Towards rapid reviews improvements: the key methodological challenges

Ariany Marques Vieira

A Thesis

In

Health and Exercise Science

Presented in Partial Fulfillment of the Requirements

For the Degree of

Doctor of Philosophy (Health and Exercise Science)

At

Concordia University

Montreal, Quebec, Canada

August 2024

CONCORDIA UNIVERSITY

SCHOOL OF GRADUATE STUDIES

This is to certify that the thesis prepared

By: Ariany Marques Vieira

Entitled: Towards rapid reviews improvements: the key methodological challenges

and submitted in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY (*Health and Exercise Science*)

complies with the regulations of the University and meets the accepted standards with respect to originality and quality.

Signed by the final examining committee:

_____ Chair
Dr. Theresa Bianco

_____ External Examiner
Dr. Ivan Dario Florez Gomez

_____ Arms-Length Examiner
Dr. Maryse Fortin

_____ Examiner
Dr. Lisa Kakinami

_____ Examiner
Dr. Angela Alberga

_____ Thesis Supervisor
Dr. Simon Bacon

Approved by

_____
Dr. Maryse Fortin, Graduate Program Director

_____
Dr. Pascale Sicotte, Dean, Faculty of Arts and Science

ABSTRACT

Towards rapid reviews improvements: the key methodological challenges

Ariany Marques Vieira
Ph.D. in Health and Exercise Science
Concordia University, 2024

This thesis aims to identify the main methodological questions around Rapid Reviews (RRs) methods and undertake methodological studies to explore the impact of time-saving methods on review results. **Study 1:** An eDelphi study and consensus meeting were conducted, involving experts and evidence synthesis knowledge users. From an initial list, participants rated (low, medium or high importance) and ranked each item's importance to improve the time-efficiency of RRs. Items rated as high by ≥75% of participants progressed to the next round, and the final list was concluded during the consensus meeting. **Study 2:** This methodological study used Cochrane cardiac rehabilitation reviews to assess how database selection impacts study inclusion and outcomes. By examining where each included study was indexed and re-running meta-analyses, we evaluated whether treatment effects varied based on different database combinations. **Study 3:** This methodological study compared single-review and peer-review (two independent reviewers with a third for discrepancies) approaches for screening titles and abstracts. We assessed the percentage of missed studies, sensitivity, specificity, time, and costs for each method. **Results:** Study 1 identified seven highly important methodological questions. Three items on search strategy, two on study selection, one on quality/bias assessment, and one on data extraction. Study 2 found that Embase plus CENTRAL was the best database combination. When considering the estimated effects on mortality, when combining the major databases in pairs (MEDLINE, Embase, or CENTRAL), only 38% of results were identical to all databases combined. This percentage increased to 66% when combining three databases. Study 3 found that a single review approach missed 4% of inclusions (sensitivity was 0.84, and specificity was 0.86) and took half the time and costs of peer-review study selections. **Conclusion:** Search strategy is an important methodological question and based on our results, using at least three databases is recommended for a meta-analysis, but one large database

may suffice depending on the review context. Regarding study selection, a single review approach can be useful when time is short. This thesis sets a research agenda to optimise RRs and has the potential to influence global literature, establish best practices, and offer replicable methods for researchers.

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my supervisor, Dr. Simon Bacon. First, for accepting me as his graduate student, and then, for all the support and valuable mentorship. His guidance and support not only made the completion of this thesis possible but also made my whole PhD journey very enjoyable. Thank you for all the great opportunities to learn, network, and attend scientific events. Most of all, thank you for all your trust, for understanding my limitations, and for not sparing any effort to be a present supervisor. I am grateful for all the confidence in myself you helped me to achieve with all your feedback and compliments.

I want to thank the Montreal Behavioural Medicine, of which I am very proud to be a member. The opportunities to learn with each one of the members were very important and valuable. A special thank you to Guillaume and Geneviève, who were there since my first steps into the PhD, thank you so much for all your assistance! Thank you, Dr. Kim Lavoie, who has provided advice, support and encouragement throughout my PhD study. Thank you, Reyhaneh, for being my peer in this journey; I am glad we had each other and that we became friends. Thank you to all the past and current members of the META Group, it was a pleasure working with all of you. Thank you so much, Drs. Paula and Jovana, who both helped me when I first arrived in Montreal. Every conversation we had on public transportation, or any meetings we had together was very important to me. I miss the time before the pandemic when we used to work together at Concordia. You both made the entire difference for me in adapting to Montreal and believing the effort would be worth it.

I would also like to sincerely thank my supervisory committee members, Drs. Kim Lavoie, Angela Alberga, and Lisa Kakinami, for all your valuable contributions and the great interactions during my PhD. They provided great support, especially for my Comprehensive exam and Thesis proposal. In addition, I would like to thank all the members of my evaluation committee who offered great insights during my thesis defense, Drs Angela Alberga, Lisa Kakinami, Ivan Florez, Maryse Fortin, and the chair of the defence, Dr Theresa Bianco.

I am very grateful to the funding agencies that have supported my work throughout my CIHR-SPOR Chair doctoral scholarship and, lately, my Fonds de recherche du Québec: Santé (FRQS) doctoral scholarship. I also would like to acknowledge all the support I received for the completion of Chapter 3 of this thesis:

This research received the support of the Mitacs Globalink Research Award in form of an intern stipend (host academic supervisor: Dr Rod Taylor). Also, thank you Cochrane Heart, for contributing to the access to the original files and database libraries, and Valerie Wells for verifying the searches and resolving issues related to database access. Thank you, both Mariam Atoui and Ivan Amorim, for working on the code and verifying the outcomes analysis using RStudio.

Thank you to all my friends and everyone that I had the pleasure of meeting throughout my journey. It was not easy to find all of you at first, but I am glad that I attended different events and tried to connect with you all. Having your company and support in Montreal made the PhD journey much more enjoyable and I do not have enough words to express how much difference you all have made to my life in the last five years. It would be hard to name each one of you, because, luckily, I have quite a few friends for whom I am thankful. Those that I connected with before coming to Montreal and shared so much information about the city and student life and opened their house to me when I had first arrived in Montreal. Those who were also living as an international student in Montreal, those I shared an apartment with, those that I met at the Orientation week at Concordia, those who came to Montreal just for a few months, the ones I met at the French course… and of course all my long-distance friends, whom I miss a lot but who showed me that some friendships can not only resist distance but can even become stronger. Thank you for all your attention and each message of encouragement, Rafa (Dr Rafael Adolfo).

I want to thank my father, Lauro Cesar (in memoriam), who always supported and encouraged me to study. You were a great example of hard work. I hope to make you proud regardless of where you are watching me from.

I am also grateful to my husband, Ivan, for supporting my dreams, believing in me, being patient, and listening to me. Thank you for all the cheerful moments and ongoing support. I am so glad to have you as a life partner. I want to extend my sincere thanks to your parents, Marcos and Maria, who both give us lots of support and love.

Lastly, I dedicate this thesis to my family. My mother, Arlete, who gives me all the love and support. She encourages me throughout my endeavours although she might not fully understand all my dedication and choices. I thank you for your support and for understanding my absence; I love you more than anything. To my sister Rarietty, who has always been an example to me and someone that I can always

count on. I can not express how much it means knowing I will always have you by my side. Thank you for being a great sister and friend and for all your encouragement. My brother-in-law Guilherme, who has provided support and even drove me twice for the required medical exams for my student visa. To my niece Manuela, who fills my life with joy. I am so grateful to have you in my life, and I love you so much. Being far away from you is probably the hardest thing I have ever done. I remember when I left, you were only two years old, and you kept calling me when you arrived at our home without understanding that I was far away. Now you are seven years old, and although it is still so hard to be far from you, we can now call each other, and you are always ready to welcome me or say goodbye when the time to leave comes; I really appreciate that, and I am sorry for my absence.

CONTRIBUTION OF AUTHORS

The present thesis comprises four manuscripts, each at a different phase of the publication process. Details of the publication status and contributions of each author are presented below.

Chapter 2: Identifying priority questions regarding rapid systematic reviews' methods: protocol for an eDelphi study

This protocol was published in 2023 (Vieira AM et al. BMJ Open 2023;13:e069856. doi:10.1136/bmjopen-2022-069856). The journal granted authorisation to reproduce the entire article as part of this thesis as the article was published under a CC BY NC license, which allows use for any non- commercial purpose (Appendix I). The only change made to the version that is part of this thesis was the number of the tables (e.g. Table 1-1-, instead of Table 1). As the lead author, I was responsible for the project development and involved in all aspects of the study: concept; design and methods; ethical application; and drafting of the manuscript. This study included a research team responsible for the administration of the project and a scientific committee involved in several aspects of the study development, as detailed below. All the authors read and accepted the last version of the protocol.

Concept: Ariany M Vieira, Geneviève Szczepanik, Jovana Stojanovic, Paula A B Ribeiro, and Simon L Bacon.

Design and Methods: Ariany M Vieira, Chiara de Waure, Andrea Tricco, Sandy Oliver, Jovana Stojanovic, Paula A B Ribeiro, Danielle Pollock, Elie Akl, John Lavis, Tanja Kuchenmuller, Peter Bragge, Laurenz Langer, Simon L Bacon.

Ethical application: Ariany M Vieira, Geneviève Szczepanik.

Drafting of the manuscript: Ariany M Vieira and Simon L Bacon.

Critical revision of the manuscript for important intellectual content: Ariany M Vieira, Geneviève Szczepanik, Chiara de Waure, Andrea Tricco, Sandy Oliver, Jovana Stojanovic, Paula A B Ribeiro, Danielle Pollock, Elie Akl, John Lavis, Tanja Kuchenmuller, Peter Bragge, Laurenz Langer, Simon L Bacon.

Supervision: Simon L Bacon.

Chapter 2: Essential methodological questions for developing time-efficient rapid reviews: results from an eDelphi study

This study reports the results of the previous protocol. As the lead author, I was involved in all phases of the study development and was responsible for coordinating each step. The version of the manuscript that is part of this thesis was not edited after receiving feedback from co-authors. The study is ready to be submitted, with additional adjustments, to Research Synthesis Methods Journal. As mentioned above, this study had a research team responsible for the administration of the project and a scientific committee involved in several aspects of the study development, as detailed below.

Concept: Ariany M Vieira, Geneviève Szczepanik, Jovana Stojanovic, Paula A B Ribeiro, and Simon L Bacon.

Design and Methods: Ariany M Vieira, Chiara de Waure, Andrea Tricco, Sandy Oliver, Jovana Stojanovic, Paula A B Ribeiro, Danielle Pollock, Elie Akl, John Lavis, Peter Bragge, Laurenz Langer, Simon L Bacon.

Ethical application: Ariany M Vieira, Geneviève Szczepanik.

Drafting of the manuscript: Ariany M Vieira and Simon L Bacon.

Critical revision of the manuscript for important intellectual content: Ariany M Vieira, Geneviève Szczepanik, Chiara de Waure, Andrea Tricco, Sandy Oliver, Jovana Stojanovic, Paula A B Ribeiro, Danielle Pollock, Elie Akl, John Lavis, Peter Bragge, Laurenz Langer, Simon L Bacon.

Additional researchers involved in the consensus meeting and invited to provide scientific contributions to the manuscript will also be invited to co-author the manuscript.

Supervision: Simon L Bacon.

Chapter 3: The impact of database choices on systematic review results: a case study using Cochrane cardiac rehabilitation reviews

For the development of this study, I had the privilege to be co-supervised by Dr Rod Taylor at the University of Glasgow, MRC/CSO Social and Public Health Sciences Unit. As the lead author, I was involved in all phases of the study development and was responsible for coordinating each step. This study is in preparation for submission.

Concept: Ariany M Vieira, Rod S Taylor, and Simon L Bacon.

Design and Methods: Ariany M Vieira, Keven Joyal-Desmarais, Rod S Taylor, and Simon L Bacon.
Data collection: Ariany M Vieira.
Code development: Ariany M Vieira, Keven Joyal-Desmarais and members of Data management team of the Montreal Behavioural Medicine Centre.
Data analysis: Ariany M Vieira.
Drafting of the manuscript: Ariany M Vieira.
Critical revision of the manuscript for important intellectual content: Ariany M Vieira, Keven Joyal-Desmarais, Rod S Taylor, Simon L Bacon.
Supervision: Rod S Taylor and Simon L Bacon.

Chapter 4: Methodological Insights: Comparing Single Review and Peer-Review Approaches in Study Selection for Evidence Synthesis

This study was developed in collaboration with the Montreal Behavioural Medicine Centre Evidence Synthesis Team members, the META Group. As the lead author, I was involved in all phases of the study development and was responsible for coordinating each step. This study is in preparation for submission.
Concept: Ariany M Vieira and Simon L Bacon.
Design and Methods: Ariany M Vieira, Laurence Paquet, and Simon L Bacon.
Data collection: Ariany M Vieira, Laurence Paquet, Camille Leger, Frederique Deslauriers and Jessica Burdick, Nana Wu.
Data analysis: Ariany M Vieira, Laurence Paquet.
Drafting of the manuscript: Ariany M Vieira.
Critical revision of the manuscript for important intellectual content: Ariany M Vieira, Laurence Paquet, Simon L Bacon.
Supervision: Simon L Bacon.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

Introduction

Evidence syntheses are a useful strategy to provide an overview of a specific topic area and are fundamental for enabling evidence-based decisions for health policy and systems.[1] Nonetheless, it is crucial to employ rigorous methodologies in producing these evidence summaries to ensure that the results are trustworthy and can inform precise clinical and policy decisions. Systematic reviews (SRs) are the most recognized type of evidence synthesis and, when based on randomized clinical trials (RCTs), are considered the highest level of evidence.[2]

SRs are a way of searching and selecting the available empirical evidence to answer a research question whilst reducing bias.[3] As SRs employ rigorous methodologies to generate a conclusion, they require notable time and resources, meaning that there are several barriers to conducting them. Performing a SR is a time-intensive process, with an estimated average time of 61 weeks from conception to completion.[4] Consequently, by the time an SR is finalised, it might not reflect the available literature anymore as it may not have new and relevant publications incorporated into the synthesis. Also, high-quality SRs need substantial resources, with an estimated cost of above 100,000 USD per SR.[5]

To address these challenges of SRs, rapid reviews (RRs) have been developed. Its mean estimated production time is two months,[6] ranging from two to 13 weeks.[7] This approach can be essential to policymakers and health systems managers who need reliable evidence to make timely decisions on healthcare. In this case, a full SR may not always be feasible or practical, so RRs are an alternative evidence synthesis approach. Table 1 lists the advantages and disadvantages when comparing RRs and SRs.[8] However, unlike SRs, to date, there is limited evidence on how best to efficiently conduct RRs.

Table 1. Comparison between advantages and disadvantages of Rapid Reviews compared to Systematic Reviews

Advantages	Disadvantages/ Limitations	Citations
Effective in synthesizing and disseminating information in a timely fashion to inform evidence-based practice	Accelerated review process often leads to a search that is less comprehensive with the potential for limited validity or generalizability of the results	Ganann et al., 2010; Haby et al., 2016; Hailey et al., 2000; Hartling et al., 2015; Hartling et al., 2017; Khangura et al., 2012; Patnode et al., 2018; Polisena et al., 2015; Tricco et al., 2015; Watt et al., 2008
No concrete methodology allows for flexibility and adaptability of the review process	Narrowing, restricting, or omitting parts of the review process may introduce bias, error, or result in missing evidence	Ganann et al., 2010; Haby et al., 2016; Hartling et al., 2017; Khangura et al., 2012; Patnode et al., 2018; Polisena et al., 2015; Schünemann & Moja, 2015; Tricco et al., 2015; Tsertsvadze et al., 2015; Watt et al., 2008
Less expensive and requires less resources or sustained investment compared to a systematic review	Research teams should engage content experts in the field to ensure accuracy of results due to limited time for training or quality control	Ganann et al., 2010; Haby et al., 2016; Khangura et al., 2012; Marshall et al., 2019; Polisena et al., 2015; Tricco et al., 2015

Credit to table given to Molcak, H. S., Appleby, C. J., Brown, J., Freeman, S., Kandola, D. K., & Banner, D. Rapid Knowledge Syntheses: Methodological and Practical Considerations. Canadian Journal of Cardiovascular Nursing. Volume 31, Issue 1, Spring 2021. No changes were made to this table.

Rapid Reviews – Definition

According to Hamel et al.,[9] RRs first appeared in the literature when a rapid health technology assessment report was described by Best et al..[9] RRs gained more attention in the early 2000s when they were not yet commonly known as “rapid reviews” as it is nowadays. Since that time, there have been different terminologies used to refer to RRs, such as “rapid evidence-based literature review,” “brief review,”

and “rapid evidence assessment of the literature.”[10] In 2018, Aronson and colleagues argued that rather than rapidity, what makes these reviews different from SRs are the methods applied, so they proposed the term “restricted review”,[11] as “rapid” implies the significance of time. However, the term Rapid Reviews had already been established in the literature, and in 2015, Tricco and colleagues showed that this was the most frequent term used.[12]

In 2021, Hamel et al. performed a systematic scoping review of the definitions of RRs.[9] From the 146 RRs that provided a definition, there were four components that were more consistently mentioned: a rapid or accelerated process; a variation in the methods from SRs; a limited scope compared to a SR; and triggered by an end-user request.[9] From this, the authors suggested the following definition: “A rapid review is a rigorous and transparent form of knowledge synthesis that accelerates the process of conducting a traditional systematic review through streamlining or omitting a variety of methods to produce evidence for stakeholders (end-users) in a resource-efficient manner.” This is the most updated definition, which is also endorsed by Cochrane, and the one that we are using for the purpose of the present thesis.

A definition of RRs was also included in a recent position statement by the Joanna Briggs Institute, which claims that RRs are not a type of evidence synthesis but rather an approach or mindset when conducting any type of review. Compared to “living” reviews, different evidence syntheses can also be rapid.[13] This contrasts the definition of the Cochrane Rapid Reviews Methods Group, which considers RRs to be a form of knowledge synthesis.[14]

Rapid Reviews – Rationale

RRs provide relevant evidence for policy-makers by using several methods for rapid searching and appraising the available literature while balancing time and mitigating against biases.[2] This product/method/approach to evidence synthesis is especially relevant when timely evidence for decision-making is needed to answer questions on urgent and emergent health issues,[14] considering the time and resources needed to develop a full systematic review.

The Cochrane Methods groups, and other authors recommend the development of RRs only to address urgent and high priority questions explicitly requested by decision makers.[4,14,15] Although it is known that engaging policymakers

and health systems managers in the development of RRs may increase their relevance and applicability,[16] it seems that with the COVID-19 pandemic, RRs may no longer follow a potential end-user request or reasoning. Hamel and collaborators identified end-user rationale as one of the eight main themes defining a RR.[9]

Rapid reviews are increasingly being used to provide evidence in both routine and emergency contexts.[17] However, authors have also argued against its use, hoping to see less RRs published in the future as timely reviews could be produced without these streamlined methods.[18] It is recommended to consider several factors when planning a RR, as the resources available,[17] communication and needs of policy makers or other end-users,[4,17,19–21] and the appropriate methods.[22,23]

Rapid Reviews – Methods

Besides the definition, there has been a lack of consistency in RR approaches, inadequate reporting, and heterogeneity of methods or processes.[24] Various procedures have been performed differently across RRs. For example, some restrict the search strategy (e.g., including only published literature or fewer databases), limit the inclusion criteria, have only one reviewer selecting the studies, do not conduct risk of bias/quality appraisal, etc..[6,10,22] The available guides for conducting RRs are vague too, suggesting balancing the methodological choices according to resources and tailoring according to the end-users' need.[17] The Cochrane recommendations also provide valuable insights into the shortcuts and tools that can facilitate the RR process.[14]

Although it is recommended for RRs to streamline its methods according to the needs of the decision-maker and the resources available to conduct the RR,[17] the impact of these timesaving methods is not well understood. Biases may be introduced through changes to scope or timeframe (e.g., selection bias, publication bias, language of publication bias). In 2023, Haby and colleagues performed a SR exploring the best methods for RRs.[22] In agreement with previous literature,[6,10,12] they found that a range of methods have been used in RRs and that some shortcuts have the potential to increase bias, such as a simplified search strategy, single versus double screening of titles and abstracts and full-texts, and machine learning approaches to aid title/abstract screening. For various streamlined methods, the authors were unable to make conclusions due to insufficient evidence. Overall, little empirical evidence is available to support the methodological choices.[6] For example,

comparing single versus peer-review screening of titles and abstracts.[25] A recent study exploring this found that single-reviewer abstract screening missed 13.4% of relevant studies compared to 2.5% lost by peer review.[26] In contrast, Cochrane recommends that one reviewer be used to include studies, with two reviewers being needed to excluded at title and abstract screening (i.e., peer-reviewed exclusions).[14] However, it is still unclear what the impact of these approaches is on the magnitudes of effects seen or the final conclusions of the review.

Considering the various timesaving methods being used, an important research agenda is the investigation on the impact of omitting or abbreviating review processes,[13] to understand if, or to what extent, biases can be introduced with these strategies.[6] However, there are too many questions about the broadly used but unjustified shortcuts that are currently used to know the best place to start.[22] For instance, what are the impacts of the different search strategies, screening approaches, or having only one reviewer performing risk of bias assessment. In 2021, Evidence Synthesis Ireland conducted a priority-setting study using the James Lind Alliance method to identify research priorities covering all the phases of a RR (question generation, how we plan, do, and disseminate findings).[27] The study resulted in a broad list of questions, and among the top 10, three of them were focused on methodological issues. One of the questions was, ‘what simplified methods of SRs could be used in a RR, and what would be the impact of these choices in general.’ The present thesis will explore this area further to generate more focused questions about the particular methods that require investigation and then conduct studies to start to address the issues raised, providing empirical evidence of the impact of such choices on outcomes.

Object and Aims of the Thesis

Considering the gaps previously shown, such as the several unanswered questions about the methods for RRs and the lack of understanding of the impact of the abbreviated processes, the ultimate aim of this thesis is to advance the science of RRs, enhancing the methodological rigor by prioritising the critical methodological questions that need to be addressed and undertaking methodological studies to explore how aspects of RR methods impact the review processes and review findings.

Based on the findings from the first study, an eDelphi process, we designed the subsequent studies to help answer some of the questions considered of high importance by experts in the fields.

The aims of this thesis (and the chapters in which they have been addressed) are:

i) To identify what are the most important methodological questions (from the generation of the question to the writing of the report) for the field to address in order to guide the effective and efficient development of RRs. (Chapter 2)

ii) To assess the impact of databases' selection on the inclusion of studies and subsequent impacts on outcome point estimates in cardiac rehabilitation-based systematic reviews. (Chapter 3)

iii) To compare two studies' selection approaches (peer review and single review) for title and abstract screening. (Chapter4)

CHAPTER 2

Identifying priority questions regarding rapid systematic reviews' methods: protocol for an eDelphi study

Ariany M Vieira[1,2], Geneviève Szczepanik[2], Chiara de Waure[3], Andrea C Tricco[4], Sandy Oliver[5], Jovana Stojanovic[1,2], Paula A B Ribeiro[2], Danielle Pollock[6], Elie A Akl[7,8], John Lavis[9,10], Tanja Kuchenmuller[11], Peter Bragge[12], Laurenz Langer[13], Simon Bacon[1,2]

[1] Department of Health, Kinesiology and Applied Physiology, Concordia University, Montreal, Quebec, Canada
[2] Montreal Behavioural Medicine Centre, CIUSSS du Nord-de-l'ile-de-Montreal, Montreal, Quebec, Canada
[3] Department of Medicine and Surgery, University of Perugia, Perugia, Italy
[4] Knowledge Translation Program, Li Ka Shing Knowledge Institute of St Michael's Hospital, Unity Health Toronto, Toronto, Ontario, Canada
[5] University College London Institute of Education, London, UK
[6] Health Evidence Synthesis, Recommendations and Impact, School of Public Health, The University of Adelaide, Adelaide, South Australia, Australia
[7] Department of Internal Medicine, American University of Beirut, Beirut, Lebanon
[8] Department of Health Research Methods, Evidence, and Impact (HEI), McMaster University, Hamilton, Ontario, Canada
[9] McMaster Health Forum and Department of Health Research Methods, Evidence and Impact, McMaster University, Hamilton, Ontario, Canada
[10] Affiliate Scientist, Institute for Better Health, Trillium Health Partners, Mississauga, Ontario, Canada
[11] Research for Health Department, Science Division, World Health Organization, Geneva, Switzerland
[12] Monash Sustainable Development Institute Evidence Review Service, Behaviour Works Australia, Monash University, Melbourne, Victoria, Australia
[13] South Africa Centre for Evidence, Johannesburg, South Africa

Publication details:

Vieira AM, Szczepanik G, de Waure C, et al. Identifying priority questions regarding rapid systematic reviews' methods: protocol for an eDelphi study. BMJ Open 2023;13:e069856. doi:10.1136/bmjopen-2022-069856.

Identifying priority questions regarding rapid systematic reviews' methods: protocol for an eDelphi study

Abstract

Introduction: Rapid systematic reviews (RRs) have the potential to provide timely information to decision-makers, thus directly impacting healthcare. However, consensus regarding the most efficient approaches to performing RRs and the presence of several unaddressed methodological issues pose challenges. With such a large potential research agenda for RRs, it is unclear what should be prioritised.

Objective: To elicit a consensus from RR experts and interested parties on what are the most important methodological questions (from the generation of the question to the writing of the report) for the field to address in order to guide the effective and efficient development of RRs.

Methods and analysis: An eDelphi study will be conducted. Researchers with experience in evidence synthesis and other interested parties (e.g., knowledge users, patients, community members, policymaker, industry, journal editors, and healthcare providers) will be invited to participate. The following steps will be taken: 1) A core group of experts in evidence synthesis will generate the first list of items based on the available literature; 2) Using LimeSurvey, participants will be invited to rate and rank the importance of suggested RR methodological questions. Questions with open format responses will allow for modifications to the wording of items or the addition of new items; 3) Survey rounds will be performed asking participants to re-rate items, with items deemed of low importance being removed at each round; 4) A list of items will be generated with items believed to be of high importance by ≥75% of participants being included; and 5) This list will be discussed at an online consensus meeting that will generate a summary document containing the final priority list. Data analysis will be performed using raw numbers, means, and frequencies.

Ethics and dissemination: This study was approved by the Concordia University Human Research Ethics Committee (#30015229). Both traditional, e.g., scientific conference presentations and publication in scientific journals, and non-traditional, e.g., lay summaries and infographics, knowledge translation products will be created.

Keywords: Rapid review; Systematic review; Delphi; Priority setting; Consensus; Evidence synthesis.

Strengths and limitations of this study:

- The eDelphi process is a well-recognised and highly structured method for consensus building.
- Understanding potential differences in research priorities will be made possible by including a variety of participant profiles, researchers, and key end users (such as policy-makers, guideline producers, healthcare professionals, etc.).
- The modified eDelphi approach, using an online format, although it may elicit challenges, can also allow for faster data collection, a broader range of individuals across the globe, is more cost-effective than in person Delphi approaches, and is less susceptible to the judgements of group members with higher status.
- Although this study is an important addition to the literature in the evidence-synthesis field, and it can serve as a 'road-map' for future RR methodological studies, it is only the first step towards refining the conduct of Rapid systematic reviews in a more time-efficient way.

1. BACKGROUND

Evidence syntheses (e.g., systematic reviews [SRs]) are a useful strategy for a number of uses and domains, notably to summarise evidence around a specific question.[28] In a health context, findings form SRs have been used to make decisions for: clinical practice, normally through clinical practice guidelines; healthcare systems; and shaping policy.[28,29] However, conducting a full SR is time-consuming, sometimes taking up to two years to conduct,[3] by which time the scientific literature may have already moved on, and expensive, with an estimated cost of at least US$100,000 needed for a high-quality SR.[5,30]

To address the challenges of SRs, the concept of rapid evidence products has been introduced, including inventories, rapid response briefs, and rapid systematic reviews (RRs).[17] RRs result from a evidence synthesis approach that use streamlined procedures,[13,31] so certain methodological elements are simplified or omitted compared to SRs.[9] Currently, RRs are being conducted to answer urgent questions and/or to support decisions where there is limited time and/or resources i.e., in situations where time- and cost-efficiency are key.[32,33] For example, RRs have been extensively used in addressing issues related to the COVID-19 pandemic.[13,34] Preliminary evidence suggests that the conclusions reached by RRs are typically consistent with those of SRs.[32] In addition, when applied to policy decision-based health technology assessment reports, RRs have been shown to positively impact the healthcare system, resulting in a reduction of expenditures.[35,36]

The use of high-quality evidence summary methods is essential to providing reliable results. For traditional SRs, there are well-defined, pre-specified methods, e.g., for conducting searches, selecting relevant studies, appraising their quality, and synthesizing the available evidence to answer the research question, which ensure quality and reduce bias.[3] However, though methodological rigor and transparency are still essential to have representative and reliable results in RRs,[13] there is a lack of standardised methodologies on how to adapt SR methods to be able to reliably perform a RR.[4,10] Several studies and reviews [4,10,37], have noted this lack of consensus in the methodological approaches being utilised for RRs, highlighting heterogeneous nomenclature and terminology being used to describe the same concepts, and the use of varied methodologies without a clear rationale behind the choices being made.

In 2017, the World Health Organization (WHO) commissioned a guide on how to perform RRs, which explored various approaches. The guide emphasised that methods can be simplified at any stage of the review process and that decisions should consider the resources at hand and be customised to the needs of the decision-makers.[17] The Cochrane Initiative has also produced some methodological guidance for RRs,[14] but the impact and costs of each approach are still unclear. Evidence Synthesis Ireland, using the James Lind Alliance method, identified RR research priorities.[27] Among the top 10 questions generated, three focused on methodological issues but in relatively broad categories.

The current study will build on the findings from Evidence Synthesis Ireland by further exploring more focused questions around RRs methods, i.e., the stages between question generation and report writing. The identification of these unanswered questions is required to design and develop methodological studies that can then inform the conduct of RRs. For example, questions about how many databases should be included, database search limitations, and if peer review is necessary for all steps have not yet been answered. Given the number of areas that still need to be explored, the small amount of current available evidence, the limited available resources to conducted methodological studies, and the lack of general consensus on where to start, the aim of this project is to elicit a consensus from RR experts and interested parties on what are the most important methodological questions to improve time-efficiency of RRs, and, ultimately, create a prioritised research agenda for the field to address.

2. OBJECTIVES

- To identify and compile the main unanswered questions related to the methods used in conducting time-efficiency RRs, specifically from the stage after generating the research question to just before writing the final report.
- To create a priority list of the most crucial questions regarding RRs methods that need to be addressed.

3. METHODS

The study will follow the general eDelphi process[38–40] and the Guidance on Conducting and REporting DElphi Studies (CREDES).[41] There will be an initial generation of potential research areas, followed by multiple rounds of an online

survey for ranking, and then a final consensus meeting. The eDelphi process is particularly useful in surveying areas of uncertainty and obtaining consensus.[38,42] This method has the advantage of enabling each participant to express views impersonally, it is low resource and flexible,[43] and it has been widely used in health research.[44] After ethical approval, the study will start in March 2022, with the first survey round starting in June 2022 and the last round in being finalized in January 2023. The consensus meeting will then occur in the summer of 2023.

