Review Article

Cutting-Edge Methodological Guidance for Authors in Conducting the Systematic Review and Meta-Analysis

Dewan Md. Sumsuzzman[1,2,3], Yonghoon Kim[2,3,4,5], Suhyeon Baek[2,3,4], Yonggeun Hong[1,2,3,4,]*

[1]Department of Physical Therapy, College of Healthcare Medical Science & Engineering, Inje University, Gimhae, Korea,
[2]Research Center for Aged-Life Redesign (RCAR), Inje University, Gimhae, Korea,
[3]Biohealth Products Research Center (BPRC), Inje University, Gimhae, Korea,
[4]Department of Rehabilitation Science, Graduate School of Inje University, Gimhae, Korea,
[5]Department of Physical Therapy, Chungdam Hospital, Seoul, Korea

The landscape of systematic reviews and meta-analyses (SRMA) in biomedicine has expanded exponentially, driven by the growing demand for evidence-based healthcare decision-making. However, the rapid increase of SRMAs has often outpaced the development of rigorous methodological standards, resulting in variability in quality and potentially limiting their effectiveness in informing healthcare practices. This gap highlights the critical need for advanced methodological guidance to enhance the quality and impact of SRMAs. Our contribution aims to provide comprehensive methodological direction for authors to conduct robust SRMAs. By effectively integrating qualitative and quantitative evidence, SRMAs can address complex healthcare questions more thoroughly than traditional reviews. Furthermore, these step-by-step guidelines will help researchers to address the challenges of synthesizing diverse types of evidence, thereby improving the rigor, relevance, and applicability of their findings in healthcare decision-making processes.

INTRODUCTION

Systematic reviews and meta-analyses (SRMA) are considered the gold standard for evidence-informed healthcare decision-making and potentially identifying research gaps to establish future research agendas [1]. Although the significance of meta-analysis findings from quantitative studies is well accepted in evidence-informed healthcare policy, researchers have also argued for the value of qualitative research in systematic reviews [2-4]. Recently, the principles of mixed methods research have been implemented in this process of systematic review [5]. Therefore, mixed methods research synthesis (MMRS) is considered a systematic review genre that combines qualitative and quantitative evidence to address complex healthcare questions [6,7]. Given the strengths and scope of this method, the MMRS method is becoming increasingly popular in healthcare decision-making [8]. The publication landscape of systematic reviews with or without meta-analysis (quantitative or qualitative) has reached epidemic proportions across the biomedical discipline in recent decades [9]. The major limitations of published SRMA include redundancy (around 33%), serious methodological flaws (around 50%), and misleading (around 17%) which will erroneously inform healthcare decision-making, increase waste in research, extend reproducibility crisis, and tarnish the prestige of these tools [9]. Despite the development of Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines for effectively reporting the systematic reviews, only 3% of SRMA are deemed to have adequate methods and be clinically useful [10], indicating the reporting guidelines alone are unable to overcome the current challenges. Therefore, methodological guidance for authors in conducting the SRMA is needed to help reduce this avoidable waste in research.

Cochrane Handbook is an excellent source of guidance for all steps of a systematic review. However, Cochrane Handbook requires a considerable time investment for the typical user, and it can be cumbersome to select among the wide range of resources available. Previous methodological guidelines for authors in conducting SRMA demonstrated profound limitations. For example, a previous guidance article by Muka et al. [11] did not offer guidance on how to validate and report the literature search strategies, when and how to update the search strategies, and how to increase the credibility of meta-analysis results and validate the methods of analysis or models used and their corresponding assumptions. Another paper comprehensively described the guidance of meta-analysis methodology but did not thoroughly guide prior steps of meta-analysis [12]. Disregarding the validation of search strategy and updating the search strategies may significantly compromise the quality of SRMA. In addition, sensitivity analysis is a quality control process of meta-analysis, but Muka et al. [11] skipped the guidance on when sensitivity analysis is essential and to what scenarios the authors need to consider for sensitivity analysis. Furthermore, the updated PRISMA (2020) statement includes new reporting guidelines such as the full search strategies for all databases, automation tools used in the study selection process, certainty assessment, and availability of data, code, and other materials that reflects advances in methods to identify, select, appraise, and synthesize studies [12]. However, most of the published author's guidance papers were based on an earlier version of PRISMA and rarely compliance with update one [11-14]. Finding the right method to conduct evidence synthesis may be challenging for authors as there are 41 alternative evidence synthesis methods available [15]. Hence, guidance is required for efficiently selecting the right methods. It is particularly important for those wishing to conduct an evidence synthesis for the first time or systematic reviewers with limited exposure to several evidence synthesis methods. Last but not the least, combining qualitative and quantitative evidence synthesis is only seldom undertaken, and most of the published systematic reviews are either qualitative or quantitative approaches [16].

The overall compliance with the quality of SRMA remains trivial because of suboptimal methodological guidance [17]. Therefore, this article aims to provide cutting-edge step-by-step methodological guidance for authors in conducting the top-quality SRMA (Fig. 1). These resources are intended to serve as indispensable tools for novice reviewers seeking comprehensive instruction.

