Ma et al. Military Medical Research (2020) 7:7
https://doi.org/10.1186/s40779-020-00238-8

MMR MILITARY MEDICAL RESEARCH

REVIEW Open Access

Methodological quality (risk of bias) assessment tools for primary and secondary medical studies: what are they and which is better?

Lin-Lu Ma[1], Yun-Yun Wang[1,2], Zhi-Hua Yang[1], Di Huang[1,2], Hong Weng[1] and Xian-Tao Zeng[1,2,3,4*]

Abstract

Methodological quality (risk of bias) assessment is an important step before study initiation usage. Therefore, accurately judging study type is the first priority, and the choosing proper tool is also important. In this review, we introduced methodological quality assessment tools for randomized controlled trial (including individual and cluster), animal study, non-randomized interventional studies (including follow-up study, controlled before-and-after study, before-after/ pre-post study, uncontrolled longitudinal study, interrupted time series study), cohort study, case-control study, cross-sectional study (including analytical and descriptive), observational case series and case reports, comparative effectiveness research, diagnostic study, health economic evaluation, prediction study (including predictor finding study, prediction model impact study, prognostic prediction model study), qualitative study, outcome measurement instruments (including patient - reported outcome measure development, content validity, structural validity, internal consistency, cross-cultural validity/ measurement invariance, reliability, measurement error, criterion validity, hypotheses testing for construct validity, and responsiveness), systematic review and meta-analysis, and clinical practice guideline. The readers of our review can distinguish the types of medical studies and choose appropriate tools. In one word, comprehensively mastering relevant knowledge and implementing more practices are basic requirements for correctly assessing the methodological quality.

Keywords: Methodological quality, Risk of bias, Quality assessment, Critical appraisal, Methodology checklist, Appraisal tool, Observational study, Qualitative study, Interventional study, Outcome measurement instrument

Background

In the twentieth century, pioneering works by distinguished professors Cochrane A [1], Guyatt GH [2], and Chalmers IG [3] have led us to the evidence-based medicine (EBM) era. In this era, how to search, critically appraise, and use the best evidence is important. Moreover, systematic review and meta-analysis is the most used tool for summarizing primary data scientifically [4–6] and also the basic for developing clinical practice guideline according to the Institute of Medicine (IOM) [7]. Hence, to perform a systematic review and/ or meta-analysis, assessing the methodological quality of based primary studies is important; naturally, it would be key to assess its own methodological quality before usage. Quality includes internal and external validity, while methodological quality usually refers to internal validity [8, 9]. Internal validity is also recommended as "risk of bias (RoB)" by the Cochrane Collaboration [9].

There are three types of tools: scales, checklists, and items [10, 11]. In 2015, Zeng et al. [11] investigated methodological quality tools for randomized controlled trial (RCT), non-randomized clinical intervention study, cohort study, case-control study, cross-sectional study, case series, diagnostic accuracy study which also called "diagnostic test accuracy (DTA)", animal study, systematic review and meta-analysis, and clinical practice

* Correspondence: zengxiantao1128@163.com; zengxiantao@whucebtm.com
[1]Center for Evidence-Based and Translational Medicine, Zhongnan Hospital, Wuhan University, 169 Donghu Road, Wuchang District, Wuhan 430071, Hubei, China
[2]Department of Evidence-Based Medicine and Clinical Epidemiology, The Second Clinical College, Wuhan University, Wuhan 430071, China
Full list of author information is available at the end of the article

© The Author(s). 2020 **Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The Creative Commons Public Domain Dedication waiver (http://creativecommons.org/publicdomain/zero/1.0/) applies to the data made available in this article, unless otherwise stated.

guideline (CPG). From then on, some changes might generate in pre-existing tools, and new tools might also emerge; moreover, the research method has also been developed in recent years. Hence, it is necessary to systematically investigate commonly-used tools for assessing methodological quality, especially those for economic evaluation, clinical prediction rule/model, and qualitative study. Therefore, this narrative review presented related methodological quality (including "RoB") assessment tools for primary and secondary medical studies up to December 2019, and Table 1 presents their basic characterizes. We hope this review can help the producers, users, and researchers of evidence.

Tools for intervention studies

Randomized controlled trial (individual or cluster)

The first RCT was designed by Hill BA (1897–1991) and became the "gold standard" for experimental study design [12, 13] up to now. Nowadays, the Cochrane risk of bias tool for randomized trials (which was introduced in 2008 and edited on March 20, 2011) is the most commonly recommended tool for RCT [9, 14], which is called "RoB". On August 22, 2019 (which was introduced in 2016), the revised revision for this tool to assess RoB in randomized trials (RoB 2.0) was published [15]. The RoB 2.0 tool is suitable for individually-randomized, parallel-group, and cluster- randomized trials, which can be found in the dedicated website https://www.riskofbias.info/welcome/rob-2-0-tool. The RoB 2.0 tool consists of five bias domains and shows major changes when compared to the original Cochrane RoB tool (Table S1A-B presents major items of both versions).

The Physiotherapy Evidence Database (PEDro) scale is a specialized methodological assessment tool for RCT in physiotherapy [16, 17] and can be found in http://www.pedro.org.au/english/downloads/pedro-scale/, covering 11 items (Table S1C). The Effective Practice and Organisation of Care (EPOC) Group is a Cochrane Review Group who also developed a tool (called as "EPOC RoB Tool") for complex interventions randomized trials. This tool has 9 items (Table S1D) and can be found in https://epoc.cochrane.org/resources/epoc-resources-review-authors. The Critical Appraisal Skills Programme (CASP) is a part of the Oxford Centre for Triple Value Healthcare Ltd. (3 V) portfolio, which provides resources and learning and development opportunities to support the development of critical appraisal skills in the UK (http://www.casp-uk.net/) [18–20]. The CASP checklist for RCT consists of three sections involving 11 items (Table S1E). The National Institutes of Health (NIH) also develops quality assessment tools for controlled intervention study (Table S1F) to assess methodological quality of RCT (https://www.nhlbi.nih.gov/health-topics/study-quality-assessment-tools).

The Joanna Briggs Institute (JBI) is an independent, international, not-for-profit researching and development organization based in the Faculty of Health and Medical Sciences at the University of Adelaide, South Australia (https://joannabriggs.org/). Hence, it also develops many critical appraisal checklists involving the feasibility, appropriateness, meaningfulness and effectiveness of healthcare interventions. Table S1G presents the JBI Critical appraisal checklist for RCT, which includes 13 items.

The Scottish Intercollegiate Guidelines Network (SIGN) was established in 1993 (https://www.sign.ac.uk/). Its objective is to improve the quality of health care for patients in Scotland via reducing variations in practices and outcomes, through developing and disseminating national clinical guidelines containing recommendations for effective practice based on current evidence. Hence, it also develops many critical appraisal checklists for assessing methodological quality of different study types, including RCT (Table S1H).

In addition, the Jadad Scale [21], Modified Jadad Scale [22, 23], Delphi List [24], Chalmers Scale [25], National Institute for Clinical Excellence (NICE) methodology checklist [11], Downs & Black checklist [26], and other tools summarized by West et al. in 2002 [27] are not commonly used or recommended nowadays.