Given the focus on efficiency, rather than just quality, the eDelphi will ask participants to answer: "How important would answering this question be to improve the time-efficiency (balance between the time taken and the quality of the final results) of a systematic RR in a particular field?".

3.1 Participants

The sample will consist of two key groups: international experts who have published RRs or undertaken methodological research in RRs and knowledge synthesis; and key end-users. To standardise the level of expertise, all experts will self-identify, answering eligibility questions, on the basis of having: verifiable experience in designing or delivering evidence summary research; participation in at least one RR; having ≥5 years of research experience; and self-rating their knowledge on evidence synthesis as ≥7 on a 0 (no expertise) to 10 (expert) point Likert-like scale. We will also include interested parties (e.g., guideline and policy developers, end-users (public and patients), industry members, and journal editors) who have had previous experiences in participating in any aspect of evidence synthesis.

A recruitment email will be distributed by our global partners through their contacts lists, e.g., the International Behavioural Trials Network (IBTN, https://www.ibtnetwork.org/), the Strategy for Patient-Oriented Research (SPOR) Evidence Alliance (https://sporevidencealliance.ca/), COVID-END (https://www.mcmasterforum.org/networks/covid-end). In addition, as performed by Tricco et. al.,[10] organisations that produce RRs, identified through the International Network of Agencies for Health Technology Assessment's (INAHTA, https://www.inahta.org/) list, will be asked to distribute the study invitation to members of their group. The recruitment email will provide a link to access the information about the study and the consent form. There are no restrictions on the

country of origin of the participants, but all study-related information will be provided in English.

3.2 Providing Consent

The informed consent forms will explain the objective, procedures, and other details that are important to participants. Participants will be asked to read the ethics board-approved information/consent forms and provide agreement by checking a box confirming that they have: reviewed the information/consent form; consent to participate in the survey, and understand that their participation is voluntary and entirely confidential. The contact details of study team members will be listed in the information/consent form in case they have queries. There will be two consent forms, one for the eDelphi rounds and one for the Consensus Meeting. Limesurvey, will be used to obtain consent, as well as to distribute the surveys.

3.3 Initial topic generation

A core group of experts in evidence synthesis, mainly within the biomedical sciences, referred to as the Central Scientific Committee (CSC), and drawn from the leadership of the SPOR Evidence Alliance, IBTN, COVID-END, and notable published scholars, generated a list of methodological questions that they think are relevant to RRs. The items are specific and focused, in order to be able to generate specific research questions rather than broad conceptual areas.

The included topics covered the period after the review question has been generated and before the creation of the final report, e.g., search strategy, studies selection (level one and two screening), data extraction, risk of bias appraisal, and synthesis. The item list was also be drawn from the WHO guide for RRs,[17] the Delphi process on RR methods,[10] and the Priority III study[27] to form the initial 'long-list' of items.

3.4 Online survey

The eDelphi process will involve approximately 50 RRs experts and end-users, who will be asked to complete at least three rounds of online questionnaires, spaced around one month apart. Each survey round will be open for about five weeks, sufficient time for participants to complete it. A system will tag data to individuals and

provide them with their scores from previous rounds, while also reporting the summated data.

3.5.1 Prior to Round 1

The initial survey will include basic demographic information, including eligibility questions (i.e., years of experience, job title, country and province of residence, age group, and sex). Once they agree to participate in the study, participants will be provided with more specific sociodemographic questions and the 'long-list' of survey items from the previous phase.[45] We will only provide the survey to those agreeing to participate to prevent attrition biases.[46]

3.5.2 Round 1

As per our previous eDelphi projects (e.g., Dragomir et. al.[47]), participants will rate the importance of suggested items ("How important would answering this question be to improve the time-efficiency - balance between the time taken and the quality of the final results - of a systematic RR in a particular field?"), focusing on the concept, rather than on the wording. Importance can be rated as: low; medium; or high (Table 1-1). For all items that an individual rates as high or medium importance, they will be asked to rank them in order of priority (1=highest priority, 2=2nd highest, etc.) until all items are ranked. Specific questions with open format responses will allow for modifications to the concept of items. Participants will also be able to add new items that they believe were missing in the initial round.

Responses will be collated and summarised.[44] Any items rated as low by 50% or more of the participants will be excluded, a consensus threshold that is similar to those adopted in other Delphi studies.[42,47] As this is the first round, the threshold will be lower than the following rounds. The CSC will review comments and make necessary changes to items or add new relevant items.

Table 1-1. Classification of the items

Importance Level	**Conceptualisation**
Low importance	Item is helpful to understand how to improve the time-efficiency (balance between the time taken and the quality of the final results) of a rapid systematic review

Medium importance	Item is desirable to understand how to improve the time-efficiency (balance between the time taken and the quality of the final results) of a rapid systematic review
High importance	Item is essential to understand how to improve the time-efficiency (balance between the time taken and the quality of the final results) of a rapid systematic review

3.5.3 Round 2

Participants will be provided with the percentage of respondents ranking each item as high priority, as well as their ratings in the previous round. They will be able to re-rate the perceived importance of each item, as well as the importance of any new items. They will also be asked whether they agree with items excluded from Round 1 or if any essential items are still missing. The items for which ≥ 75% of people disagree with the exclusion of will remain on the main list for the next round. For all items that an individual rates as high importance, they will be asked to rank them in order of priority (1=highest priority, 2=2nd highest, etc.) until all items are ranked. Items rated as low by 75% or more of the participants in Round 2 will be excluded.[47]

As in Round 1, open-format questions will allow suggestions for modifications to the items or the addition of new items. The comments will be reviewed by the CSC and changes or additions will be made as needed.

3.5.4 Round 3

A summary of round 2 will be provided, including the percentage of respondents rating each item as high priority, as well as their own rating. Participants will re-rate and re-rank the remaining items. After Round 3, we will generate a final list of items for discussion at the consensus meeting (those believed to be of high importance by ≥75% of participants). Three rounds should allow us to reach stability and agreement about most items.[46,48] Information about deviant cases will be shared with the consensus group.[45]

3.5 Security of the data

All data that we capture will be stored on secure servers located within Canada, with only information necessary for the research study being collected. All information

obtained will be kept strictly confidential, within the limits of the law. To preserve the confidentiality of the data, a code number known only to those directly involved with this research project will be assigned to each participant, and any personally identifiable information will be stored in a secured computer file.

3.6 Consensus meeting

This step will aim to detail the final items to be included in the priority list.

3.6.1 Participants

Participants will be invited from the eDelphi phase and selected purposively by the Research team to include individuals with a variety of backgrounds (e.g., country, academic level, research context), and that had selected the box showing their interest in participating in the consensus meeting. Approximately 25 people will be invited to an online meeting, a size that balances diversity of opinion with meaningful opportunities for interaction,[49] and maximizes the ability to achieve consensus.

The individuals selected will be contacted by email, with a link that provides access to the Information and Consent Form of the Consensus Meeting. After accepting, participants will access the Zoom platform with an invitation link sent by email.

The meeting will be recorded to aid with the generation of the final report. Zoom's inbuilt anonymous voting system will be used for people to be able to vote on the inclusion or exclusion of items.

3.6.2 Meeting structure

Established nominal group technique methods will guide the consensus meeting.[44,50] The summary of the results of the previous work will be provided in advance to ground conversations on empirical information and to facilitate cohesive discussion during the meeting.[45] The meeting will start with formal presentations. Using a triangulation approach,[51,52] we will then lead a structured discussion of each proposed item.[53] An experienced, independent facilitator will conduct the discussions.[45] Participants will discuss and vote (using anonymous e-ballots), with the potential for a re-vote if needed,[46] with only items supported by at least 75% of participants being adopted.[45]

3.6.3 Anticipated output

The consensus meeting will generate a summary document detailing the questions that will generate the final priority list. This list draft will be circulated to the consensus group participants who will be asked to check if the document accurately represents the discussions and decisions made during the meeting.[53] Then, we will distribute a final version of the document to all eDelphi participants to seek feedback on its wording and content and to assess whether the consensus meeting accurately captured their opinions.[45]

3.7 Data analysis

The research team will analyze the sociodemographic characteristics of the participants using raw numbers, means, and percentages. For each round of data collection, the frequency of participant ratings for each item will be used to determine the percentage of low, medium, or high for each item. For the ranking question, each ranking position will receive a score with the highest position receiving the lowest score. The average score of each item will be calculated by dividing the sum of scores attributed to that item by the number of participants that ranked it. An ascending order will be presented, with the first item, considered the most important one, i.e., the one with the lowest score. Data on average rank and the number of individuals providing data will be included in summary tables.

3.8 Team members

The project will be organized and developed by two main groups: the Central Scientific Committee and the Coordinating Research Team. The full list of members is available on the website (https://mbmc-cmcm.ca/projects/edelphi/). The Central Scientific Committee will be responsible for: the review and editing of the initial list of methodological items; providing feedback on the survey structure and project plan; providing feedback on the results of each survey round (agreeing on the items that participants may suggest, dropping of items, etc.); and helping to share the eDelphi with their networks. The research team, the Montreal Behavioural Medicine Centre, will be responsible for: creating and delivering on the project timelines; creating project documents; setting up and organising the surveys; and managing the public partner involvement in the project.

3.9 Patient and public involvement

Given the emphasis on the methodological aspects of the RR process, with researchers being the primary target end-user of this work, we decided to not include patients in the CSC. The eDelphi does include interested parties, e.g., guideline and policy developers, end-users (public and patients), journal editors, from whom we will draw upon for the final consensus meeting, to ensure that the final document will have direct input from all related groups. In addition, we will leverage interested parties in the creation of a variety of knowledge translation products, e.g., lay summaries, public-facing presentations, infographics, etc.

3.10 Expected outcomes and limitations

The Delphi process is a well-established consensus-building process that will provide us with a good picture of the priority questions that need to be answered regarding the methodological conduct of RRs. The present study will generate a list of specific and focused questions, which can be used to prioritise research questions and to design future methodological studies that will answer those questions. These will ultimately create an evidence base for evidence synthesis researchers when deciding the best approaches to perform a RR.

While this research represents an important initial stage towards refining the conduction of RRs in a more time-efficient way, it will not provide definitive answers on the conduct of RRs. In addition, the response rates and representation of different profiles, perspectives, and experiences of participant's can not be guaranteed. However, the breadth and diversity of the recruitment strategy will likely help mitigate this issue. Finally, the terminology used might be interpreted differently across individuals from different domains and backgrounds. To try and mitigate against this an extensive list of definitions will be used and we will emphasise that items need to be evaluated based on the concept, rather than on the wording.

4. ETHICS AND DISSEMINATION

This study was approved by the Concordia University Human Research Ethics Committee under the Certification Number 30015229.

The dissemination plan includes both traditional academic knowledge products,

e.g., presentations and scientific meetings and publication in peer-reviewed journals, as well as other knowledge dissemination products, e.g., lay summaries, public-facing presentations, and infographics. We will also leverage social media, via the members of the CSC and related organisations, to disseminate results and information as broadly as possible. We will specifically target potential funders, as these will be the bodies that will be targeted for the future methodological studies that will be needed to address the final priority list.

AUTHORS' CONTRIBUTIONS:

Concept: Ariany M Vieira, Geneviève Szczepanik, Jovana Stojanovic, Paula A B Ribeiro, and Simon L Bacon.
Design and Methods: Ariany M Vieira, Chiara de Waure, Andrea Tricco, Sandy Oliver, Jovana Stojanovic, Paula A B Ribeiro, Danielle Pollock, Elie Akl, John Lavis, Tanja Kuchenmuller, Peter Bragge, Laurenz Langer, Simon L Bacon.
Drafting of the manuscript: Ariany M Vieira and Simon L Bacon.
Critical revision of the manuscript for important intellectual content: Ariany M Vieira, Geneviève Szczepanik, Chiara de Waure, Andrea Tricco, Sandy Oliver, Jovana Stojanovic, Paula A B Ribeiro, Danielle Pollock, Elie Akl, John Lavis, Tanja Kuchenmuller, Peter Bragge, Laurenz Langer, Simon L Bacon.
Supervision. Simon L Bacon.
All the authors read and accept the last version of the protocol.

FUNDING STATEMENT: This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.
Ariany Marques Vieira is supported by Fonds de recherche du Québec: Santé (FRQS) doctoral scholarship. Andrea Tricco is supported by a research grant, Tier 2 Canada Research Chair in Knowledge Synthesis. Sandy Oliver is supported by a Research Grant to institution, The International Development Research Centre has funded Prof Sandy Oliver for her PEERSS partnership work (https:// peerss. org/), in the context of which she contributed to this paper. Tanja Kuchenmüller holds a leadership or fiduciary role in the WHO EVIPNet Steering Group. Simon L Bacon is supported by the CIHR-SPOR initiative through the Mentoring Chair program (SMC-151518) and by the FRQS through the Chaire de recherche double en Intelligence Artificielle / Santé Numérique ET sciences de la vie program (309811).

COMPETING INTERESTS STATEMENT:

The authors alone are responsible for the views expressed in this paper and they do not necessarily represent the views, decisions or policies of the institutions with which they are affiliated. The authors have no conflicts of interest to declare.

Essential methodological questions for developing time-efficient rapid reviews: results from an eDelphi study

Ariany M Vieira[1,2], Geneviève Szczepanik[2], Chiara de Waure[3], Andrea C Tricco[4], Sandy Oliver[5], Jovana Stojanovic[1,2], Paula A B Ribeiro[2], Danielle Pollock[6], Elie A Akl[7,8], John Lavis[9,10], Peter Bragge[11], Laurenz Langer[12], Simon L. Bacon[1,2]

[1] Department of Health, Kinesiology and Applied Physiology, Concordia University, Montreal, Quebec, Canada
[2] Montreal Behavioural Medicine Centre, CIUSSS du Nord-de-l'ile-de-Montreal, Montreal, Quebec, Canada
[3] Department of Medicine and Surgery, University of Perugia, Perugia, Italy
[4] Knowledge Translation Program, Li Ka Shing Knowledge Institute of St Michael's Hospital, Unity Health Toronto, Toronto, Ontario, Canada
[5] University College London Institute of Education, London, UK
[6] Health Evidence Synthesis, Recommendations and Impact, School of Public Health, The University of Adelaide, Adelaide, South Australia, Australia
[7] Department of Internal Medicine, American University of Beirut, Beirut, Lebanon
[8] Department of Health Research Methods, Evidence, and Impact (HEI), McMaster University, Hamilton, Ontario, Canada
[9] McMaster Health Forum and Department of Health Research Methods, Evidence and Impact, McMaster University, Hamilton, Ontario, Canada
[10] Affiliate Scientist, Institute for Better Health, Trillium Health Partners, Mississauga, Ontario, Canada
[11] Monash Sustainable Development Institute Evidence Review Service, BehaviourWorks Australia, Monash University, Melbourne, Victoria, Australia
[12] South Africa Centre for Evidence, Johannesburg, South Africa

Identifying Essential Questions in Rapid Reviews Methods: Results from an eDelphi Study

Abstract

Introduction: Rapid Reviews (RR) aim to efficiently synthesize evidence for decision making. However, there is little consensus on the most time-efficient methods that should be employed to achieve this, meaning that we need specific methodological studies to address this.

Objective: To generate a priority list of key unanswered methodological questions for conducting RRs (i.e., between generating the research question and report writing).

Methods: A three-survey round eDelphi study and consensus meeting was conducted, involving experts and evidence synthesis knowledge users (n=52, 70, and 60 per round). From an expert-generated initial list, participants rated (low-high importance) and ranked each item's importance to improve the time-efficiency of a RR. Items of low importance were removed, and those rated as highly important by ≥75% of participants progressed to the next round.

Results: In round 1, from 29 initial items, 15 were excluded, 14 continued, and 12 items were added. In round 2, all 26 items achieved consensus. In round 3, 6 items were rated highly important, 2 were close to the inclusion criteria, and 18 were excluded. The consensus meeting generated a final list of 7 items.

Conclusion: Seven research areas covering the search strategy, study selection, quality/bias assessment, and data extraction were included. This list should drive the future methodological research agenda to improve the quality of RRs.

Keywords: Rapid review; Systematic review; Delphi; Priority setting; Consensus; Evidence synthesis.

Strengths and limitations of this study:

- A rigorous and well-established Delphi process was used for consensus building.
- Although this study included researchers from multiple countries with high experience in evidence synthesis, recruitment was not able to include representatives from public and community partners.
- This study addresses an important gap around the multiple questions available on RR methods, as it can serve as a ‘road map’ for future RR methodological studies to understand better the impact of RR methodological choices on time efficiency.

1. BACKGROUND

According to Hamel et al.,[9] a rapid systematic review, or rapid review (RR), is "a form of knowledge synthesis that accelerates the process of conducting a traditional systematic review through streamlining or omitting a variety of methods to produce evidence for end-users in a resource-efficient manner.'' This method/approach to evidence synthesis is especially relevant when timely evidence for decision making is needed to answer questions on urgent and emergent health issues,[14] considering the time and resources needed to develop a full systematic review.[17]

There are relatively well-defined methods for full systematic reviews. However, in the case of RRs, there are no clear and standardised methodologies to guide their execution and ensure representative and reliable results.[10] This has led to inconsistent nomenclature and terminology to describe similar concepts within RRs, and the adoption of various approaches and methodologies in conducting RRs without clearly acknowledging the rationale behind these choices.[4,10,37,54] These issues make evaluating these evidence products difficult.[24,55] Although there is a general agreement on the different steps involved in a RR, the degree to which these are executed varies, including practices that limit search strategy, narrowing the inclusion criteria, relying on a single reviewer for study selection, and omitting the assessment of risk of bias or study quality [6,10,22,56], with no clear justification or understanding of the potential negative impacts on these decisions on the quality of the final product.

Although RRs are recommended to streamline its methods according to the needs of the decision-maker and the resources available to conduct the RR,[17] the impact of these timesaving methods are poorly understood. In 2024, Cochrane updated a list of recommendations that provide valuable insights on the streamlining strategies and tools that can facilitate the RR process.[14,57] However, little empirical evidence is available to understand if, or to what extent, biases can be introduced with these strategies.[6] Exploring the consequences of omitting or abbreviating review processes in RRs is a crucial area for future methodological research.[13]

As highlighted by the James Lind Alliance, the identification of methodological questions is needed to tailor the design and development of methodological studies about the conduct of RRs.[27] For instance, what are the impacts of the different search strategies, screening approaches, or bias assessments on the efficiency of RRs. As specific questions are needed to guide methodological studies, the objective

of this study was to identify and collate the key unanswered questions regarding the methods for conducting time-efficient RRs (i.e., after the generation of the research question to just before the report writing), and to rank, in order of priority, the most important questions to be answered.

2. METHODS

The study protocol has been published elsewhere.[58] This electronic Delphi (e-Delphi) study was conducted and is reported following the available literature on Delphi processes[38–40] and the Guidance on Conducting and REporting DElphi Studies (CREDES).[41]

After the initial generation of research question items, three rounds of an online survey were conducted followed by a final consensus meeting. The eDelphi asked participants to answer: "how important would answering this question be to improve the time-efficiency (balance between the time taken and the quality of the final results) of a systematic RR in a particular field?". Inspired by the methodology used in previous eDelphi projects (e.g., Dragomir et. al.,[47]) participants had to rate the importance of suggested items according to Table 2-1. For all items that an individual rated as high importance, they were asked to rank them in order of priority (1=highest priority, 2=2nd highest, etc.) until all items were ranked.

Table 2-1. Classification of the items

Importance Level	Conceptualisation
Low importance	Item is helpful to understand how to improve the time-efficiency* of a systematic RR
Medium importance	Item is desirable to understand how to improve the time-efficiency* of a systematic RR
High importance	Item is essential to understand how to improve the time-efficiency* of a systematic RR

*Balance between the time taken and the quality of the final results.

2.1 Recruitment and Participants

An invitation to collaborate was distributed by our international partners through their contacts lists, e.g., the International Behavioural Trials Network (IBTN,

https://www.ibtnetwork.org/), the Strategy for Patient-Oriented Research (SPOR) Evidence Alliance (https://sporevidencealliance.ca/), and COVID-END (https://www.mcmasterforum.org/networks/covid-end). In addition, 84 organisations that produce RRs, identified through the International Network of Agencies for Health Technology Assessment's (INAHTA, https://www.inahta.org/) were contacted by email asking to distribute the study invitation to members of their group. The recruitment email provided a link to access the information about the study and the Information and Consent form.

The target population consisted of international experts who had published RRs or undertaken methodological research in RRs and knowledge synthesis as well as key end-users (i.e., key interested and affected parties). To standardise the level of expertise, all potential participants had to answer eligibility questions before having access to the study. Experts had to self-identify as having verifiable experience in designing or conducting evidence summary research; having been engaged in at least one RR; having had a minimum of 5 years of research experience; and self-assessing their knowledge in evidence synthesis as ≥7 on a 0 (no expertise) to 10 (expert) point Likert-like scale. Additionally, we welcomed individuals such as guideline and policy developers, members of the public and/or patients, industry members, and journal editors who had prior experience participating in any facet of evidence synthesis.

After confirming eligibility participants were asked to review and sign the Information & Consent Form. All electronic aspects of the study utilised LimeSurvey (LimeSurvey Project Team 2012, Germany).

2.2 Prior to Round 1

A central team of nine experts specialising in evidence synthesis, primarily in the health field, known as the Central Scientific Committee (CSC, more information available in section I of appendix II), compiled a set of methodological questions pertinent to RRs. The topics covered the period between formulating the review question and the final report creation, including aspects such as search strategy, study selection (both level one and two screening), data extraction, risk of bias assessment, and synthesis. The list of items was based on available sources, such as the WHO guide for RRs,[17] a former Delphi on RR methods,[10] and the Priority III study [27].

The first list of items was structured in questions asking “what is the optimal…” number, approach, or individual criteria, for a specific review step. For instance, “What is the optimal number of databases used for the search strategy to improve time-efficiency of a RR in a particular field?”.

2.3 Survey Rounds

Prior to starting the survey, participants provided basic demographic information (such as years of experience, job title, country and province of residence, age group, and sex).

Participants were presented with a list of items separated by categories: high-level/conceptual; search strategy; study selection; data extraction; quality/bias assessment; and synthesis. After every category there was an open question to encourage suggestions on alterations to items and the opportunity to suggest new items.

Responses were collected and summarised for analyses.[44] Items rated low by more than 50% of the participants were excluded from the final list, a threshold akin to those employed in previous eDelphi studies.[42,47] The CSC reviewed participants’ comments, leading to adjustments in existing items or the inclusion of new items.

For Rounds 2 and 3, participants received feedback on the percentage of respondents rating each item as high priority, as well as their own ratings in the previous round, and were given the opportunity to re-rate the importance of each item, including any new additions in Round 2. Items in Round 2 were retained when not rated as low by 75% or more of the participants. In addition, participants were also presented with the list of excluded items from Round 1 and could express disagreement with their exclusions; items where at least 75% of participants disagreed that the item should be excluded returned to the list. Items in Round 3 were retained when considered to be of high importance by ≥75% of participants.

2.4 Consensus meeting

The goal of the consensus meeting was to finalize the list of the most important methodological questions to consider when conducting a RR. Participants that answered the eDelphi survey rounds and expressed their willingness to attend the meeting by selecting a checkbox in the Information and Consent Form, were later

contacted through email. Those who confirmed their intent to participate in the consensus meeting signed a second Information and Consent Form.

Two meetings were held to accommodate different time zones. Both were conducted online using the Zoom platform. Before the meetings, participants received a package of documents including the pre-read document, the technical document, and supplementary material with all the results from the eDelphi phase of the study (available at: https://osf.io/zpxm4/).

Established Nominal Group Technique process guided the consensus meeting.[44,50] An experienced moderator facilitated the discussion and the voting process. Three different groups of items were discussed: (1) items included in the final list following the survey rounds; (2) items that did not reach consensus during the eDelphi surveys but were close to the cut-off point; and (3) items that were excluded from the final list during the surveys. For each block, participants were presented the items and had an opportunity to express any opinions about their inclusion or exclusion. Items were discussed and when a vote was necessary, the Zoom poll function was used to record their answers anonymously. Consensus was reached when at least 75% of participants agreed on an inclusion or exclusion.

Since two meetings were offered, the final list of items was prepared after the second meeting. In the second meeting, we followed the same procedures as the first and left space for participants to express themselves or their desire to discuss an item further. After this discussion the moderator added any discussion points that had been raised in the first meeting. If participants in the second meeting decided to vote on an item, provided additional elements to a discussion, or suggested changes that were in disagreement with decisions made in the first meeting, all participants were contacted by email after the meeting, and a formal voting process took place outside the meeting.

The consensus meeting produced a summary document outlining the final priority list of questions. This draft was shared with participants of the consensus meeting to verify its alignment with the discussions and decisions. Subsequently, a final document was distributed to all eDelphi participants to gather feedback on wording and content, ensuring the consensus meeting accurately reflected their opinions (Available on OSF: https://osf.io/utx8e).

2.5 Data analysis

Only complete responses were included in the analysis, as partial answers rarely provided relevant information beyond the initial survey questions.

As explained in the protocol,[58] raw numbers, means, and percentages were used to analyse the sociodemographic characteristics of our participants. In each data collection round, the distribution of participant ratings for each item was analysed to ascertain the percentage categorised as low, medium, or high for each item. For the ranking question, an average score was calculated. Each item in the first position (highest priority) received one point, the item in the second position (2nd highest priority) received two points, and so on. The sum of points was divided by the number of participants that rated that item as high (that had the item included in their ranking question). So, the lower the average score, the item was considered to be more important.

In addition to what it was described in the protocol,[58] the CSC also requested extra analyses. Data collected was analysed considering the overall sample (all participants) and individuals who completed all the three survey rounds. Data was also analysed stratified by participants' profile, i.e., researcher, healthcare practitioner, policymaker, and community member. For the purpose of this publication, only the results for the third survey round are presented by participants' profile.

3. RESULTS

The first eDelphi round was launched in June 2022 and the last round closed in January 2023. The initial list included 29 items, categorized into 6 conceptual/high-level sections, with 6 pertaining to search strategy, 6 addressing study selection, 3 focusing on data extraction, 5 concerning quality/bias assessment, and 3 centered on synthesis (Appendix II – section II).

3.1 Participants

In total, 78 participants answered one of the three survey rounds. From those, 41 participants answered all three rounds. In Table 2-2, you can find the characteristics of the participants. The majority self-identified as researchers (76.9%), were between 36 and 45 years old (39.7%), and were female (60.2%). Regarding years

of experience with evidence synthesis, the majority of the sample reported more than 9 years of experience, with 34.6% of participants reporting 15 years or more. The sample was made of participants from diverse corners of the globe and representing every continent.

Table 2-2. Characteristics of the eDelphi participants

		OVERALL PARTICIPANTS	**CONSISTENT PARTICIPANTS**
		n (%)	n (%)
Profile			
TOTAL		*78 (100%)*	*41 (100%)*
Healthcare Practitioner		11 (14.1%)	5 (12.1%)
Researcher		60 (76.9%)	33 (80.4%)
Policymaker		7 (8.9%)	3 (7.3%)
Patient / Community member / Caregiver		0 (0.0%)	0 (0.0%)
Sociodemographic Information			
Age Group	18 – 25 years	1 (1.2%)	0 (0.0%)
	26 – 35 years	16 (20.5%)	8 (19.5%)
	36 – 45 years	31 (39.7%)	14 (34.1%)
	46 – 55 years	17 (21.7%)	10 (24.3%)
	56 – 65 years	12 (15.3%)	8 (19.5%)
	66 years or more	1 (1.2%)	1 (2.4%)
Sex	Female	47 (60.2%)	23 (56%)
	Male	28 (35.8%)	16 (39%)
	Prefer not to answer	3 (3.8%)	3 (7.3%)
Country of work	Argentina	2 (2.56%)	0 (0.0%)
	Australia	6 (7.69%)	5 (12.19%)
	Austria	1 (1.28%)	0 (0.0%)
	Belgium	1 (1.28%)	1 (2.43%)
	Brazil	3 (3.85%)	0 (0.0%)
	Canada	26 (33.33%)	17 (41.46%)
	Canada/Colombia	1 (1.28%)	0 (0.0%)
	Colombia	3 (3.85%)	1 (2.43%)
	Ethiopia	1 (1.28%)	0 (0.0%)
	Greece	1 (1.28%)	1 (2.43%)
	India	2 (2.56%)	0 (0.0%)
	India/New Zealand	1 (1.28%)	0 (0.0%)
	Ireland	1 (1.28%)	1 (2.43%)
	Italy	3 (3.85%)	2 (4.87%)
	Kyrgyzstan	1 (1.28%)	0 (0.0%)
	Lebanon	1 (1.28%)	1 (2.43%)
	Mexico	1 (1.28%)	0 (0.0%)
	Romania	2 (2.56%)	1 (2.43%)
	Slovenia	1 (1.28)	0 (0.0%)
	South Africa	2 (2.56)	1 (2.43%)
	Spain	3 (3.85%)	3 (7.31%)
	Switzerland	2 (2.56%)	0 (0.0%)
	Turkey	1 (1.28%)	0 (0.0%)
	United Kingdom	6 (7.69%)	5 (12.19%)
	United States	6 (7.69%)	2 (4.87%)

Experience			
Years of experience with evidence syntheses	≤ 4 years 5-6 years 7-8 years 9-10 years 11-12 years 13-14 years ≥ 15 years	3 (3.8%) 16 (20.5%) 9 (11.5%) 8 (10.2%) 8 (10.2%) 7 (8.9%) 27 (34.6%)	1 (2.4%) 6 (14.6%) 6 (14.6%) 5 (12.1%) 4 (9.7%) 4 (9.7%) 15 (36.5%)
Aspects of evidence synthesis that previously participated	Conceptualization/Research question development	78 (100%)	41 (100%)
	Undertaking literature searches	69 (88.4%)	38 (92.6%)
	Study screening and selection	71 (91%)	39 (95.1%)
	Data extraction	71 (91%)	38 (92.6%)
	Quality appraisal	66 (84.6%)	34 (82.9%)
	Data synthesis	70 (89.7%)	35 (85.3%)
	Interpretation of results	72 (92.3%)	37 (90.2%)
	Knowledge translation	63 (80.7%)	35 (85.3%)
	Other	9 (11.5%)	8 (19.5%)
Knowledge in evidence syntheses (0 = no expertise to 10 = very strong expertise)	*(Mean, SD)*	8.3 (±1.43)	8.5 (±0.9)
Fields/areas of research predominantly perform evidence syntheses	Clinical	39 (50%)	25 (60.9%)
	Public Health	54 (69.2%)	32 (78%)
	Health Systems	45 (57.6%)	28 (68.2%)
	Other	15 (19.2%)	10 (24.3%)

SD: standard deviation.

3.2 Overall Results

3.2.1 Round 1

The Round 1 survey was active for 6 weeks and 129 participants provided responses, 52 participants completed the whole survey, while 77 provided partial responses and were excluded. The characteristics of the eDelphi participants separated by partial and complete answers are available in the Appendix II – section III. From the initial 29 items, 14 items reached the level to move forward to Round 2 and 15 were excluded.