STEP 1: FORMULATING THE RESEARCH QUESTION

A well-formulated research question lead to the most appropriate study design and methodology. As with any research, a well-defined review question is the first step of the systematic review process. The best way to accomplish it is by clearly framing the research question. The advantages of a well-formulated research question are provided in Supplementary Box 1. The clinical question can be categorized into background and foreground questions [18]. Background questions ask for general questions about clinical problems or a disease. These types of questions have two essential components (Supplementary Box 1). For example: What causes migraines? Foreground questions ask specific questions about clinical problems or diseases to inform clinical decisions. These questions are typically more specific, and complex compared to background questions. Most frequently, foreground questions investigate compari-
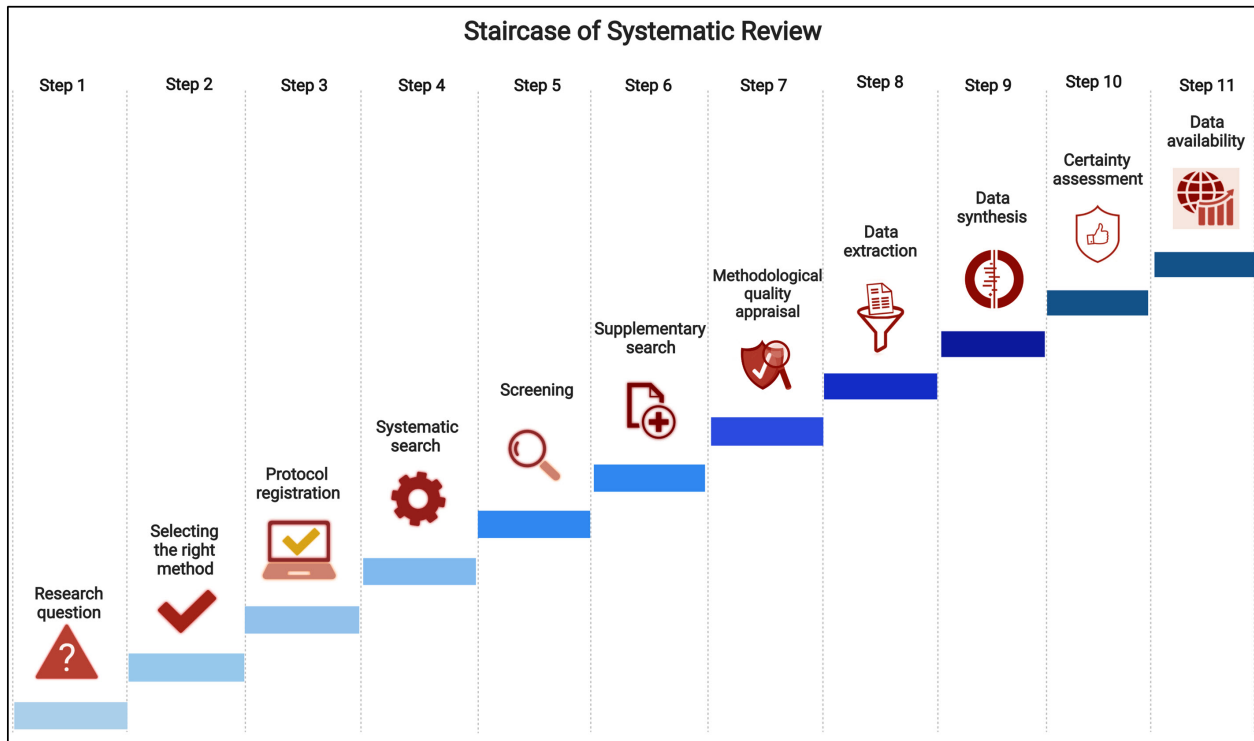
Fig. 1. Step-by-step methodological guidance for authors in conducting the top-quality systematic review and meta-analysis.

sons, such as two interventions, two diagnostic tests, etc. Foreground questions can be further categorized into four major types (Supplementary Box 1). For example: Is melatonin monotherapy effective when compared with placebo in improving cognition both in mild and moderate Alzheimer's disease patients? Developing a good foreground clinical question and defining the specific objectives of the study requires scoping search of the literature to identify gaps in the field. There are several tools available that may facilitate the formulating and analyzing of foreground questions.

Formulating a clear and focused research question required compliance with two frameworks. Firstly, the population, intervention, control, and outcomes (PICO) framework is considered a widely known strategy for framing a well-defined foreground research question (Supplementary Box 2). However, a well-defined research question may not necessarily be a good question. The proposed study may not be feasible in terms of time and resources, interesting to clinical practice, capable to generate new hypotheses, ethical appropriateness, and realistic perspectives (relevant to translating the findings to inform clinical decisions making) [19]. Cummings et al. [20] proposed another framework, feasible, interesting, novel, ethical and relevant (FINER), about how to build a good research question effectively. Supplementary Box 2 highlights the main characteristics of FINER criteria for systematic review. At the end of this step, the author should finalize the objectives, and selection cri-

teria of the study.

STEP 2: SELECTING THE RIGHT METHOD

The field of evidence-based medicine has adapted systematic review methods to address a wide variety of research questions. With over 20 alternative evidence synthesis methods available, selecting the appropriate method can be daunting, especially for those new to evidence synthesis or with limited exposure to various methods [21]. To address this challenge, the web-based "Right Review" tool has been developed. This tool guides systematic reviewers in selecting the most suitable method through five straightforward questions related to the study objective, interventions/diagnostics, types of evidence, types of analysis, and time/cost considerations [15]. After framing the research question, we, therefore, recommended the "Right Review" tool for choosing an appropriate evidence synthesis method, which could be the second step in the systematic review process. The "Right Review" tool can be found online: https://rightreview. knowledgetranslation.net/map/form.

STEP 3: PROTOCOL AND REGISTRATION

Systematic reviews ideally include a protocol outlining

predefined eligibility criteria and methodological approaches to maintain focus and address the research question. This protocol typically encompasses the study rationale, research question, primary and secondary aims, inclusion/exclusion criteria, electronic search strategy, data extraction plan, synthesis strategy, and timeline. An example protocol can be referenced elsewhere [22]. PRISMA-Protocols 17-item checklist is highly recommended to systematic reviewers to verify that each component of a protocol is completely reported and therefore reducing selective reporting bias [23]. The protocol should be published ideally at the same time as the systematic review is registered. PROSPERO, the international prospective register of systematic reviews (www.crd.york.ac.uk/PROSPERO/), is the most popular platform for systematic review registration.

STEP 4: SYSTEMATIC SEARCH OF THE RELATED LITERATURE

1. The comprehensiveness of the literature search

Systematic reviews require priori strategies to search the literature from the well-defined research question. Without a well-formulated search strategy, the identification of all relevant studies from electronic databases can be arduous. Advanced tips to develop search strategies in major bibliographic databases are shown in Appendix 1. Furthermore, unpublished data and grey literature search are now becoming more accessible to the public, and searching this literature adds value to the systematic review. For example, searching trial registries and grey literature reduce the publication bias of a systematic review [24]. To reduce the risk of reporting bias, the search of all relevant studies for a systematic review should be comprehensive enough. Detailed information on the different items of comprehensiveness of

the literature search is provided in Appendix 2.