Animal study

Before starting clinical trials, the safety and effectiveness of new drugs are usually tested in animal models [28], so animal study is considered as preclinical research, possessing important significance [29, 30]. Likewise, the methodological quality of animal study also needs to be assessed [30]. In 1999, the initial "Stroke Therapy Academic Industry Roundtable (STAIR)" recommended their criteria for assessing the quality of stroke animal studies [31] and this tool is also called "STAIR". In 2009, the STAIR Group updated their criteria and developed "Recommendations for Ensuring Good Scientific Inquiry" [32]. Besides, Macleod et al. [33] proposed a 10-point tool based on STAIR to assess methodological quality of animal study in 2004, which is also called "CAMARADES (The Collaborative Approach to Meta-Analysis and Review of Animal Data from Experimental Studies)"; with "S" presenting "Stroke" at that time and now standing for "Studies" (http://www.camarades.info/). In CAMARADES tool, every item could reach a highest score of one point and the total score for this tool could achieve 10 points (Table S1J).

In 2008, the Systematic Review Center for Laboratory animal Experimentation (SYRCLE) was established in Netherlands and this team developed and released an RoB tool for animal intervention studies - SYRCLE's RoB tool in 2014, based on the original Cochrane RoB

Table 1 The basic characteristics of the included methodological quality (risk of bias) assessment tools

No.	Development Organization	Tool's name	Type of study
1	The Cochrane Collaboration	Cochrane RoB tool and RoB 2.0 tool	Randomized controlled trial Diagnostic accuracy study
2	The Physiotherapy Evidence Database (PEDro)	PEDro scale	Randomized controlled trial
3	The Effective Practice and Organisation of Care (EPOC) Group	EPOC RoB tool	Randomized controlled trial Clinical controlled trials Controlled before-and-after study Interrupted time series studies
4	The Critical Appraisal Skills Programme (CASP)	CASP checklist	Randomized controlled trial Cohort study Case-control study Cross-sectional study Diagnostic test study Clinical prediction rule Economic evaluation Qualitative study Systematic review
5	The National Institutes of Health (NIH)	NIH quality assessment tool	Controlled intervention study Cohort study Cross-sectional study Case-control study Before-after (Pre-post) study with no control group Case-series (Interventional) Systematic review and meta-analysis
6	The Joanna Briggs Institute (JBI)	JBI critical appraisal checklist	Randomized controlled trial Non-randomized experimental study Cohort study Case-control study Cross-sectional study Prevalence data Case reports Economic evaluation Qualitative study Text and expert opinion papers Systematic reviews and research syntheses
7	The Scottish Intercollegiate Guidelines Network (SIGN)	SIGN methodology checklist	Randomized controlled trial Cohort study Case-control study Diagnostic study Economic evaluation Systematic reviews and meta-analyses
8	The Stroke Therapy Academic Industry Roundtable (STAIR) Group	CAMARADES tool	Animal study
9	The SYstematic Review Center for Laboratory animal Experimentation (SYRCLE)	SYRCLE's RoB tool	Animal study
10	Sterne JAC et al.	ROBINS-I tool	Non-randomised interventional study
11	Slim K et al.	MINORS tool	Non-randomised interventional study
12	The Canada Institute of Health Economics (IHE)	IHE quality appraisal tool	Case-series (Interventional)
13	Wells GA et al.	Newcastle-Ottawa Scale (NOS)	Cohort study Case-control study
14	Downes MJ et al.	AXIS tool	Cross-sectional study
15	The Agency for Healthcare Research and Quality (AHRQ)	AHRQ methodology checklist	Cross-sectional/ Prevalence study
16	Crombie I	Crombie's items	Cross-sectional study
17	The Good Research for Comparative Effectiveness (GRACE) Initiative	GRACE checklist	Comparative effectiveness research
18	Whiting PF et al.	QUADAS tool and QUADAS-2 tool	Diagnostic accuracy study

Table 1 The basic characteristics of the included methodological quality (risk of bias) assessment tools *(Continued)*

No.	Development Organization	Tool's name	Type of study
19	The National Institute for Clinical Excellence (NICE)	NICE methodology checklist	Economic evaluation
20	The Cabinet Office	The Quality Framework: Cabinet Office checklist	Qualitative study (social research)
21	Hayden JA et al.	QIPS tool	Prediction study (predictor finding study)
22	Wolff RF et al.	PROBAST	Prediction study (prediction model study)
23	The (COnsensus-based Standards for the selection of health Measurement INstruments) initiative	COSMIN RoB checklist	Patient-reported outcome measure development Content validity Structural validity Internal consistency Cross-cultural validity/ measurement invariance Reliability Measurement error Criterion validity Hypotheses testing for construct validity Responsiveness
24	Shea BJ et al.	AMSTAR and AMSTAR 2	Systematic review
25	The Decision Support Unit (DSU)	DSU network meta-analysis (NMA) methodology checklist	Network meta-analysis
26	Whiting P et al.	ROBIS tool	Systematic review
27	Brouwers MC et al.	AGREE instrument and AGREE II instrument	Clinical practice guideline

AMSTAR A measurement tool to assess systematic reviews, *AHRQ* Agency for healthcare research and quality, *AXIS* Appraisal tool for cross-sectional studies, *CASP* Critical appraisal skills programme, *CAMARADES* The collaborative approach to meta-analysis and review of animal data from experimental studies, *COSMIN* Consensus-based standards for the selection of health measurement instruments, *DSU* Decision support unit, *EPOC* the effective practice and organisation of care group, *GRACE* The god research for comparative effectiveness initiative, *IHE* Canada institute of health economics, *JBI* Joanna Briggs Institute, *MINORS* Methodological index for non-randomized studies, *NOS* Newcastle-Ottawa scale, *NMA* network meta-analysis, *NIH* national institutes of health, *NICE* National institute for clinical excellence, *PEDro* physiotherapy evidence database, *PROBAST* The prediction model risk of bias assessment tool, *QUADAS* Quality assessment of diagnostic accuracy studies, *QIPS* Quality in prognosis studies, *RoB* Risk of bias, *ROBINS-I* Risk of bias in non-randomised studies - of interventions, *ROBIS* Risk of bias in systematic review, *SYRCLE* Systematic review center for laboratory animal experimentation, *STAIR* Stroke therapy academic industry roundtable, *SIGN* The Scottish intercollegiate guidelines network

Tool [34]. This new tool contained 10 items which had become the most recommended tool for assessing the methodological quality of animal intervention studies (Table S1I).