The CSC reviewed all the comments by the participants in Round 1 and voted around the structure of the items. The main concern raised by participants was the use of the concept of an absolute “optimal” method or approach in the

anchoring question. After discussion, the anchoring of questions was re-written to identify the relative importance of an item for the time-efficiency of a RR. The items were re-structured to the following format “What is the optimal method/approach … to improve the time-efficiency of a RR in a particular field?”. The CSC was also in favour of having a general question for team composition and expertise (e.g. What is the optimal method for balancing the skills and experience of the team to improve the time-efficiency of a RR in a particular field?), instead of by specific tasks (e.g. the optimal number of people to perform screening and optimal number of people needed to perform data extraction).

Twelve newly suggested questions were added by the CSC with some of the questions suggested by participants being considered out of scope and not being incorporated into the survey, bringing the total to: 1 conceptual item; 4 search strategy items; 2 studies selection items; 2 data extraction items; 2 quality assessment items; and 1 synthesis item.

3.2.2 Round 2

During Round 2, which lasted three weeks, new participants were welcomed and a combined total of 70 complete responses were collected, with 27 from new participants and 43 from returning participants. At the beginning of the round, there were a total of 26 items (14 carried over from Round 1 and 12 new items). After analysis, all 26 items achieved the importance level to move forward to Round 3. From the items that were previously excluded, none received agreement for future inclusion.

3.2.3 Round 3

Sixty participants answered the Round 3 survey which was active for four weeks. From the 26 initial items, 6 items were rated as ‘high importance’ by at least 75% of participants and were included in the final list of items (Table 2-3) and two items were rated as ‘high importance’ by 73% and 65% of participants (Table 2-4). The ranking question was analyzed, and items are presented in order of relative importance in Tables 2-3 and 2-4.

Table 2-3. Items that achieved consensus in Round 3.

Item	Classification	Frequency	Average rating Score
6. What is the optimal method for developing search terms for the key elements in the rapid review...? (Search strategy)	High importance	96%	4.37
	Middle importance	2%	
	Low importance	2%	
12. What is the optimal method for determining the inclusion and exclusion criteria for studies...? (Studies Selection)	High importance	83%	4.46
	Middle importance	13%	
	Low importance	3%	
7. What is the optimal method for defining restrictions and search limits (e.g., years of inclusion, language, phase of study, study design)..? (Search strategy)	High importance	88%	4.96
	Middle importance	10%	
	Low importance	2%	
21. What is the optimal method for determining how to assess the quality of included studies and/or risk of bias...? (Quality/bias assessment)	High importance	86%	5.28
	Middle importance	12%	
	Low importance	1%	
14. What is the optimal method for determining how to perform screening (e.g., independent screening, 1+1 approach/partial peer review, etc.)..? (Studies Selection)	High importance	77%	6.10
	Middle importance	21%	
	Low importance	2%	
20. What is the optimal method for determining the dimensions of quality (e.g., trustworthiness, relevance) that should be considered when appraising studies...? (Quality/bias assessment)	High importance	88%	6.67
	Middle importance	8%	
	Low importance	4%	

All items are finalized by the following sentence: “to improve the time-efficiency of a RR in a particular field?”.

Table 2-4. Round 3 ratings' results of items close to cut-off point.

Item	Classification	Frequency	Average rating Score
4. What is the optimal method for defining the core set of databases that should be searched...?) (Search Strategy)	High importance	73%	6.02
	Middle importance	25%	
	Low importance	2%	
17. What is the optimal method for defining the data extraction approach (e.g. peer review with independent extraction and comparison of discrepancies, one reviewer extracting and the other double checking)...? (Data Extraction)	High importance	65%	7.42
	Middle importance	34%	
	Low importance	1%	

3.2.4 Consensus Meeting

Twenty-four participants were invited to the consensus meeting, 18 accepted the invitation and 12 were present at one of the two meetings. In addition, a moderator, the project lead, a notetaker, and a technical support professional also participated in both meetings; none of these individuals voted on the items.

The meetings were separated into three blocks. The objective of the first block was to confirm inclusion of the 6 items from the 3 rounds of eDelphi surveys. Across the 2 meetings, five items reached consensus for inclusion, while one item (item 20) was discussed and excluded (What is the optimal method for determining the dimensions of quality (e.g., trustworthiness, relevance) that should be considered when appraising studies...? - Quality/bias assessment). Participants expressed concerns that this item could allow misinterpretation and selective modification of the tools used to perform the evaluation of quality and risk of bias (i.e., researchers selectively choosing items from the tools to their benefit).

For the second block, participants discussed and voted on the 2 items that did not reach consensus during the eDelphi surveys but were close to the cut-off percentage. Both items were voted in to the final list of items (items 4 and 17). Item 4 generated substantial discussion and was compared against item 5, which was in the list of excluded items and covered the aspect around gray literature (What is the optimal method for determining the non-peer-reviewed publication databases (i.e., pre-prints servers and high-volume producers) that should be included...?). It was

decided to broaden the original wording of item 4, replacing “core set of databases” for “evidence/information sources”.

The third block covered the remaining 18 items that were excluded following the eDelphi surveys. Participants decided to open discussion for 3 items (item 5 - What is the optimal method for determining the non-peer-reviewed publication databases (i.e., pre-prints servers and high-volume producers) that should be included...? – see above; item 19 - What is the optimal method to decide on the core sets of variables everyone should aim to extract (e.g., basic sample info from each sample)...?; and item 22 - What is the optimal method for determining the usage of quality assessment results (e.g., to further exclude studies, to allow a more tailored synthesis)...?). A formal voting process was opened for item 22 and 50% of participants voted for inclusion and 50% for exclusion, thus not reaching consensus for inclusion. Table 2-5 presents the final list of methodological questions needed to be answered to improve time-efficiency of RRs resulting from this meeting.

Table 2-5. Final list of items presented in order of highest average rating.

Item	Average rating
6. What is the optimal method for developing search terms and how to use them for the key elements in the RR...? * (Search strategy)	4.37
12. What is the optimal method for determining the inclusion and exclusion criteria for studies...? (Studies Selection)	4.46
7. What is the optimal method for defining restrictions and search limits (e.g., years of inclusion, language, phase of study, study design) ...? (Search strategy)	4.96
21. What is the optimal method for determining how to assess the methodological quality (e.g., RoB) …? ** (Quality/bias assessment)	5.28
4.1 (New Item): What is the optimal method for defining the evidence/information sources that should be searched...? (Search Strategy)	6.02[#]
14. What is the optimal method for determining how to perform screening (e.g., independent screening, 1+1 approach/partial peer review, etc.)...? (Studies Selection)	6.10
17. What is the optimal method for defining the data extraction approach (e.g., peer review with independent extraction and comparison of discrepancies, one reviewer extracting and the other double checking) ...? (Data Extraction)	7.42

All items are completed with the following words: "… to improve the time-efficiency of a RR in a particular field?".

* This item has been edited from its original form, which was "What is the optimal method for developing search terms for the key elements in the rapid review...?".

** This item has been edited from its original form, which was "What is the optimal method for determining how to assess the quality of included studies and/or risk of bias...?".

[#] The average score considered was the one from the original item 4 (original item: What is the optimal method for defining the core set of databases that should be searched...?). However, it should be noted that the new item also incorporates elements of the original item 5 (original item: What is the optimal method for determining the non-peer-reviewed publication databases (i.e., pre-prints servers and high-volume producers) that should be included...?)

3.3 Results by participants' profile

As presented in Table 2-2, 41 participants answered all three survey rounds. In the case of the 6 items having made it on the inclusion list for the consensus meeting, these were also rated as highly important by all consistent participants and had achieved consensus with ≥75% agreement. Results are presented in more details in Table S11 (Appendix II – section V).

From the 60 participants that answered Round 3 of the survey, 47 self-identified as researchers (79%), 7 as healthcare professionals (11%), and 6 as policymakers (10%). No participant self-identified as a patient/community member/caregiver. Detailed results are presented in Tables S12-S14 (Appendix II – section V). For most of the included items (4 out of 6) in the list generated by Round 3, all participant's profiles agreed on their classification as highly important (≥75%: Table 2-5). For item 14, 1.4% of healthcare professionals and 80.9% of researchers rated it as highly important, while only 50% of policymakers did. For item 21, 85.7% of healthcare professionals and 89.4% of researchers rated it as highly important, while only 66.7% of policymakers rated it in the same way.

4. DISCUSSION

This study generated a list of the most meaningful methodological questions to improve time-efficiency of RRs. Specialists in the field generated 29 initial items that were rated and ranked by experts in RRs, and after three rounds of surveys and a consensus meeting, 7 items reached a consensus as the most important questions. These items covered aspects of search strategy (3 items), study selection (2 items), quality/bias assessment, and data extraction (1 item each).

Aspects around search strategy were consistently identified as priority questions within this study, notably the evidence/information sources, search terms, and search limits. The selection of databases, can impact the time invested to perform the review as well as the reviews' results.[59,60] Most health-based RRs already use this step to save time by performing searches on the PubMed/MEDLINE, Cochrane Library, and Embase databases, using limits such as date, language and study design[56]. Currently, based on available literature, the Cochrane RRs Methods Group recommends limiting main database searching to CENTRAL, MEDLINE (e.g., via PubMed), and Embase.[14,57] However, comprehensiveness and efficiency can vary depending on the review question, so the optimal search strategy may depend on

the field, and for some research topics a smaller number of database searches combined with supplementary search methods might prove more valuable or noninferior compared to searching numerous bibliographic databases.[61,62] In addition, it is recommended to peer-review the search strategy,[14,56] and while the use of gray literature searching is common, it needs to be considered carefully leveraging the available resources and the review topic.[14,56] Gray literature can provide important contributions to reviews findings, but might be challenging because of time constraints.[63] For full systematic reviews, the mean time taken to conduct gray literature searches is approximately 7 hours, which represents around 27% of the total time taken to finalise the search step.[64] Empirical research in the future is essential to understand how to decide on the evidence/information sources that should be searched in different fields, especially for time-efficiency.

Another methodological research topic identified as a high priority in this study, was the study selection phase. For instance, how to determine the inclusion and exclusion criteria for study selection and how to perform screening to improve the time-efficiency of a RR in a particular field. The inclusion/exclusion question is not widely discussed but has also the potential to impact time-efficiency. As RRs are recommended to answer key questions from end-users, their involvement is vital to make sure the review question is refined and inclusion criteria are appropriate.[14,23,65] RRs often also streamline the process of screening studies, with about 40-50% using a single reviewer at each stage.[12,56] A study that compared single versus peer-review screening of titles and abstracts found that single-reviewer abstract screening missed 13.4% of relevant studies compared to 2.5% lost by peer-review.[26] In contrast, Cochrane recommends one reviewer be used to include and two reviewers to exclude studies at title and abstract screening.[14] However, it is still unclear what the impact of these approaches are on the magnitudes of effects seen in the final analyses or conclusions of the review. According to our results, there is a consensus that exploring theses questions is of high importance.

When data was analysed by participants' profiles, there was a high agreement on the importance of items, with only differences found in two items, one on studies selection and the other on quality/bias assessment. A lack of comparable literature does not allow for a comparison with other results. However, it would seem that whilst policymakers may be willing to sacrifice some certainty for the sake of speed

and efficiency, they expect that RR will mirror the validity of systematic reviews,[56,62] which researchers might not seen in the same way. It is crucial for leading evidence synthesis organisations to advocate and inform funders about the significance of maintaining credible approaches to evidence synthesis.[18] In our study, only 50% of policymakers, compared to 86% of researchers, considered highly important the question around the method for determining how to perform screening. A similar gap was seen in the question on the optimal method for determining how to assess the methodological quality. These differences highlight the importance of further exploring how policymakers and researchers value and consider each review step.

As per the feedback received during rounds 1 and 2, the discussion around how to frame and word the items led to certain clarifications. In the second round of the survey, all items were rewritten, and additional details were provided on the survey welcome page. We wanted to highlight that there is not a single “optimal” approach to all RRs, nor a “one size fit all” approach, by switching to asking “what is the optimal method for defining the core set of databases…” instead of “what is the optimal core set of databases…”, for example. We believe that there are procedures that can help select the optimal methods, or that can help understand the impact of those choices. Methodological challenges will vary depending on the type of RR,[14,66] end-users needs, time and resources available.[17] Researchers running RRs should include the rationale for tailoring their methods,[18] and evidence-informed decisions should be followed.

The final list of items that was generated, as mentioned previously, includes three items around search strategy, two items on study selection, one item for quality/bias assessment, and another one around data extraction. Comparing with the updated Cochrane recommendations,[57] we found some consistencies and inconsistencies. Notably, for search strategy, there was no recommendation around the development and use of the search terms in the Cochrane recommendations, but it was highlighted as a key concern in our study. The only Cochrane recommendation was to involve an information specialist in developing the search strategy, which was based on expert opinion. Regarding the other search strategy-related items, ranked third and fifth, Cochrane provided evidence-informed guidelines on search limits, including limitations on databases and language, and expert opinion on the need for grey literature and supplemental searches. In terms

of study selection, Cochrane offered recommendations aligned with our identified items, offering expert-based recommendations on inclusion/exclusion criteria and screening methods. Finally, for assessing methodological quality and data extraction, Cochrane suggested using validated tools, prioritizing key outcomes, and having one person perform the task with a second to verify, all based on expert opinion. However, it is important to note that our list of items not only has the potential to guide future recommendations but also serves to help understand the methods and how to explore each one of those questions to improve the time-efficiency of an RR in a particular field.

Our study had some limitations that need to be considered when interpreting the results. Our recruitment strategy, though international in nature, was not sensitive enough to include participants who self-identified as a patient/community member/ caregiver. However, we were able to include a good representation of the researcher group, which is considered to be the main knowledge-user for this study. The group that participated in the consensus meeting was generally representative of the survey participants and brought diverse perspectives and expertise. The varied background of the group members allowed for discussions that considered a variety of point of views, enhancing the validity and applicability of the final list of items, making the findings more relevant. However, to be inclusive, we held two consensus meetings, which limited the interaction between some of the participants and may have influenced the nature of some discussions of items. However, all the discussions held in the first day were brought up by the moderator at the second meeting and any discussion at the second meeting were circulated to members of the first meeting, which partially mitigates this. Due to the nature of the recruitment process, the CSC and participants generally represented the biomedical field, with experience in different areas of research such as clinical, health systems, and public health. Though the conceptualisation of the responses should translate to other fields, this can’t be confirmed in the current study. Despite all the limitations, the eDelphi process is a well-established consensus-building process and provided us with the list of priority questions that need to be answered regarding the methodological conduct of RRs.

5. FUTURE DIRECTIONS

Our results can be used as an initial step towards investigating the different timesaving methods followed by RRs. Our final list of items can be used as a research agenda that has the potential to impact how we develop RRs, contributing to the empirical evidence that raise the scientific rigour of this evidence synthesis approach. Methodologists can design studies using this list of items, knowing that there is the need to explore these issues according to RRs experts.

As per our dissemination plan,[58] besides academic knowledge products, additional knowledge dissemination products will be produced, such as lay summaries, public-facing presentations, and infographics. We will also leverage social media, via the members of the CSC and related organisations, to disseminate results and information as broadly as possible. We will specifically target potential funders, as these will be the bodies that will be targeted for the future methodological studies that will be needed to address the final priority list.

6. CONCLUSION

Following a well-recognised and highly structured method, seven methodological questions were identified by RR experts as the most important questions to be answered to improve the time-efficiency of a RR in a particular field, outlined in Table 2-5. This list should be used as a prioritised research agenda for exploring the methodological aspects of RRs. Methodologists can use this list to explore issues aiming to improve the time-efficiency of RRs, ideally using good methodological research designs. Ultimately, we hope this will provide reviewers with better evidence to inform decisions, allowing them to better understand the impact of their methodological choices when timesaving methods are needed.

CHAPTER 3

The impact of database choices on systematic review results: a case study using Cochrane cardiac rehabilitation reviews

Ariany M Vieira[1,2], Keven Joyal-Desmarais[2,3], Laurence Paquet[2], Rod S Taylor[4], Simon Bacon[1,2].

[1] Department of Health, Kinesiology and Applied Physiology, Concordia University, Montreal, Quebec, Canada

[2] META Grroup, Montreal Behavioural Medicine Centre, CIUSSS du Nord-de-l'ile-de-Montreal, Montreal, Quebec, Canada

[3] School of Psychology, University of Leeds, Leeds, England, United Kingdom

[4] MRC/CSO Social & Public Health Sciences Unit, University of Glasgow, Glasgow, Scotland, United Kingdom

Abstract

Background: The necessity of delivering evidence synthesis in a time-effective and resource-limited way led to an increased interest in Rapid Reviews (RRs). Several approaches to streamlining the review methods exist including restricted search strategies. However, it is unclear what the impact of searching in fewer databases is.

Objectives: To assess the impact of varying databases in the search strategy for cardiac rehabilitation (CR) reviews on the list of included studies and primary outcome findings.

Methods: A methodological study used Cochrane CR reviews to evaluate the indexing rates across different database combinations. Using RStudio, we performed a multiverse analysis to understand how these combinations and missed trials influenced the overall effect sizes of the main outcomes.

Results: Reviews included 10 to 145 studies, and 6 to 85 trials. CENTRAL indexed the greatest number of studies and was the database that varied least between reviews, indexing from 84.2 to 100.0% of the trials. The best combination was Embase plus CENTRAL for study (97.2%, ±2.1) and trials-level (96.8%, ±3.5). When considering the mortality meta-analyses mortality, when combining the major databases in pairs (MEDLINE, Embase, or CENTRAL), only 37.8% of estimated effects were identical to when all databases were combined. When considering trios of databases, treatment effects were identical in 65.9% of the analyses.

Conclusion: In the case of CR reviews, searching on CENTRAL plus two major databases such as MEDLINE and Embase is recommended in case a streamlined approach to evidence identification is necessary.

Keywords: Rapid review; Systematic review; Search Strategy; Cardiac rehabilitation; Databases selection.

Strengths and Limitations

- There is limited empirical evidence available to support the streamlined methods applied for the conduct of rapid reviews. This study helps to understand the impact of bibliographic database choices on the results and conclusions of systematic reviews in cardiac rehabilitation.
- For researchers in the field of cardiac rehabilitation, this study offers valuable insights into how to employ a resource-efficient approach when crafting search strategies for evidence synthesis.
- This is a small methodological study using Cochrane Cardiac Rehabilitation Reviews as a case study. Findings may not be translated to other fields.
- Other researchers can employ our methods to explore whether similar results can be observed in different research fields.

1. INTRODUCTION

Rapid Reviews (RRs) have been developed in order to produce evidence for knowledge users in a more resource-efficient way by accelerating the process of conducting a traditional systematic review through streamlining or omitting a variety of methods.[9] The conduct of RRs saves an average of 75% of the time compared to a standard systematic review.[37] High-quality evidence summary methods are essential for providing reliable results, which are then used to make policy and develop clinical practice guidelines.[28] However, it is not clear what are the best methods to conduct a RR, balancing speed with accuracy, e.g., the number of databases that are needed, the process of study selection, methods for data extraction, etc.[10]

There is a large number of RRs types and no common rationale for methods choices.[37] The Cochrane Rapid Reviews Methods Group recommends a few approaches, such as limiting the search strategy and including a restricted list of outcomes.[14] An important program for future methodological research is the investigation of the impact of omitting or abbreviating review processes.[67] Identifying and understanding the impact of the methodological choices for RRs is critical to guarantee the validity and robustness of any findings they might generate, which can ultimately impact the policy and clinical decisions that they support.

An eDelphi study developed by our own group, explored the main uncertainties around the methods to develop RRs.[58] One of the main areas identified as high-priority was search strategy. The fifth most important question that experts had identified was: “What is the optimal method for defining the evidence/information sources that should be searched...?”. The search strategy, more specifically the choice of databases, are usually streamlined when performing a RR.[6,10,17] This can potentially impact the results included and introduce bias. Depending on the selection of databases, treatment effect estimates from meta-analysis may vary, both in magnitude and direction.[61]

Previous methodological studies have been developed to understand the impact of different databases combinations or how important comprehensive literature searches are. Different databases combinations may lead to various recall rates depending on the field. Field-specificity seems to be important when comparing search strategies, coverage of Pubmed, for instance, varies across specialties and over time.[68] For references included in systematic reviews of qualitative research

regarding diabetes mellitus, MEDLINE/PubMed and CINAHL and MEDLINE/PubMed, CINAHL, and Embase had the highest overall recall rates.[60] For various Cochrane reviews on clinical interventions, several abbreviated search approaches led to the same conclusion. However, combining three regularly used electronic databases (MEDLINE, CENTRAL, and Embase) did not allow authors to have the same level of certainty when drawing conclusions as a comprehensive search that included specialised databases.[62] We are not aware of any study that explored the impact of different databases combinations on cardiac rehabilitation.

Considering the limited empirical evidence available to support the streamlined methods applied for the conduct of RRs, the critical question around databases' selection this study aims to assess the impact of database selection on the inclusion of studies and subsequent implications on outcome point-estimates in cardiac rehabilitation-based systematic reviews. Cochrane Heart, Stroke, and Circulation, formerly the Cochrane Heart Group, has published several systematic reviews, including in cardiac rehabilitation. These reviews are usually updated every five years. And although an information specialist is usually part of the team, the reviews take around a year to be developed. Understanding their methods could potentially help to leverage a time-saving approach that can allow for more frequent efficient updates.

2. METHODS

Seven Cochrane reviews on cardiac rehabilitation were explored as a case study (Table 3-1). One review was considered as two independent reviews because its update used a new search strategy with different outcomes. The original searches for these reviews were predominantly done in: (1) five primary databases (Cochrane Central Register of Controlled Trials (CENTRAL), MEDLINE, Embase, CINAHL, and Science Citation Index Expanded); (2) two trial registers (World Health Organisation's International Clinical Trials Registry Platform and Clinicaltrials.gov); and (3) hand-searches of reference lists of retrieved articles and recent systematic reviews. As most of the reviews are not in their first version, the most updated version of the review, its list of included studies, and its search strategy were considered. For the purpose of this study, we considered the main electronic databases only (CENTRAL, MEDLINE, Embase, CINAHL, and any other included in the review).

The primary outcome of this methodological study is the number of missed studies

and trials for each combination of databases, while the secondary outcomes are the changes in treatment effect sizes of the primary outcome, mortality, for each combination of databases.

Table 3-1. List of cardiac rehabilitation Cochrane reviews

CR1	Exercise-based cardiac rehabilitation for coronary heart disease.
CR2	Exercise-based cardiac rehabilitation for adults with heart failure.
CR3	Exercise-based cardiac rehabilitation for adults with atrial fibrillation-version 1
CR4	Exercise-based cardiac rehabilitation for adults with atrial fibrillation-version 2
CR5	Psychological interventions for coronary heart disease.
CR6	Psychological interventions for depression and anxiety in patients with coronary heart disease, heart failure or atrial fibrillation
CR7	Home-based versus centre-based cardiac rehabilitation.
CR8	Effectiveness of social network interventions to support cardiac rehabilitation and secondary prevention in the management of people with heart disease.

2.1 Data collection and extraction

To determine where each included study was initially indexed, we used the original libraries of each review for every database before de-duplication. If access to the original libraries files was not possible, the search strategy was reverse engineered and re-run, considering the most updated version available of the search strategy. All the references reported in the “References of studies included in this review” section of each Cochrane review were used and exported to an Excel Sheet (Microsoft). The number of included studies (all randomised controlled trials) in each review varied from 10 to 145.

After replicating the searches, or based on the original libraries, we recorded whether each reference/record included was identifiable from each database. This allowed us to understand where each reference is indexed. If a study was not indexed in any of the databases, it was considered that it was derived using the hand search/grey literature strategy.

Data from each study was recorded considering the data already extracted in the existing reviews: author; year; publication information; country; language; and design. The quality of each included study was also recorded, using the Risk of Bias evaluation already performed on each one of the reviews. Trials included in each review with ≥ 300 participants randomised and with low risk of bias in < 3 out of 5

domains were categorized as influential trials.

We also mapped the outcomes that each reference contributed to in the review, extracting whether the study was included in the qualitative or quantitative analyses. We extracted the main values for the studies included in the meta-analysis as the main raw number (the number of people per group, broken down by whether they experienced a given outcome), the effect sizes (e.g., risk ratio), and confidence intervals.

2.2 Data analysis

Excel Sheet (Microsoft) equations were used to test different search approaches (i.e., the different possible combinations of databases, for instance, MEDLINE, Embase, and CENTRAL, individually and in combinations of two and three – but at least with one of these main databases). We quantified the loss in the number of records using an abbreviated search, for instance, when only two databases were combined. To perform this analysis, we considered the reference level, where each manuscript is considered individually, and at the trial level, where we assessed the indexing of any record from a group of references related to a single trial. This is because each trial often leads to multiple publications, sometimes in different journals, and, in some cases, exploring different outcomes or updated analyses. When a publication from a trial is retrieved from the databases' search, the trial is already known by the reviewers and its protocol, and additional publications are searched by hand. So, one trial publication may be sufficient for inclusion in the review. Exploring the trial-level data becomes an interesting approach, as it allows for grouping different publications associated with a trial, recognising that the inclusion of one publication alone warrants the consideration of the entire trial in the review.

References not indexed in any of the databases were considered derived from hand searches. Those were considered for the primary outcome but not excluded when analysing the change in treatment effect sizes. Using the sample size and risk of bias of each study, we assessed if the trials lost in any particular database combination were influential trials or not. For this, the classification of each trial was considered, and we calculated the number and percentages of lost trials, understanding the distribution of high and low-quality trials among the missed studies

Using RStudio, version 4.2.2, the meta-analysis for the primary outcome (i.e., overall mortality) was re-run according to the list of references indexed in each database's combination (combining from one to three databases). This list considered the indexing of the databases at trial-level, following the original reviews. We assessed whether the treatment effect varies depending on the database combinations, examining if the point estimate remains consistent in direction and statistical significance level. We also investigated cases where the point estimate retains its direction but differs in statistical significance. For this, as our analyses excluded hand search, the results used as reference were the values when all databases were combined, which differs from the results of the original reviews because they also included the trials derived from the hand search. This comparison between treatment effects in direction and significance was performed previously by Ewald et. al.[61] Additionally, depending on the databases used, we registered how often the treatment effect estimates could not be calculated (i.e., the abbreviated search failed to retrieve any of the included trials or fewer than two trials).

3. RESULTS

3.1 Indexing rates and loss of studies and trials

The eight original reviews included a total of 584 studies (547 excluding hand search) and 305 trials (293 excluding hand search). Only 6.3% and 3.9% of studies and trials, respectively, were derived from hand searches exclusively. Most of those references were trial protocols, and some others were corrigendum's to full reports. Table 3-2 presents the number of studies and trials indexed and missed for each database. Indexing rates varied from 0 to 100%, considering all the databases. CENTRAL had the highest indexing rate, with a mean of 93.5% of included studies and 92.8% of included trials, and it was the one that varied least in indexing rates across the eight reviews. LILACS was searched only by one review (CR3) and did not index any included trial. Across the other databases, PsycINFO had the lowest percentage of included studies (8.1% ± 6.1) and trials (16.7% ± 12.9).

Figure 3-1 illustrates the percentage of indexed trials per database for each review when excluding hand search. CENTRAL (mean 92.8% ± 5.7), MEDLINE (82.2% ± 12.5), and Embase (72.9% ± 14.1) indexed the greatest number of included trials. In the Appendix III, we present detailed study and trial levels results, including and excluding retrieves from hand search.

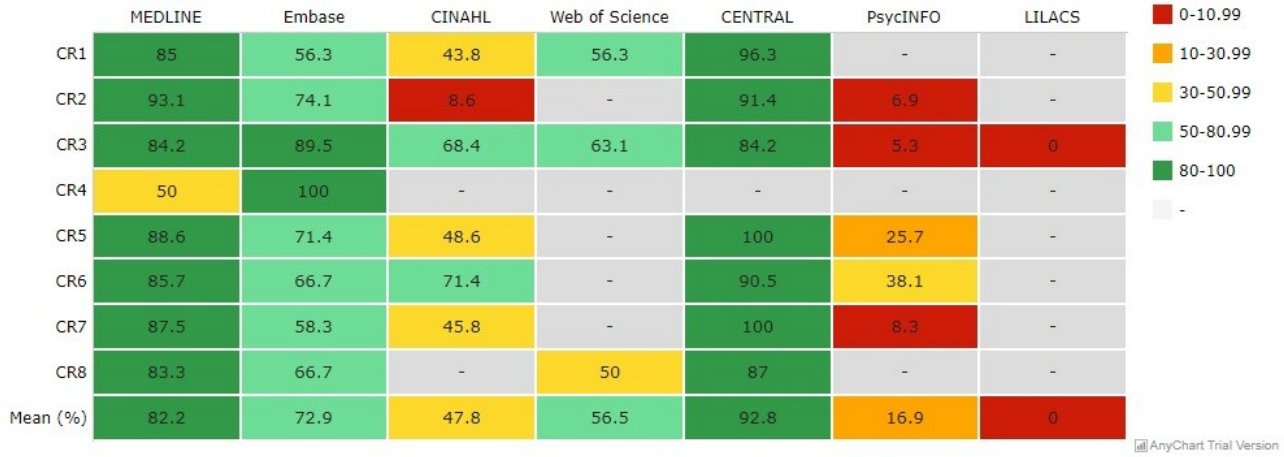
Table 3-2. Number of included and missed studies and trials for each database compared to the full search.