2. Identifying sources of relevant literature

Numerous electronic databases serve as vital resources for retrieving primary studies, categorized into bibliographic, subject-specific, regional, clinical trial registries, grey literature, and web sources [25]. The choice of databases depends largely on the clinical question due to their varying scopes. Essential bibliographic databases include MEDLINE, Embase, Cochrane Library, Scopus, and Web of Science, while CINAHL caters specifically to nursing. A comprehensive search strategy typically integrates both bibliographic and subject-specific databases, as depicted in Supplementary Tables 1-2. For questions pertaining to specific regions or countries, identifying and searching regional databases may be necessary, as detailed in Supplementary Table 3. Accessing trial registries for unpublished data is critical for enhancing systematic reviews by reducing bias and improving conclusions, exemplified by updated Cochrane reviews on neuraminidase inhibitors [26]. The search sources to obtain unpublished data from trial registries are shown in Supplementary Table 4. Instructions on searching clinical trials and formulating effective search strategies, screening records, obtaining data, and updating searches are available elsewhere [27]. Similarly, including grey literature enriches review findings and minimizes publication bias, reflecting current practices [28]. Sources for grey literature databases are outlined in Supplementary Table 5. Fig. 2 illustrates potential approaches for locating both published and unpublished literature.

3. Updating search strategies

To maximize the currency of a review, an updated search of all relevant databases is recommended before submis-
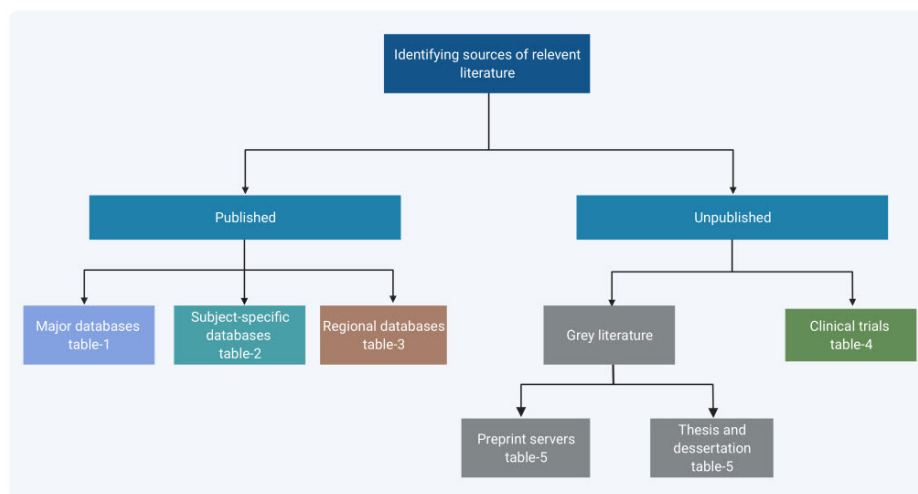

Fig. 2. Identification of source of relevant literature.

sion for publication [29]. Several ways the authors can update the searches, including setting auto-alerts, running the full search, and date limitations. Firstly, by setting auto-alerts in different databases, the corresponding authors will regularly receive the new results in an email. However, all the databases may not have an automatic alert service. Secondly, running the full search and de-duplicating the new search results against the old search results by using reference management software would be an attractive alternative strategy. The details for implementing this method in updating the searches can be found elsewhere [30]. Lastly, running the update search using publication date limitation options in each database, the authors purposefully update the searches. In the above discussion, to maximize the currency of a systematic review and reduce the waste in research, we suggest peer-reviewers routinely check the last search date during reviewing a systematic review manuscript.

4. Validating and reporting literature search

The literature search strategies should be validated because they can affect the overall quality of systematic reviews. To validate the search strategies for evidence syntheses, the authors and peers can follow the PRESS 2015 evidence-based checklist [31].

Suboptimal reporting of relevant literature searches is unable to reproduce how information was retrieved in a systematic review, which introduces bias in the final systematic review conclusions. Hence, PRISMA-S checklist is highly recommended to systematic reviewers to verify that each component of a search is completely reported and therefore reproducible [32].

STEP 5: SCREENING

Once the search strategy is executed across databases, importing retrieved records into reference management software initiates the screening process. Deduplication is crucial to streamline this phase, removing duplicates to create a unique library for efficient citation screening. According to the PRISMA guidelines, the study selection process should be done by at least two authors independently [33]. It reduces the workload and bias in the systematic review because it ensures all the records are screened more than once. Furthermore, any disagreement regarding study selection can be settled by discussion among the reviewers. Screening of relevant literature is further categorized into primary screening (title and abstract screening) and secondary screening (full-text screening). The title and abstract screening are labor-intensive and time-consuming

processes. Hence, there are several software tools that have been developed and widely used to facilitate these screening processes. Abstrackr, Colandr, Covidence, DRAGON, EPPI-Reviewer, and Rayyan are ranking higher acceptability in terms of their efficiency of screening. But all are not free, and the scope of these tools is different from each other. The selection of the most appropriate software to support the screening process will depend on the specific skill set and processes of the local research environment. Recent studies highlight Covidence and Rayyan as popular choices in healthcare research [34]. The reviewers should keep in mind that often not all outcome measures are described in the abstract therefore article must not be excluded based on specific outcome measures during the primary screening phase. Conflict of study selection can be solved either through full-text screening or both primary and secondary phases. The reviewers should explicitly mention in the manuscript how they resolved any disagreement during the study selection process. Nowadays, several machine learning-based tools are available for accelerating the screening process by sorting out irrelevant and relevant literature. Semi-automated or automated title and abstract screening with machine learning-based apps have the potential to save time and reduce research waste as these tools replaced the second reviewer. Abstrackr is a free, open-source, semi-automated online tool that can correctly identify all relevant citations [35]. There are other machine learning-based tools available for the title and abstract screening such as DistillerSR (paid software), and ASReview (open source, https://asreview.nl/about/). However, concerns about the reliability and false-negative rates of such tools persist, suggesting the need for further evaluation. Assessing the impact of these tools on review outcomes and conclusions remains an ongoing area of research.