Non-randomised studies

In clinical research, RCT is not always feasible [35]; therefore, non-randomized design remains considerable. In non-randomised study (also called quasi-experimental studies), investigators control the allocation of participants into groups, but do not attempt to adopt randomized operation [36], including follow-up study. According to with or without comparison, non-randomized clinical intervention study can be divided into comparative and non-comparative sub-types, the Risk Of Bias In Non-randomised Studies - of Interventions (ROBINS-I) tool [37] is the preferentially recommended tool. This tool is developed to evaluate risk of bias in estimating comparative effectiveness (harm or benefit) of interventions in studies not adopting randomization in allocating units (individuals or clusters of individuals) into comparison groups. Besides, the JBI critical appraisal checklist for quasi-experimental studies (non-randomized experimental studies) is also suitable, which includes 9 items. Moreover, the methodological index for non-randomized studies (MINORS) [38] tool can also be used, which contains a total of 12 methodological points; the first 8 items could be applied for both non-comparative and comparative studies, while the last 4 items appropriate for studies with two or more groups. Every item is scored from 0 to 2, and the total scores over 16 or 24 give an overall quality score. Table S1K-L-M presented the major items of these three tools.

Non-randomized study with a separate control group could also be called clinical controlled trial or controlled before-and-after study. For this design type, the EPOC RoB tool is suitable (see Table S1D). When using this tool, the "random sequence generation" and "allocation concealment" should be scored as "High risk", while grading for other items could be the same as that for randomized trial.

Non-randomized study without a separate control group could be a before-after (Pre-Post) study, a case series (uncontrolled longitudinal study), or an interrupted time series study. A case series is described a series of individuals, who usually receive the same intervention, and contains non control group [9]. There are several tools for assessing the methodological quality of case series study. The latest one was developed by Moga C et al. [39] in 2012 using a modified Delphi technique, which was developed by the Canada Institute of Health Economics (IHE); hence, it is also called "IHE Quality Appraisal Tool" (Table S1N). Moreover, NIH also develops a quality assessment tool for case series study, including 9 items (Table S1O). For interrupted time series studies, the "EPOC RoB tool for interrupted time series studies" is recommended (Table S1P). For the before-after study, we recommend the NIH quality assessment tool for before-after (Pre-Post) study without control group (Table S1Q).

In addition, for non-randomized intervention study, the Reisch tool (Check List for Assessing Therapeutic Studies) [11, 40], Downs & Black checklist [26], and other tools summarized by Deeks et al. [36] are not commonly used or recommended nowadays.

Tools for observational studies and diagnostic study

Observational studies include cohort study, case-control study, cross-sectional study, case series, case reports, and comparative effectiveness research [41], and can be divided into analytical and descriptive studies [42].

Cohort study

Cohort study includes prospective cohort study, retrospective cohort study, and ambidirectional cohort study [43]. There are some tools for assessing the quality of cohort study, such as the CASP cohort study checklist (Table S2A), SIGN critical appraisal checklists for cohort study (Table S2B), NIH quality assessment tool for observational cohort and cross-sectional studies (Table S2C), Newcastle-Ottawa Scale (NOS; Table S2D) for cohort study, and JBI critical appraisal checklist for cohort study (Table S2E). However, the Downs & Black checklist [26] and the NICE methodology checklist for cohort study [11] are not commonly used or recommended nowadays.

The NOS [44, 45] came from an ongoing collaboration between the Universities of Newcastle, Australia and Ottawa, Canada. Among all above mentioned tools, the NOS is the most commonly used tool nowadays which also allows to be modified based on a special subject.

Case-control study

Case-control study selects participants based on the presence of a specific disease or condition, and seeks earlier exposures that may lead to the disease or outcome [42]. It has an advantage over cohort study, that is the issue of "drop out" or "loss in follow up" of participants as seen in cohort study would not arise in such study. Nowadays, there are some acceptable tools for assessing the methodological quality of case-control study, including CASP case-control study checklist (Table S2F), SIGN critical appraisal checklists for case-control study (Table S2G), NIH quality assessment tool of case-control study (Table S2H), JBI critical appraisal checklist for case-control study (Table S2I), and the NOS for case-control study (Table S2J). Among them,

the NOS for case-control study is also the most frequently used tool nowadays and allows to be modified by users.

In addition, the Downs & Black checklist [26] and the NICE methodology checklist for case-control study [11] are also not commonly used or recommended nowadays.

Cross-sectional study (analytical or descriptive)

Cross-sectional study is used to provide a snapshot of a disease and other variables in a defined population at a time point. It can be divided into analytical and purely descriptive types. Descriptive cross-sectional study merely describes the number of cases or events in a particular population at a time point or during a period of time; whereas analytic cross-sectional study can be used to infer relationships between a disease and other variables [46].

For assessing the quality of analytical cross-sectional study, the NIH quality assessment tool for observational cohort and cross-sectional studies (Table S2C), JBI critical appraisal checklist for analytical cross-sectional study (Table S2K), and the Appraisal tool for Cross-Sectional Studies (AXIS tool; Table S2L) [47] are recommended tools. The AXIS tool is a critical appraisal tool that addresses study design and reporting quality as well as the risk of bias in cross-sectional study, which was developed in 2016 and contains 20 items. Among these three tools, the JBI checklist is the most preferred one.

Purely descriptive cross-sectional study is usually used to measure disease prevalence and incidence. Hence, the critical appraisal tool for analytic cross-sectional study is not proper for the assessment. Only few quality assessment tools are suitable for descriptive cross-sectional study, like the JBI critical appraisal checklist for studies reporting prevalence data [48] (Table S2M), Agency for Healthcare Research and Quality (AHRQ) methodology checklist for assessing the quality of cross-sectional/prevalence study (Table S2N), and Crombie's items for assessing the quality of cross-sectional study [49] (Table S2O). Among them, the JBI tool is the newest.

Case series and case reports

Unlike above mentioned interventional case series, case reports and case series are used to report novel occurrences of a disease or a unique finding [50]. Hence, they belong to descriptive studies. There is only one tool – the JBI critical appraisal checklist for case reports (Table S2P).

Comparative effectiveness research

Comparative effectiveness research (CER) compares real-world outcomes [51] resulting from alternative treatment options that are available for a given medical condition. Its key elements include the study of effectiveness (effect in the real world), rather than efficacy (ideal effect), and the comparisons among alternative strategies [52]. In 2010, the Good Research for Comparative Effectiveness (GRACE) Initiative was established and developed principles to help healthcare providers, researchers, journal readers, and editors evaluate inherent quality for observational research studies of comparative effectiveness [41]. And in 2016, a validated assessment tool – the GRACE Checklist v5.0 (Table S2Q) was released for assessing the quality of CER.

Diagnostic study

Diagnostic tests, also called "Diagnostic Test Accuracy (DTA)", are used by clinicians to identify whether a condition exists in a patient or not, so as to develop an appropriate treatment plan [53]. DTA has several unique features in terms of its design which differ from standard intervention and observational evaluations. In 2003, Penny et al. [53, 54] developed a tool for assessing the quality of DTA, namely Quality Assessment of Diagnostic Accuracy Studies (QUADAS) tool. In 2011, a revised "QUADAS-2" tool (Table S2R) was launched [55, 56]. Besides, the CASP diagnostic checklist (Table S2S), SIGN critical appraisal checklists for diagnostic study (Table S2T), JBI critical appraisal checklist for diagnostic test accuracy studies (Table S2U), and the Cochrane risk of bias assessing tool for diagnostic test accuracy (Table S2V) are also common useful tools in this field.