	N of Included Studies	N of Included Studies Without Hand Search	N of Included Trials	N of Included Trials Without Hand Search
	Number of inclusions (% of total) Number missing (% of total)			
CR1				
All databases	145	135	85	80
M	105 (72.41) 40 (27.59)	105 (77.7) 30 (22.22)	68 (80) 17 (20)	68 (85) 12 (15)
E	67 (46.20) 78 (53.79)	67 (49.62) 68 (50.37)	45 (52.94) 40 (47.06)	45 (56.25) 35 (43.75)
CE	127 (87.58) 18 (12.41)	45 (33.33) 90 (66.67)	35 (41.17) 50 (58.82)	35 (43.75) 45 (56.25)
CI	45 (31.03) 100 (68.97)	60 (44.44) 75 (55.56)	45 (52.94) 40 (47.06)	45 (56.25) 35 (43.75)
W	60 (41.37) 85 (58.62)	127 (94.07) 8 (5.93)	77 (90.58) 8 (9.41)	77 (96.25) 3 (3.75)
CR2				
All databases	104	97	60	58
M	87 (83.65) 17 (16.35)	87 (89.69) 10 (10.31)	54 (90) 6 (10)	54 (93.10) 4 (6.90)
E	72 (69.23) 32 (30.77)	72 (74.22) 25 (25.77)	43 (71.66) 17 (28.33)	43 (74.13) 15 (25.86)
CE	92 (88.46) 12 (11.54)	5 (5.15) 92 (94.85)	5 (8.33) 55 (91.67)	5 (8.62) 53 (91.38)
CI	5 (4.8) 99 (95.19)	5 (5.15) 92 (94.84)	5 (8.33) 55 (91.66)	5 (8.62) 53 (91.37)
P	4 (3.84) 100 (96.15)	4 (4.12) 93 (95.88)	4 (6.66) 56 (93.33)	4 (6.89) 54 (93.10)
CR3				
All databases	24	23	20	19
M	19 (79.16) 5 (20.83)	19 (82.60) 4 (17.39)	16 (80) 4 (20)	16 (84.21) 3 (15.79)
E	21 (87.5) 3 (12.50)	21 (91.30) 2 (8.70)	17 (85) 3 (15)	17 (89.47) 2 (10.53)
CE	20 (83.33) 4 (16.67)	13 (56.52) 10 (43.48)	13 (65) 7 (35)	13 (68.42) 6 (31.58)
CI	13 (54.16) 11 (45.83)	13 (56.52) 10 (43.48)	12 (60) 8 (40)	12 (63.15) 7 (36.84)
W	13 (54.16) 11 (45.83)	20 (86.95) 3 (13.04)	16 (80) 4 (20)	16 (84.21) 3 (15.79)
P	1 (4.16) 23 (95.83)	1 (4.34) 22 (95.65)	1 (5) 19 (95)	1 (5.26) 18 (94.74)
L	0 24 (100)	0 23 (100)	0 20 (100)	0 19 (100)
CR4				
All databases	10	3	6	2
M	1 (10) 9 (90)	1 (33.33) 2 (66.67)	1 (16.66) 5 (83.33)	1 (50) 1 (50)
E	3 (30) 7 (70)	3 (100) 0	2 (33.33) 4 (66.67)	2 (100) 0
CR5				
All databases	81	79	35	35

M	61 (75.30) 20 (24.69)	61 (77.21) 18 (22.78)	31 (88.57) 4 (11.43)	31 (88.57) 4 (11.43)
E	59 (72.83) 22 (27.16)	59 (74.68) 20 (25.32)	25 (71.42) 10 (28.57)	25 (71.42) 10 (28.57)
CE	74 (91.35) 7 (8.64)	31 (39.24) 48 (60.76)	17 (48.57) 18 (51.43)	17 (48.57) 18 (51.43)
CI	31 (38.27) 50 (61.73)	74 (93.67) 5 (6.33)	35 (100) 0	35 (100) 0
P	10 (12.34) 71 (87.65)	10 (12.65) 69 (87.34)	9 (25.71) 26 (74.29)	9 (25.71) 26 (74.29)
CR6				
All databases	44	40	21	21
M	28 (63.63) 16 (36.36)	28 (70) 12 (30)	18 (85.71) 3 (14.29)	18 (85.71) 3 (14.29)
E	21 (47.72) 23 (52.27)	21 (51.5) 19 (47.50)	14 (66.66) 7 (33.33)	14 (66.66) 7 (33.33)
CE	38 (86.36) 6 (13.64)	20 (50) 20 (50)	15 (71.42) 6 (28.57)	15 (71.42) 6 (28.57)
CI	20 (45.45) 24 (54.55)	38 (95) 2 (5)	19 (90.47) 2 (9.52)	19 (90.47) 2 (9.52)
P	8 (18.18) 36 (81.82)	8 (20) 32 (80)	8 (38.09) 13 (61.90)	8 (38.09) 13 (61.90)
CR7				
All databases	50	47	24	24
M	37 (74) 13 (26)	37 (78.72) 10 (21.28)	21 (87.5) 3 (12.50)	21 (87.5) 3 (12.50)
E	19 (38) 31 (62)	19 (40.42) 28 (59.57)	14 (58.33) 10 (41.67)	14 (58.33) 10 (41.67)
CE	47 (94) 3 (6)	15 (31.91) 32 (68.09)	11 (45.83) 13 (54.17)	11 (45.83) 13 (54.17)
CI	15 (30) 35 (70)	47 (100) 0	24 (100) 0	24 (100) 0
P	1 (2) 49 (98)	1 (2.12) 46 (97.87)	2 (8.33) 22 (91.66)	2 (8.33) 22 (91.66)-
CR8				
All databases	126	123	54	54
M	77 (61.11) 49 (38.89)	77 (62.6) 46 (37.40)	45 (83.33) 9 (16.66)	45 (83.33) 9 (16.66)
E	55 (43.65) 71 (56.35)	55 (44.71) 68 (55.28)	36 (66.66) 18 (33.33)	36 (66.66) 18 (33.33)
CE	111 (88) 15 (11.90)	39 (31.70) 84 (68.29)	47 (87.03) 7 (12.96)	47 (87.03) 7 (12.96)
W	39 (30.95) 87 (69.05)	111 (90.24) 12 (9.76)	27 (50) 27 (50)	27 (50) 27 (50)

M, MEDLINE; E, Embase; CE, CENTRAL; CI: CINAHL; W, Web of Science; P, PsycINFO ; L, LILACS.

Figure 3-1. Trial indexing rates per database



Gray: Not applicable – The review did not include this database as a resource.

The number of influential trials included in each review (considering sample size and risk of bias evaluation) was low. CR1 included 9 influential trials; CR2, 5; CR3,1; CR4, none; CR5, 6; and CR7, 1. This information was unavailable for either CR6 or CR8 because of the stage of the review update or the way the risk of bias evaluation was registered. There were no differences in the numbers when excluding trials from hand search. However, one trial that was missed when excluding hand search was not considered influential because of the risk of bias but included more than 3 thousand participants and contributed to the reviews' meta-analysis weight by about 30%. In the Appendix III, we present the indexing rates of influential trials per databases.

3.2 Combination of databases in pairs

Embase + CENTRAL had the highest indexing rates for both studies (92.0%, ± 2.7 with hand search and 97.2% ± 2.7 excluding hand search studies) and trials-level (94.8% ± 3.7 with hand search and 96.8% ± 3.5 excluding hand search studies). Figure 3-2 shows the study indexing percentages for each pair of databases that includes at least one of the main databases (MEDLINE, Embase, or CENTRAL). Rates were lower when considering the total number of included studies, including those derived from hand searches, compared to ratings from the total with no hand search studies (Figure 3-2B). Rates that include those derived from hand search

reflect the coverage of each pair of databases on the final list of studies included in the original reviews.

Figure 3-3 shows the percentage of trial coverage for each pair of databases (that include at least one of the main databases). When CENTRAL is combined with other databases besides LILACS (used in one review only), it has an indexing rate higher than 90%. When considering all the trials included in the reviews, including the ones derived from hand search (Figure 3-2B), pairs of databases can still retrieve a high percentage of the trials as a low percentage came from hand search. Embase plus PsycINFO showed the lowest indexing rate for both study and trial levels from the explored combinations.

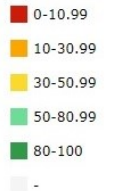
Figure 3-2. Study indexing rates per pair of databases

A

	M + E	M + CI	M + W	M + P	M + L	M + CE	E + CI	E + W	E + P	E + L	E + CE	CE + CI	CE + W	CE + P	CE + L
CR1	83	83.7	84.4	-	-	95.5	63.7	73.3	-	-	96.3	96.3	97.8	-	-
CR2	96.9	90.7	-	89.7	-	100	77.3	-	76.3	-	95.9	94.8	-	94.8	-
CR3	100	87	87	82.6	82.6	87	100	100	95.7	91.3	100	91.3	87	87	87
CR4	100	-	-	-	-	-	-	-	-	-	-	-	-	-	-
CR5	94.9	78.5	-	77.2	-	97.5	82.3	-	75.9	-	97.5	97.5	-	93.7	-
CR6	72.5	77.5	-	70	-	95	72.5	-	57.5	-	97.5	97.5	-	95	-
CR7	80.9	80.9	-	80.9	-	100	55.3	-	44.7	-	100	100	-	100	-
CR8	66.7	-	68.3	-	-	93.5	-	53.6	-	-	93.5	-	94.3	-	-
Mean (%)	86.8	83	79.9	80.1	82.6	95.5	75.2	75.7	70	91.3	97.2	96.2	93	94.1	87

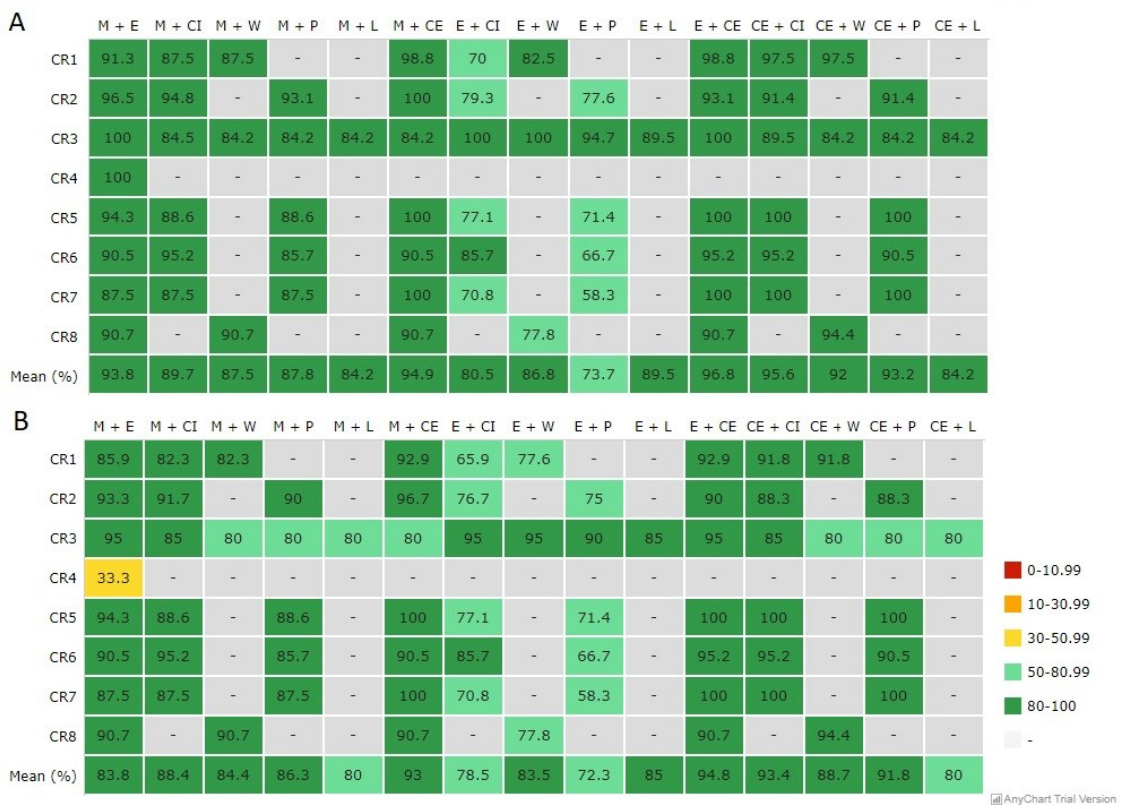
B

	M + E	M + CI	M + W	M + P	M + L	M + CE	E + CI	E + W	E + P	E + L	E + CE	CE + CI	CE + W	CE + P	CE + L
CR1	77.2	77.9	78.6	-	-	89	59.3	68.3	-	-	89.7	89.7	91	-	-
CR2	90.4	84.6	-	83.7	-	93.3	72.1	-	71.2	-	89.4	88.5	-	88.5	-
CR3	95.8	83.3	83.3	79.2	79.2	83.3	95.8	95.8	91.7	87.5	95.8	87.5	83.3	83.3	83.3
CR4	30	-	-	-	-	-	-	-	-	-	-	-	-	-	-
CR5	92.6	76.5	-	75.3	-	95.1	80.2	-	74.1	-	95.1	95.1	-	91.3	-
CR6	65.9	70.5	-	63.6	-	86.4	86.4	-	52.3	-	88.6	88.6	-	86.4	-
CR7	76	76	-	76	-	94	52	-	42	-	94	94	-	94	-
CR8	65.1	-	66.7	-	-	91.3	-	52.4	-	-	91.3	-	92.1	-	-
Mean (%)	74.1	78.1	76.2	75.5	79.2	90.3	74.3	72.2	66.2	87.5	92	90.5	88.8	88.7	83.3



2A: Percentage of studies indexed in each pair of databases corresponding to each review, excluding from the total trials the ones derived from hand search. 2B: Percentage of studies indexed in each pair of databases corresponding to each review, considering in the total trials derived from hand search. In grey, no results were available as the combination of databases did not apply to that review. M, MEDLINE; E, Embase; CI: CINAHL; W, Web of Science; P, PsycINFO; L, LILACS; CE: CENTRAL.

Figure 3-3. Trial indexing rates per pair of databases



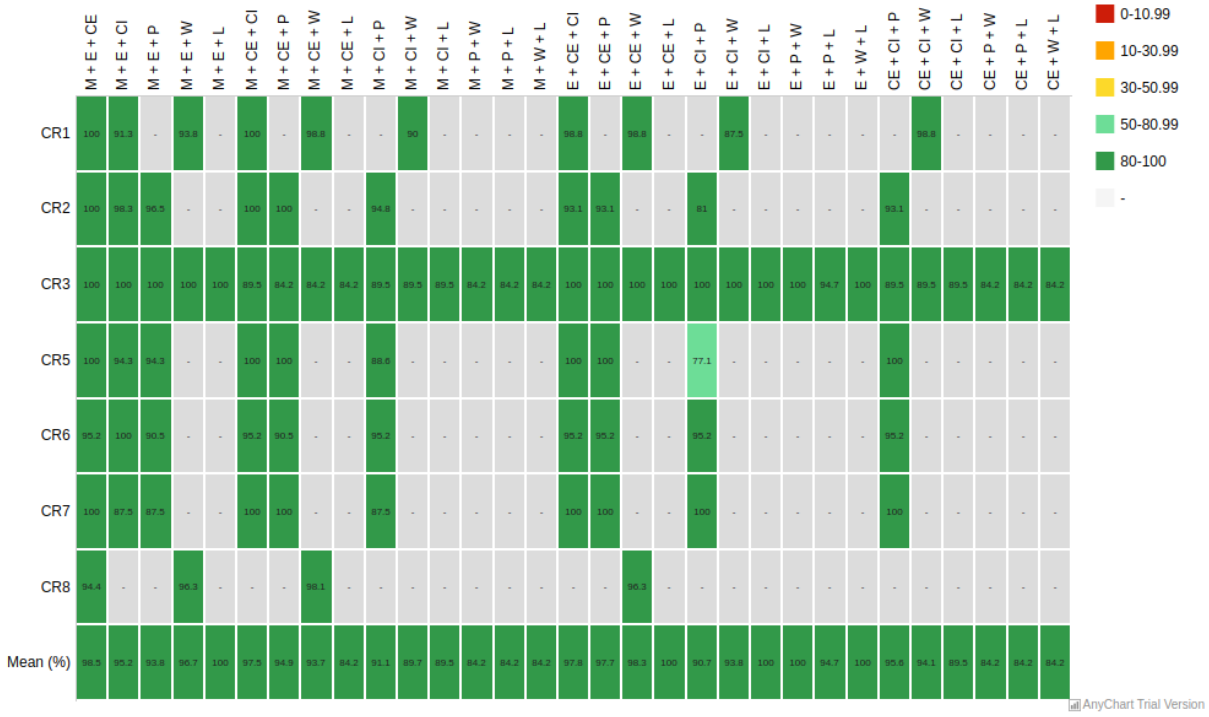
3A: Percentage of trials covered by each pair of databases corresponding to each review, excluding from the total trials derived from hand search. 3B: Percentage of trials covered by each pair of databases corresponding to each review, considering in the total trials derived from hand search. In grey, no results were available as the combination of databases did not apply to that review. M, MEDLINE; E, Embase; CI: CINAHL; W, Web of Science; P, PsycINFO; L, LILACS; CE: CENTRAL.

3.3 Combination of three databases

Figure 3-4 illustrates the indexing rates for trials, without hand search, when combining three databases with at least one of the primary databases (MEDLINE, Embase, or CENTRAL). All the combinations pertinent to each review showed high indexing rates. The only combination relevant to all the reviews and with several cases of 100% coverage was MEDLINE plus Embase and CENTRAL.

From the 31 combinations explored, the only combination of databases with a mean of 100% (k = 19 trials) were those only applicable to CR3. This seems to be driven by Embase, that although it covers 98% of the trials indexed in this review and only one more than CENTRAL, the trials indexed by each differ, with no total overlap.

Figure 3-4. Trial indexing rates per combination of three databases



M, MEDLINE; E, Embase; CI: CINAHL; W, Web of Science; P, PsycINFO; L, LILACS; CE: CENTRAL.

3.4 Comparison of effect estimates

The analysis of effect estimates was only possible for five reviews (CR1, CR2, CR5, CR7, and CR8). For CR3 and CR4, not enough trials were included without a hand search with estimable effect, so it was not possible to perform a meta-analysis with different sets of included trials (i.e., different databases' combinations). This review was considered separate from the original review because it was a previous version with a different search strategy. In addition, because of the ongoing update, the meta-analysis of CR6 was not available.

Figure 3-5 presents Forrest Plots of effect estimates for mortality up to 12 months of follow-up time for combinations of one to three and all databases for each review. The results for all databases are the reference for comparison. For CR1, CR4, and CR5, CENTRAL alone found the same results as all combinations of pairs or trios that included CENTRAL. Overall, results were fairly stable for less than 12 months as long as MEDLINE, Embase, or CENTRAL were included. For the ones mentioned above, CENTRAL or any combination with this database was enough,

while CR2 and CR6 needed to have the combination of MEDLINE + Embase or CENTRAL to be able to find the same effect estimates as all databases combined. For CR6, any combination that included Embase or CENTRAL, but not MEDLINE combined with one of the main ones, results were very close to the reference (0.9 (1.08-0.72) vs. 0.89 or 0.9 (1.09-0.74)). In the Appendix III, we present data for additional follow-up times.

3.4.1 Same effect estimates as all studies

Treatment effect estimates based on pairs of databases were identical to those based on all databases in 34 cases of the 90 meta-analyses (37.8%: Figure 3-6). When considering trios of databases, the treatment effect was identical in 60 cases of the 90 meta-analyses (66.7%). Notably, even with only one database, an identical effect was estimated in nine cases (20.0%, out of the 45 meta-analyses that could be performed with only one database). Five of these cases included only CENTRAL. Only four of the meta-analyses had insufficient data to estimate treatment effects when testing the inclusion of trials derived from only one database.

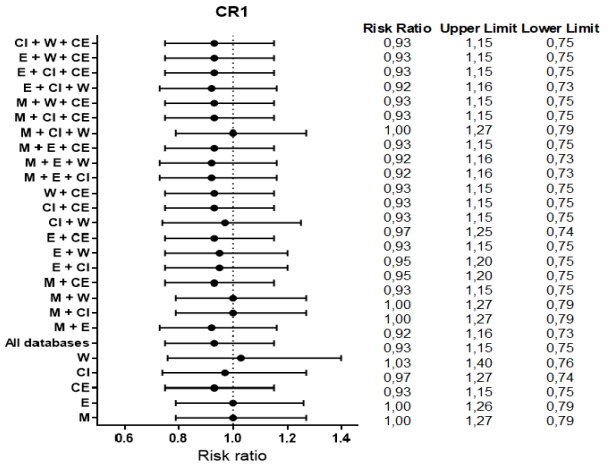
3.4.2 Same direction but different magnitude as all studies

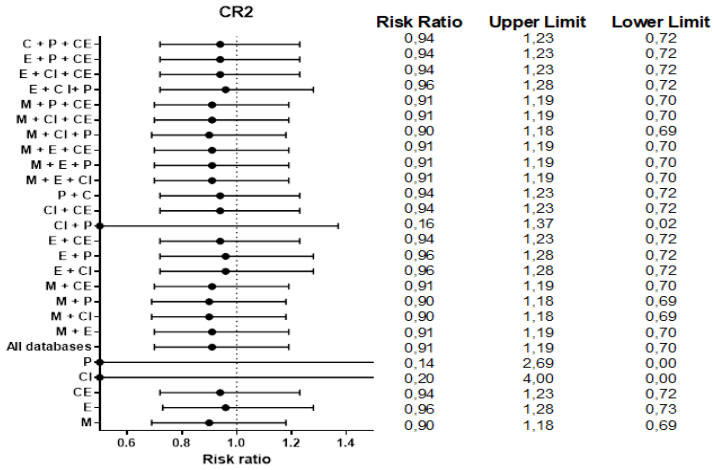
When considering only one database, 27 out of the 45 meta-analyses (60.0%) found the same direction but a different magnitude. In 46 cases of the 90 meta-analyses with pairs of databases (51.1%), and in 28 of the 90 meta-analyses (31.1%) with trios of databases they were not identical but in the same direction and had the same level of statistical significance.

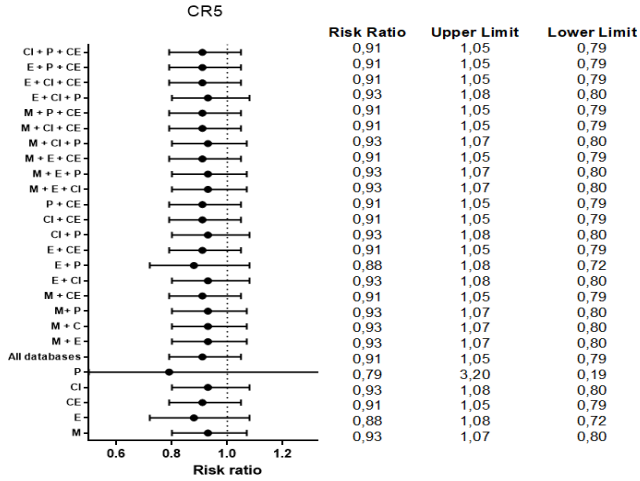
3.4.3 Opposite to as all studies

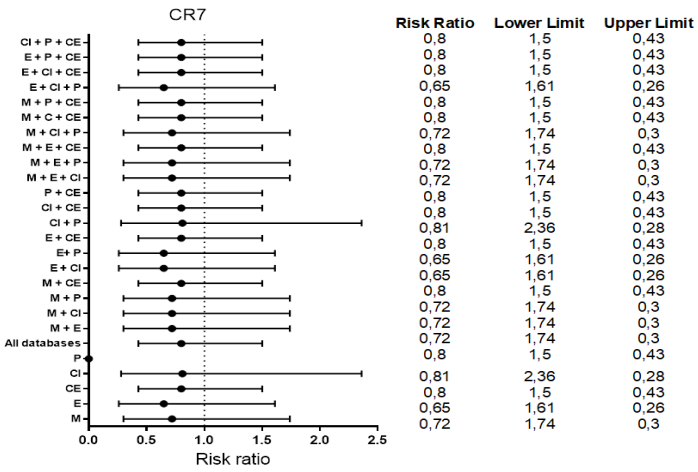
Treatment effect estimates moved in opposite directions in five cases for both when only one database was included (out of 45, 11.1%) and for pairs of databases (out of 90, 5.6%), and in two cases (out of 90, 2.0%) for the combinations of databases in trios.

Figure 3-5. Forest plots of effect estimates at up to 12 months follow-up using various database combinations









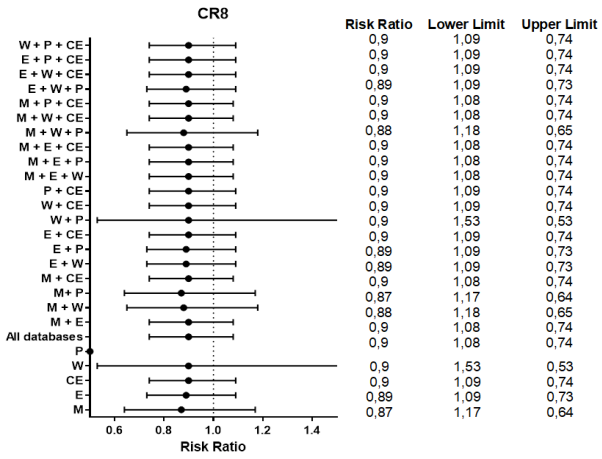
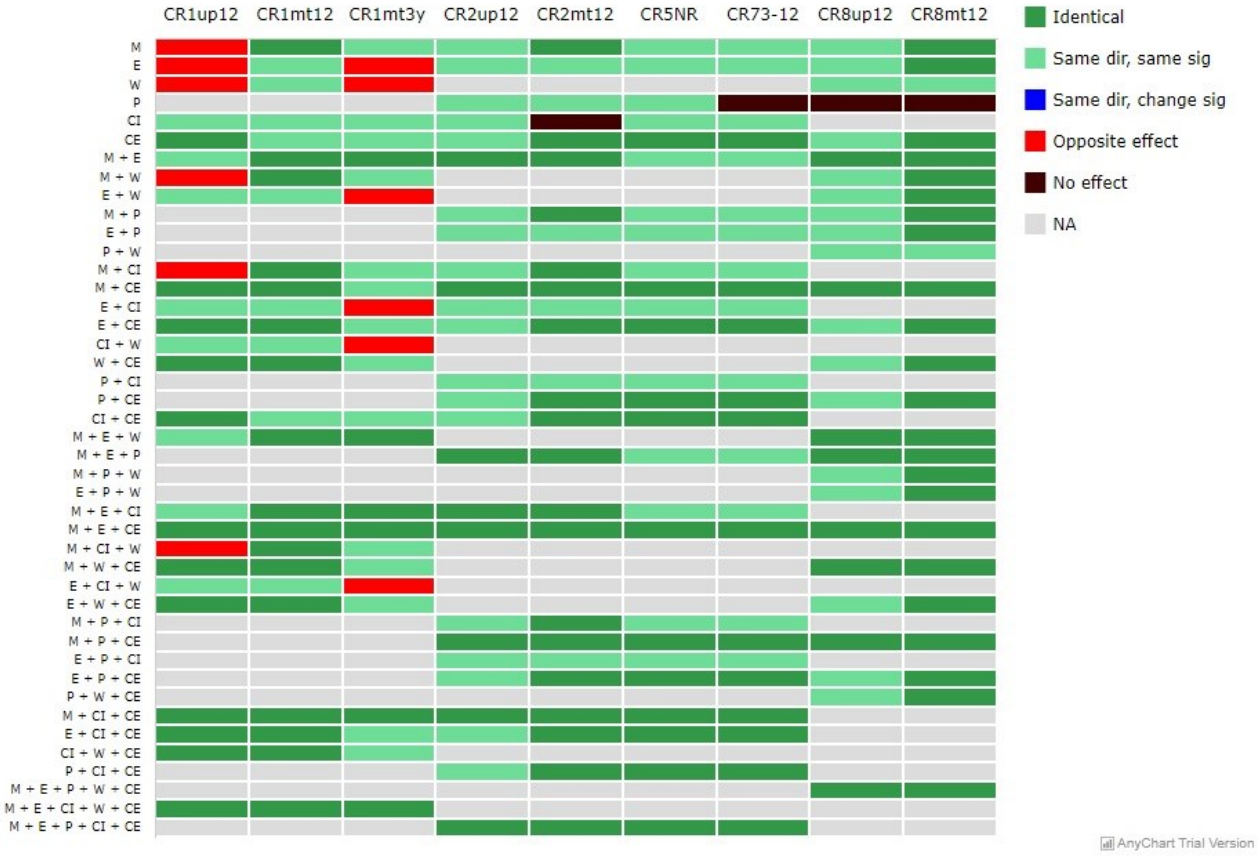


Figure 3-6. Comparison between treatment effects estimates



The figure shows the agreement of the treatment effect estimates resulting from a meta-analysis based on trials found with different combinations of databases compared with that based on all included trials from the original searches (last three rows in the figure). Each row corresponds to a different database combination, and the columns reflect each review and follow-up time. The last three lines reflect the reference, the most comprehensive search.

Up: up to 6 or 12 months; mt: more than 12 months or more than three years. Or in the case of CR7, from 3 to 12 months.

Color legend: 1: full concordance between treatment effect estimates; 2: estimates without changes in the direction of the point estimate and with the same statistical significance; 3: estimates without changes in the direction of the point estimate with gain/loss of the statistical significance; 4: estimates with changes in the direction of the point estimate; 5: estimates unavailable because of number of included trials. NA, not applicable; M, MEDLINE; E, Embase; W, Web of Science; P, PsycINFO; CI: CINAHL; CE: CENTRAL.

4. DISCUSSION

After exploring various bibliographic database choices, including single, pairs, and trios of database combinations, our results indicate that CENTRAL, MEDLINE, and Embase have the highest indexing rates. Only CENTRAL indexed more than 90% of the trials and 97.8% of the influential trials throughout all reviews. In addition, estimate effects only from CENTRAL frequently achieved the same values as when

all databases were included. CENTRAL was enough to find the same results for three of the five available reviews (CR1, CR4, and CR5) for the outcome analysis. For the other two reviews, CR2 and CR6, it was necessary to combine MEDLINE with Embase or CENTRAL to see the same effect estimates as for all databases combined.

When comparing with the available literature, we can notice that the type of review question can influence the impact of different databases. In the present study, although all eight reviews used as a case study are about cardiac rehabilitation, the nature of their questions differed. Half of the reviews were focused on exercise interventions and the other half on other rehabilitation components or modalities (e.g., home-based, psychological, and social network interventions). However, the indexing rates, per single database or different combinations, seemed to be very similar throughout all the reviews independent of this. When exploring a simplified search strategy for review updates, a previous study found that a simplified search strategy using MEDLINE performed better for clinically focused topics compared to complex or broader topics.[69] Another study reported that abbreviated searches affected Cochrane reviews of nonpharmacological topics more and for pharmacological topics. Studies of pharmacological interventions were usually indexed in the major databases, while studies of psychological interventions, for instance, were more frequently indexed in specialised databases.[62] Of note, in our study for cardiac rehabilitation, recall rates were high in major databases.

In the present study, we found that CENTRAL alone was enough to acquire the same results as the original broader search strategy in most cases. Similarly, when exploring Cochrane reviews on therapeutic interventions, Halladay et al. found that searching only Pubmed alone was enough to agree with meta-analysis results in 98% of the cases.[70] All the Cochrane reviews explored in the present study are based on randomised controlled trials, which are the only bibliographic type of report indexed by CENTRAL. Most of the records they index come from bibliographic databases, such as Pubmed, Embase, and CINAHL. However, searching one database is not recommended for RRs.[62] Our results showed that only one database does not consistently retrieve the same results. For example, searching CENTRAL + Embase or CENTRAL + CINAHL had additional value as it achieved the highest indexing rates (96.8% and 95.6%, respectively, compared to 92.8% for CENTRAL only). In some cases, even if it is a major database that is being searched (such as

MEDLINE or Embase), it can also lead to opposite estimated treatment effects, as was seen in the current study. Our results support the recommendation by the Cochrane Rapid Reviews Methods Group to limit main database searching to CENTRAL, MEDLINE, and Embase, when developing a RR in the medical field.[14] Using this time-saving approach could also be explored as an option when planning a review update instead of initiating a review, although literature around that is scarce. However, a simplified search strategy with fewer search terms and only for MEDLINE may not be enough for broader topics even when updating reviews.[69]

Our results show that the outcomes of meta-analyses vary slightly depending on the combination of databases used. Including more databases tends to align the estimated effects closer to those found when searching five databases. However, significant time savings can be achieved by limiting searches to just two or three databases. This confirms what previous research has found, particularly for respiratory infection-related reviews; restricting meta-analysis based on MEDLINE + Embase led to the fewest changes in statistical significance. This previous methodological study also explored Infectious Diseases and Developmental Psychosocial and Learning Problems reviews and concluded that although most relevant studies appear in a limited number of databases, database choice is topic-specific.[71] Trials that are more difficult to find tend to be smaller and with less methodological quality, and their importance seems to vary within different health fields.[59] In the context of cardiac rehabilitation, we found a low number of influential trials included in the reviews, so the importance of other trials seem to be high. For dementia care research, for instance, instead of exploring eight original databases, searching more specific databases (a combination of CINAHL, MEDLINE, Web of Science Core Collection, and citation tracking) seemed necessary to retrieve all studies.[72] For a review of reviews, MEDLINE, Epistemonikos, and reference checking was the best combination, with a mean indexing rate of 97.7% of the systematic reviews.[73]

Reference checking and additional searches beyond traditional databases are key components of comprehensive search strategies. Currently, Cochrane recommends to “assess the need for grey literature and supplemental searching, and justify the sources to be searched”, based on expert opinion.[57] When exploring 60 reviews, half pharmacological and half nonpharmacological, Nussbaumer-Streit et al. found that combining the major databases (MEDLINE, Embase, and CENTRAL) with

searches of references lists generated the lowest proportion of changes in conclusions.[62] While this might be time-consuming,[63,64] it might also be a crucial step, especially when authors anticipate whether the main body of literature is more recent. In the future, empirical research is essential to understand how to decide on the evidence/information sources that should be searched in different fields, especially for time efficiency. Not only which databases to search but what is the best combination of databases and supplementary searches. Depending on the field, a well-executed grey literature search may be more cost-effective and a valuable complement to more standardised search strategies.

