that a specific method should be explored based on the GRADE approach and the characteristics of acupuncture therapy in the guideline development for clinical practice with acupuncture and moxibustion.

However, traditional, complementary and integrative medicine therapies are not well represented in majority of clinical practice guidelines. For example, guidelines for musculoskeletal pain excluded traditional healing/medicine or those that required payment from patient's self-pocket.¹⁴ On the other hand, integrative Chinese-Western medicine and Chinese herbal medicine were recommended for the management of Covid-19 based on current research evidence but without evidence grading and quality assessment.^{15,16,17,18} Similar guidelines for traditional Chinese medicine for Covid-19 adopted GRADE methods to grade evidence and make recommendation.¹⁹ Facing the unique features of traditional and integrative medicine, some experts advised to adapt consensus statements to develop clinical guidelines.²⁰

Below is an example of using GRADE to develop recommendations.²¹

Question: Should acupuncture versus usual care be used for knee OA?

GRADE evidence profile (Table 2)

Another example is a rapid development of traditional Chinese medicine guidelines for coronavirus disease 2019 by a panel of experts, who used GRADE methods for grading the evidence and making the recommendations. In addition, the guidelines were developed in accordance with the WHO rapid guideline process.²² The evidence on TCM for COVID-19 from published guide-

Table 2
Evidence summary on acupuncture versus usual care for knee osteoarthritis.

No of studies	Quality assessment										Importance of outcome	
	Summary of findings										Quality	Importance of outcome
	Design	Limitations	Inconsistency	Indirectness	Imprecision	Publication bias	Acupuncture	Usual care	Effect Relative (95% CI)	Absolute		
2	Pain (follow-up median 6 months; measured with: WOMAC; range of scores: 0–100; Better indicated by less)	not serious	not serious	not serious	not serious	none	520	531	NA	SMD -0.52 (-0.66 to -0.39)	⊕⊕⊕⊕ HIGH	CRITICAL
2	Function (follow-up 6 months; measured with: WOMAC; range of scores: 0–100; Better indicated by more)	not serious	not serious	not serious	not serious	none	520	531	NA	SMD -0.45 (-0.59 to -0.32)	⊕⊕⊕⊕ HIGH	CRITICAL
1	Hematoma (follow-up 6 months; number of patients reporting reaction)	not serious	not serious	not serious	not serious	none	12/326 (3.7%)	1/316 (0.3%)	11.63 (1.52 to 88.93) ⁵	3% more (0.2% to 26% more)	⊕⊕⊕⊕ HIGH	IMPORTANT

NA, not available.
Note: Limitations refer to methodological flaws and conduct; Hematoma refers to adverse events related to acupuncture.

lines, direct and indirect published clinical evidence, first hand clinical data, and expert experience and consensus were incorporated. Based on the available evidence, the guidelines recommended 17 Chinese medicines for COVID-19 including 2 Chinese herbal granules, 7 Chinese patent medicines, and 8 Chinese herbal injections.^{19,22}

In summary, some challenges exist when using GRADE for development of recommendations. One is lack of high-quality evidence in the field of traditional, complementary and integrative medicine. Second, there are huge variations among the interventions such as herbal remedy, acupuncture, tuina (therapeutic massage), taichi/qigong, Ayurveda and other traditional healing systems. Third, an obvious gap between research evidence and practice as research try to find out efficacy or effectiveness from standardized therapy while in daily practice, it is highly individualized, tailored treatment. Therefore, expert consensus is suggested to develop the recommendations not only relying on existing evidence.^{4,19,23}

7. Conclusions

GRADE system is an international recognized approach to evaluate quality of evidence and strength of recommendation specifically for development of clinical practice guideline. It has been widely adapted by many international organizations such as WHO, the Cochrane Collaboration, NICE, and US CDC. One of the key steps in guideline development by GRADE is to classify the outcomes into critical, important, and less important in accordance with relevant disease, and evaluate overall quality of evidence body. However, when adapting GRADE in guideline of traditional medicine or integrative medicine, five domains for downgrading of evidence should be carefully considered, which should include study design and execution, inconsistency, indirectness, imprecision and publication bias, while three domains for upgrading of evidence, that is, large effect size, dose-response and control of confounding factors. We recommend stakeholders involvement and context being considered as well as mixed approaches such as expert consensus incorporated with diversity of evidence for traditional medicine guidelines.

Author contribution

This is the sole author's work.

Conflict of interests

JP Liu is an editorial board member of the journal and this article was invited, but it was externally peer reviewed. There are no other conflicts of interests.

Funding

JP Liu was supported by The Key project of the [National Natural Science Foundation of China](#) (No. 81830115) "Key techniques and outcome research for therapeutic effect of traditional Chinese medicine as complex intervention based on holistic system and pattern differentiation & prescription". This work was partially supported by the [NCCIH](#) grant (AT001293 with sub-award No. 020468C).

Ethical statement

Not applicable.

Data availability

Not applicable.