STEP 6: SUPPLEMENTARY SEARCH

In the field of evidence synthesis, supplementary search is also termed "citation chasing". It involves using the citation network surrounding a source study to identify similar studies. Citation chasing can be classified into two types: i) forward citation chasing is referred to the process of finding all records citing one or more articles of known relevance; ii) backward citation chasing looks for the references of included studies that meet the inclusion criteria of a systematic review. In line with the Methodological Expectations of Cochrane Intervention Reviews group, backward citation searching is mandatory for conducting Cochrane reviews [36]. However, there is no guidance for forwarding citation chasing. Further study is required to develop recommendations for the use of forwarding citation chasing in terms of

comprehensiveness, and transparency of the systematic literature search process. Traditionally, this citation chasing process was done manually. For forward citation chasing, Google Scholar can be used to identify potentially relevant records from included studies of a systematic review. According to reverse citation chasing, the reference list of all included studies of a systematic review is usually checked manually to find out relevant articles. To accelerate the citation chasing process the "citationchaser" (https://estech.shinyapps.io/citationchaser/) application can be used.

STEP 7: APPRAISAL OF STUDY QUALITY AND RISK OF BIAS

Appraisal of the methodological quality (internal validity) in each included primary study is crucial as it makes the conclusion of a systematic review credible and trustworthy. Primary studies such as randomized controlled trials (RCTs), non-RCTs (NRCTs), cohort studies, case-control studies, cross-sectional studies, case series, case reports, diagnostic studies, and animal studies are most frequently included in a systematic review. Furthermore, secondary studies like systematic review and meta-analysis are often included in an umbrella review. Hence, appropriately judging study type is the priority, and choosing the proper tool for quality assessment is also equally important. A comprehensive list of tools for methodological quality assessment is included in Supplementary Table 6. However, all quality appraisal tools are subjective in nature, therefore authors of systematic reviews must receive training, and a minimum of two authors should be engaged in appraising and cross-checking to nullify the performance bias.

STEP 8: DATA EXTRACTION

1. Data extraction form and tools

Data extraction is the systematic process of gathering essential details on population characteristics, intervention specifics, and study outcomes from the selected research. The relevance of information extracted hinges on the specific research question, necessitating tailored data extraction forms. Planning how extracted data will be analyzed and presented in the manuscript is the initial phase of this process. During protocol development, a preliminary data extraction form and process may be outlined, requiring pilot testing to prevent extraction of irrelevant data. Utilizing standardized data extraction forms can mitigate bias, thereby enhancing the validity and reliability of the findings [37]. Therefore, enough time should be invested in the protocol

development phase to reduce the waste of resources. The authors should extract as much of the reported information as is likely to be needed so that the data synthesis process will be faster and easier. Representative information required for data extraction is shown in Supplementary Table 7. Several data extraction tools are now available to assist the authors in conducting a systematic review. The selection of optimal data extraction tools for systematic reviews depends on resources and review complexity including paper and pencil, spreadsheets, web-based surveys, electronic databases, and web-based specialized software [38]. The advantages and disadvantages of different data extraction tools are shown in Supplementary Table 8. Reporting of the tools used for data extraction in a systematic review is limited, which can make review findings questionable in terms of ambiguity, reproducibility, and applicability. Publishing of data extraction forms and reporting steps and tools used for data extraction in the method section are strongly recommended, which can significantly increase the transparency of the review process and make review findings more reliable.

2. Process of data extraction

Minimal two authors should be independently involved in data extraction to make the systematic review reliable and free from bias. Kappa statistic (a statistic that is used to measure inter-rater reliability) can be used to measure the level of inter-rater agreement. The liberal accelerated method can be used as an acceptable minimum, where one reviewer extracts all relevant data and the second reviewer verifies the data extraction forms for accuracy and completeness [39]. However, single author-driven data extraction generated more errors than two authors independently performing data extraction in systematic reviews. Any disagreement should be resolved by discussion among reviewers or by a third reviewer.

3. Extracting data from figures

Most often published preclinical studies (or sometimes RCTs) contain numerical data that are presented only in figures. Systematic reviewers may request the data from the corresponding or first author, which seldom yields results. Systematic reviewers could then exclude the data from the analysis, which introduces bias to the systematic review. In this scenario, the authors should try to extract graphical data from figures. Recently, Cochrane Croatia recommended Plot Digitizer (https://sourceforge.net/projects/plotdigitizer/) software for data extraction from figures [40]. This open-source software has a simple operating interface, faster data extraction process capability, and higher interra-

ter reliability than manual extraction. Therefore, along with Plot Digitizer, we suggested two additional open-source software for efficiently extracting data from figures: *Web-PlotDigitizer* and *GetDataGraphDigitizer*.

STEP 9: DATA SYNTHESIS

After data extraction, all the raw data from the newly created database need conversion into tidy data for data synthesis. Data syntheses can be qualitative (structured summary of non-numerical information) or quantitative (meta-analysis). When both these synthesis methods are used in a systematic review, in this guidance paper, we defined its MMRS.

1. Qualitative synthesis

Qualitative synthesis depends on primarily non-numerical information such as words and text to summarize and explain findings. Regardless of including a meta-analysis or not, reviewers should draw a PRISMA flow diagram of the systematic review by summarizing the number of references they found from the different databases, the number of abstracts and full texts they screened, the reasons for excluding studies, the number of studies included from other sources, and the final number of primary studies they included in the review. Reviewers should also tabulate the study characteristics such as the author's name, year of publication, location, population characteristics, intervention characteristics, follow-up, measurement scale, and outcomes of the included studies. Meta-analysis is not always feasible owing to several reasons such as incompletely reported outcomes, different effect measures used across studies, high risk of bias in the evidence, and too much clinical/methodological/statistical heterogeneity. In such a scenario, reviewers might consider presenting a harvest plot, effect direction plot, or albatross plot to present the results of included studies [41]. Presentation findings are especially important for transparent reporting in reviews without meta-analysis, displaying the data in a structured tabulation format that conveys detailed information is more efficient than the text format [41]. For example, a structured table of results across studies can be ordered by the risk of bias, authors, or certainty of evidence.

2. Quantitative synthesis: meta-analysis

Meta-analysis refers to the statistically synthesizing of quantitative data from two or more studies. During protocol development, the authors should specify reasonable details regarding pre-planned meta-analysis, including objectives,

effect size, model, method, statistical testing for investigating heterogeneity and interpretation of results, sensitivity analysis, meta-regression, and subgroup analysis. Supplementary Box 3 recommended minimum criteria during protocol development and reporting the results of the meta-analysis. Forest plot is the keyway researchers can summarize quantitative data from multiple papers in a single figure. For a novice in the field of evidence synthesis, the interpretation of forest plots is somewhat challenging. How these authors can read and interpret the forest plot is illustrated in Fig. 3.