To the best of our knowledge, this is the first study to assess the impact of different database combinations on the results of reviews in cardiac rehabilitation. However, our study has some limitations. First, we had a limited number of the original searches that were available. When re-running the searchers, even when considering the date range that the original review searched, the electronic databases are generally updated retrospectively, which may influence the indexing rates encountered. Also, both CINAHL and LILAS had updated the way that the databases are searched, which meant that exact replication of prior searches was not possible. Additional issues were encountered when re-running the searches, as we only had access to PsycINFO APA, not OVID, as the reviews had used. Second, we assumed that the most updated search for each review was the best strategy, and we used this version to replicate the search, but the old versions of the review may have used different search terms. We also assumed that the Cochrane reviews were the best quality reference in the field and that no errors were made when selecting the studies, and that the relevant studies that fit the criteria for each review were identified and included. Third, as mentioned previously, the Cochrane reviews also included searches on two trial registers (World Health Organisation's International Clinical Trials Registry Platform and Clinicaltrials.gov). We decided not to check if the included studies were indexed in those as only ongoing trials are generally included (outside those that have been published). Finally, when excluding hand search retrieves from the outcomes analysis, the estimated effects would not be the same as the ones in the original reviews that include hand search. Hand searching is an important consideration when exploring evidence sources, as searching reference lists of included studies is usually a step followed. However, our study aimed to focus

on the question of electronic databases and did not explore the impact of hand search.

5. CONCLUSION

Our study offers valuable insights for researchers in the field of cardiac rehabilitation into how to employ a resource-efficient approach when crafting search strategies for evidence synthesis. This study also has the potential to inspire other researchers to reproduce our methods and further explore the question of the impact of searching fewer databases on reviews' results in other fields.

For reviews on cardiac rehabilitation, searching only CENTRAL plus a major database such as MEDLINE or Embase is viable when a time-saving approach to evidence identification is needed. However, it would seem that a minimum of three databases would be recommended when developing a meta-analysis. The bibliographic database choices need to be based on the resources available, and the review commissioners' needs and results should be interpreted taking into consideration the influence of these choices.

CHAPTER 4

Methodological Insights: Comparing Single Review and Peer-Review Approaches in Study Selection for Evidence Synthesis

Ariany M Vieira[1,2], Laurence Paquet[2], Simon Bacon[1,2].

[1] Department of Health, Kinesiology and Applied Physiology, Concordia University, Montreal, Quebec, Canada

[2] META Group, Montreal Behavioural Medicine Centre, CIUSSS du Nord-de-l'ile-de-Montreal, Montreal, Quebec, Canada

Abstract

Background: Rapid Reviews (RRs) are used to perform time-efficient evidence syntheses. Time-saving methods have been applied, including single-review for the study selection process. Peer review (i.e., two reviewers independently assess studies for eligibility, with a third solving discrepancies) is usually recommended. However, a single review approach has the potential to save time and resources.

Objectives: To understand the impact on inclusions, time, and costs of having single review compared to peer review for title and abstract screening.

Methods: Using a concluded review, title and abstract screening was performed following the single review approach. Peer review was then performed to assess the differences in time invested and potential costs. Accuracy (sensitivity and specificity) was calculated.

Results: 2,526 retrieves were divided equally across the five reviewers. Peer-reviewing all titles and abstracts led to 647 retrieves being included, compared to 797 retrieves from single review. The sensitivity and specificity of single review was 0.84 and 0.86, missing 99 (3.9%) retrieves (false negatives) and including 249 (9.9%) false positives. Peer review took twice the amount of time (87.57 hours) and twice the estimated cost (CA$ 2,479.96).

Conclusion: The single reviewer approach missed around 4% of inclusions and included more false positives (9.9%), potentially leading to longer full-text assessments. However, single review can be an alternative approach that takes half the time and costs. Larger methodological studies in various fields are needed to explore further the impact of single reviewer on studies selection and review results.

Keywords: Rapid reviews, Systematic reviews, Study selection, Methodology

Strengths and limitations of this study:

- By providing empirical data on the differences between single- and peer-review study selection, this study contributes to the body of evidence on the impact of review methods. Insights from this study are valuable for understanding how study selection approaches can impact the inclusion of studies, costs, and time.
- As this is a case study based on a specific context, generalizability is limited.
- The impact of the two study selection approaches was not evaluated on the final list of studies included in the review and on the review outcomes.
- Future studies can replicate this study's methods to further evaluate the impact of study selection approaches. It would be important to consider additional outcomes, such as the review results, across different disciplines, fields of study, or contexts.

1. INTRODUCTION

Evidence summary papers (e.g., systematic reviews [SRs]) are an effective method for providing an overview of a specific topic or research area. When conducting an SR, it is crucial to employ rigorous methods to ensure that the results are trustworthy and can inform precise clinical and policy decisions. SRs are a way of searching and selecting the available empirical evidence to answer a research question while reducing bias.[3] As SRs employ rigorous methodologies to generate a conclusion, they require notable time and resources. Performing a SR is a time-intensive process, taking from several months to over two years, from conception to completion. Consequently, by the time an SR is finalised, it might not reflect the available literature anymore as it may not have new and relevant publications incorporated into the synthesis. To address these challenges, rapid reviews (RRs) have been developed.

Rapid Reviews are ‘‘a rigorous and transparent form of knowledge synthesis that accelerates the process of conducting a traditional systematic review through streamlining or omitting a variety of methods to produce evidence for end-users in a resource-efficient manner.’’[9] This approach can be essential to policymakers and health systems managers who need reliable evidence to make timely decisions on healthcare. In this case, a full SR may not always be feasible or practical, so RRs are an alternative evidence synthesis approach. However, there is no standardized methodology applied to RRs,[32] and limited evidence on how best to conduct such reviews efficiently and reliably**.**

In a RR, each step of a systematic review can be omitted or abbreviated, depending on the resources and time available.[74] This includes a more focused question, a narrower search strategy, and streamlined methods for study selection, data extraction, quality assessment, and reporting.[54] Studies selection is a step that can take a long time during the development of a review, depending on the number of retrieves, reviewers, and resources available. According to a scoping review of RRs methods, 18% (15 out of 82 reviews) used one reviewer only for title and abstract screening, and 34% (28 out of 82 reviews) performed screening with two or more independent reviewers. Of note, the other 48% of reviews used one reviewer and one verifier; did not clearly report the number of reviewers; did not do this step; or did not reported the methods.[12] The Selecting Approaches for Rapid Reviews (STARR) decision tool supports RR researchers in deciding which RR

approach to use when planning a review.[65] However, this tool provides no specific information on study selection. On the other hand, the Guide for RRs, developed by McMaster University, recommends a peer-review approach (two reviewers independently review the references) to minimise bias.[75]

Cochrane also recommends using two reviewers for dual screening of at least 20% of abstracts for the selection of studies, with conflict resolution. Then, one reviewer screens the remaining abstracts, and a second reviewer screens all excluded abstracts and, if needed, resolve conflicts.[14] This approach is based on a study from 2021 that aimed to understand the accuracy of single- and dual-reviewer screening. Using two reviews and a group of trained participants, they found that single-reviewer abstract screening missed 13.4% of relevant studies, while dual-reviewer abstract screening missed 2.5% of relevant studies.[26]

Empirical evidence is still limited to support the decisions on the study selection approach, especially when efficiency is key. Researchers need to consider several aspects to make the choice, including available resources, the size of the review, and the potential for bias.[74] To help inform researchers, this study aims to understand the impact on inclusions, time and costs, of having a single review compared to peer review process for titles and abstracts screening.

1.1 Hypothesis

This study aimed to answer: “What is the difference in accuracy and the impact on study selection during title and abstract screening between single and peer review approaches?” Our hypothesis was that there will be no significant difference between single and peer review approaches in the accuracy of study selection during title and abstract screening.

2. METHODS

A methodological study was developed using a rapid review recently completed by the META Group's, the evidence synthesis team of the Montreal Behavioural Medicine Centre (MBMC). Concordia University Human Research Ethics Committee approved this study under the certification number 30020328.

2.1 Data source

The META Group worked on developing a rapid review for the Public Health Agency of Canada, via the COVID-END Initiative. This review aimed to understand the effectiveness of different lengths of quarantine and isolation in reducing transmission of COVID-19 in non-healthcare community-based settings (review questions on Appendix IV).

Before reviewing titles and abstracts, all individuals involved received training on the specific criteria for including or excluding studies in the review. In this training, criteria for inclusion and exclusion were discussed, and six examples of titles and abstracts were screened together. In addition, reviewers received a decision tree and a table with inclusion and exclusion criteria to be used during the actual screening. Three rounds of pilot were then performed, summing a total of 110 studies screened independently by each reviewer and discussed afterward. After reaching an 80% agreement level, the official screening started.

2.2 Participants

The sample consisted of members of the META Group, including students (one Master's and three doctoral students), a postdoctoral fellow, and a research assistant (who was responsible for the project management and did not take part in the study selection).

2.3 Single Review

Following an RR approach, which already has a standard protocol in the group, titles and abstracts were screened using Rayyan (https://www.rayyan.ai/), followed by the full-text screening. For titles and abstract screening, all the retrieves from the database search were divided equally among each reviewer. A separate Rayyan project was created for each reviewer. They all contained the same screening information, such as keywords for inclusion and exclusion, reasons for exclusion, or labels. (Appendix IV). These keywords were only highlighted (either in green or in red) and used as a tool to help decision making as the reviewer still needed to read the entire abstract before making a decision.

2.4 Peer-review

We considered peer-review when two independent reviewers assessed each title

and abstract separately, with discrepancies between reviewers being resolved by a third reviewer. For this, all the titles and abstracts reviewed by one person during single review were equally distributed to each of the other four reviewers. For instance, 1/5th of the reviewer one's (R1) articles was sent to reviewer 2 (R2), R3, R4 and R5, respectively. We followed this approach to make sure every reviewer was paired with one another, so that every pair of reviewers had a different combination of experience levels.

Again, each reviewer received a single Rayan project with their own set of titles and abstracts to screen and the same screening information—keywords, reasons for exclusion, or labels.

2.5 Data collection and Informed Consent

The data collected included the decisions on inclusion during the single and peer-review title and abstract screening. To explore the cost-benefit of each screening approach, the time invested in the process was recorded using Rayyan's time control feature and a time-tracking sheet.

An Information and Consent Form was sent to all reviewers to clarify the study objective and procedures and invite them to answer a short survey to gather information about their level of knowledge and experience. After participants provided agreement, a link to the survey using Concordia Microsoft Forms was shared by email. The survey consisted of eight questions, taking around 15 minutes to be completed. Questions asked about their job title, level of experience in evidence synthesis, and level of knowledge in the research topics.

2.6 Statistical analysis

Microsoft Excel (version 2406, Microsoft Corporation) was used for data organisation and analysis. The level of agreement between all the reviewers was calculated using the proportional agreement (Po), expected agreement (Pe), and Cohen's Kappa. The peer-review screening results were compared to the single review screening, and the percentage of missed studies was calculated. The primary outcome of interest was the accuracy (sensitivity and specificity) of study selection during single (already performed to the review development) compared to peer-review title and abstract screening. We followed the analyses used by previous

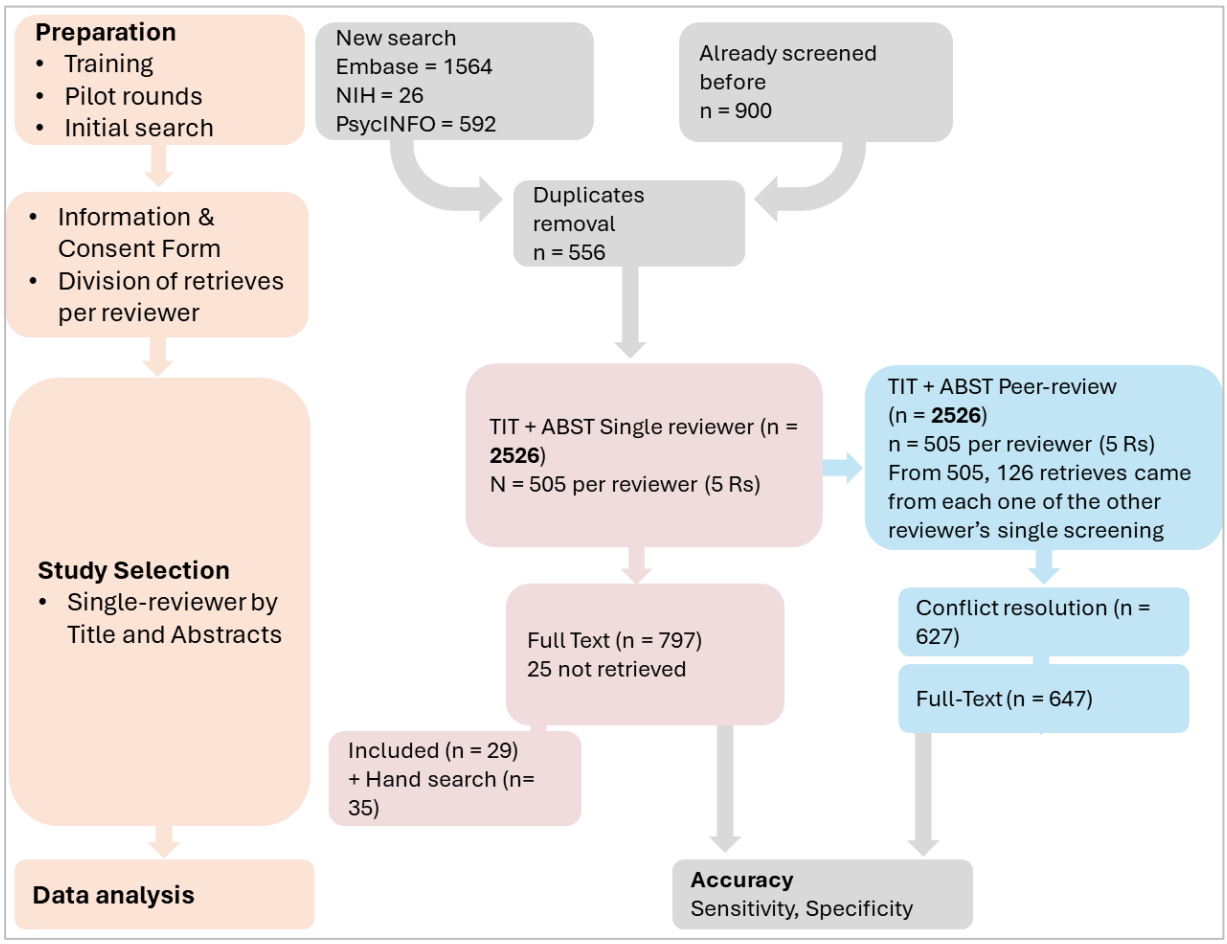
research.[26] For this, we calculated the following outcomes: the number of studies identified as relevant by the screening process (true positives); the number of studies actually relevant but missed by the screening process (false negatives); the number of studies identified as non-relevant by the screening process (true negatives); and the number of studies actually non-relevant but incorrectly included (false positives). Sensitivity was calculated as True Positives / (True Positives + False Negatives). And Specificity as True Negatives / (True Negatives + False Positives). The reference was the results obtained by the peer-review approach.

To explore the cost-benefit of each screening approach, the potential cost was calculated based on the working hours of each member. All costs were reported in Canadian dollars (CA$) and were based on the average salary in Canada. For students, the reference was how much a standard scholarship pays (https://src.uqam.ca/subvention/remuneration/). Human resources were considered as the number of reviewers needed and their invested working hours for the title and abstract screening only, excluding any time and costs involved in project management or prospective review steps.

3. RESULTS

The search yielded a total of 2,526 retrieves needing to be screened for this review. Figure 4-1 shows the overall process followed in this methodological study. The total number of titles and abstracts was divided between five reviewers. After single review screening, the same titles and abstracts were divided between the same five reviewers for the peer-review process, ensuring each reviewer was paired with every other reviewer equally.

Figure 4-1. Overview of the two screening processes



The single review process led to 797 articles being sent to full text screening (31.6% of the retrieves) as this is the standard method used to develop the review. After peer-reviewing all titles and abstracts, there were 627 conflicts (24.8%) and 1,899 agreements (75.2%). This led to 647 studies that would have been sent to full-text screening (25.6% of the retrieves).

Table 4-1 presents the accuracy, cost, and time results for the two title and abstract screening approaches. Single review screening included 150 (6% of the total number of studies), more retrieves, missing 99 (3.91%) retrieves (false negatives), and including 249 (9.85%) false positives. Sensitivity was 0.84 (± 0.08), and specificity was 0.86 (± 0.03). Overall, the single review screening took half the time peer review took, although peer-review accounts for the additional screening plus conflict resolution. This is also true for costs; the screening phase (no other related steps, such as training, search, export of libraries, or creation of Rayyan projects) was estimated to cost CA$ 1,246.48 for a single review screening and CA$ 2,479.96 for peer review.

Table 4-1. Impact and costs of the two studies' selection approaches

		Single reviewer (n = 2,526)	**Peer-review** (n = 2,526)
Results	Conflict	NA	627 (24.8%)
	Agreement	NA	1,899 (75.2%)
	Included	797 (31.6%)	647 (25.6%)
Accuracy	TP	548 (21.7%)	Reference
	FP	249 (9.9%)	Reference
	FN	99 (3.9%)	Reference
	TN	1,630 (64.5%)	Reference
	Sensitivity	0.84	Reference
	Specificity	0.86	Reference
Resources	Time (hours)	43.82	87.57
	Cost (CAD)	1,246.48	2,479.96

NA: not applicable; TP: true positive (the number of studies identified as relevant by the screening process); FP: false positives (the number of studies actually non-relevant but incorrectly included); FN: false negatives (the number of studies relevant but missed by the screening process); TN: true negatives (the number of studies identified as non-relevant by the screening process).

The reviewers' experience levels varied (Table 4-2). Self-rated knowledge in evidence synthesis varied from 2 to 10 (mean 6.0 ± 3.0) on a 10-point Likert-like scale, while knowledge on the review topics varied from 2 to 8 (mean 5.0 ± 2.2 on respiratory illness, 6.6 ± 1.0 on social isolation/quarantine, and 6.8 ± 0.7 on public health measures). R4 and R5 were highly experienced, with ≥ 7 years of experience and 8 and 10 self-rated knowledge in evidence synthesis.

The level of agreement between reviewers evaluated by Cohen's Kappa was moderate overall (mean 0.40 ± 0.07), varying from 0.27 to 0.53 (Table 4-3). The highest agreement level came from pairing the most experienced reviewer with a master's student, with an average experience level (2 years of experience and self-rated knowledge in evidence synthesis as 7).

Table 4-2. Level of knowledge and experience of reviewers

Reviewer	Job title	Years of experience	Evidence synthesis knowledge*	No. of SRs	No. of RRs	Field of research	Respiratory illness knowledge*	Social isolation/ quarantine knowledge*	Public health measures knowledge*
R1	Master's student	2	7	0	2	HP & S	5	7	7
R2	PhD student	0	3	0	1	NA	2	6	6
R3	PhD student	1	2	1	0	HP	3	5	7
R4	PhD student	7	8	4	5	Covid-19, rehabilitation, patient assessment tools	7	7	6
R5	Postdoctoral fellow	12	10	10	35	HP & S, infectious disease treatment, health	8	8	8

*Self-rated knowledge using a Likert-scale, from 0 (no expertise) to 10 (expert).

HP: health policy; HP & S: health policy and systems.

Table 4-3. Agreement between the different pairs of reviewers

Pair	Po	Pe	Cohen's Kappa
R1+R2	0.78	0.58	0.47
R1+R3	0.71	0.57	0.32
R1+R4	0.69	0.58	0.27
R1+R5	0.79	0.55	0.53
R2+R3	0.77	0.61	0.41
R2+R4	0.77	0.62	0.40
R2+R5	0.71	0.56	0.34
R3+R4	0.77	0.57	0.47
R3+R5	0.73	0.56	0.38
R4+R5	0.75	0.58	0.42

Po: proportional agreement; Pe: expected agreement.

Tables 4-4 and 4-5 present the results by members of the evidence synthesis

team. As can be observed, the mean time for each reviewer to screen 505 retrieves was 526 minutes / 8.76 hours (±3.39). A similar time was taken when peer-reviewing the additional 505 retrieves, leading to an average time for the entire peer review of 1,051 min / 17.51 hours (±5.59) per reviewer. The average cost per reviewer was CA$249.26 for single review and CA$495.99 for peer review.

Sensitivity across individual reviewers varied from 0.74 to 0.92, while specificity ranged from 0.84 to 0.92. High values of true positives (TP) and true negatives (TN) and low values of false positives (FP) and false negatives (FN) indicate better performance in study selection. R2, the reviewer with the lowest experience level, had the lowest sensitivity, missing a higher proportion of relevant studies (high number of false negatives), but the highest specificity (with the lowest number of false positives), indicating a good ability to exclude non-relevant studies from the selection correctly. R5, the more experienced reviewer, achieved the highest sensitivity among all the reviewers, identifying 117 TP out of a total of 126 relevant studies (117 TP + 9 FN), and also performed well in excluding studies that did not meet the inclusion criteria.

Table 4-4. Time and costs by each member of the evidence synthesis team

Reviewer	Reviewer job title	Time invested Single reviewer (min)	Time invested Second - reviewer (min)	Time invested Peer-review (min)	Hourly rate (CAD)	Single reviewer Total Cost (CAD)	Peer-review Additional Cost (CAD)	Peer-review Total Cost (CAD)
R1	Master's student	389	389	778	25.51	165.30	165.30	330.60
R2	Doctoral student	824	760	1,584	27.73	380.73	351.06	731.79
R3	Doctoral student	483	557	1,040	27.73	223.22	257.33	480.56
R4	Doctoral student	309	443	752	27.73	142.80	204.64	347.45
R5	Postdoctoral fellow	625	477*	1,102*	32.11	334.26	255.27	589.53

*This reviewer did not participate in the conflict resolution, so the peer-review time did not account for that.

Table 4-5. Accuracy results by members of the evidence synthesis team

			Conflict					No conflict			Total						
	Years of experience	Self-rated knowledge	n	TP	TN	FP	FN	n	TP	TN	n	TP	FP	FN	TN	Sensitivity	Specificity
R1	2	7	115	26	32	44	13	383	105	285	505	131	44	13	317	0.910	0.878
R2	0	3	122	14	49	29	30	390	72	311	505	86	29	30	360	0.741	0.925
R3	1	2	139	21	41	63	14	392	73	293	505	94	63	14	334	0.870	0.841
R4	7	8	138	28	22	55	33	366	92	276	506	120	55	33	298	0.784	0.844
R5	12	10	113	26	20	58	9	368	91	301	505	117	58	9	321	0.929	0.847

TP: true positive (the number of studies identified as relevant by the screening process); TN: true negatives (the number of studies identified as non-relevant by the screening process); FP: false positives (the number of studies actually non-relevant but incorrectly included); FN: false negatives (the number of studies relevant but missed by the screening process).

4. DISCUSSION

In the present study, we found that when performing title and abstract screening, a single review approach included 797 retrieves, compared to 647 retrieves included by a peer-review approach. The single review included 548 (21.7%) true positives and 249 (9.9%) false positives, while it excluded 1,630 (64.5%) true negatives but missed 99 (3.9%) retrieves, i.e., false negatives. The single-review approach's sensitivity was 0.84, and its specificity was 0.86. The level of agreement and accuracy varied among the five reviewers, with the highest sensitivity and specificity of 0.92 coming from different reviewers with divergent levels of knowledge and experience. Peer review took twice the estimated time and cost.

The single review approach during title and abstracts missed almost 4% of the relevant studies. In a more extensive methodological study, Gartlehner et al. found that single reviewer abstract screening missed 13%, with a sensitivity similar to the one found by the present study of 86.6% and a specificity of 79.2%.[26] However, the reviewers' experience level seems to play an important role. When synthesising four studies that compared single screening vs peer review screening, the median proportion of missed studies was 3% for more experienced reviewers and 13% for less experienced reviewers.[25] It is recommended that producers of RRs keep highly experienced staff in their teams.[6,55] In the present study, 2 out of 5 (40%) reviewers were highly experienced in evidence synthesis, while 3 out of 5 (60%) reviewers had

high domain-knowledge. So, missing 3.9% of eligible studies seemed to be better than expected according to experience level when looking at the reports available in the literature. Two studies compared the performance of students and experts and found that performance was highly variable among students, below that of experienced reviewers in one study,[76] but with the higher inter-rater agreement and no difference for abstract sentencing in another one.[77]

Although the single review approach missed a small percentage of studies, it included 249 false positives, which would greatly impact the time taken to complete the full-text screening phase. This contrasts with previous findings indicating that a single review approach typically identifies most of the relevant studies, but peer review can enhance the identification of relevant records by an average of 9%.[78] Our findings should be leveraged depending on the time and budget available. We agree with the available literature,[78–80] and recommend the peer-review approach whenever possible. Still, the single review can be a valid alternative method when time or cost-saving methods are necessary.

Despite the training offered and the pilot screenings performed, the agreement between reviewers was lower than expected. A pilot with 30 to 50 abstracts is recommended so that the entire screening team can calibrate and test the review material.[14] In the present study, this was not enough to guarantee a good agreement level during full screening. Although the reviewers were part of the same team, most of them were not used to working together and were only available for this particular review. Although piloting is part of good practices when developing a review, the question of its structure and impact before screening is still open.[25]

The moderate agreement between reviewers may be due to the nature of the review question and the topic. This methodological study is based on a review commissioned by knowledge users to answer a few questions on the consequences of isolation and quarantine related to infectious diseases. There is not only one question but three, and definitions of isolation and quarantine are often confounded in the literature and the remits for what constitutes these measures is inconsistent across studies.[81–83] For instance, lockdown or any measure of mass quarantine was an exclusion criterion for the review. However, this is usually unclear in the abstracts. Similarly in other parallel domains, the lack of information in the abstract was shown to impact the accuracy in identifying studies that meet eligibility criteria in studies on diagnostic tests.[79]

Research questions that are too vague may make the identification of studies difficult. For example, Waffenschmidt et al. found the best results for single screening with the most specific research questions.[25] Also, as highlighted by various authors, deciding on methodological strategies for conducting RRs may be challenging due to the inherent variability across different research topics.[32,80] Single-reviewer screening may seem to miss more relevant studies for public health than pharmacological topics.[26] Although an optimal approach seems to not exist, developing methods to explore this question may be relevant for researchers in various fields who want to investigate the same methodological question.

Although the present study can offer valuable insights into the RR field, there are a few limitations. The generalisability is limited because the review question and inclusion criteria were very specific, as well as the size of the present study (only one review domain and five reviewers). However, we believe the transferability of methods can help assess how these results vary across different disciplines, fields of study, or contexts. Previous studies have highlighted the limitations of using the kappa coefficient of agreement.[84,85] One key point is that the kappa statistic can vary significantly based on the proportion of records that are considered eligible, it is sample size sensible. In the present study, more than 20% of the retrieves were considered eligible. Such a higher disproportion of eligible:ineligible records may lead to increased variability in kappa values. In addition, there is no universally accepted standard for what constitutes a "good" kappa value, as it can depend on the context and field of study.[85] Another limitation of our study is that the review results were not taken into consideration as an outcome to better understand the impact of having a different set of studies included.

Future studies need to explore further the impact of performing single review versus peer review approaches, to understand the impact on the end results of the reviews. More extensive methodological studies that control for several factors, such as experience level, field of study, training, and piloting, are necessary to create substantial evidence for decisions around RR methods.

5. CONCLUSION

By examining the advantages and drawbacks of single and peer review for study selection during the title and abstracts phase, we found that the single reviewer approach missed around 4% of inclusions and included many false positives (9.9%),

potentially leading to longer full-text assessments. However, the single review approach took half the time and cost of peer-review title and abstract study selection. As such, a single review approach can be a useful alternative when time-saving methods are necessary.

This study offers valuable insights to evidence synthesis teams as they develop their selection methods by comparing two study selection approaches. Information on accuracy and costs is useful for researchers and organisations conducting reviews, and guideline developers. These findings can inform decisions about human resource allocation in RR projects.

CHAPTER 5

Discussion

Summary of Findings and Thesis Implications

The primary purpose of this thesis was to identify critical methodological questions and undertake methodological studies to explore how aspects of RR methods impact the review processes and review findings. In order to do this, we first conducted an eDelphi study to build consensus around the main methodological questions on RRs development. Based on the list generated from this study, we designed and conducted two methodological studies. The first one aimed to explore the impact of the number of databases on included studies and results. The second methodological study explored the differences between single review and peer review approaches for titles and abstract screening.

In chapter 2, 78 experts in RRs and evidence synthesis knowledge users were surveyed about the main methodological questions on RR development. The majority had more than nine years of experience in evidence synthesis and were from various countries around the world. After three survey rounds and a consensus meeting, the list that started with 29 items was reduced to seven key methodological questions. Those items are essential to understanding how to improve an RR's time efficiency. Of the seven items, three were related to search strategy, while two were on study selection, one on quality/bias assessment, and one on data extraction.

Considering the increasing use of RRs in policy and decision-making and the numerous unresolved methodological questions, this eDelphi study advances the science of RRs by building a research agenda and serving as a “road map” for researchers. The questions were structured to be specific and pertinent to one review step, so they are ready for researchers to design methodological studies. The generated list can also help to prioritise the issues that experts consider more important to be explored. This can support resource allocation, ensuring that funding focuses primarily on the key areas.