References

- Sackett DL, Rosenberg WM, Gray JA, Haynes RB, Richardson WS. Evidence based medicine: what it is and what it isn't. *BMJ*. 1996;312(7023):71–72. doi:10.1136/bmj.312.7023.71.
- Yu WY, Xu JL, Shi NN, et al. Assessing the quality of the first batch of evidence-based clinical practice guidelines in traditional Chinese medicine. *J Tradit Chin Med*. 2011;31(4):376–381. doi:10.1016/s0254-6272(12)60021-1.
- Atkins D, Best D, Briss PA, GRADE Working Group, et al. Grading quality of evidence and strength of recommendations. *BMJ*. 2004;328(7454):1490. doi:10.1136/bmj.328.7454.1490.
- Guyatt GH, Oxman AD, Vist GE, GRADE Working Group, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ*. 2008;336(7650):924–926. doi:10.1136/bmj.39489.470347.AD.
- Higgins JP, Altman DG, Gøtzsche PC, Cochrane Bias Methods Group; Cochrane Statistical Methods Group, et al. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *BMJ*. 2011;343:d5928. doi:10.1136/bmj.d5928.
- Sterne JAC, Savović J, Page MJ, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. *BMJ*. 2019;366:14898. doi:10.1136/bmj.14898.
- Guyatt G, Oxman AD, Akl EA, et al. GRADE guidelines: 1. Introduction- GRADE evidence profiles and summary of findings tables. *J Clin Epidemiol*. 2011;64(4):383–394. doi:10.1016/j.jclinepi.2010.04.026.
- Guyatt G, Oxman AD, Sultan S, et al. GRADE guidelines: 11. Making an overall rating of confidence in effect estimates for a single outcome and for all outcomes. *J Clin Epidemiol*. 2013;66(2):151–157. doi:10.1016/j.jclinepi.2012.01.006.
- Alonso-Coello P, Schünemann HJ, Moberg J, GRADE Working Group, et al. GRADE Evidence to Decision (EtD) frameworks: a systematic and transparent approach to making well informed healthcare choices. 1: Introduction. *BMJ*. 2016;353:i2016. doi:10.1136/bmj.i2016.
- Ren J, Li X, Sun J, et al. Is traditional Chinese medicine recommended in Western medicine clinical practice guidelines in China? A systematic analysis. *BMJ Open*. 2015;5(6):e006572. doi:10.1136/bmjopen-2014-006572.
- Xie YM, Yuwen Y, Dong FH, et al. Clinical practice guideline of traditional medicine for primary osteoporosis. *Chin J Integr Med*. 2011;17(1):52–63. doi:10.1007/s11655-011-0613-6.
- Li J, Li B, Zhao XK, Tu JY, Li Y. A critical review to grading systems and recommendations of traditional Chinese medicine guidelines. *Health Qual Life Outcomes*. 2020;18(1):174. doi:10.1186/s12955-020-01432-x.
- Zhao H, Liang F, Fang Y, Liu B. Application of Grading of Recommendations Assessment, Development, and Evaluation (GRADE) to the guideline development for clinical practice with acupuncture and moxibustion. *Front Med*. 2017;11(4):590–594. doi:10.1007/s11684-017-0537-4.
- Lin I, Wiles L, Waller R, et al. What does best practice care for musculoskeletal pain look like? Eleven consistent recommendations from high-quality clinical practice guidelines: systematic review. *Br J Sports Med*. 2020;54(2):79–86. doi:10.1136/bjsports-2018-099878.
- Chan KW, Wong VT, Tang SCW. COVID-19: an Update on the Epidemiological, Clinical, Preventive and Therapeutic Evidence and Guidelines of Integrative Chinese-Western Medicine for the Management of 2019 Novel Coronavirus Disease. *Am J Chin Med*. 2020;48(3):737–762. doi:10.1142/S0192415X20500378.
- Ang L, Lee HW, Kim A, Lee JA, Zhang J, Lee MS. Herbal medicine for treatment of children diagnosed with COVID-19: a review of guidelines. *Complement Ther Clin Pract*. 2020;39:101174. doi:10.1016/j.ctcp.2020.101174.
- Ang L, Lee HW, Choi JY, Zhang J, Lee MS. Herbal medicine and pattern identification for treating COVID-19: a rapid review of guidelines. *Integr Med Res*. 2020;9(2):100407. doi:10.1016/j.imr.2020.100407.
- Ge L, Zhu H, Wang Q, et al. Integrating Chinese and western medicine for COVID-19: a living evidence-based guideline (version 1). *J Evid Based Med*. 2021;14(4):313–332. doi:10.1111/jebm.12444.
- Liang N, Ma Y, Wang J, et al. Traditional Chinese Medicine guidelines for coronavirus disease 2019. *J Tradit Chin Med*. 2020;40(6):891–896. doi:10.19852/j.cnki.jtcm.20200902.001.
- Hunter J, Leach M, Braun L, Bensoussan A. An interpretive review of consensus statements on clinical guideline development and their application in the field of traditional and complementary medicine. *BMC Complement Altern Med*. 2017;17(1):116. doi:10.1186/s12906-017-1613-7.
- Manheimer E, Linde K, Lao L, Bouter LM, Berman BM. Meta-analysis: acupuncture for osteoarthritis of the knee. *Ann Intern Med*. 2007;146(12):868–877. doi:10.7326/0003-4819-146-12-200706190-00008.
- Liang N, Li H, Wang J, et al. Development of Rapid Advice Guidelines for the Treatment of Coronavirus Disease 2019 with Traditional Chinese Medicine. *Am J Chin Med*. 2020;48(7):1511–1521. doi:10.1142/S0192415X20500743.
- Schünemann HJ, Zhang Y, Oxman AD. Expert Evidence in Guidelines Group Distinguishing opinion from evidence in guidelines. *BMJ*. 2019;366:14606. doi:10.1136/bmj.l4606.