Of them, the Cochrane risk of bias tool (https://methods.cochrane.org/sdt/) is based on the QUADAS tool, and the SIGN and JBI tools are based on the QUADAS-2 tool. Of course, the QUADAS-2 tool is the first recommended tool. Other relevant tools reviewed by Whiting et al. [53] in 2004 are not used nowadays.

Tools for other primary medical studies

Health economic evaluation

Health economic evaluation research comparatively analyses alternative interventions with regard to their resource uses, costs and health effects [57]. It focuses on identifying, measuring, valuing and comparing resource use, costs and benefit/effect consequences for two or more alternative intervention options [58]. Nowadays, health economic study is increasingly popular. Of course, its methodological quality also needs to be assessed before its initiation. The first tool for such assessment was developed by Drummond and Jefferson in 1996 [59], and then many tools have been developed based on the Drummond's items or its revision [60], such as the SIGN critical appraisal checklists for economic evaluations (Table S3A), CASP economic evaluation checklist (Table S3B), and the JBI critical appraisal checklist for economic evaluations (Table S3C). The

NICE only retains one methodology checklist for economic evaluation (Table S3D).

However, we regard the Consolidated Health Economic Evaluation Reporting Standards (CHEERS) statement [61] as a reporting tool rather than a methodological quality assessment tool, so we do not recommend it to assess the methodological quality of health economic evaluation.

Qualitative study

In healthcare, qualitative research aims to understand and interpret individual experiences, behaviours, interactions, and social contexts, so as to explain interested phenomena, such as the attitudes, beliefs, and perspectives of patients and clinicians; the interpersonal nature of caregiver and patient relationships; illness experience; and the impact of human sufferings [62]. Compared with quantitative studies, assessment tools for qualitative studies are fewer. Nowadays, the CASP qualitative research checklist (Table S3E) is the most frequently recommended tool for this issue. Besides, the JBI critical appraisal checklist for qualitative research [63, 64] (Table S3F) and the Quality Framework: Cabinet Office checklist for social research [65] (Table S3G) are also suitable.

Prediction studies

Clinical prediction study includes predictor finding (prognostic factor) studies, prediction model studies (development, validation, and extending or updating), and prediction model impact studies [66]. For predictor finding study, the Quality In Prognosis Studies (QIPS) tool [67] can be used for assessing its methodological quality (Table S3H). For prediction model impact studies, if it uses a randomized comparative design, tools for RCT can be used, especially the RoB 2.0 tool; if it uses a non-randomized comparative design, tools for non-randomized studies can be used, especially the ROBINS-I tool. For diagnostic and prognostic prediction model studies, the Prediction model Risk Of Bias Assessment Tool (PROBAST; Table S3I) [68] and CASP clinical prediction rule checklist (Table S3J) are suitable.

Text and expert opinion papers

Text and expert opinion-based evidence (also called "non-research evidence") comes from expert opinions, consensus, current discourse, comments, and assumptions or assertions that appear in various journals, magazines, monographs and reports [69–71]. Nowadays, only the JBI has a critical appraisal checklist for the assessment of text and expert opinion papers (Table S3K).

Outcome measurement instruments

An outcome measurement instrument is a "device" used to collect a measurement. The range embraced by the term 'instrument' is broad, and can refer to questionnaire (e.g. patient-reported outcome such as quality of life), observation (e.g. the result of a clinical examination), scale (e.g. a visual analogue scale), laboratory test (e.g. blood test) and images (e.g. ultrasound or other medical imaging) [72, 73]. Measurements can be subjective or objective, and either unidimensional (e.g. attitude) or multidimensional. Nowadays, only one tool - the COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) Risk of Bias checklist [74–76] (www.cosmin.nl/) is proper for assessing the methodological quality of outcome measurement instrument, and Table S3L presents its major items, including patient - reported outcome measure (PROM) development (Table S3LA), content validity (Table S3LB), structural validity (Table S3LC), internal consistency (Table S3LD), cross-cultural validity/ measurement invariance (Table S3LE), reliability (Table S3LF), measurement error (Table S3LG), criterion validity (Table S3LH), hypotheses testing for construct validity (Table S3LI), and responsiveness (Table S3LJ).

Tools for secondary medical studies

Systematic review and meta-analysis

Systematic review and meta-analysis are popular methods to keep up with current medical literature [4–6]. Their ultimate purposes and values lie in promoting healthcare [6, 77, 78]. Meta-analysis is a statistical process of combining results from several studies, commonly a part of a systematic review [11]. Of course, critical appraisal would be necessary before using systematic review and meta-analysis.

In 1988, Sacks et al. developed the first tool for assessing the quality of meta-analysis on RCTs - the Sack's Quality Assessment Checklist (SQAC) [79]; And then in 1991, Oxman and Guyatt developed another tool – the Overview Quality Assessment Questionnaire (OQAQ) [80, 81]. To overcome the shortcomings of these two tools, in 2007 the A Measurement Tool to Assess Systematic Reviews (AMSTAR) was developed based on them [82] (http://www.amstar.ca/). However, this original AMSTAR instrument did not include an assessment on the risk of bias for non-randomised studies, and the expert group thought revisions should address all aspects of the conduct of a systematic review. Hence, the new instrument for randomised or non-randomised studies on healthcare interventions - AMSTAR 2 was released in 2017 [83], and Table S4A presents its major items.

Besides, the CASP systematic review checklist (Table S4B), SIGN critical appraisal checklists for systematic reviews and meta-analyses (Table S4C), JBI critical appraisal checklist for systematic reviews and research syntheses (Table S4D), NIH quality assessment tool for

systematic reviews and meta-analyses (Table S4E), The Decision Support Unit (DSU) network meta-analysis (NMA) methodology checklist (Table S4F), and the Risk of Bias in Systematic Review (ROBIS) [84] tool (Table S4G) are all suitable. Among them, the AMSTAR 2 is the most commonly used and the ROIBS is the most frequently recommended.

Among those tools, the AMSTAR 2 is suitable for assessing systematic review and meta-analysis based on randomised or non-randomised interventional studies, the DSU NMA methodology checklist for network meta-analysis, while the ROBIS for meta-analysis based on interventional, diagnostic test accuracy, clinical prediction, and prognostic studies.

Clinical practice guidelines

Clinical practice guideline (CPG) is integrated well into the thinking of practicing clinicians and professional clinical organizations [85–87]; and also make scientific evidence incorporated into clinical practice [88]. However, not all CPGs are evidence-based [89, 90] and their qualities are uneven [91–93]. Until now there were more than 20 appraisal tools have been developed [94]. Among them, the Appraisal of Guidelines for Research and Evaluation (AGREE) instrument has the greatest potential in serving as a basis to develop an appraisal tool for clinical pathways [94]. The AGREE instrument was first released in 2003 [95] and updated to AGREE II instrument in 2009 [96] (www.agreetrust.org/). Now the AGREE II instrument is the most recommended tool for CPG (Table S4H).

Besides, based on the AGREE II, the AGREE Global Rating Scale (AGREE GRS) Instrument [97] was developed as a short item tool to evaluate the quality and reporting of CPGs.