Next, in Chapter 3, we developed a methodological study exploring how different combinations of databases searched would impact cardiac rehabilitation-based reviews. For this, seven reviews were used as a case study, and each included study's possible indexing was investigated. The meta-analyses of each

review were re-run depending on the list of studies included by each combination of databases. The results from pairs and trios of databases were compared to the results from all databases included in the review combined. We found that CENTRAL had the highest indexing rates, and Embase plus CENTRAL was the best combination of databases. However, when looking at effect estimates for meta-analyses, combining the major databases (MEDLINE, Embase, and CENTRAL) in pairs, only 38% of results were identical to all databases combined. While with three major databases combined, 66.7% of results were identical to when all databases were combined.

An additional meta-analysis could be conducted to explore the characteristics of included trials across different database combinations. For example, a sensitivity analysis could be performed, considering various study characteristics, such as risk of bias and sample size, to assess their impact on effect estimates. Furthermore, a qualitative assessment comparing the characteristics of studies indexed in each database combination could provide a deeper understanding of how these factors influence the effect estimates. These analyses could also include an examination of the geographic location of studies retrieved from each database combination, which may reveal whether certain methods capture a broader or more diverse set of evidence. Additionally, coverage of key journals and redundancy analysis — examining the overlap of studies retrieved from different database combinations — could further enrich the discussion by highlighting the comprehensiveness and uniqueness of the evidence gathered.

This study offers valuable insights into an essential question about RR methods. Researchers in the field of cardiac rehabilitation now have some guidance on the best combination of databases if a more streamlined approach is necessary. Our findings indicate that major databases can be utilised to identify potentially eligible studies, though additional databases are required for comprehensive meta-analyses. Evidence synthesis producers can consider these findings to understand how different database combinations can impact reviews, the list of included studies, and estimated effects. Furthermore, this study shows the importance of selecting appropriate databases to enhance the quality and reliability of reviews. By adopting the recommended database combinations, researchers can improve the accuracy and comprehensiveness of their reviews. To build on these findings, the methods of this study need to be replicated in other fields and also be used in bigger studies with

review questions of a different nature to further understand the impact of various databases' combinations.

In Chapter 4, we present the results of a methodological study using a rapid review as a case study. We explored the impact of single vs. peer-review approaches for title and abstract screening. The single review process included more studies, 797 compared to 647 retrieves included by peer review. Single review took half the time and cost but missed 3.9% of retrieves (false negatives) and included 9.9% false positives.

Our study was based on a review with complex questions and results should be interpreted taking this into consideration. We provided empirical evidence on the impact of a single reviewer and how this approach only missed a small percentage of eligible studies while saving time and resources. These results contribute to a more comprehensive understanding of the advantages and disadvantages of each study selection approach. However, the time-benefits during screening for the single review approach need to be weighed against the time that might be taken to review more studies during full-text screening, which could increase the total time taken to complete the review. Depending on the time and resources available, in general, our results suggest recommending a single review approach for studies selection, as it missed a low percentage of included studies and had the potential to save resources during initial screening.

It is relevant to highlight the differences in the scope and eligibility criteria between studies 2 and 3. The database study was developed in the context of Cochrane reviews, which are known for their rigorous evidence synthesis methods. They gather data on RCTs, while the studies selection in study 3 focused on a RR of observational and modeling studies. The findings of each study are context-dependent, as previously mentioned, and potentially influenced by these differing scopes. As discussed in Chapter 3, the identification of trials also varies depending on the study size and topics. If study 2 had focused on indexing rates of observational studies, the results might have shown a higher degree of variability due to the inherent limitations of observational data. The conclusion could have been different, also emphasizing the need for a more cautious interpretations of findings and the importance of RR methods tailored to handle the proper identification of observational studies. Conversely, if study 3 had been conducted using a Cochrane review, screening title and abstracts of RCTs, the results would likely have been

more robust in terms of accuracy for study selections, as RCTs usually follow more structured abstract, with more specific questions in terms of population, intervention, and outcomes. In this situation, the appropriateness of single review selection might be improved. This limits the generalizability of studies 2 and 3, as the findings from a methodological study in one context may not be directly applicable to others, but understanding and confirming the impact of these differences would also be important to reviewers to better understand the methods choices.

Taken together, the results of these studies advance the science of RRs, contribute to building the research agenda on RR methods, and provide specific insights and empirical evidence, encouraging further exploration of these questions.

Strengths and Limitations of this thesis

As a collective thesis there are several limitations which need to be considered when interpreting the results of the studies (beyond those previously mentioned in chapters 2-4).

Lack of consensus on RRs: RRs are a useful strategy for informing decision-making more efficiently. However, their accelerated process can have several implications that are still poorly understood nowadays. Although the present thesis aimed to contribute to a better understanding of the questions surrounding this and the impact of these time-saving methods, the heterogeneity and lack of consensus around RRs still exist. Heterogeneity in nomenclature, methods, and reporting have already been observed.[12,22,86,87] During the development of the eDelphi study, we received several comments and discussed how to phrase each methodological question to avoid any misunderstanding about the search for a single optimal method for each review stage. Asking about an “optimal” approach would leave unclear whether the answer should be generalizable to different reviews. Authors have argued that standards and expectations for systematic reviews are nonnegotiable and RRs could potentially be placed into a spectrum of reviews, from non-systematic to a full systematic review.[18] RRs are also viewed as an alternative until a full systematic review can be developed,[32,86] but it should not be used as a ‘faster’ substitute for a comprehensive systematic review.[32] Therefore, it is essential to carefully assess whether a RR is suitable, particularly when an urgent need for a summary of the evidence arises.[88] When comparing the outcomes of RRs and SRs addressing the same questions, Watt et al. found that RRs were able to reach the

appropriate conclusions, but their scope was narrower, and their appropriateness may be better for well-defined questions.[32,89] Although this restricts the generalisability of the review findings to other contexts, it enables RRs to provide timely information for specific policy or practice decisions.[80]

*The limitations of methodological studies: "*Research-on-research," also known as methodological studies, are important for health research and can greatly impact the quality of research conducted and decrease research waste. Experimental studies that compare different methods also play an important role.[90] A few limitations have been previously highlighted in chapters 3 and 4, as the complexity of topics, reviewer variability, and team expertise. However, we want to emphasise some limitations, such as sample size and generalizability as limitations. The development of those studies is resource-intensive, and funding is essential to develop larger methodological studies. Moreover, given the specificity of the studies developed in the present thesis, it is not possible to generalise the findings to various research fields as they might be context specific, with context covering the field of research, the nature of the research question, the team composition, and the available resources and methods in place. The fact that the reviews chosen for comparison might not represent the full diversity of reviews can introduce selection bias. However, inconsistent metrics across fields may be an issue. For example, how accuracy, time, and cost are estimated can make comparisons difficult. In a methodological systematic review of studies comparing study selection methods, authors found that the evaluations were inconsistent and incomplete.[25]

Inclusion of knowledge-users: Although the eDelphi study included a broadly representative group of participants, our recruitment strategy was not good enough to capture community members and public partners. This might be a reflection of the absence of a representative of this group on our research team and scientific committee. Although the participation of patient/community member/caregiver was considered since the beginning, a proactive recruitment plan specifically targeting this group was not developed. The Delphi survey was shared by multiple partners that have this group of people as members and by social media, which we expected to generate some participation from this group. Our inclusion criteria may also have been a barrier to participation, as we looked for people with some experience in evidence synthesis. More accessible language with better-defined concepts or examples would potentially have influenced people's interest in the study. The same

issue is true for the methodical studies, in which we only have researchers in the health field with similar backgrounds. In the case of these studies, we could have included policymakers and decision-makers. As all the studies developed were methodological, one could argue that the primary knowledge users were the researchers themselves. However, it is known that the inclusion of other knowledge users can enhance the quality of the research and the potential uptake of its findings.[91,92]

Strengths: Despite these limitations, some overall strengths should be emphasised in addition to the ones already described in each study, notably, the uniqueness of this thesis. RRs are still considered a new review approach,[67] and if they are now better described the concept of RRs is clarified,[9] there are still considerable gaps in our understanding of RRs. It is an emerging topic that caught great attention during the COVID-19 pandemic due to the urgency of answering health-related questions. The present thesis addresses some of the gaps in the field, including understanding the essential methods questions on RRs' development. A key strength was the inclusion of highly experienced researchers from across the globe who were part of the Scientific Committee, thus providing a broader perspective on the results that were generated. Another strength is the potential replication of the methods we applied in this thesis. Although the eDelphi approach is already a well-established consensus-building method, this thesis shows how it can be a useful process for creating research agendas in the context of evidence synthesis. In addition, the methods followed by our methodological studies can also be also replicated in other evidence synthesis types, e.g., scoping reviews. A final strength is how the three main components of this thesis (i.e., the Delphi and the two methodological studies) build upon each other and provide unique and valuable results in a complementary and coherent way.

Future Directions

This thesis generated a list of the essential questions regarding methods to develop more time efficient RRs. Future studies can be designed to answer these questions for different research fields. Conducting methodological studies with rigorous procedures across multiple fields can help to understand the impact and potential bias caused by the time-saving methods followed in RRs. The goal is to create a substantial body of evidence on the influences of methodological choices

across a range of areas in order to inform researchers and end-users when selecting the reviews' methods or when considering the reviews' results.

A large body of evidence would also serve as a model when researchers need to explore a methodological question in another field. Exploring the same questions in different fields can help to identify commonalities and differences, which can help to improve the understanding of how RRs perform or whether they could or should be developed in various contexts. The methods followed in our methodological studies can also serve this purpose. Future methodological studies can adapt and expand those questions by examining the effects of other methodological choices or the same ones in additional outcomes. For instance, questions about the impact of databases should be further explored, considering the available hand search strategies. Exploring the same question for other fields is also appropriate, such as reviews on pharmacological interventions. This can help to develop more universally applicable methods for RRs development.

There are a variety of tools that apply artificial intelligence (AI) to the development of evidence synthesis, especially for study selection and data extraction. These tools are not yet widely available, sometimes not considered user-friendly, and still need further exploration around their accuracy and sensitivity, as they can help reduce human biases but can also increase errors. For this reason, exploring the more traditional methods that lack empirical evidence is still important. However, the great potential of AI to speed up the evidence synthesis process needs to be considered, especially in the context of RRs.[93,94] AI has the potential to further advance the science of RRs, and the available methods should also be compared to the available tools to further understand their efficiency and limitations.

Concluding Remarks

The results from this thesis elucidate the essential questions that need to be answered about RR methods. Questions around search strategy and study selection should be prioritised when developing methodological studies, followed by quality/bias assessment and data extraction approaches. Our results also showed the impact of database choices in cardiac rehabilitation reviews and the impact of two study selection approaches.

For database combinations, a minimum of three databases would be recommended when developing a meta-analysis, but depending on the review

question, and resources available, one large database may be enough for identifying the most relevant studies. Considering study selection approaches, single review can be an alternative method for title and abstract screening, as it saves resources, but considerations for the potential additional screening time in the full-text phase need to be considered.

The present thesis advances the field by providing a comprehensive analysis of RR methods and some basis for improved practices and methodologies in future research. Thus, further exploration of the most critical questions and advances in understanding the impact of RR methods in different contexts are needed.

REFERENCES

1. Langlois EV, Daniels K. *Evidence Synthesis for Health Policy and Systems: A Methods Guide*. World Health Organization; 2018. Accessed July 19, 2024. https://www.ncbi.nlm.nih.gov/books/NBK569589/

2. Moosapour H, Saeidifard F, Aalaa M, Soltani A, Larijani B. The rationale behind systematic reviews in clinical medicine: a conceptual framework. *J Diabetes Metab Disord*. 2021;20(1):919-929. doi:10.1007/s40200-021-00773-8

3. Higgins J, Thomas J, Chandler J, et al. Cochrane Handbook for Systematic Reviews of Interventions version 6.3. Published online February 2022. www.training.cochrane.org/handbook

4. Kelly SE, Moher D, Clifford TJ. Defining rapid reviews: a modified delphi consensus approach. *Int J Technol Assess Health Care*. 2016;32(4):265-275. doi:10.1017/S0266462316000489

5. Michelson M, Reuter K. The significant cost of systematic reviews and meta-analyses: A call for greater involvement of machine learning to assess the promise of clinical trials. *Contemporary Clinical Trials Communications*. 2019;16:100443. doi:10.1016/j.conctc.2019.100443

6. Haby MM, Chapman E, Clark R, Barreto J, Reveiz L, Lavis JN. What are the best methodologies for rapid reviews of the research evidence for evidence-informed decision making in health policy and practice: a rapid review. *Health Res Policy Sys*. 2016;14(1):83. doi:10.1186/s12961-016-0155-7

7. Griebler U, Dobrescu A, Ledinger D, et al. Evaluation of the interim Cochrane rapid review methods guidance—A mixed-methods study on the understanding of and adherence to the guidance. *Research Synthesis Methods*. 2023;14(6):824-846. doi:10.1002/jrsm.1656

8. Molcak HS, Appleby CJ, Brown J, Freeman S, Kandola DK, Banner D. Rapid Knowledge Syntheses: Methodological and Practical Considerations. *Canadian Journal of Cardiovascular Nursing*. 2021;31(1):6-12.

9. Hamel C, Michaud A, Thuku M, et al. Defining Rapid Reviews: a systematic scoping review and thematic analysis of definitions and defining characteristics of rapid reviews. *Journal of Clinical Epidemiology*. 2021;129:74-85. doi:10.1016/j.jclinepi.2020.09.041

10. Tricco AC, Zarin W, Antony J, et al. An international survey and modified Delphi approach revealed numerous rapid review methods. *Journal of Clinical Epidemiology*. 2016;70:61-67. doi:10.1016/j.jclinepi.2015.08.012

11. Aronson JK, Heneghan C, Mahtani KR, Plüddemann A. A word about evidence: 'rapid reviews' or 'restricted reviews'? *BMJ EBM*. 2018;23(6):204-205. doi:10.1136/bmjebm-2018-111025

12. Tricco AC, Antony J, Zarin W, et al. A scoping review of rapid review methods. *BMC Med*. 2015;13(1):224. doi:10.1186/s12916-015-0465-6

13. Tricco AC, Khalil H, Holly C, et al. Rapid reviews and the methodological rigor of evidence synthesis: a JBI position statement. *JBI Evidence Synthesis*. 2022;20(4):944-949. doi:10.11124/JBIES-21-00371

14. Garritty C, Gartlehner G, Nussbaumer-Streit B, et al. Cochrane Rapid Reviews Methods Group offers evidence-informed guidance to conduct rapid reviews. *Journal of Clinical Epidemiology*. 2021;130:13-22. doi:10.1016/j.jclinepi.2020.10.007

15. Garritty CM, Norris SL, Moher D. Developing WHO rapid advice guidelines in the setting of a public health emergency. *Journal of Clinical Epidemiology*. 2017;82:47-60. doi:10.1016/j.jclinepi.2016.08.010

16. Langlois EV, Straus SE, Antony J, King VJ, Tricco AC. Using rapid reviews to strengthen health policy and systems and progress towards universal health coverage. *BMJ Glob Health*. 2019;4(1):e001178. doi:10.1136/bmjgh-2018-001178

17. Tricco AC, Langlois EV, Straus SE. *Rapid Reviews to Strengthen Health Policy and Systems: A Practical Guide*. World Health Organization; 2017. Accessed November 1, 2022. https://apps.who.int/iris/handle/10665/327019

18. Munn Z, Pollock D, Barker TH, et al. The Dark Side of Rapid Reviews: A Retreat From Systematic Approaches and the Need for Clear Expectations and Reporting. *Ann Intern Med*. 2023;176(2):266-267. doi:10.7326/M22-2603

19. Kaltenthaler E, Cooper K, Pandor A, Martyn-St. James M, Chatters R, Wong R. The use of rapid review methods in health technology assessments: 3 case studies. *BMC Med Res Methodol*. 2016;16(1):108. doi:10.1186/s12874-016-0216-1

20. Featherstone RM, Dryden DM, Foisy M, et al. Advancing knowledge of rapid reviews: an analysis of results, conclusions and recommendations from published review articles examining rapid reviews. *Syst Rev*. 2015;4(1):50. doi:10.1186/s13643-015-0040-4

21. Khangura S, Konnyu K, Cushman R, Grimshaw J, Moher D. Evidence summaries: the evolution of a rapid review approach. *Syst Rev*. 2012;1(1):10. doi:10.1186/2046-4053-1-10

22. Haby MM, Barreto JOM, Kim JYH, et al. What are the best methods for rapid reviews of the research evidence? A systematic review of reviews and primary studies. *Research Synthesis Methods*. Published online September 11, 2023:jrsm.1664. doi:10.1002/jrsm.1664

23. Wilson MG, Oliver S, Melendez-Torres GJ, Lavis JN, Waddell K, Dickson K. Paper 3: Selecting rapid review methods for complex questions related to health policy and system issues. *Syst Rev*. 2021;10(1):286. doi:10.1186/s13643-021-01834-y

24. Kelly SE, Moher D, Clifford TJ. Quality of conduct and reporting in rapid reviews: an exploration of compliance with PRISMA and AMSTAR guidelines. *Syst Rev*. 2016;5(1):79. doi:10.1186/s13643-016-0258-9

25. Waffenschmidt S, Knelangen M, Sieben W, Bühn S, Pieper D. Single screening versus conventional double screening for study selection in systematic reviews: a methodological systematic review. *BMC Med Res Methodol*. 2019;19(1):132. doi:10.1186/s12874-019-0782-0

26. Gartlehner G, Affengruber L, Titscher V, et al. Single-reviewer abstract screening missed 13 percent of relevant studies: a crowd-based, randomized controlled trial. *Journal of Clinical Epidemiology*. 2020;121:20-28. doi:10.1016/j.jclinepi.2020.01.005

27. Beecher C, Toomey E, Maeso B, et al. Priority III: top 10 rapid review methodology research priorities identified using a James Lind Alliance Priority Setting Partnership. *Journal of Clinical Epidemiology*. 2022;151:151-160. doi:10.1016/j.jclinepi.2022.08.002

28. Liberati A, Altman DG, Tetzlaff J, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. *BMJ*. 2009;339(jul21 1):b2700-b2700. doi:10.1136/bmj.b2700

29. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*. Published online March 29, 2021:n71. doi:10.1136/bmj.n71

30. Petticrew M, Roberts H. *Systematic Reviews in the Social Sciences : A Practical Guide*. 1st edition. Blackwell Pub; 2006.

31. Khangura S, Polisena J, Clifford TJ, Farrah K, Kamel C. Rapid review: an emerging approach to evidence synthesis in health technology assessment. *Int J Technol Assess Health Care*. 2014;30:20-27. doi:10.1017/S0266462313000664

32. Watt A, Cameron A, Sturm L, et al. Rapid reviews versus full systematic reviews: An inventory of current methods and practice in health technology assessment. *Int J Technol Assess Health Care*. 2008;24(02):133-139. doi:10.1017/S0266462308080185

33. Akl EA, Haddaway NR, Rada G, Lotfi T. Future of Evidence Ecosystem Series: Evidence synthesis 2.0: when systematic, scoping, rapid, living, and overviews of reviews come together. *Journal of Clinical Epidemiology*. 2020;123:162-165. doi:10.1016/j.jclinepi.2020.01.025

34. Tricco AC, Garritty CM, Boulos L, et al. Rapid review methods more challenging during COVID-19: commentary with a focus on 8 knowledge synthesis steps. *Journal of Clinical Epidemiology*. 2020;126:177-183. doi:10.1016/j.jclinepi.2020.06.029

35. McGregor M, Brophy JM. End-user involvement in health technology assessment (HTA) development: A way to increase impact. *Int J Technol Assess Health Care*. 2005;21(2):263-267. doi:10.1017/S026646230505035X

36. Zechmeister I, Schumacher I. The impact of health technology assessment reports on decision making in Austria. *Int J Technol Assess Health Care*. 2012;28(1):77-84. doi:10.1017/S0266462311000729

37. Speckemeier C, Niemann A, Wasem J, Buchberger B, Neusser S. Methodological guidance for rapid reviews in healthcare: A scoping review. *Research Synthesis Methods*. 2022;13(4):394-404. doi:10.1002/jrsm.1555

38. Hsu CC, Sandford BA. The Delphi Technique: Making Sense of Consensus. doi:10.7275/PDZ9-TH90

39. Donohoe H, Stellefson M, Tennant B. Advantages and Limitations of the e-Delphi Technique: Implications for Health Education Researchers. *American Journal of Health Education*. 2012;43(1):38-46. doi:10.1080/19325037.2012.10599216

40. McMillan SS, King M, Tully MP. How to use the nominal group and Delphi techniques. *Int J Clin Pharm*. Published online February 5, 2016. doi:10.1007/s11096-016-0257-x

41. Jünger S, Payne SA, Brine J, Radbruch L, Brearley SG. Guidance on Conducting and REporting DElphi Studies (CREDES) in palliative care: Recommendations based on a methodological systematic review. *Palliat Med*. 2017;31(8):684-706. doi:10.1177/0269216317690685

42. Byrne M, O'Connell A, Egan AM, et al. A core outcomes set for clinical trials of interventions for young adults with type 1 diabetes: an international, multi-perspective Delphi consensus study. *Trials*. 2017;18(1):602. doi:10.1186/s13063-017-2364-y

43. Fink-Hafner D, Dagen T, Doušak M, Novak M, Hafner-Fink M. Delphi method: Strengths and weaknesses. *Adv Meth Stat*. 2019;16(2). doi:10.51936/fcfm6982

44. Jones J, Hunter D. Qualitative Research: Consensus methods for medical and health services research. *BMJ*. 1995;311(7001):376-380. doi:10.1136/bmj.311.7001.376

45. Murphy M, Black N, Lamping D, et al. Consensus development methods, and their use in clinical guideline development. *Health Technology Assessment*. 1998;2.

46. Sinha IP, Smyth RL, Williamson PR. Using the Delphi technique to determine which outcomes to measure in clinical trials: recommendations for the future based on a systematic review of existing studies. *PLoS Medicine*. 2011;8:e1000393. doi:10.1371/journal.pmed.1000393

47. Dragomir AI, Boucher VG, Bacon SL, et al. An international Delphi consensus study to define motivational communication in the context of developing a training program for physicians. *Translational Behavioral Medicine*. 2021;11(2):642-652. doi:10.1093/tbm/ibaa015

48. von der Gracht HA. Consensus measurement in Delphi studies. *Technological Forecasting and Social Change*. 2012;79(8):1525-1536. doi:10.1016/j.techfore.2012.04.013

49. Moher D, Schulz KF, Simera I, Altman DG. Guidance for Developers of Health Research Reporting Guidelines. *PLoS Med*. 2010;7(2):e1000217. doi:10.1371/journal.pmed.1000217

50. Cantrill JA, Sibbald B, Buetow S. The Delphi and nominal group techniques in health services research. *International Journal of Pharmacy Practice*. 2011;4(2):67-74. doi:10.1111/j.2042-7174.1996.tb00844.x

51. O'Cathain A, Murphy E, Nicholl J. Three techniques for integrating data in mixed methods studies. *BMJ*. 2010;341:c4587. doi:https://doi.org/10.1136/bmj.c4587

52. Atkins L, Francis J, Islam R, et al. A guide to using the Theoretical Domains Framework of behaviour change to investigate implementation problems. *Implementation Sci*. 2017;12(1):77. doi:10.1186/s13012-017-0605-9

53. Boutron I, Moher D, Altman DG, Schulz KF, Ravaud P, for the CONSORT Group. Methods and Processes of the CONSORT Group: Example of an Extension for Trials Assessing Nonpharmacologic Treatments. *Ann Intern Med*. 2008;148(4):W-60. doi:10.7326/0003-4819-148-4-200802190-00008-w1

54. Smela B, Toumi M, Świerk K, et al. Rapid literature review: definition and methodology. *Journal of Market Access & Health Policy*. 2023;11(1):2241234. doi:10.1080/20016689.2023.2241234

55. Hartling L, Guise JM, Kato E, et al. *EPC Methods: An Exploration of Methods and Context for the Production of Rapid Reviews*. Agency for Healthcare Research and Quality; 2015.

56. King VJ, Stevens A, Nussbaumer-Streit B, Kamel C, Garritty C. Paper 2: Performing rapid reviews. *Syst Rev*. 2022;11(1):151. doi:10.1186/s13643-022-02011-5

57. Garritty C, Hamel C, Trivella M, et al. Updated recommendations for the Cochrane rapid review methods guidance for rapid reviews of effectiveness. *BMJ*. Published online February 6, 2024:e076335. doi:10.1136/bmj-2023-076335

58. Vieira AM, Szczepanik G, De Waure C, et al. Identifying priority questions regarding rapid systematic reviews' methods: protocol for an eDelphi study. *BMJ Open*. 2023;13(7):e069856. doi:10.1136/bmjopen-2022-069856

59. Egger M, Jüni P, Bartlett C, Holenstein F, Sterne J. How important are comprehensive literature searches and the assessment of trial quality in systematic reviews? Empirical study. *Health Technol Assess*. 2003;7(1):1-82. doi:10.3310/hta7010

60. Justesen T, Freyberg J, Schultz ANØ. Database selection and data gathering methods in systematic reviews of qualitative research regarding diabetes mellitus

- an explorative study. *BMC Med Res Methodol*. 2021;21(1):94. doi:10.1186/s12874-021-01281-2

61. Ewald H, Klerings I, Wagner G, et al. Abbreviated and comprehensive literature searches led to identical or very similar effect estimates: a meta-epidemiological study. *Journal of Clinical Epidemiology*. 2020;128:1-12. doi:10.1016/j.jclinepi.2020.08.002

62. Nussbaumer-Streit B, Klerings I, Wagner G, et al. Abbreviated literature searches were viable alternatives to comprehensive searches: a meta-epidemiological study. *Journal of Clinical Epidemiology*. 2018;102:1-11. doi:10.1016/j.jclinepi.2018.05.022

63. Paez A. Gray literature: An important resource in systematic reviews. *J Evidence Based Medicine*. 2017;10(3):233-240. doi:10.1111/jebm.12266

64. Saleh AA, Ratajeski MA, Bertolet M. Grey Literature Searching for Health Sciences Systematic Reviews: A Prospective Study of Time Spent and Resources Utilized. *EBLIP*. 2014;9(3):28-50. doi:10.18438/B8DW3K

65. Pandor A, Kaltenthaler E, Martyn-St James M, et al. Delphi consensus reached to produce a decision tool for SelecTing Approaches for Rapid Reviews (STARR). *Journal of Clinical Epidemiology*. 2019;114:22-29. doi:10.1016/j.jclinepi.2019.06.005

66. Tricco AC, Garritty CM, Boulos L, et al. Rapid review methods more challenging during COVID-19: commentary with a focus on 8 knowledge synthesis steps. *Journal of Clinical Epidemiology*. 2020;126:177-183. doi:10.1016/j.jclinepi.2020.06.029

67. Tricco AC, Khalil H, Holly C, et al. Rapid reviews and the methodological rigor of evidence synthesis: a JBI position statement. *JBI Evidence Synthesis*. 2022;20(4):944-949. doi:10.11124/JBIES-21-00371

68. Frandsen TF, Eriksen MB, Hammer DMG, Christensen JB. PubMed coverage varied across specialties and over time: a large-scale study of included studies in Cochrane reviews. *Journal of Clinical Epidemiology*. 2019;112:59-66. doi:10.1016/j.jclinepi.2019.04.015

69. Rice M, Ali MU, Fitzpatrick-Lewis D, Kenny M, Raina P, Sherifali D. Testing the effectiveness of simplified search strategies for updating systematic reviews. *Journal of Clinical Epidemiology*. 2017;88:148-153. doi:10.1016/j.jclinepi.2017.06.005

70. Halladay CW, Trikalinos TA, Schmid IT, Schmid CH, Dahabreh IJ. Using data sources beyond PubMed has a modest impact on the results of systematic reviews of therapeutic interventions. *Journal of Clinical Epidemiology*. 2015;68(9):1076-1084. doi:10.1016/j.jclinepi.2014.12.017

71. Hartling L, Featherstone R, Nuspl M, Shave K, Dryden DM, Vandermeer B. The contribution of databases to the results of systematic reviews: a cross-sectional study. *BMC Med Res Methodol*. 2016;16(1):127. doi:10.1186/s12874-016-0232-1

72. Hirt J, Bergmann J, Karrer M. Overlaps of multiple database retrieval and citation tracking in dementia care research: a methodological study. *jmla*. 2021;109(2). doi:10.5195/jmla.2021.1129

73. Goossen K, Hess S, Lunny C, Pieper D. Database combinations to retrieve systematic reviews in overviews of reviews: a methodological study. *BMC Med Res Methodol*. 2020;20(1):138. doi:10.1186/s12874-020-00983-3

74. Tricco AC, Langlois EtienneV, Straus SE, Alliance for Health Policy and Systems Research, World Health Organization. *Rapid Reviews to Strengthen Health Policy and Systems: A Practical Guide*. World Health Organization; 2017. Accessed May 29, 2023. https://apps.who.int/iris/handle/10665/258698

75. Dobbins M. *Rapid Review Guidebook*. Vol 1.3. 1.3. (McMaster U, ed.). https://www.nccmt.ca/uploads/media/media/0001/02/800fe34eaedbad09edf80ad5081b9291acf1c0c2.pdf

76. Ng L, Pitt V, Huckvale K, et al. Title and Abstract Screening and Evaluation in Systematic Reviews (TASER): a pilot randomised controlled trial of title and abstract screening by medical students. *Syst Rev*. 2014;3(1):121. doi:10.1186/2046-4053-3-121

77. Cooper M, Ungar W, Zlotkin S. An assessment of inter-rater agreement of the literature filtering process in the development of evidence-based dietary guidelines. *Public Health Nutr*. 2006;9(4):494-500. doi:10.1079/PHN2005877

78. Edwards P, Clarke M, DiGuiseppi C, Pratap S, Roberts I, Wentz R. Identification of randomized controlled trials in systematic reviews: accuracy and reliability of screening records. *Statistics in Medicine*. 2002;21(11):1635-1640. doi:10.1002/sim.1190

79. Doust JA, Pietrzak E, Sanders S, Glasziou PP. Identifying studies for systematic reviews of diagnostic tests was difficult due to the poor sensitivity and precision of methodologic filters and the lack of information in the abstract. *Journal of Clinical Epidemiology*. 2005;58(5):444-449. doi:10.1016/j.jclinepi.2004.09.011

80. Pham MT, Waddell L, Rajić A, Sargeant JM, Papadopoulos A, McEwen SA. Implications of applying methodological shortcuts to expedite systematic reviews: three case studies using systematic reviews from agri-food public health. *Research Synthesis Methods*. 2016;7(4):433-446. doi:10.1002/jrsm.1215

81. Bacon SL, Wu N, Joyal-Desmarais K, et al. *COVID-19 Living Evidence Synthesis 13.2a: Effectiveness of Isolation on the Reduction of the Transmission of Respiratory Infectious Diseases (RIDs: I.e., COVID-19, H1N1, SARS, MERS)*. Montreal Behavioural Medicine Centre, CIUSSS-NIM; 2024.

82. Bacon SL, Wu N, Joyal-Desmarais K, et al. *COVID-19 Living Evidence Synthesis 13.2b: Unintended Health and Social Consequences of Isolation and Quarantine for Respiratory Infectious Diseases (RIDs: I.e., COVID-19, H1N1, SARS, and MERS).* Montreal Behavioural Medicine Centre, CIUSSS-NIM; 2024.