3. Evaluation and investigation of heterogeneity (inconsistency)

Heterogeneity in meta-analysis refers to diversity among the results of individual studies. Heterogeneity will be a serious concern if the variation is substantial and unable to be explained through random variation or noticeable differences in PICO or study methodology. Considerable heterogeneity is more common in continuous than binary outcomes and authors should adduce a priori hypotheses to potentially explain variation in study results [42]. Authors can test such hypotheses by subgroup analyses, meta-regression, or sensitivity analyses. In this section, we discuss how to evaluate and investigate the source of heterogeneity. Both qualitative and statistical methods can be used to assess heterogeneity. Inspection of forest plots to determine the extent of variation in point estimates and the extent to which confidence intervals overlap refers to a qualitative approach for the identification of heterogeneity. On the other hand, Chi-square (x^2) statistic and I^2 are the two most common statistical approaches for the assessment of heterogeneity. The x^2 test assumes that all the studies are homogeneous (null-hypothesis), or each study is measuring an identical effect (alternative hypothesis) and gives us a p-value to test this hypothesis (if the p-value is low, we can reject the hypothesis and heterogeneity is present). However, the x^2 test has low power to detect heterogeneity when meta-analysis involved in small sample size. Additionally, if meta-analysis is involved in many studies, the x^2 test has high power to detect clinically unimportant heterogeneity [43]. Arguably, statistical heterogeneity due to clinical and methodological diversity is unavoidable [44]. Thus, quantifying heterogeneity across studies is crucial. In this scenario, the I^2 statistic is a useful tool to quantify heterogeneity. The I^2 statistic quantifies the percentage of the inconsistency in point estimates due to between-study differences; a low I^2 score suggests that included studies are considered homogeneous and a high indicates included studies are substantially heterogeneous. Heterogeneity can be categorized into low, moderate, and high heterogeneity corresponding to I^2
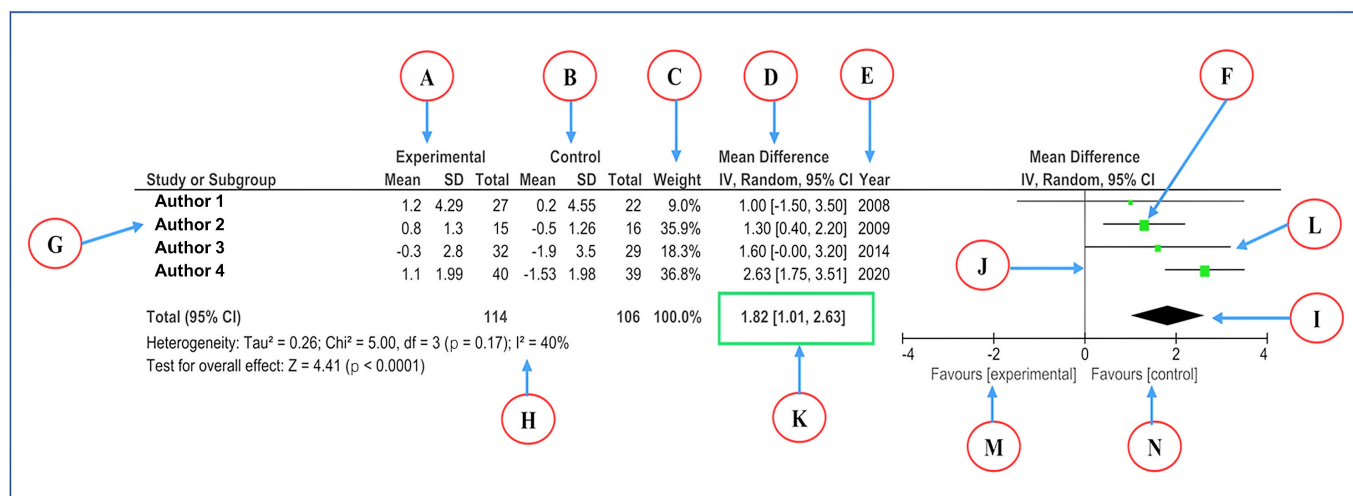
Fig. 3. An Example of a forest plot. Results from individual studies are plotted horizontally along a vertical line of no effect. The small squares indicate point estimates; the extending lines indicate confidence intervals. The black diamond represents the combined overall result calculated by meta-analysis, suggesting that the intervention is effective, lying as it does on the "favors experimental" side rather than the "favors control" side of the line of no effect. This conclusion could not be drawn from any of the individual studies alone, each of which showed no statistically significant effect. (A, B) Two distinct groups that are being compared in the analysis to assess their differences and similarities. (C) Weights are assigned to studies based on their contribution to the overall estimate. (D) Study effect measure in numeric forms. (E) Publication year. (F) The central square represents the mean treatment effect, with its size indicating the weight of that effect in the analysis. (G) Study identification by first author name. (H) Heterogeneity. (I) The diamond represents the overall effect estimate from all studies pooled together in the meta-analysis. (J) The vertical line of no effect. (K) Overall effect size and corresponding 95% confidence intervals. (L) The length of the lines the 95% confidence interval. (M) Values to the left of the null line favors the experimental group. (N) Values to the right of the null line favors the experimental group.

values around 25%, 50%, and 75%, respectively [25,36,43]. Recently, the Cochrane Collaboration categorized I^2 into four categories: unimportant (0-40%), moderate (30-60%), substantial (50-90%), and considerable heterogeneity (75-100%) [25,36,43].

4. Subgroup analysis

Subgroup analysis is involved in making a comparison between all participant data by splitting the categorical variable. Subgroup analyses can be used either to investigate inconsistency of results or hypothesis testing to answer specific research questions. Importantly, to address specific research questions, the subgroup analysis should be prespecified in the protocol and should not be undertaken in a post-hoc manner [45]. However, the authors are highly encouraged to investigate a covariate by post-hoc subgroup analysis when its importance was overlooked in the protocol. Appropriate interpretation of subgroup analysis results is crucial in hypothesis testing. The interpretation of subgroup analyses may be varied, and it depends upon different results produced from subgroup analyses. In this context, Richardson et al. [46] nicely demonstrated and recommended how to interpret subgroup analyses in the systematic review using five theoretical scenarios that cover almost all types of subgroup analysis results. These five theoretical scenarios and how the authors should interpret

each scenario are shown in Supplementary Table 9.