Discussion and conclusions

Currently, the EBM is widely accepted and the major attention of healthcare workers lies in "Going from evidence to recommendations" [98, 99]. Hence, critical appraisal of evidence before using is a key point in this process [100, 101]. In 1987, Mulrow CD [102] pointed out that medical reviews needed routinely use scientific methods to identify, assess, and synthesize information. Hence, perform methodological quality assessment is necessary before using the study. However, although there are more than 20 years have been passed since the first tool emergence, many users remain misunderstand the methodological quality and reporting quality. Of them, someone used the reporting checklist to assess the methodological quality, such as used the Consolidated Standards of Reporting Trials (CONSORT) statement [103] to assess methodological quality of RCT, used the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement [104] to methodological quality of cohort study. This phenomenon indicates more universal education of clinical epidemiology is needed for medical students and professionals.

The methodological quality tool development should according to the characteristics of different study types. In this review, we used "methodological quality", "risk of bias", "critical appraisal", "checklist", "scale", "items", and "assessment tool" to search in the NICE website, SIGN website, Cochrane Library website and JBI website, and on the basis of them, added "systematic review", "meta-analysis", "overview" and "clinical practice guideline" to search in PubMed. Compared with our previous systematic review [11], we found some tools are recommended and remain used, some are used without recommendation, and some are eliminated [10, 29, 30, 36, 53, 94, 105–107]. These tools produce a significant impetus for clinical practice [108, 109].

In addition, compared with our previous systematic review [11], this review stated more tools, especially those developed after 2014, and the latest revisions. Of course, we also adjusted the method of study type classification. Firstly, in 2014, the NICE provided 7 methodology checklists but only retains and updated the checklist for economic evaluation now. Besides, the Cochrane RoB 2.0 tool, AMSTAR 2 tool, CASP checklist, and most of JBI critical appraisal checklists are all the newest revisions; the NIH quality assessment tool, ROBINS-I tool, EPOC RoB tool, AXIS tool, GRACE Checklist, PROBAST, COSMIN Risk of Bias checklist, and ROBIS tool are all newly released tools. Secondly, we also introduced tools for network meta-analysis, outcome measurement instruments, text and expert opinion papers, prediction studies, qualitative study, health economic evaluation, and CER. Thirdly, we classified interventional studies into randomized and non-randomized sub-types, and then further classified non-randomized studies into with and without controlled group. Moreover, we also classified cross-sectional study into analytic and purely descriptive sub-types, and case-series into interventional and observational sub-types. These processing courses were more objective and comprehensive.

Obviously, the number of appropriate tools is the largest for RCT, followed by cohort study; the applicable range of JBI is widest [63, 64], with CASP following closely. However, further efforts remain necessary to develop appraisal tools. For some study types, only one assessment tool is suitable, such as CER, outcome measurement instruments, text and expert opinion papers, case report, and CPG. Besides, there is no proper assessment tool for many study types, such as overview, genetic association study, and cell study. Moreover, existing tools have not been fully accepted. In the future, how to develop well accepted tools remains a significant and important work [11].

Our review can help the professionals of systematic review, meta-analysis, guidelines, and evidence users to choose the best tool when producing or using evidence. Moreover, methodologists can obtain the research topics for developing new tools. Most importantly, we must remember that all assessment tools are subjective, and actual yields of wielding them would be influenced by user's skills and knowledge level. Therefore, users must receive formal training (relevant epidemiological knowledge is necessary), and hold rigorous academic attitude, and at least two independent reviewers should be involved in evaluation and cross-checking to avoid performance bias [110].

Supplementary information

Supplementary information accompanies this paper at https://doi.org/10.1186/s40779-020-00238-8.

Additional file 1: Table S1. Major components of the tools for assessing intervention studies

Additional file 2: Table S2. Major components of the tools for assessing observational studies and diagnostic study

Additional file 3: Table S3. Major components of the tools for assessing other primary medical studies

Additional file 4: Table S4. Major components of the tools for assessing secondary medical studies

Abbreviations

AGREE GRS: AGREE Global rating scale; AGREE: Appraisal of guidelines for research and evaluation; AHRQ: Agency for healthcare research and quality; AMSTAR: A measurement tool to assess systematic reviews; AXIS: Appraisal tool for cross-sectional studies; CAMARADES: The collaborative approach to meta-analysis and review of animal data from experimental studies; CASP: Critical appraisal skills programme; CER: Comparative effectiveness research; CHEERS: Consolidated health economic evaluation reporting standards; CONSORT: Consolidated standards of reporting trials; COSMIN: Consensus-based standards for the selection of health measurement instruments; CPG: Clinical practice guideline; DSU: Decision support unit; DTA: Diagnostic test accuracy; EBM: Evidence-based medicine; EPOC: The effective practice and organisation of care group; GRACE: The good research for comparative effectiveness initiative; IHE: Canada institute of health economics; IOM: Institute of medicine; JBI: Joanna Briggs Institute; MINORS: Methodological index for non-randomized studies; NICE: National institute for clinical excellence; NIH: National institutes of health; NMA: Network meta-analysis; NOS: Newcastle-Ottawa scale; OQAQ: Overview quality assessment questionnaire; PEDro: Physiotherapy evidence database; PROBAST: The prediction model risk of bias assessment tool; PROM: Patient - reported outcome measure; QIPS: Quality in prognosis studies; QUADAS: Quality assessment of diagnostic accuracy studies; RCT: Randomized controlled trial; RoB: Risk of bias; ROBINS-I: Risk of bias in non-randomised studies - of interventions; ROBIS: Risk of bias in systematic review; SIGN: The Scottish intercollegiate guidelines network; SQAC: Sack's quality assessment checklist; STAIR: Stroke therapy academic industry roundtable; STROBE: Strengthening the reporting of observational studies in epidemiology; SYRCLE: Systematic review center for laboratory animal experimentation

Acknowledgements

The authors thank all the authors and technicians for their hard field work for development methodological quality assessment tools.

Authors' contributions

XTZ is responsible for the design of the study and review of the manuscript; LLM, ZHY, YYW, and DH contributed to the data collection; LLM, YYW, and HW contributed to the preparation of the article. All authors read and approved the final manuscript.

Funding

This work was supported (in part) by the Entrusted Project of National commission on health and health of China (No. [2019]099), the National Key Research and Development Plan of China (2016YFC0106300), and the Nature Science Foundation of Hubei Province (2019FFB03902). The funder had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. The authors declare that there are no conflicts of interest in this study.

Availability of data and materials

The data and materials used during the current review are all available in this review.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

[1]Center for Evidence-Based and Translational Medicine, Zhongnan Hospital, Wuhan University, 169 Donghu Road, Wuchang District, Wuhan 430071, Hubei, China. [2]Department of Evidence-Based Medicine and Clinical Epidemiology, The Second Clinical College, Wuhan University, Wuhan 430071, China. [3]Center for Evidence-Based and Translational Medicine, Wuhan University, Wuhan 430071, China. [4]Global Health Institute, Wuhan University, Wuhan 430072, China.