83. Bacon SL, Wu N, Joyal-Desmarais K, et al. *COVID-19 Living Evidence Synthesis 13.2c: Effectiveness of Quarantine on the Reduction of the Transmission of Respiratory Infectious Diseases (RIDs: I.e., COVID-19, H1N1, SARS, and MERS)*. Montreal Behavioural Medicine Centre, CIUSSS-NIM; 2024.

84. Sim J, Wright CC. The Kappa Statistic in Reliability Studies: Use, Interpretation, and Sample Size Requirements. *Physical Therapy*. 2005;85(3):257-268. doi:10.1093/ptj/85.3.257

85. Viera AJ, Garrett JM. Understanding Interobserver Agreement: The Kappa Statistic. *Family Medicine*.

86. Ganann R, Ciliska D, Thomas H. Expediting systematic reviews: methods and implications of rapid reviews. *Implementation Science*. 2010;5:56.

87. Abrami PC, Borokhovski E, Bernard RM, et al. Issues in conducting and disseminating brief reviews of evidence. *Evidence & Policy*. 2010;6(3):371-389. doi:10.1332/174426410X524866

88. Moons P, Goossens E, Thompson DR. Rapid reviews: the pros and cons of an accelerated review process. *European Journal of Cardiovascular Nursing*. 2021;20(5):515-519. doi:10.1093/eurjcn/zvab041

89. Watt A, Cameron A, Sturm L, et al. Rapid versus full systematic reviews: validity in clinical practice? *ANZ Journal of Surgery*. 2008;78(11):1037-1040. doi:10.1111/j.1445-2197.2008.04730.x

90. Mbuagbaw L, Lawson DO, Puljak L, Allison DB, Thabane L. A tutorial on methodological studies: the what, when, how and why. *BMC Med Res Methodol*. 2020;20(1):226. doi:10.1186/s12874-020-01107-7

91. Domecq JP, Prutsky G, Elraiyah T, et al. Patient engagement in research: a systematic review. *BMC Health Serv Res*. 2014;14(1):89. doi:10.1186/1472-6963-14-89

92. Wiles LK, Kay D, Luker JA, et al. Consumer engagement in health care policy, research and services: A systematic review and meta-analysis of methods and effects. Gholipour K, ed. *PLoS ONE*. 2022;17(1):e0261808. doi:10.1371/journal.pone.0261808

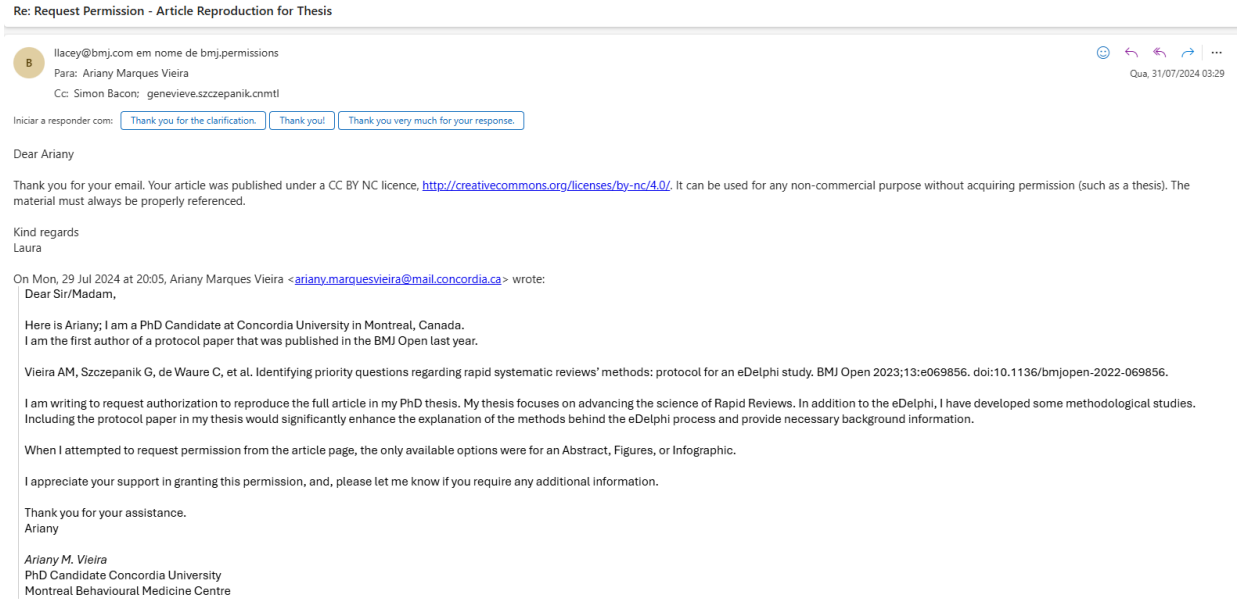
93. Fabiano N, Gupta A, Bhambra N, et al. How to optimize the systematic review process using AI tools. *JCPP Advances*. 2024;4(2):e12234. doi:10.1002/jcv2.12234

94. Malik VS, Singh M. Machine Learning and Artificial Intelligence in Evidence Generation and Evidence Synthesis. *Journal of Medical Evidence*. Published online April 24, 2024. doi:10.4103/JME.JME_95_23

APPENDIX I: Chapter 2

Identifying priority questions regarding rapid systematic reviews' methods: protocol for an eDelphi study

Permission and license to reproduce the manuscript



Re: Request Permission - Article Reproduction for Thesis

B llacey@bmj.com em nome de bmj.permissions
Para: Ariany Marques Vieira
Cc: Simon Bacon; genevieve.szczepanik.cnmtl
Qua, 31/07/2024 03:29

Iniciar a responder com: Thank you for the clarification. | Thank you! | Thank you very much for your response.

Dear Ariany

Thank you for your email. Your article was published under a CC BY NC licence, http://creativecommons.org/licenses/by-nc/4.0/. It can be used for any non-commercial purpose without acquiring permission (such as a thesis). The material must always be properly referenced.

Kind regards
Laura

On Mon, 29 Jul 2024 at 20:05, Ariany Marques Vieira <ariany.marquesvieira@mail.concordia.ca> wrote:

> Dear Sir/Madam,
>
> Here is Ariany; I am a PhD Candidate at Concordia University in Montreal, Canada.
> I am the first author of a protocol paper that was published in the BMJ Open last year.
>
> Vieira AM, Szczepanik G, de Waure C, et al. Identifying priority questions regarding rapid systematic reviews' methods: protocol for an eDelphi study. BMJ Open 2023;13:e069856. doi:10.1136/bmjopen-2022-069856.
>
> I am writing to request authorization to reproduce the full article in my PhD thesis. My thesis focuses on advancing the science of Rapid Reviews. In addition to the eDelphi, I have developed some methodological studies. Including the protocol paper in my thesis would significantly enhance the explanation of the methods behind the eDelphi process and provide necessary background information.
>
> When I attempted to request permission from the article page, the only available options were for an Abstract, Figures, or Infographic.
>
> I appreciate your support in granting this permission, and, please let me know if you require any additional information.
>
> Thank you for your assistance.
> Ariany
>
> *Ariany M. Vieira*
> PhD Candidate Concordia University
> Montreal Behavioural Medicine Centre

Ethics Certification

CERTIFICATION OF ETHICAL ACCEPTABILITY FOR RESEARCH INVOLVING HUMAN SUBJECTS

Name of Applicant:	Dr. Simon Bacon
Department:	Faculty of Arts and Science\Health, Kinesiology and Applied Physiology
Agency:	Canadian Diabetes Association Canadian Institutes of Health Research
Title of Project:	Identifying priority questions regarding rapid reviews methodology: an eDelphi study
Certification Number:	30015229

Valid From: May 02, 2022 To: May 01, 2023

The members of the University Human Research Ethics Committee have examined the application for a grant to support the above-named project, and consider the experimental procedures, as outlined by the applicant, to be acceptable on ethical grounds for research involving human subjects.

Dr. Richard DeMont, Chair, University Human Research Ethics Committee

Sociodemographic Information Questions

This project aims to include responses from a wide range of people, including people with a variety of backgrounds considered experts in evidence-synthesis. For that, we would like to ask you for some general information about you. Your answers will be confidential, and no individual will be identified when the results are presented. Your contact is requested to send you the next rounds of the survey. This project aims to include responses from a wide range of people, including people with a variety of backgrounds providing valuable expertise in evidence-synthesis. To this end, we would like to ask questions about your personal background. Your answers will be confidential, and no individual will be identified when the results are presented. Your contact information is only requested to send you the next rounds of the survey.

1. In which age group do you better fit?

66 years or more
56-65 years
46-55 years
36-45 years
26-35 years
18-25 years
Less than 18 years
Prefer not to answer

2. With which sex do you most strongly identify?

Female
Male
Prefer not to answer
Other

3. What is your job title?

This information will help to understand the profile of the participants. You can write in a few words your current position. For example, Graduate student, Research Assistant, Managing director.

4. In which country do you currently work?

This question will help to understand the demographics of the participants. You can write the name of the country where you hold a position. For example: Canada, Australia, Nigeria.

5. In which city do you currently work?

6. In what field/area or research do you predominantly perform your evidence syntheses (please select all that apply)?

Evidence syntheses are studies developed to gather evidence available to answer a specific question. This includes systematic reviews, scoping reviews, and rapid reviews, for example.

Clinical
Public Health
Health system
Prefer not to answer
Other

7. What is your role in evidence synthesis (lead reviewer, coordinator, field expert, contributor to study selection and data extraction, responsible for results interpretation,…) ?

Evidence syntheses are studies developed to gather evidence available to answer a specific question. This includes systematic reviews, scoping reviews, and rapid reviews, for example.

Glossary of terms/List of definitions

Data analysis is the process of taking data and turning it into a useful material to answer a research question. There are different methods, such as qualitative and quantitative approaches.

Data abstraction/extraction is related to the act of separating, withdrawing, and taking data of interest from included studies or different sources. Usually, information about study characteristics, descriptive data, and findings (outcome data) are part of data extraction (Munn *et al*., 2014).

Efficiency is the ability to perform something well, successfully, and without waste (e.g. time, money). Balance between quality and resource consumption.

Evidence synthesis is a type of study developed to gather available evidence to answer a specific question. This includes SRs, scoping reviews, living reviews, overview of reviews and RRs for example.

Grey literature is materials and research produced outside of the traditional commercial or academic publishing and distribution channels. Common grey literature publication types include pre-prints, reports, working papers, government documents, white papers and evaluation (Simon Fraser Library, accessed in 2022).

Interested and affected parties/end users (previous referred to as stakeholders): the parties who will engage in, benefit from or be affected by the procedure (Tricco AC, et al. WHO Practical Guide, 2017). For this study, end users of a rapid review process include decision-makers, guideline and policy developers, healthcare providers, health system managers, end-users (public and patients), and journal editors.

Methods: Research methods are particular processes for collecting and analyzing data. For evidence syntheses, it usually covers the methods for: acquisition of evidence (search strategy, inclusion criteria, selection process), data extraction, data

analysis, data appraisal/risk of bias/quality assessment strategy, and data synthesis process.

Rapid systematic reviews (RRs) are another evidence synthesis method that accelerates the process of conducting a traditional systematic review through streamlining or omitting a variety of methods to produce evidence in a resource-efficient manner (Hamel *et. al.*, 2021). The kinds of methods that this study will include are: search strategy, studies selection (level one and two of the screening), data extraction, risk of bias appraisal and data analysis. It is also referred in this project as **Rapid Reviews**.

Report: ***"****A document (paper or electronic) supplying information about a particular study. It could be a journal article, preprint, conference abstract, study register entry, clinical study report, dissertation, unpublished manuscript, government report, or any other document providing relevant information"* (Page *et al*., 2021).

Risk of bias appraisal/assessment: "*The purpose of study quality assessment is to capture and analyze variations among the included studies—those that met initial inclusion criteria— in terms of their credibility and vulnerability to various sources of bias"* (Littell *et al.*, 2008, Chapter 4).

Screening is part of the studies selection process for a review, checking if the references fit or not the inclusion criteria. It includes different levels, such as Title and Abstract and Full text screening.

Search Strategy, in the context of evidence syntheses, is the structured plan of how to find studies of interest. The search strategy includes the terms that are going to be used and also the sources that will be consulted (e.g. databases, repositories).

Synthesis: In the context of evidence syntheses, the synthesis is the summarization of the data that were collected. "*In systematic reviews of quantitative (numerical) data, data synthesis usually appears as a meta-analysis, a statistical method that*

combines the results of a number of studies to calculate a single summary effect" (Munn *et al*., 2014).

Systematic reviews (SRs) are the most common type of evidence synthesis. It is a way of searching, selecting, appraising, and synthesising the available evidence to answer a research question. It organises all empirical evidence that fits in pre-specified eligibility criteria and aim to reduce bias (Higgins *et. al.*, 2022).

References

Hamel C, et al. Defining Rapid Reviews: a systematic scoping review and thematic analysis of definitions and defining characteristics of rapid reviews. Journal of Clinical Epidemiology 129 (2021) 74e85.

Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA (editors). Cochrane Handbook for Systematic Reviews of Interventions version 6.3 (updated February 2022). Cochrane, 2022. Available from www.training.cochrane.org/handbook

Littell JH, Corcoran J, Pillai V. Systematic Reviews and Meta-Analysis. Published to Oxford Scholarship Online: January 2009. DOI: 10.1093/acprof:oso/9780195326543.001.0001

Munn Z, Tufanaru C, Aromataris E. Data Extraction and Synthesis: The steps following study selection in a systematic review. AJN 2014 vol. 114 (7).

Page MJ, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021;372:n71. DOI: 10.1136/bmj.n71

Simon Fraser Library, accessed in 2022. https://www.lib.sfu.ca/help/research-assistance/format-type/grey-literature

Tricco AC, Langlois EV, Straus SE, editors. Rapid reviews to strengthen health policy and systems: a practical guide. Geneva: World Health Organization; 2017. Licence: CC BY-NC-SA 3.0 IGO.

APPENDIX II: Chapter 2

Essential methodological questions for developing time-efficient rapid reviews: results from an eDelphi study

Section I: Summary of the Central Scientific Committee roles and expertise

The Central Scientific Committee (CSC) was formed initially with twelve members, with one member resigning due to other commitments after the start of the project, leaving a final group of eleven. The CSC included members with a variety of expertise, including pioneering researchers in the field of evidence synthesis, members who have been part of multiple evidence synthesis teams, and policymakers. Detailed information is presented in Table S1. This group was responsible for reviewing, editing, and agreeing with the first list of methodological items, sharing the survey with all their networks, providing feedback on the survey structure and project plan, and providing feedback on the results of each survey round (agreeing on the items that participants may suggest, dropping off items). They were also invited to participate in the consensus meeting and had the opportunity to vote on the inclusion and exclusion of items and discuss items' wording.

Table S1. Members of the Central Scientific Committee

Central Scientific Committee (in alphabetical order)		
Andrea Tricco	Lead of the Strategy for Patient-Oriented Research (SPOR) Evidence Alliance	Canada
Ariany Marques Vieira *(Chair of the Committee)*	Concordia University / MBMC MBMC Meta Group – Evidence synthesis team	Canada
Chiara de Waure	University of Perugia/ Chair HTA section of EUPHA	Italy
Danielle Pollock	HESRI (Health Evidence Synthesis Recommendations and Impact); University of Adelaide	Australia
Elie Akl	Clinical Research Institute, AUB GRADE center, and SPARK, American University of Beirut (AUB)	Lebanon
John Lavis	Lead of the McMaster Health Forum and Department of Health Research Methods, Evidence and Impact, McMaster University;	Canada

	Institute for Better Health, Trillium Health Partners	
Jovana Stojanovic	Canadian Agency for Drugs and Technologies in Health	Canada
Laurenz Langer	Africa Centre for Evidence (UJ)	South Africa
Peter Bragge	Monash Sustainable Development Institute Evidence Review Service, Monash University, Melbourne Australia	Australia
Sandy Oliver	EPPI-Centre, University College London. Africa Centre of Evidence, University of Johannesburg	United Kingdom
Simon L Bacon, PhD	Concordia University / Co-director MBMC Lead MBMC Meta Group – Evidence synthesis team	Canada
Tanja Kuchenmuller*	WHO Regional Office for Europe	Switzerland

*This member resigned participation after ethical approval and after the first eDelphi round due to other commitments.

Section II: First list of items

List of items: **Questions on rapid reviews methods**

This list was created based on the following references:

- Results of the Priority III Study: Setting priorities for rapid review research. https://evidencesynthesisireland.ie/priority-iii/
- Tricco, AC; et al. An international survey and modified Delphi approach revealed numerous rapid review methods. *Journal of Clinical Epidemiology* 2015; 70:61-7.
- Tricco AC, Langlois EV, Straus SE, editors. *Rapid reviews to strengthen health policy and systems: a practical guide.* Geneva: World Health Organization; 2017. Licence: CC BY-NC-SA 3.0 IGO.

Conceptual/high level questions

1. What is the optimal composition of a rapid review team (e.g., number of team members, expertise, stakeholders, etc.)...?
2. What is the optimal way of registering a rapid review or making the protocol available on a freely accessible platform (such as PROSPERO or Open Science Framework)...?
3. What is the optimal time frame that a rapid review should take (from search strategy to data analysis)...?
4. What is the optimal use of guidelines (including checklists and templates)...?

5. What are the optimal methods for involving stakeholders...?
6. When updating a completed rapid review, what is the optimal level of consistency in the reviewers (e.g., the same reviewers vs. different reviewers)...?

Search Strategy

7. What is the optimal number of databases used for the search strategy...?
8. What are the optimal core set of databases that should be searched...?
9. What are the optimal non-peer-reviewed publication databases (i.e., pre-prints) that should be included...?
10. What are the optimal kinds of grey literature (other than pre-prints) that should be included...?
11. What are the optimal approaches for developing search terms for the key elements in the rapid review...?
12. What are the optimal restrictions and search limits (e.g., years of inclusion, language, phase of study, study design)...?

Studies selection and data extraction

13. What are the optimal approaches for determining the inclusion and exclusion criteria for studies...?
14. What is the optimal balance between automated and manual screening...?
15. What are the optimal automated processes (e.g., machine learning algorithms, specific SR programs, etc.) for screening...?
16. What is the optimal way to leverage online collaboration platforms (e.g., Cochrane Task Exchange) when doing screening...?
17. If using peer-reviewing, what is the optimal method to perform screening (e.g., independent screening, partial peer review, etc.)...?
18. What is the optimal number of people needed to perform screening...?

Data Extraction

19. What is the optimal balance between automated and manual data extraction techniques...?
20. If using peer-reviewing, what is the optimal approach to perform data extraction (e.g. independent extraction and comparison of discrepancies, one reviewer extracting and the other double checking)...?
21. What is the optimal number of people needed to perform data extraction...?

Quality/bias assessment

22. What are the optimal approaches to assess the quality of included studies and/or risk of bias...?
23. What are the optimal dimensions of quality (e.g., trustworthiness, relevance) that should be considered when appraising studies...?
24. What is the optimal usage of quality assessment results (e.g., to further exclude studies, to allow a more tailored synthesis)...?

25. What is the optimal balance between automated and manual quality assessments...?
26. What is the optimal number of people needed to perform quality assessment...?

Synthesis

27. What are the optimal synthesis methods to use (e.g., narrative, thematic, meta-analysis)...?
28. What are the optimal methods for assessing the strength of individual synthesis findings and recommendations (e.g., GRADE)...?
29. What factors need to be used when choosing how to disaggregate synthesis findings for priority populations (e.g., PROGRESS-Plus groups)...?

Section III: Characteristics of the eDelphi participants

Table S2. Characteristics of all survey respondents

		NUMBER OF RESPONDENTS	PARTICIPANTS WITH COMPLETED ANSWER N (%)	NUMBER OF RESPONDENTS	PARTICIPANTS WITH PARTIAL ANSWERS N (%)
Profile					
Healthcare Practitioner		52	11 (14.1%)	50	4 (8%)
Researcher		52	60 (76.9%)	50	45 (90%)
Policymaker		52	7 (8.9%)	50	0
Patient / community member / caregiver		52	0	50	1 (2%)
Sociodemographic Information					
Age Group	18-25 years	52	1 (1.2%)	9	0
	26 – 35 years	52	16 (20.5%)	9	3 (33.3%)
	36 – 45 years		31 (39.7%)		
	46 – 55 years	52	17 (21.7%)	9	1 (11.1%)
	56 – 65 years		12 (15.3%)		
	66 years or more	52	1 (1.2%)	9	5 (55.5%)
		52		9	0
		52		9	0
Sex	Female	52	47 (60.2%)	9	5 (55.5%)
	Male	52	28 (35.8%)	9	4 (44.44%)
	Prefer not to answer	52	3 (3.8%)	9	0
Country of work	Argentina	52	2 (2.56%)	9	0
	Australia	52	6 (7.69%)	9	0
	Austria	52	1 (1.28%)	9	0
	Belgium	52	1 (1.28%)	9	0
	Brazil	52	3 (3.85%)	9	0

	Canada	52	26 (33.33%)	9	6 (66.6%)
	Canada/Colo	52	1 (1.28%)	9	0
	mbia	52	3 (3.85%)	9	0
	Colombia	52	1 (1.28%)	9	0
	Ethiopia	52	1 (1.28%)	9	0
	Greece	52	2 (2.56%)	9	0
	India	52	1 (1.28%)	9	0
	India/New	52	1 (1.28%)	9	0
	Zealand	52	3 (3.85%)	9	0
	Ireland	52	1 (1.28%)	9	0
	Italy	52	1 (1.28%)	9	1 (11.1%)
	Kyrgyzstan	52	1 (1.28%)	9	0
	Lebanon	52	2 (2.56%)	9	0
	Mexico	52	1 (1.28)	9	0
	Romania	52	2 (2.56)	9	0
	Slovenia	52	3 (3.85%)	9	0
	South Africa	52	2 (2.56%)	9	0
	Spain	52	1 (1.28%)	9	0
	Switzerland	52	6 (7.69%)	9	0
	Turkey	52	6 (7.69%)	9	0
	United Kingdom	52		9	2 (22.2%)
	United States	52		9	0
Experience					
Years of experience with evidence syntheses	None	52	0	51	2 (4%)
	≤ 4 years	52	3 (3.8%)	51	21 (41.2%)
	5-6 years	52	16 (20.5%)	51	12 (23.5%)
	7-8 years	52	9 (11.5%)	51	2 (4%)
	9-10 years	52	8 (10.2%)	51	4 (8%)
	11-12 years	52	8 (10.2%)	51	1 (2%)
	13-14 years	52	7 (8.9%)	51	1 (2%)
	≥ 15 years	52	27 (34.6%)	51	8 (15.6%)
Aspects of evidence synthesis that previously participated	Conceptualization/Research question development	52	78 (100%)	56	42 (75%)
	Undertaking literature searches	52	69 (88.4%)	56	41 (73.2%)
	Study screening and selection	52	71 (91%)	56	41 (73.2%)
	Data extraction	52	71 (91%)	56	42 (75%)
	Quality appraisal	52	66 (84.6%)	56	38 (67.8%)
	Data synthesis	52	70 (89.7%)	56	34 (60.7%)
	Interpretation of results	52	72 (92.3%)	56	38 (67.8%)
	Knowledge translation	52	63 (80.7%)	56	32 (57.14%)
	Other	52	9 (11.5%)	56	4 (7.14%
Knowledge in evidence syntheses (0 = no expertise to 10 = very strong expertise)	*(Mean, SD)*	52	8.3 (±1.43)	50	7.2 (±2.12)
	Clinical	52	39 (50%)	9	4 (44.5%)

Fields/areas of research predominantly perform evidence syntheses	Public Health	52	54 (69.2%)	9	7 (77.8%)
	Health Systems	52	45 (57.6%)	9	4 (44.5%)
	Other	52	15 (19.2%)	9	1 (11.1%)

Section IV: Detailed Results

The first eDelphi round was launched in June 2022 and the last round closed in January 2023. The specific dates and duration of each round is listed int the table below (table S3).

Table S3. Date and duration of each eDelphi survey round

	START DATE	END DATE	DURATION
Round 1	02/06/2022	13/07/2022	6 weeks
Round 2	20/10/2022	09/11/2023	3 weeks
Round 3	06/12/2022	01/01/2023	4 weeks

In total, 78 participants answered one of the three rounds. From those, 41 participants were consistent and answered all the three rounds. In Table S4, you can find the characteristics of participants. Although only 76 participants answered the question on where they currently work, the sample represent all continents.

Table S4. Characteristics of participants

		OVERALL PARTICIPANTS	CONSISTENT PARTICIPANTS
		n (%)	n (%)
Profile			
Healthcare Practitioner		11 (14.1%)	5 (12.1%)
Researcher		60 (76.9%)	33 (80.4%)
Policymaker		7 (8.9%)	3 (7.3%)
Sociodemographic Information			
Age Group	18-25 years	1 (1.2%)	0
	26 – 35 years	16 (20.5%)	8 (19.5%)
	36 – 45 years	31 (39.7%)	14 (34.1%)
	46 – 55 years	17 (21.7%)	10 (24.3%)
	56 – 65 years	12 (15.3%)	8 (19.5%)
	66 years or more	1 (1.2%)	1 (2.4%)
Sex	Female	47 (60.2%)	23 (56%)
	Male	28 (35.8%)	16 (39%)
	Prefer not to answer	3 (3.8%)	3 (7.3%)
Country of work	Argentina	2 (2.56%)	0
	Australia	6 (7.69%)	5 (12.19%)
	Austria	1 (1.28%)	0
	Belgium	1 (1.28%)	1 (2.43%)
	Brazil	3 (3.85%)	0
	Canada	26 (33.33%)	17 (41.46%)
	Canada/Colombia	1 (1.28%)	0
	Colombia	3 (3.85%)	1 (2.43%)
	Ethiopia	1 (1.28%)	0
	Greece	1 (1.28%)	1 (2.43%)
	India	2 (2.56%)	0
	India/New Zealand	1 (1.28%)	0

	Ireland	1 (1.28%)	1 (2.43%)
	Italy	3 (3.85%)	2 (4.87%)
	Kyrgyzstan	1 (1.28%)	0
	Lebanon	1 (1.28%)	1 (2.43%)
	Mexico	1 (1.28%)	0
	Romania	2 (2.56%)	1 (2.43%)
	Slovenia	1 (1.28)	0
	South Africa	2 (2.56)	1 (2.43%)
	Spain	3 (3.85%)	3 (7.31%)
	Switzerland	2 (2.56%)	0
	Turkey	1 (1.28%)	0
	United Kingdom	6 (7.69%)	5 (12.19%)
	United States	6 (7.69%)	2 (4.87%)
Experience			
Years of experience with evidence syntheses	≤ 4 years	3 (3.8%)	1 (2.4%)
	5-6 years	16 (20.5%)	6 (14.6%)
	7-8 years	9 (11.5%)	6 (14.6%)
	9-10 years	8 (10.2%)	5 (12.1%)
	11-12 years	8 (10.2%)	4 (9.7%)
	13-14 years	7 (8.9%)	4 (9.7%)
	≥ 15 years	27 (34.6%)	15 (36.5%)
Aspects of evidence synthesis that previously participated	Conceptualization/Research question development	78 (100%)	41 (100%)
	Undertaking literature searches	69 (88.4%)	38 (92.6%)
	Study screening and selection	71 (91%)	39 (95.1%)
	Data extraction	71 (91%)	38 (92.6%)
	Quality appraisal	66 (84.6%)	34 (82.9%)
	Data synthesis	70 (89.7%)	35 (85.3%)
	Interpretation of results	72 (92.3%)	37 (90.2%)
	Knowledge translation	63 (80.7%)	35 (85.3%)
	Other	9 (11.5%)	8 (19.5%)
Knowledge in evidence syntheses (0 = no expertise to 10 = very strong expertise)	*(Mean, SD)*	8.3 (±1.43)	8.5 (±0.9)
Field/area of research predominantly perform evidence syntheses	Clinical	39 (50%)	25 (60.9%)
	Public Health	54 (69.2%)	32 (78%)
	Health Systems	45 (57.6%)	28 (68.2%)
	Other	15 (19.2%)	10 (24.3%)

SD: standard deviation.

Round 1 – Overall Results

The results of round 1 are presented in the table S5. Overall, in this round 129 answers were collected, 52 were complete and 77 partials.

Table S5. Ratings results of Round 1.

ITEMS	CLASSIFICATION	n (%)
Conceptual/high-level questions		
What is the optimal composition of a rapid review team (e.g., number of team members, expertise, stakeholders, etc.)...?	High importance	25 (20%)
	Medium importance	21(39%)
	Low importance	7 (13%)
What is the optimal way of registering a rapid review or making the protocol available on a freely accessible platform (such as prospero or open science framework)...?	High importance	8 (15%)
	Medium importance	24 (45%)
	Low importance	21 (39%)
What is the optimal time frame that a rapid review should take (from search strategy to data analysis)...?	High importance	25 (47%)
	Medium importance	15 (28%)
	Low importance	13 (24%)
What is the optimal use of guidelines (including checklists and templates)...?	High importance	30 (56%)
	Medium importance	21 (39%)
	Low importance	2 (3%)
What are the optimal methods for involving stakeholders...?	High importance	23 (43%)
	Medium importance	24 (45%)
	Low importance	6 (11%)
When updating a completed rapid review, what is the optimal level of consistency in the reviewers (e.g., the same reviewers vs. Different reviewers)...?	High importance	13 (24%)
	Medium importance	19 (35%)
	Low importance	21 (39%)
Search strategy		
What is the optimal number of databases used for the search strategy...?	High importance	21 (39%)
	Medium importance	27 (50%)
	Low importance	5 (9%)
What are the optimal core set of databases that should be searched...?	High importance	31 (58%)
	Medium importance	16 (30%)
	Low importance	6 (11%)
What are the optimal non-peer-reviewed publication databases (i.e., pre-prints) that should be included...?	High importance	13 (24%)
	Medium importance	27 (50%)
	Low importance	13 (24%)
What are the optimal kinds of grey literature (other than pre-prints) that should be included...?	High importance	20 (37%)
	Medium importance	22 (41%)
	Low importance	11 (20%)
What are the optimal approaches for developing search terms for the key elements in the rapid review...?	High importance	40 (75%)
	Medium importance	9 (16%)
	Low importance	4 (7%)
What are the optimal restrictions and search limits (e.g., years of inclusion, language, phase of study, study design)...?	High importance	41 (77%)
	Medium importance	8 (15%)
	Low importance	4 (7%)
Studies selection		
What are the optimal approaches for determining the inclusion and exclusion criteria for studies...?	High importance	38 (71%)
	Medium importance	10 (18%)
	Low importance	5 (9%)
What is the optimal balance between automated and manual screening...?	High importance	28 (52%)
	Medium importance	19 (35%)

	Low importance	6 (11%)
What are the optimal automated processes (e.g., machine learning algorithms, specific SR programs, etc.) for screening...?	High importance	25 (47%)
	Medium importance	20 (37%)
	Low importance	8 (15%)
What is the optimal way to leverage online collaboration platforms (e.g., Cochrane Task Exchange) when doing screening...?	High importance	14 (26%)
	Medium importance	24 (45%)
	Low importance	15 (28%)
If using peer-reviewing, what is the optimal method to perform screening (e.g., independent screening, partial peer review, etc.)...?	High importance	29 (54%)
	Medium importance	21 (39%)
	Low importance	3 (5%)
What is the optimal number of people needed to perform screening...?	High importance	18 (33%)
	Medium importance	23 (43%)
	Low importance	12 (22%)
Data extraction		
What is the optimal balance between automated and manual data extraction techniques...?	High importance	24 (45%)
	Medium importance	21 (39%)
	Low importance	8 (15%)
If using peer-reviewing, what is the optimal approach to perform data extraction (e.g. independent extraction and comparison of discrepancies, one reviewer extracting and the other double checking)...?	High importance	30 (56%)
	Medium importance	21 (39%)
	Low importance	2 (3%)
What is the optimal number of people needed to perform data extraction...?	High importance	14 (26%)
	Medium importance	25 (47%)
	Low importance	14 (26%)
Quality/bias assessment		
What are the optimal approaches to assess the quality of included studies and/or risk of bias...?	High importance	41 (77%)
	Medium importance	9 (16%)
	Low importance	3 (5%)
What are the optimal dimensions of quality (e.g., trustworthiness, relevance) that should be considered when appraising studies...?	High importance	36 (67%)
	Medium importance	15 (28%)
	Low importance	2 (3%)
What is the optimal usage of quality assessment results (e.g., to further exclude studies, to allow a more tailored synthesis)...?	High importance	31 (58%)
	Medium importance	17 (32%)
	Low importance	5 (9%)
What is the optimal balance between automated and manual quality assessments...?	High importance	12 (22%)
	Medium importance	24 (45%)
	Low importance	17 (32%)
What is the optimal number of people needed to perform quality assessment...?	High importance	14 (26%)
	Medium importance	25 (47%)
	Low importance	14 (26%)
Synthesis		
What are the optimal synthesis methods to use (e.g., narrative, thematic, meta-analysis)...?	High importance	28 (52%)
	Medium importance	15 (28%)
	Low importance	10 (18%)
What are the optimal methods for assessing the strength of individual synthesis findings and recommendations (e.g., GRADE)...?	High importance	26 (49%)
	Medium importance	21 (39%)
	Low importance	6 (11%)
What factors need to be used when choosing how to disaggregate synthesis findings for priority populations (e.g., PROGRESS-Plus groups)...?	High importance	18 (33%)
	Medium importance	24 (45%)
	Low importance	11 (20%)

After round 1 data collection, 15 items from the initial list of 29 items were rated high or medium by less than 50% of the participants and added to a list of exclusion, while

14 items remained in the main list. The Central Scientific Committee reviewed all the comments, and a survey was circulated among the members asking for agreement on the modification of the items and the inclusion of new suggested questions. The format in the way the items were first suggested was repeatedly commented on. The main concerned was around asking about an “optimal” approach, which left unclear whether the answer should be generalizable to different reviews. After discussion, all the items were re-written, and members of the CSC provided agreement and comments.