5. Meta-regression analysis

Meta-regression involves exploring whether a linear association exists between study level characteristics and treatment effect, along with the direction of that association. Ideally, clinical, or methodological diversity should be determined before the results are pooled in a meta-analysis and should be a sound rationale when discerning to undertake a meta-regression. However, overfitted regression model and aggregation bias may lead to invalid conclusions due to the insufficient number of studies, and aggregate data, respectively [47]. Cochrane Handbook suggests a minimum of 10 studies is necessary to undertake meta-regression to avoid overfitting regression models. Aggregate data will be misleading when authors explore the heterogeneity of patient-level factors (e.g., age, blood pressure) [48]. On the other hand, meta-regression based on aggregate data will be worthwhile, if authors explore heterogeneity within study-level factors (e.g., methodological quality, follow-up time) [48]. The key points to conduct and interpret the meta-regression analysis are shown in Supplementary Table 10.

6. Sensitivity analysis

Sensitivity analysis is a quality control process of meta-

analysis, which increases the credibility of meta-analysis results and validates the methods of analysis, or models used and their corresponding assumptions. The principle of sensitivity analysis is iteratively analyzing the primary analysis by removing a subset of studies (e.g., removing poor quality studies) or changing the statistical methods (e.g., fixed-effect to random effects) to determine whether these alterations have any effect on the combined outcome estimate [49]. There is no guidance available on when sensitivity analysis is essential and to what scenarios the authors need to consider for sensitivity analysis. In the above discussion, as we understand the clinical diversity, variation of intervention doses, methodological quality, study design, the time point of data analysis, missing data, inputted data, statistical methods, and considerable heterogeneity can significantly change the meta-analysis conclusion. Therefore, the systematic review authors should routinely consider these scenarios when conducting sensitivity analysis.

7. Publication bias

The presence of publication bias in a meta-analysis is considered a serious problem that can significantly affect the validity and generalizability of the conclusion. Suboptimal study design or execution, such as sample size and the method of reporting data, may introduce publication bias. Besides, the researcher's personal beliefs and expectations may also influence the results. Theses and dissertations were less likely to be published when they were negative findings than positive [50]. A possible explanation for this might be that researchers decide not to submit their negative results for publication because journal editors do not want to publish negative results. Several methods have been developed for detecting and adjusting correction of the publication bias are shown in Supplementary Table 11. Ideally, an assessment of publication bias should be included both the graphical (e.g., funnel plot) and statistical test (e.g., Egger's test, Begg's test). Egger's test is extensively used to detect publication bias, but its applicability might be limited to the binary outcome, not to continuous outcomes [51]. Further work is required to appraise the strengths and weaknesses of each method and recommend which one is most appropriate for the continuous outcome. Several methods have been developed to adjust/correct meta-analytic estimates for the possible effects of publication bias. The most widely used adjustment method is the nonparametric trim-and-fill method owing to its less complexity and availability in different software [52]. When statistical tests and funnel plots detect publication bias, the authors should perform trim-and-fill analysis or other suitable methods for adjustment of meta-analysis results. Caution should be taken before conducting publication bias when several studies in a funnel plot are less than ten as there are fewer studies the power of these tests is too low to distinguish chance from real asymmetry.

8. Software for meta-analysis

A brief timeline of the development of meta-analysis software concerning biomedical research is depicted in Fig. 4. Advantages and disadvantages of diverse meta-analysis software were provided in Supplementary Box 4.

STEP 10: CERTAINTY ASSESSMENT

Translating all available evidence to clinical practice recommendations needs a framework that assesses the certainty of the body of evidence. The systematic review authors can rate the certainty of evidence by utilizing the Grades of Recommendation, Assessment, Development, and Evaluation (GRADE) approach. GRADE tool has been adopted by more than 100 organizations worldwide including the Cochrane Collaboration, the World Health Organization, and National Institute for Health and Care Excellence of United Kingdom (National Institute for Health and Care Excellence), which can facilitate the systematic reviewers in assessing the certainty of the evidence and determining the strength of recommendations [53]. A brief description of GRADE and other certainty assessment ap-
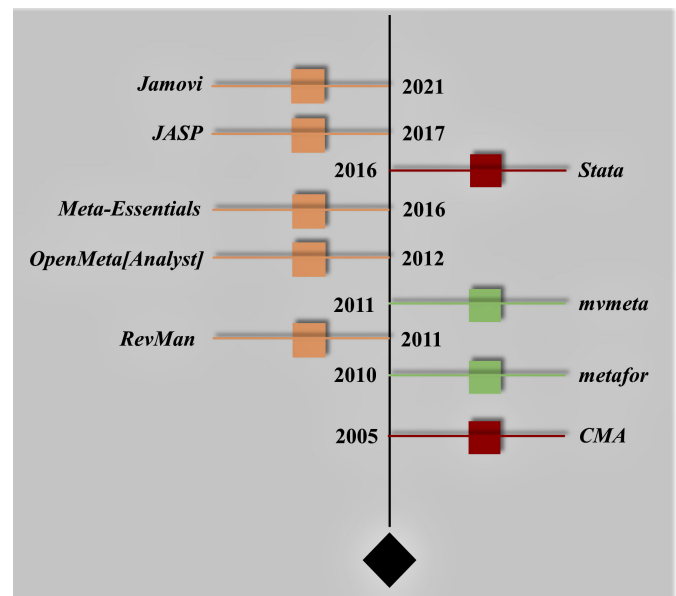

Fig. 4. Brief timeline of the development of meta-analysis software with respect to biomedical research. Orange-color: free and open-source software, red-color: paid software, green-color: R packages that required coding skills, RevMan: review manager, CMA: comprehensive meta-analysis.

proaches both in qualitative and quantitative synthesis was provided in Supplementary Box 5.