Received: 17 January 2020 Accepted: 18 February 2020
Published online: 29 February 2020

References

1. Stavrou A, Challoumas D, Dimitrakakis G. Archibald Cochrane (1909-1988): the father of evidence-based medicine. Interact Cardiovasc Thorac Surg. 2013;18(1):121–4.
2. Group E-BMW. Evidence-based medicine. A new approach to teaching the practice of medicine. JAMA. 1992;268(17):2420–5.
3. Levin A. The Cochrane collaboration. Ann Intern Med. 2001;135(4):309–12.
4. Lau J, Ioannidis JP, Schmid CH. Summing up evidence: one answer is not always enough. Lancet. 1998;351(9096):123–7.
5. Clarke M, Chalmers I. Meta-analyses, multivariate analyses, and coping with the play of chance. Lancet. 1998;351(9108):1062–3.
6. Oxman AD, Schunemann HJ, Fretheim A. Improving the use of research evidence in guideline development: 8. Synthesis and presentation of evidence. Health Res Policy Syst. 2006;4:20.
7. Zhang J, Wang Y, Weng H, Wang D, Han F, Huang Q, et al. Management of non-muscle-invasive bladder cancer: quality of clinical practice guidelines and variations in recommendations. BMC Cancer. 2019;19(1):1054.
8. Campbell DT. Factors relevant to the validity of experiments in social settings. Psychol Bull. 1957;54(4):297–312.
9. Higgins J, Green S. Cochrane handbook for systematic reviews of interventions version 5.1.0 [updated March 2011]. The Cochrane Collaboration; 2011.
10. Juni P, Altman DG, Egger M. Systematic reviews in health care: assessing the quality of controlled clinical trials. BMJ. 2001;323(7303):42–6.
11. Zeng X, Zhang Y, Kwong JS, Zhang C, Li S, Sun F, et al. The methodological quality assessment tools for preclinical and clinical studies, systematic review and meta-analysis, and clinical practice guideline: a systematic review. J Evid Based Med. 2015;8(1):2–10.
12. A Medical Research Council Investigation. Treatment of pulmonary tuberculosis with streptomycin and Para-aminosalicylic acid. Br Med J. 1950; 2(4688):1073–85.
13. Armitage P. Fisher, Bradford Hill, and randomization. Int J Epidemiol. 2003; 32(6):925–8.

14. Higgins JP, Altman DG, Gotzsche PC, Juni P, Moher D, Oxman AD, et al. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. BMJ. 2011;343:d5928.
15. Sterne JAC, Savovic J, Page MJ, Elbers RG, Blencowe NS, Boutron I, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. BMJ. 2019;366: l4898.
16. Maher CG, Sherrington C, Herbert RD, Moseley AM, Elkins M. Reliability of the PEDro scale for rating quality of randomized controlled trials. Phys Ther. 2003;83(8):713–21.
17. Shiwa SR, Costa LO, Costa Lda C, Moseley A, Hespanhol Junior LC, Venancio R, et al. Reproducibility of the Portuguese version of the PEDro scale. Cad Saude Publica. 2011;27(10):2063–8.
18. Ibbotson T, Grimshaw J, Grant A. Evaluation of a programme of workshops for promoting the teaching of critical appraisal skills. Med Educ. 1998;32(5): 486–91.
19. Singh J. Critical appraisal skills programme. J Pharmacol Pharmacother. 2013;4(1):76.
20. Taylor R, Reeves B, Ewings P, Binns S, Keast J, Mears R. A systematic review of the effectiveness of critical appraisal skills training for clinicians. Med Educ. 2000;34(2):120–5.
21. Jadad AR, Moore RA, Carroll D, Jenkinson C, Reynolds DJ, Gavaghan DJ, et al. Assessing the quality of reports of randomized clinical trials: is blinding necessary? Control Clin Trials. 1996;17(1):1–12.
22. Schulz KF, Chalmers I, Hayes RJ, Altman DG. Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. JAMA. 1995;273(5):408–12.
23. Hartling L, Ospina M, Liang Y, Dryden DM, Hooton N, Krebs Seida J, et al. Risk of bias versus quality assessment of randomised controlled trials: cross sectional study. BMJ. 2009;339:b4012.
24. Verhagen AP, de Vet HC, de Bie RA, Kessels AG, Boers M, Bouter LM, et al. The Delphi list: a criteria list for quality assessment of randomized clinical trials for conducting systematic reviews developed by Delphi consensus. J Clin Epidemiol. 1998;51(12):1235–41.
25. Chalmers TC, Smith H Jr, Blackburn B, Silverman B, Schroeder B, Reitman D, et al. A method for assessing the quality of a randomized control trial. Control Clin Trials. 1981;2(1):31–49.
26. Downs SH, Black N. The feasibility of creating a checklist for the assessment of the methodological quality both of randomised and non-randomised studies of health care interventions. J Epidemiol Community Health. 1998; 52(6):377–84.
27. West S, King V, Carey TS, Lohr KN, McKoy N, Sutton SF, et al. Systems to rate the strength of scientific evidence. Evid Rep Technol Assess (Summ). 2002; 47:1–11.
28. Sibbald WJ. An alternative pathway for preclinical research in fluid management. Crit Care. 2000;4(Suppl 2):S8–15.
29. Perel P, Roberts I, Sena E, Wheble P, Briscoe C, Sandercock P, et al. Comparison of treatment effects between animal experiments and clinical trials: systematic review. BMJ. 2007;334(7586):197.
30. Hooijmans CR, Ritskes-Hoitinga M. Progress in using systematic reviews of animal studies to improve translational research. PLoS Med. 2013;10(7): e1001482.
31. Stroke Therapy Academic Industry R. Recommendations for standards regarding preclinical neuroprotective and restorative drug development. Stroke. 1999;30(12):2752–8.