STEP 11: AVAILABILITY OF DATA, CODE, AND OTHER MATERIALS

Making publicly available all the extracted data throughout the systematic review may reduce redundancy and costs of research, improve the transparency and reproducibility, support reanalysis to answer secondary research questions, and facilitate meta-research. Reporting of template data collection forms, data extracted from included studies, clean datasets used for all analyses, metadata (complete description of the variable name), analytic code (complete description of steps implemented in software to run analyses or command-line), any other materials used in the systematic review are therefore recommended in updated PRISMA guidelines [33]. The authors should report where this information will be publicly available (e.g., a link to files deposited in a public repository). If the author states that these materials will be made available upon request, then the authors should provide the contact details of the author responsible for sharing the materials and describe in which circumstances such materials will be shared.

CONCLUSION

This article outlines essential steps and advanced methods for conducting SRMA. Our guide aims to support researchers in performing methodologically rigorous SRMA, thereby improving publication standards in evidence synthesis and facilitating its integration into healthcare decision-making. Furthermore, it aims to empower readers, reviewers, and healthcare providers to make informed judgments based on systematic review findings.

NOTES

- **ORCID**
  Dewan Md. Sumsuzzman,
  https://orcid.org/0000-0002-4468-1202
  Yonghoon Kim, https://orcid.org/0000-0001-8724-745X
  Suhyeon Baek, https://orcid.org/0009-0001-7958-0196
  Yonggeun Hong, https://orcid.org/0000-0003-1288-0546
- **Authors' contributions:** Y.H. participated in conceptualization, funding acquisition, providing resources, validation, and supervised the study. D.M.S., Y.K., and S.B. participated in data curation, drafting methodology, and providing software. D.M.S. participated in conducting the formal analysis, investigation, project administration, and visualization. Y.H. and D.M.S. wrote the original draft of the manuscript. All authors reviewed, edited, and approved the final version of the manuscript.

SUPPLEMENTARY MATERIALS

Supplementary data is available at https://doi.org/10.15280/jlm.2024.14.2.57

REFERENCES

1. Page MJ, Shamseer L, Altman DG, Tetzlaff J, Sampson M, Tricco AC, et al. Epidemiology and reporting characteristics of systematic reviews of biomedical research: A cross-sectional study. PLoS Med 2016;13(5):e1002028.
2. Heyvaert M, Hannes K, Onghena P. Using mixed methods research synthesis for literature reviews. SAGE Publications, Inc. 2016;1-344.
3. Oliver S, Harden A, Rees R, Shepherd J, Brunton G, Garcia J, et al. An emerging framework for including different types of evidence in systematic reviews for public policy. Eval Int J Theory Res Practice 2005:11(4);428-46.
4. Petticrew M, Rehfuess E, Noyes J, Higgins JP, Mayhew A, Pantoja T, et al. Synthesizing evidence on complex interventions: How meta-analytical, qualitative, and mixed-method approaches can contribute. J Clin Epidemiol 2013;66(11):1230-43.
5. Ferguson SL, Kerrigan MR, Hovey KA. Leveraging the opportunities of mixed methods in research synthesis: Key decisions in systematic mixed studies review methodology. Res Synth Methods 2020;11(5):580-93.
6. Heyvaert M, Maes B, Onghena P. Mixed methods research synthesis: Definition, framework, and potential. Qual Quant 2013;47:659-76.
7. Pluye P, Hong QN. Combining the power of stories and the power of numbers: Mixed methods research and mixed studies reviews. Annu Rev Public Health 2014;35:29-45.
8. Noyes J, Booth A, Moore G, Flemming K, Tunçalp Ö, Shakibazadeh E. Synthesising quantitative and qualitative evidence to inform guidelines on complex interventions: Clarifying the purposes, designs and outlining some methods. BMJ Glob Health 2019;4(Suppl 1):e000893.
9. Ioannidis JP. The mass production of redundant, misleading, and conflicted systematic reviews and meta-analyses. Milbank Q 2016;94(3):485-514.
10. Moore RA, Fisher E, Eccleston C. Systematic reviews do not

(yet) represent the 'gold standard' of evidence: A position paper. Eur J Pain 2022;26(3):557-66.

11. Muka T, Glisic M, Milic J, Verhoog S, Bohlius J, Bramer W, et al. A 24-step guide on how to design, conduct, and successfully publish a systematic review and meta-analysis in medical research. Eur J Epidemiol 2020;35(1):49-60.

12. Pigott TD, Polanin JR. Methodological guidance paper: High-quality meta-analysis in a systematic review. Rev Educ Res 2020;90(1):24-46.

13. Gopalakrishnan S, Ganeshkumar P. Systematic reviews and meta-analysis: Understanding the best evidence in primary healthcare. J Family Med Prim Care 2013;2(1):9-14.

14. Tawfik GM, Dila KAS, Mohamed MYF, Tam DNH, Kien ND, Ahmed AM, et al. A step by step guide for conducting a systematic review and meta-analysis with simulation data. Trop Med Health 2019;47:46.

15. Amog K, Pham B, Courvoisier M, Mak M, Booth A, Godfrey C, et al. The web-based "Right Review" tool asks reviewers simple questions to suggest methods from 41 knowledge synthesis methods. J Clin Epidemiol 2022;147:42-51.

16. Heyvaert M, Maes B, Onghena P. Applying mixed methods research at the synthesis level: An overview. Res Sch 2011;18:12-24.

17. Mahtta D, Altibi A, Gad MM, Samara A, Barakat AF, Bagur R, et al. Methodological rigor and temporal trends of cardiovascular medicine meta-analyses in highest-impact journals. J Am Heart Assoc 2021;10(18):e021367.

18. Seguin A, Haynes RB, Carballo S, Iorio A, Perrier A, Agoritsas T. Translating clinical questions by physicians into searchable queries: Analytical survey study. JMIR Med Educ 2020;6(1):e16777.

19. Fandino W. Formulating a good research question: Pearls and pitfalls. Indian J Anaesth 2019;63(8):611-6.

20. Cummings SR, Browner WS, Hulley SB. Conceiving the research question and developing the study plan. In: Hulley SB, Cummings SR, Browner WS, Grady DG, Newman TB, editors. Designing Clinical Research. 4th ed. Lippincott Williams and Wilkins;2013;14-22.

21. Tricco AC, Soobiah C, Antony J, Cogo E, MacDonald H, Lillie E, et al. A scoping review identifies multiple emerging knowledge synthesis methods, but few studies operationalize the method. J Clin Epidemiol 2016;73:19-28.

22. Greenfield K, Holley S, Schoth DE, Bayliss J, Anderson AK, Jassal S, et al. A protocol for a systematic review and meta-analysis to identify measures of breakthrough pain and evaluate their psychometric properties. BMJ Open 2020;10(3):e035541.

23. Moher D, Shamseer L, Clarke M, Ghersi D, Liberati A, Petticrew M, et al.; PRISMA-P Group. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. Syst Rev 2015;4(1):1.

24. Song F, Parekh S, Hooper L, Loke YK, Ryder J, Sutton AJ, et al. Dissemination and publication of research findings: An updated review of related biases. Health Technol Assess 2010;14(8):iii, ix-xi, 1-193.

25. Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, et al. Cochrane handbook for systematic reviews of interventions version 6.2. The Cochrane Collaboration, 2021.

26. Jefferson T, Jones MA, Doshi P, Del Mar CB, Hama R, Thompson MJ, et al. Neuraminidase inhibitors for preventing and treating influenza in adults and children. Cochrane Database Syst Rev 2014;2014(4):CD008965.

27. Hunter KE, Webster AC, Page MJ, Willson M, McDonald S, Berber S, et al. Searching clinical trials registers: Guide for systematic reviewers. BMJ 2022;377:e068791.

28. Paez A. Gray literature: An important resource in systematic reviews. J Evid Based Med 2017;10(3):233-40.

29. Chandler J, Churchill R, Higgins J, Lasserson T, Tovey D. Methodological standards for the conduct of new Cochrane Intervention Reviews. Sl: Cochrane Collaboration 2013;3(2):1-14.

30. Bramer W, Bain P. Updating search strategies for systematic reviews using EndNote. J Med Libr Assoc 2017;105(3):285-9.

31. McGowan J, Sampson M, Salzwedel DM, Cogo E, Foerster V, Lefebvre C. PRESS peer review of electronic search strategies: 2015 guideline statement. J Clin Epidemiol 2016;75:40-6.

32. Rethlefsen ML, Kirtley S, Waffenschmidt S, Ayala AP, Moher D, Page MJ, et al.; PRISMA-S Group. PRISMA-S: An extension to the PRISMA statement for reporting literature searches in systematic reviews. Syst Rev 2021;10(1):39.

33. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. BMJ 2021;372:n71.

34. Harrison H, Griffin SJ, Kuhn I, Usher-Smith JA. Software tools to support title and abstract screening for systematic reviews in healthcare: An evaluation. BMC Med Res Methodol 2020;20(1):7.

35. Rathbone J, Hoffmann T, Glasziou P. Faster title and abstract screening? Evaluating Abstrackr, a semi-automated online screening program for systematic reviewers. Syst Rev 2015;4:80.

36. Higgins J, Churchill R, Lasserson T, Chandler J, Tovey D. Update from the methodological expectations of Cochrane intervention reviews (MECIR) project. Cochrane Methods. Cochrane DB Syst Rev 2012;(Suppl 1):1-56.

37. Higgins JPT, Green S. Cochrane handbook for systematic reviews of interventions version 5.0.0. The Cochrane Collaboration, 2008.

38. Elamin MB, Flynn DN, Bassler D, Briel M, Alonso-Coello P, Karanicolas PJ, et al. Choice of data extraction tools for systematic reviews depends on resources and review complexity. J Clin Epidemiol 2009;62(5):506-10.

39. Khangura S, Konnyu K, Cushman R, Grimshaw J, Moher D. Evidence summaries: the evolution of a rapid review approach. Syst Rev 2012;1:10.

40. Jelicic Kadic A, Vucic K, Dosenovic S, Sapunar D, Puljak L. Extracting data from figures with software was faster, with higher interrater reliability than manual extraction. J Clin Epidemiol 2016;74:119-23.

41. McKenzie JE, Brennan SE. Chapter 12: Synthesizing and presenting findings using other methods. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA, editors. Cochrane handbook for systematic reviews of interventions version 6.0. The Cochrane Collaboration;2019;321-47.

42. Alba AC, Alexander PE, Chang J, MacIsaac J, DeFry S, Guyatt GH. High statistical heterogeneity is more frequent in meta-

analysis of continuous than binary outcomes. J Clin Epidemiol 2016;70:129-35.

43. Higgins JPT, Altman DG, Sterne JAC. Assessing risk of bias in included studies. In: Higgins JPT, Green S, editors. Cochrane handbook for systematic reviews of interventions version 5.1.0. The Cochrane Collaboration;2011:243-96.

44. Higgins JPT, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. BMJ 2003;327:557.

45. Burke JF, Sussman JB, Kent DM, Hayward RA. Three simple rules to ensure reasonably credible subgroup analyses. BMJ 2015;351:h5651.

46. Richardson M, Garner P, Donegan S. Interpretation of subgroup analyses in systematic reviews: A tutorial. Clin Epidemiol Glob Health 2019;7(2):192-8.

47. Geissbühler M, Hincapié CA, Aghlmandi S, Zwahlen M, Jüni P, da Costa BR. Most published meta-regression analyses based on aggregate data suffer from methodological pitfalls: A meta-epidemiological study. BMC Med Res Methodol 2021;21(1):123.

48. Schmid CH, Stark PC, Berlin JA, Landais P, Lau J. Meta-regression detected associations between heterogeneous treatment effects and study-level, but not patient-level, factors. J Clin Epidemiol 2004;57(7):683-97.

49. Bown MJ, Sutton AJ. Quality control in systematic reviews and meta-analyses. Eur J Vasc Endovasc Surg 2010;40(5):669-77.

50. Thornton A, Lee P. Publication bias in meta-analysis: its causes and consequences. J Clin Epidemiol 2000;53(2):207-16.

51. Freeman S, Sutton A. Identifying publication bias in meta-analyses of continuous outcomes [Internet]. Cochrane Training; 2020 [cited 2024 Jun 20]. Available from: https://training.cochrane.org/resource/identifying-publication-bias-meta-analyses-continuous-outcomes

52. Duval S, Tweedie R. A nonparametric "Trim and Fill" method of accounting for publication bias in meta-analysis. J American Stat Assoc 2000;95(449):89-98.

53. Goldet G, Howick J. Understanding GRADE: An introduction. J Evid Based Med 2013;6(1):50-54.