32. Fisher M, Feuerstein G, Howells DW, Hurn PD, Kent TA, Savitz SI, et al. Update of the stroke therapy academic industry roundtable preclinical recommendations. Stroke. 2009;40(6):2244–50.
33. Macleod MR, O'Collins T, Howells DW, Donnan GA. Pooling of animal experimental data reveals influence of study design and publication bias. Stroke. 2004;35(5):1203–8.
34. Hooijmans CR, Rovers MM, de Vries RB, Leenaars M, Ritskes-Hoitinga M, Langendam MW. SYRCLE's risk of bias tool for animal studies. BMC Med Res Methodol. 2014;14:43.
35. McCulloch P, Taylor I, Sasako M, Lovett B, Griffin D. Randomised trials in surgery: problems and possible solutions. BMJ. 2002;324(7351):1448–51.
36. Deeks JJ, Dinnes J, D'Amico R, Sowden AJ, Sakarovitch C, Song F, et al. Evaluating non-randomised intervention studies. Health Technol Assess. 2003;7(27):1–173.
37. Sterne JA, Hernan MA, Reeves BC, Savovic J, Berkman ND, Viswanathan M, et al. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. BMJ. 2016;355:i4919.
38. Slim K, Nini E, Forestier D, Kwiatkowski F, Panis Y, Chipponi J. Methodological index for non-randomized studies (minors): development and validation of a new instrument. ANZ J Surg. 2003;73(9):712–6.
39. Moga C, Guo B, Schopflocher D, Harstall C. Development of a quality appraisal tool for case series studies using a modified delphi technique 2012. http://www.ihe.ca/documents/Case%20series%20studies%20using %20a%20modified%20Delphi%20technique.pdf .(Accept 15 Januray 2020).
40. Reisch JS, Tyson JE, Mize SG. Aid to the evaluation of therapeutic studies. Pediatrics. 1989;84(5):815–27.
41. Dreyer NA, Schneeweiss S, McNeil BJ, Berger ML, Walker AM, Ollendorf DA, et al. GRACE principles: recognizing high-quality observational studies of comparative effectiveness. Am J Manag Care. 2010;16(6):467–71.
42. Grimes DA, Schulz KF. An overview of clinical research: the lay of the land. Lancet. 2002;359(9300):57–61.
43. Grimes DA, Schulz KF. Cohort studies: marching towards outcomes. Lancet. 2002;359(9303):341–5.
44. Wells G, Shea B, O'Connell D, Peterson J, Welch V, Losos M, *et al.* The Newcastle-Ottawa Scale (NOS) for assessing the quality of nonrandomised studies in meta-analyses. http://www.ohri.ca/programs/clinical_ epidemiology/oxford.asp (Accessed 16 Jan 2020).
45. Stang A. Critical evaluation of the Newcastle-Ottawa scale for the assessment of the quality of nonrandomized studies in meta-analyses. Eur J Epidemiol. 2010;25(9):603–5.
46. Wu L, Li BH, Wang YY, Wang CY, Zi H, Weng H, et al. Periodontal disease and risk of benign prostate hyperplasia: a cross-sectional study. Mil Med Res. 2019;6(1):34.
47. Downes MJ, Brennan ML, Williams HC, Dean RS. Development of a critical appraisal tool to assess the quality of cross-sectional studies (AXIS). BMJ Open. 2016;6(12):e011458.
48. Munn Z, Moola S, Lisy K, Riitano D, Tufanaru C. Methodological guidance for systematic reviews of observational epidemiological studies reporting prevalence and cumulative incidence data. Int J Evid Based Healthc. 2015; 13(3):147–53.
49. Crombie I. Pocket guide to critical appraisal: Oxford. UK: John Wiley & Sons, Ltd; 1996.
50. Gagnier JJ, Kienle G, Altman DG, Moher D, Sox H, Riley D, et al. The CARE guidelines: consensus-based clinical case report guideline development. J Clin Epidemiol. 2014;67(1):46–51.
51. Li BH, Yu ZJ, Wang CY, Zi H, Li XD, Wang XH, et al. A preliminary, multicenter, prospective and real world study on the hemostasis, coagulation, and safety of hemocoagulase bothrops atrox in patients undergoing transurethral bipolar plasmakinetic prostatectomy. Front Pharmacol. 2019;10:1426.
52. Strom BL, Schinnar R, Hennessy S. Comparative effectiveness research. Pharmacoepidemiology. Oxford, UK: John Wiley & Sons, Ltd; 2012. p. 561–79.
53. Whiting P, Rutjes AW, Dinnes J, Reitsma J, Bossuyt PM, Kleijnen J. Development and validation of methods for assessing the quality of diagnostic accuracy studies. Health Technol Assess. 2004;8(25):1–234.
54. Whiting P, Rutjes AW, Reitsma JB, Bossuyt PM, Kleijnen J. The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. BMC Med Res Methodol. 2003;3:25.
55. Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med. 2011;155(8):529–36.
56. Schueler S, Schuetz GM, Dewey M. The revised QUADAS-2 tool. Ann Intern Med. 2012;156(4):323.
57. Hoch JS, Dewa CS. An introduction to economic evaluation: what's in a name? Can J Psychiatr. 2005;50(3):159–66.
58. Donaldson C, Vale L, Mugford M. Evidence based health economics: from effectiveness to efficiency in systematic review. UK: Oxford University Press; 2002.
59. Drummond MF, Jefferson TO. Guidelines for authors and peer reviewers of economic submissions to the BMJ. The BMJ economic evaluation working party. BMJ. 1996;313(7052):275–83.
60. Drummond MF, Richardson WS, O'Brien BJ, Levine M, Heyland D. Users' guides to the medical literature. XIII. How to use an article on economic analysis of clinical practice. A. Are the results of the study valid? Evidence-based medicine working group. JAMA. 1997;277(19):1552–7.
61. Husereau D, Drummond M, Petrou S, Carswell C, Moher D, Greenberg D, et al. Consolidated health economic evaluation reporting standards (CHEERS) statement. Value Health. 2013;16(2):e1–5.

62. Wong SS, Wilczynski NL, Haynes RB, Hedges T. Developing optimal search strategies for detecting clinically relevant qualitative studies in MEDLINE. Stud Health Technol Inform. 2004;107(Pt 1):311–6.
63. Vardell E, Malloy M. Joanna briggs institute: an evidence-based practice database. Med Ref Serv Q. 2013;32(4):434–42.
64. Hannes K, Lockwood C. Pragmatism as the philosophical foundation for the Joanna Briggs meta-aggregative approach to qualitative evidence synthesis. J Adv Nurs. 2011;67(7):1632–42.
65. Spencer L, Ritchie J, Lewis J, Dillon L. Quality in qualitative evaluation: a framework for assessing research evidence. UK: Government Chief Social Researcher's office; 2003.
66. Bouwmeester W, Zuithoff NP, Mallett S, Geerlings MI, Vergouwe Y, Steyerberg EW, et al. Reporting and methods in clinical prediction research: a systematic review. PLoS Med. 2012;9(5):1–12.
67. Hayden JA, van der Windt DA, Cartwright JL, Cote P, Bombardier C. Assessing bias in studies of prognostic factors. Ann Intern Med. 2013;158(4):280–6.
68. Wolff RF, Moons KGM, Riley RD, Whiting PF, Westwood M, Collins GS, et al. PROBAST: a tool to assess the risk of bias and applicability of prediction model studies. Ann Intern Med. 2019;170(1):51–8.
69. 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):71–2.
70. Tonelli MR. Integrating evidence into clinical practice: an alternative to evidence-based approaches. J Eval Clin Pract. 2006;12(3):248–56.
71. Woolf SH. Evidence-based medicine and practice guidelines: an overview. Cancer Control. 2000;7(4):362–7.
72. Polit DF. Assessing measurement in health: beyond reliability and validity. Int J Nurs Stud. 2015;52(11):1746–53.
73. Polit DF, Beck CT. Essentials of nursing research: appraising evidence for nursing practice, ninth edition: Lippincott Williams & Wilkins, north American; 2017.
74. Mokkink LB, de Vet HCW, Prinsen CAC, Patrick DL, Alonso J, Bouter LM, et al. COSMIN risk of bias checklist for systematic reviews of patient-reported outcome measures. Qual Life Res. 2018;27(5):1171–9.
75. Mokkink LB, Prinsen CA, Bouter LM, Vet HC, Terwee CB. The consensus-based standards for the selection of health measurement instruments (COSMIN) and how to select an outcome measurement instrument. Braz J Phys Ther. 2016;20(2):105–13.
76. Prinsen CAC, Mokkink LB, Bouter LM, Alonso J, Patrick DL, de Vet HCW, et al. COSMIN guideline for systematic reviews of patient-reported outcome measures. Qual Life Res. 2018;27(5):1147–57.
77. Swennen MH, van der Heijden GJ, Boeije HR, van Rheenen N, Verheul FJ, van der Graaf Y, et al. Doctors' perceptions and use of evidence-based medicine: a systematic review and thematic synthesis of qualitative studies. Acad Med. 2013;88(9):1384–96.
78. Gallagher EJ. Systematic reviews: a logical methodological extension of evidence-based medicine. Acad Emerg Med. 1999;6(12):1255–60.
79. Sacks HS, Berrier J, Reitman D, Ancona-Berk VA, Chalmers TC. Meta-analyses of randomized controlled trials. N Engl J Med. 1987;316(8):450–5.
80. Oxman AD. Checklists for review articles. BMJ. 1994;309(6955):648–51.
81. Oxman AD, Guyatt GH. Validation of an index of the quality of review articles. J Clin Epidemiol. 1991;44(11):1271–8.
82. Shea BJ, Grimshaw JM, Wells GA, Boers M, Andersson N, Hamel C, et al. Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews. BMC Med Res Methodol. 2007;7:10.
83. Shea BJ, Reeves BC, Wells G, Thuku M, Hamel C, Moran J, et al. AMSTAR 2: a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. BMJ. 2017;358:j4008.
84. Whiting P, Savovic J, Higgins JP, Caldwell DM, Reeves BC, Shea B, et al. ROBIS: a new tool to assess risk of bias in systematic reviews was developed. J Clin Epidemiol. 2016;69:225–34.
85. Davis DA, Taylor-Vaisey A. Translating guidelines into practice. A systematic review of theoretic concepts, practical experience and research evidence in the adoption of clinical practice guidelines. CMAJ. 1997;157(4):408–16.
86. Neely JG, Graboyes E, Paniello RC, Sequeira SM, Grindler DJ. Practical guide to understanding the need for clinical practice guidelines. Otolaryngol Head Neck Surg. 2013;149(1):1–7.
87. Browman GP, Levine MN, Mohide EA, Hayward RS, Pritchard KI, Gafni A, et al. The practice guidelines development cycle: a conceptual tool for practice guidelines development and implementation. J Clin Oncol. 1995; 13(2):502–12.
88. Tracy SL. From bench-top to chair-side: how scientific evidence is incorporated into clinical practice. Dent Mater. 2013;30(1):1–15.
89. Chapa D, Hartung MK, Mayberry LJ, Pintz C. Using preappraised evidence sources to guide practice decisions. J Am Assoc Nurse Pract. 2013;25(5):234–43.
90. Eibling D, Fried M, Blitzer A, Postma G. Commentary on the role of expert opinion in developing evidence-based guidelines. Laryngoscope. 2013; 124(2):355–7.
91. Chen YL, Yao L, Xiao XJ, Wang Q, Wang ZH, Liang FX, et al. Quality assessment of clinical guidelines in China: 1993–2010. Chin Med J. 2012; 125(20):3660–4.
92. Hu J, Chen R, Wu S, Tang J, Leng G, Kunnamo I, et al. The quality of clinical practice guidelines in China: a systematic assessment. J Eval Clin Pract. 2013; 19(5):961–7.
93. Henig O, Yahav D, Leibovici L, Paul M. Guidelines for the treatment of pneumonia and urinary tract infections: evaluation of methodological quality using the appraisal of guidelines, research and evaluation ii instrument. Clin Microbiol Infect. 2013;19(12):1106–14.
94. Vlayen J, Aertgeerts B, Hannes K, Sermeus W, Ramaekers D. A systematic review of appraisal tools for clinical practice guidelines: multiple similarities and one common deficit. Int J Qual Health Care. 2005;17(3):235–42.
95. Collaboration A. Development and validation of an international appraisal instrument for assessing the quality of clinical practice guidelines: the AGREE project. Qual Saf Health Care. 2003;12(1):18–23.
96. Brouwers MC, Kho ME, Browman GP, Burgers JS, Cluzeau F, Feder G, et al. AGREE II: advancing guideline development, reporting and evaluation in health care. CMAJ. 2010;182(18):E839–42.
97. Brouwers MC, Kho ME, Browman GP, Burgers JS, Cluzeau F, Feder G, et al. The global rating scale complements the AGREE II in advancing the quality of practice guidelines. J Clin Epidemiol. 2012;65(5):526–34.
98. Guyatt GH, Oxman AD, Kunz R, Falck-Ytter Y, Vist GE, Liberati A, et al. Going from evidence to recommendations. BMJ. 2008;336(7652):1049–51.
99. Andrews J, Guyatt G, Oxman AD, Alderson P, Dahm P, Falck-Ytter Y, et al. GRADE guidelines: 14. Going from evidence to recommendations: the significance and presentation of recommendations. J Clin Epidemiol. 2013; 66(7):719–25.
100. Tunguy-Desmarais GP. Evidence-based medicine should be based on science. S Afr Med J. 2013;103(10):700.
101. Muckart DJ. Evidence-based medicine - are we boiling the frog? S Afr Med J. 2013;103(7):447–8.
102. Mulrow CD. The medical review article: state of the science. Ann Intern Med. 1987;106(3):485–8.
103. Moher D, Schulz KF, Altman D, Group C. The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomized trials. JAMA. 2001;285(15):1987–91.
104. von Elm E, Altman DG, Egger M, Pocock SJ, Gotzsche PC, Vandenbroucke JP, et al. The strengthening the reporting of observational studies in epidemiology (STROBE) statement: guidelines for reporting observational studies. Lancet. 2007;370(9596):1453–7.
105. Sanderson S, Tatt ID, Higgins JP. Tools for assessing quality and susceptibility to bias in observational studies in epidemiology: a systematic review and annotated bibliography. Int J Epidemiol. 2007;36(3):666–76.
106. Willis BH, Quigley M. Uptake of newer methodological developments and the deployment of meta-analysis in diagnostic test research: a systematic review. BMC Med Res Methodol. 2011;11:27.
107. Whiting PF, Rutjes AW, Westwood ME, Mallett S, Group Q-S. A systematic review classifies sources of bias and variation in diagnostic test accuracy studies. J Clin Epidemiol. 2013;66(10):1093–104.
108. Swanson JA, Schmitz D, Chung KC. How to practice evidence-based medicine. Plast Reconstr Surg. 2010;126(1):286–94.
109. Manchikanti L. Evidence-based medicine, systematic reviews, and guidelines in interventional pain management, part I: introduction and general considerations. Pain Physician. 2008;11(2):161–86.
110. Gold C, Erkkila J, Crawford MJ. Shifting effects in randomised controlled trials of complex interventions: a new kind of performance bias? Acta Psychiatr Scand. 2012;126(5):307–14.