In addition, there were different items covering aspects of team composition and team expertise. So, the CSC voted on whether we could structure those items: as a general concept (e.g. What is the optimal method for balancing the skills and experience of the team...?), or by specific tasks (e.g. What is the optimal method for selecting the team needed to perform data extraction...?). Most of the members preferred to have it as a general question.

Some of the suggested questions were considered to be out of scope and twelve new questions were voted in favor to be added between the members of the CSC (conceptual – 1, search strategy – 4, studies selection – 2, data extraction – 2, quality assessment – 2, synthesis – 1).

Table S6. Rewording and inclusion of items to Round 2.

OLD WORDING	NEW WORDING
Conceptual/high-level questions	
What is the optimal composition of a rapid review team (e.g., number of team members, expertise, stakeholders, etc.)...?	*Excluded- This item was grouped with a new item.*
What is the optimal way of registering a rapid review or making the protocol available on a freely accessible platform (such as prospero or open science framework)...?	*Excluded*
What is the optimal time frame that a rapid review should take (from search strategy to data analysis)...?	*Excluded*
What is the optimal use of guidelines (including checklists and templates)...?	What is the optimal method for using guidelines (including checklists and templates)..?
What are the optimal methods for involving stakeholders...?	*Excluded*
When updating a completed rapid review, what is the optimal level of consistency in the reviewers (e.g., the same reviewers vs. Different reviewers)...?	*Excluded*
NEW ITEM	What is the optimal method for balancing the skills and experience of the review team...?
Search strategy	

What is the optimal number of databases used for the search strategy...?	What is the optimal method for selecting the minimum number of databases for the search strategy...?
What are the optimal core set of databases that should be searched...?	What is the optimal method for defining the core set of databases that should be searched...?
What are the optimal non-peer-reviewed publication databases (i.e., pre-prints) that should be included...?	What is the optimal method for determining the non-peer-reviewed publication databases (i.e., pre-prints servers and high-volume producers) that should be included...?
What are the optimal kinds of grey literature (other than pre-prints) that should be included...?	*Excluded – This item was grouped with another one.*
What are the optimal approaches for developing search terms for the key elements in the rapid review...?	What is the optimal method for developing search terms for the key elements in the rapid review...?
What are the optimal restrictions and search limits (e.g., years of inclusion, language, phase of study, study design)...?	What is the optimal method for defining restrictions and search limits (e.g., years of inclusion, language, phase of study, study design)...?
NEW ITEM	What is the optimal method for gauging the sensitivity vs. specificity of your search strategy...?
NEW ITEM	What is the optimal method for evaluating the quality of your search (i.e., estimating coverage rate)…?
NEW ITEM	What is the optimal method for using different types of search fields (e.g., controlled vocab vs. title/abstract vs. full text)…?
NEW ITEM	If you search in multiple languages, what is the optimal method for adapting your search for each language…?
Studies selection	
What are the optimal approaches for determining the inclusion and exclusion criteria for studies...?	What is the optimal method for determining the inclusion and exclusion criteria for studies...?
What is the optimal balance between automated and manual screening...?	What is the optimal method for determining the use of automated and manual screening...?
What are the optimal automated processes (e.g., machine learning algorithms, specific SR programs, etc.) for screening...?	*Excluded*
What is the optimal way to leverage online collaboration platforms (e.g., Cochrane Task Exchange) when doing screening...?	*Excluded*
If using peer-reviewing, what is the optimal method to perform screening (e.g., independent screening, partial peer review, etc.)...?	What is the optimal method for determining how to perform screening (e.g., independent screening, 1+1 approach/partial peer review, etc.)...?
What is the optimal number of people needed to perform screening...?	*Excluded – This is was grouped to the question on team composition.*
NEW ITEM	When duplicating screening, what is the optimal method to determine the screening approach (e.g. 10% duplication of title and abstract, full

	duplication of title and abstract with single reviewer for full-text review, and vice versa)…?
NEW ITEM	What is the optimal method to determine the level of reliability between coders should we aim for during screening…?
Data extraction	
What is the optimal balance between automated and manual data extraction techniques...?	*Excluded*
If using peer-reviewing, what is the optimal approach to perform data extraction (e.g. independent extraction and comparison of discrepancies, one reviewer extracting and the other double checking)...?	What is the optimal method for defining the data extraction approach (e.g. peer review with independent extraction and comparison of discrepancies, one reviewer extracting and the other double checking)...?
What is the optimal number of people needed to perform data extraction...?	*Excluded*
NEW ITEM	What is the optimal method to decide on the development of valid and reliable data extraction sheet...?
NEW ITEM	What is the optimal method to decide on the core sets of variables everyone should aim to extract (e.g., basic sample info from each sample)...?
Quality/bias assessment	
What are the optimal approaches to assess the quality of included studies and/or risk of bias...?	What is the optimal method for determining how to assess the quality of included studies and/or risk of bias...?
What are the optimal dimensions of quality (e.g., trustworthiness, relevance) that should be considered when appraising studies...?	What is the optimal method for determining the dimensions of quality (e.g., trustworthiness, relevance) that should be considered when appraising studies...?
What is the optimal usage of quality assessment results (e.g., to further exclude studies, to allow a more tailored synthesis)...?	What is the optimal method for determining the usage of quality assessment results (e.g., to further exclude studies, to allow a more tailored synthesis)...?
What is the optimal balance between automated and manual quality assessments...?	*Excluded*
What is the optimal number of people needed to perform quality assessment...?	*Excluded*
NEW ITEM	What is the optimal method to decide on the standard set of quantitative analyses you can conduct to supplement reviews to examine various forms of bias (e.g., funnel plots, p-curves, Z-curves)…?
NEW ITEM	What is the optimal method to decide between having ratings of quality assessments, or having a set of RoB variables…?
Synthesis	
What are the optimal synthesis methods to use (e.g., narrative, thematic, meta-analysis)...?	What is the optimal method for deciding on the synthesis methods to use (e.g., narrative, thematic, meta-analysis)...?
What are the optimal methods for assessing the strength of individual	*Excluded*

synthesis findings and recommendations (e.g., GRADE)...?	
What factors need to be used when choosing how to disaggregate synthesis findings for priority populations (e.g., PROGRESS-Plus groups)...?	*Excluded*
NEW ITEM	What is the optimal method for deciding the data synthesis quantitative procedures...?

Round 2 – Overall Results

The second round's results are displayed in Table S7. On this round, new participants were also welcomed, bringing the total number of collected answers to 127. Of these, 81 responses were obtained from new participants, while 45 were contributed by participants who had already taken part in the previous round. Among these answers, a combined total of 70 consisted of complete responses, with 27 from new participants and 43 from those who had previously participated.

At the beginning of the round, there were a total of 26 items (14 carried over from Round 1 and 12 new items). All items listed for exclusion received agreement from over 75% of the participants, resulting in their removal. After analysis, all the 26 items were kept on the list to Round 3. All the comments received are available on Appendix VI.

Table S7. Ratings results of Round 2.

ITEMS	CLASSIFICATION	N (%)
Conceptual/high-level questions		
What is the optimal method for using guidelines (including checklists and templates)..?	High importance	34 (47%)
	Medium importance	29 (40%)
	Low importance	8 (11%)
What is the optimal method for balancing the skills and experience of the review team...?	High importance	25 (35%)
	Medium importance	30 (42%)
	Low importance	16 (22%)
Search strategy		
What is the optimal method for selecting the minimum number of databases for the search strategy...?	High importance	23 (32%)
	Medium importance	33 (46%)
	Low importance	14 (19%)
What is the optimal method for defining the core set of databases that should be searched...?	High importance	43 (60%)
	Medium importance	21 (29%)
	Low importance	6 (8%)
What is the optimal method for determining the non-peer-reviewed publication databases (i.e., pre-prints servers and high-volume producers) that should be included...?	High importance	14 (19%)
	Medium importance	32 (45%)
	Low importance	24 (33%)
What is the optimal method for developing search terms for the key elements in the rapid review...?	High importance	57 (80%)
	Medium importance	10 (14%)
	Low importance	3 (4%)
What is the optimal method for defining restrictions and search limits (e.g., years of inclusion, language, phase of study, study design)...?	High importance	55 (77%)
	Medium importance	14 (19%)
	Low importance	1 (1%)

What is the optimal method for gauging the sensitivity vs. specificity of your search strategy...?	High importance	24 (33%)
	Medium importance	33 (46%)
	Low importance	13 (18%)
What is the optimal method for evaluating the quality of your search (i.e., estimating coverage rate)	High importance	28 (39%)
	Medium importance	23 (32%)
	Low importance	19 (26%)
What is the optimal method for using different types of search fields (e.g., controlled vocab vs. Title/abstract vs. Full text)	High importance	22 (30%)
	Medium importance	27 (38%)
	Low importance	21 (29%)
If you search in multiple languages, what is the optimal method for adapting your search for each language	High importance	13 (18%)
	Medium importance	31 (43%)
	Low importance	26 (36%)
Studies selection		
What is the optimal method for determining the inclusion and exclusion criteria for studies...?	High importance	57 (80%)
	Medium importance	5 (7%)
	Low importance	8 (11%)
What is the optimal method for determining the use of automated and manual screening...?	High importance	31 (43%)
	Medium importance	32 (45%)
	Low importance	7 (9%)
What is the optimal method for determining how to perform screening (e.g., independent screening, 1+1 approach/partial peer review, etc.)...?	High importance	42 (59%)
	Medium importance	27 (38%)
	Low importance	1 (1%)
When duplicating screening, what is the optimal method to determine the screening approach (e.g. 10% duplication of title and abstract, full duplication of title and abstract with single reviewer for full-text review, and vice versa)...?	High importance	30 (42%)
	Medium importance	29 (40%)
	Low importance	11 (15%)
What is the optimal method to determine the level of reliability between coders should we aim for during screening...?	High importance	16 (22%)
	Medium importance	39 (54%)
	Low importance	15 (21%)
Data extraction		
What is the optimal method for defining the data extraction approach (e.g. peer review with independent extraction and comparison of discrepancies, one reviewer extracting and the other double checking)...?	High importance	37 (52%)
	Medium importance	30 (42%)
	Low importance	2 (2%)
What is the optimal method to decide on the development of valid and reliable data extraction sheets...?	High importance	28 (39%)
	Medium importance	22 (30%)
	Low importance	19 (26%)
What is the optimal method to decide on the core sets of variables everyone should aim to extract (e.g., basic sample info from each sample)...?	High importance	34 (47%)
	Medium importance	24 (33%)
	Low importance	11 (15%)
Quality/bias assessment		
What is the optimal method for determining the dimensions of quality (e.g., trustworthiness, relevance) that should be considered when appraising studies...?	High importance	50 (70%)
	Medium importance	16 (22%)
	Low importance	3 (4%)
What is the optimal method for determining how to assess the quality of included studies and/or risk of bias...?	High importance	50 (70%)
	Medium importance	16 (22%)
	Low importance	3 (4%)
What is the optimal method for determining the usage of quality assessment results (e.g., to further exclude studies, to allow a more tailored synthesis)...?	High importance	37 (52%)
	Medium importance	26 (36%)
	Low importance	6 (8%)
	High importance	17 (23%)

What is the optimal method to decide on the standard set of quantitative analyses you can conduct to supplement reviews to examine various forms of bias (e.g., funnel plots, p-curves, Z-curves)...?	Medium importance Low importance	32 (45%) 20 (28%)
What is the optimal method to decide between having ratings of quality assessments, or having a set of RoB variables...?	High importance Medium importance Low importance	19 (26%) 31 (43%) 19 (26%)
Synthesis		
What is the optimal method for deciding on the synthesis methods to use (e.g., narrative, thematic, meta-analysis)...?	High importance Medium importance Low importance	31 (43%) 28 (39%) 10 (14%)
What is the optimal method for deciding the data synthesis quantitative procedures...?	High importance Medium importance Low importance	25 (35%) 26 (36%) 18 (25%)

Round 3 – Overall Results

Sixty participants answered this round. From the 26 initial items, 6 items were rated as high important by at least 75% of participants and are included in the final list of items. The ranking question was analyzed, and an average score was calculated. Each item in the first position received one point, the item in the second position received two points and so on. The sum of points was divided by the number of participants that rated that item as high (that had the item included in their ranking question). So, lowest the average score, more important the item is. Items are presented in order of importance in Table S8.

Table S8. Round 3 ratings' results of items that met the cut-off point (included in the final list of items).

Item	Classification	Frequency	Average Score
What is the optimal method for developing search terms for the key elements in the rapid review...? (Search strategy)	High importance Middle importance Low importance	96% 2% 2%	4.37
What is the optimal method for determining the inclusion and exclusion criteria for studies...? (Studies Selection)	High importance Middle importance Low importance	83% 13% 3%	4.46
What is the optimal method for defining restrictions and search limits (e.g., years of inclusion, language, phase of study, study design)..? (Search strategy)	High importance Middle importance Low importance	88% 10% 2%	4.96
What is the optimal method for determining how to assess the quality	High importance Middle importance	86% 12%	5.28

of included studies and/or risk of bias...? (Quality/bias assessment)	Low importance	1%	
What is the optimal method for determining how to perform screening (e.g., independent screening, 1+1 approach/partial peer review, etc.)..? (Studies Selection)	High importance	77%	6.10
	Middle importance	21%	
	Low importance	2%	
What is the optimal method for determining the dimensions of quality (e.g., trustworthiness, relevance) that should be considered when appraising studies...? (Quality/bias assessment)	High importance	88%	6.67
	Middle importance	8%	
	Low importance	4%	

All items are finalized by the following sentence: “to improve the time-efficiency of a RR in a particular field?”.

Two items were rated as high important by 73% and 65% of participants (Table S9). These items will be discussed and voted on the consensus meeting for potential inclusion in the final list.

Table S9. Round 3 rating results of items close to cut-off point.

Item	Classification	Frequency	Average Score
17. What is the optimal method for defining the data extraction approach (e.g. peer review with independent extraction and comparison of discrepancies, one reviewer extracting and the other double checking)...? (Data Extraction)	High importance	65%	7.42
	Middle importance	34%	
	Low importance	1	
4. What is the optimal method for defining the core set of databases that should be searched...?) (Search Strategy)	High importance	73%	6.02
	Middle importance	25%	
	Low importance	2%	

All items are finalized by the following sentence: “to improve the time-efficiency of a RR in a particular field?”.

All the items that did not meet the cut-off point or are not close by (18 items) are presented in the table below (Table S10).

Table S10. Round 3 rating results of excluded items.

Item	Classification	Frequency
Conceptual/high-level questions		
1. What is the optimal method for using guidelines (including checklists and templates)..?	High importance Medium importance Low importance	43% 49% 8%
2. What is the optimal method for balancing the skills and experience of the review team...?	High importance Medium importance Low importance	29% 50% 21%
Search strategy		
3. What is the optimal method for selecting the minimum number of databases for the search strategy...?	High importance Medium importance Low importance	15% 60% 25%
5. What is the optimal method for determining the non-peer-reviewed publication databases (i.e., pre-prints servers and high-volume producers) that should be included...?	High importance Medium importance Low importance	10% 47% 43%
8. What is the optimal method for gauging the sensitivity vs. specificity of your search strategy...?	High importance Medium importance Low importance	22% 58% 20%
9. What is the optimal method for evaluating the quality of your search (i.e., estimating coverage rate)	High importance Medium importance Low importance	30% 55% 15%
10. What is the optimal method for using different types of search fields (e.g., controlled vocab vs. Title/abstract vs. Full text)	High importance Medium importance Low importance	14% 50% 36%
11. If you search in multiple languages, what is the optimal method for adapting your search for each language	High importance Medium importance Low importance	6% 32% 62%
Studies selection		
13. What is the optimal method for determining the use of automated and manual screening...?	High importance Medium importance Low importance	33% 55% 12%
15. When duplicating screening, what is the optimal method to determine the screening approach (e.g. 10% duplication of title and abstract, full duplication of title and abstract with single reviewer for full-text review, and vice versa)...?	High importance Medium importance Low importance	32% 58% 10%
16. What is the optimal method to determine the level of reliability between coders should we aim for during screening...?	High importance Medium importance Low importance	8% 49% 43%
Data extraction		
18. What is the optimal method to decide on the development of valid	High importance Medium importance	25% 52%

and reliable data extraction sheets...?	Low importance	23%
19. What is the optimal method to decide on the core sets of variables everyone should aim to extract (e.g., basic sample info from each sample)...?	High importance Medium importance Low importance	47% 48% 5%
Quality/bias assessment		
22. What is the optimal method for determining the usage of quality assessment results (e.g., to further exclude studies, to allow a more tailored synthesis)...?	High importance Medium importance Low importance	53% 38% 9%
23. What is the optimal method to decide on the standard set of quantitative analyses you can conduct to supplement reviews to examine various forms of bias (e.g., funnel plots, p-curves, Z-curves)...?	High importance Medium importance Low importance	6% 50% 44%
24. What is the optimal method to decide between having ratings of quality assessments, or having a set of RoB variables...?	High importance Medium importance Low importance	12% 52% 36%
Synthesis		
25. What is the optimal method for deciding on the synthesis methods to use (e.g., narrative, thematic, meta-analysis)...?	High importance Medium importance Low importance	33% 57% 10%
26. What is the optimal method for deciding the data synthesis quantitative procedures...?	High importance Medium importance Low importance	14% 61% 25%

All items are finalized by the following sentence: “to improve the time-efficiency of a RR in a particular field?”.

Section V: Additional analysis – overall and consistent participants

An additional analysis was performed considering only the participants that answered all three survey rounds (consistent participants). The results are presented in Table S11, showing the frequency for each rate of overall participants (all participants in round 3) and consistent participants (that had participated in all the three rounds).

Table S11. Ratings' results of included items broken down by consistent participants and others.

ITEMS	CLASSIFICATION	OVERALL N (%)			CONSISTENT PARTICIPANTS N (%)		
		Rounds			Rounds		
Conceptual/high-level questions		1*	2	3	1*	2	3
What is the optimal method for using guidelines (including checklists and templates)..?	High importance	30 (56%)	34 (47%)	26 (43%)	23 (56%)	24 (58%)	17 (41%)
	Medium importance	21 (39%)	29 (40%)	29 (48%)	17 (41%)	16 (39%)	23 (56%)
	Low importance	2 (3%)	8 (11%)	5 (8%)	1 (2.4%)	1 (2%)	1 (2%)
What is the optimal method for balancing the skills and experience of the review team...?	High importance	NA	25 (35%)	17 (28%)	NA	9 (21%)	11 (26%)
	Medium importance	NA	30 (42%)	30 (50%)	NA	26 (63%)	22 (53%)
	Low importance	NA	16 (22%)	13 (21%)	NA	6 (14%)	8 (19%)
Search strategy							
What is the optimal method for selecting the minimum number of databases for the search strategy...?	High importance	21 (39%)	23 (32%)	9 (15%)	15 (36%)	10 (24%)	6 (14%)
	Medium importance	27 (50%)	33 (46%)	36 (60%)	22 (53%)	24 (58%)	27 (65%)
	Low importance	5 (9%)	14 (19%)	15 (25%)	4 (9%)	7 (10%)	8 (19%)
What is the optimal method for defining the core set of databases that should be searched...?	High importance	31 (58%)	43 (60%)	44 (73%)	25 (60%)	29 (70%)	31 (75%)
	Medium importance	16 (30%)	21 (29%)	15 (25%)	11 (26%)	10 (24%)	9 (21%)
	Low importance	6 (11%)	6 (8%)	1 (1.6%)	5 (12%)	2 (4.8%)	1 (2%)
What is the optimal method for determining the non-peer-reviewed publication databases (i.e., pre-prints servers and high-volume producers) that should be included...?	High importance	13 (24%)	14 (19%)	6 (10%)	11 (26%)	7 (17%)	4 (9%)
	Medium importance	27 (50%)	32 (45%)	28 (46%)	21 (51%)	18 (43%)	18 (43%)
	Low importance	13 (24%)	24 (33%)	26 (43%)	9 (21%)	16 (39%)	19 (46%)
What is the optimal method	High importance	40 (75%)	57 (80%)	58 (96%)	30 (73%)	35 (85%)	40 (97%)

for developing search terms for the key elements in the rapid review...?	Medium importance	9 (16%)	10 (14%)	1 (1.6%)	8 (19%)	5 (12%)	0
	Low importance	4 (7%)	3 (4%)	1 (1.6%)	3 (7%)	1 (2.4%)	1 (2%)
What is the optimal method for defining restrictions and search limits (e.g., years of inclusion, language, phase of study, study design)...?	High importance	41 (77%)	55 (77%)	53 (88%)	33 (80%)	34 (82%)	37 (90%)
	Medium importance	8 (15%)	14 (19%)	6 (10%)	6 (14%)	6 (14%)	3 (7%)
	Low importance	4 (7%)	1 (1.4%)	1 (1.6%)	2 (4%)	1 (2.4%)	1 (2%)
What is the optimal method for gauging the sensitivity vs. specificity of your search strategy...?	High importance	NA	24 (33%)	13 (21%)	NA	15 (36%)	9 (21%)
	Medium importance	NA	33 (46%)	35 (58%)	NA	20 (48%)	26 (63%)
	Low importance	NA	13 (18%)	12 (20%)	NA	6 (14%)	6 (14%)
What is the optimal method for evaluating the quality of your search (i.e., estimating coverage rate)	High importance	NA	28 (39%)	18 (30%)	NA	15 (36%)	11 (26%)
	Medium importance	NA	23 (32%)	33 (55%)	NA	15 (36%)	25 (60%)
	Low importance	NA	19 (26%)	9 (15%)	NA	11 (26%)	5 (12%)
What is the optimal method for using different types of search fields (e.g., controlled vocab vs. Title/abstract vs. Full text)	High importance	NA	22 (30%)	8 (13%)	NA	12 (29%)	4 (9%)
	Medium importance	NA	27 (38%)	30 (50%)	NA	15 (36%)	23 (56%)
	Low importance	NA	21 (29%)	22 (36%)	NA	14 (34%)	14 (34%)
If you search in multiple languages, what is the optimal method for adapting your search for each language	High importance	NA	13 (18%)	4 (6%)	NA	7 (17%)	2 (5%)
	Medium importance	NA	31 (43%)	19 (31%)	NA	18 (43%)	15 (36%)
	Low importance	NA	26 (36%)	37 (62%)	NA	16 (39%)	24 (58%)
Studies Selection							
What is the optimal method for determining the inclusion and exclusion criteria for studies...?	High importance	38 (71%)	57 (80%)	50 (83%)	30 (73%)	37 (90%)	36 (87%)
	Medium importance	10 (18%)	5 (7%)	8 (13%)	8 (19%)	1 (2%)	4 (9%)
	Low importance	5 (9%)	8 (11%)	2 (3%)	3 (7%)	3 (7%)	1 (2%)
What is the optimal method for determining the use of automated and manual screening...?	High importance	28 (52%)	31 (43%)	20 (33%)	23 (56%)	22 (53%)	14 (34%)
	Medium importance	19 (35%)	32 (45%)	33 (55%)	14 (34%)	16 (39%)	23 (56%)
	Low importance	6 (11%)	7 (9%)	7 (11%)	4 (9%)	3 (7%)	4 (9%)

What is the optimal method for determining how to perform screening (e.g., independent screening, 1+1 approach/partial peer review, etc.)...?	High importance	29 (54%)	42 (59%)	46 (77%)	24 (58%)	24 (58%)	33 (80%)
	Medium importance	21 (39%)	27 (38%)	13 (21%)	15 (36%)	16 (39%)	8 (20%)
	Low importance	3 (5%)	1 (1.4%)	1 (1.6%)	2 (4.8%)	1 (2%)	0
When duplicating screening (e.g., peer-reviewing), what is the optimal method to determine the screening approach (e.g., 10% duplication of title and abstract, full duplication of title and abstract with single reviewer for full-text review, and vice versa)...?	High importance	NA	30 (42%)	19 (32%)	NA	14 (34%)	13 (32%)
	Medium importance	NA	29 (40%)	35 (58%)	NA	18 (43%)	24 (58%)
	Low importance	NA	11 (15%)	6 (10%)	NA	9 (21%)	4 (9.7%)
What is the optimal method to determine the level of reliability between coders should we aim for during screening...?	High importance	NA	16 (22%)	5 (8%)	NA	9 (21%)	2 (5%)
	Medium importance	NA	39 (54%)	29 (49%)	NA	22 (53%)	19 (46%)
	Low importance	NA	15 (21%)	26 (43%)	NA	10 (24%)	20 (49%)
Data extraction							
What is the optimal method for defining the data extraction approach (e.g. peer review with independent extraction and comparison of discrepancies, one reviewer extracting and the other double checking)...?	High importance	30 (56%)	37 (52%)	39 (65%)	24 (58%)	24 (58%)	26 (63%)
	Medium importance	21 (39%)	30 (42%)	20 (33%)	16 (39%)	15 (36%)	14 (34%)
	Low importance	2 (3%)	2 (2%)	1 (1.6%)	1 (2%)	1 (2%)	1 (2%)
What is the optimal method to decide on the development of valid and reliable data extraction sheets...?	High importance	NA	28 (39%)	15 (25%)	NA	17 (41%)	11 (26%)
	Medium importance	NA	22 (30%)	31 (52%)	NA	11 (26%)	18 (44%)
	Low importance	NA	19 (26%)	14 (23%)	NA	12 (29%)	12 (30%)
What is the optimal method to	High importance	NA	34 (47%)	47	NA	22 (53%)	18 (44%)

decide on the core sets of variables everyone should aim to extract (e.g., basic sample info from each sample)...?	Medium importance	NA	24 (33%)	48	NA	14 (34%)	20 (49%)
	Low importance	NA	11 (15%)	5	NA	4 (9%)	3 (7%)
Quality/bias assessment							
What is the optimal method for determining the dimensions of quality (e.g., trustworthiness, relevance) that should be considered when appraising studies...?	High importance	36 (67%)	50 (70%)	53 (88%)	32 (78%)	34 (82%)	39 (95%)
	Medium importance	15 (28%)	16 (22%)	5 (8%)	9 (21%)	6 (14%)	2 (4.8%)
	Low importance	2 (3%)	3 (4%)	2 (4%)	0	0	0
What is the optimal method for determining how to assess the quality of included studies and/or risk of bias...?	High importance	41 (77%)	50 (70%)	52 (86%)	34 (82%)	33 (80%)	38 (93%)
	Medium importance	9 (16%)	16 (22%)	7 (12%)	5 (12%)	7 (17%)	3 (7%)
	Low importance	3 (5%)	3 (4%)	1 (1.6%)	2 (4%)	0	0
What is the optimal method for determining the usage of quality assessment results (e.g., to further exclude studies, to allow a more tailored synthesis)...?	High importance	31 (58%)	37 (52%)	32 (53%)	26 (63%)	26 (63%)	24 (59%)
	Medium importance	17 (32%)	26 (36%)	23 (38%)	12 (29%)	13 (31%)	16 (40%)
	Low importance	5 (9%)	6 (8%)	5 (8%)	3 (7%)	1 (2%)	1 (2%)
What is the optimal method to decide on the standard set of quantitative analyses you can conduct to supplement reviews to examine various forms of bias (e.g., funnel plots, p-curves, Z-curves)...?	High importance	NA	17 (23%)	4 (6%)	NA	7 (17%)	3 (7%)
	Medium importance	NA	32 (45%)	30 (50%)	NA	20 (48%)	20 (49%)
	Low importance	NA	20 (28%)	26 (43%)	NA	13 (31%)	18 (44%)
What is the optimal method to decide between having ratings of quality assessments, or	High importance	NA	19 (26%)	7 (12%)	NA	8 (19%)	5 (12%)
	Medium importance	NA	31 (43%)	31 (52%)	NA	18 (43%)	18 (44%)
	Low importance	NA	19 (26%)	22 (36%)	NA	14 (34%)	18 (44%)

having a set of RoB variables...?							
Synthesis							
What is the optimal method for deciding on the synthesis methods to use (e.g., narrative, thematic, meta-analysis)...?	High importance	28 (52%)	31 (43%)	20 (33%)	22 (51%)	19 (46%)	14 (34%)
	Medium importance	15 (28%)	28 (39%)	34 (57%)	16 (39%)	18 (43%)	25 (60%)
	Low importance	10 (18%)	10 (14%)	6 (10%)	4 (9%)	3 (7%)	2 (5%)
What is the optimal method for deciding the data synthesis quantitative procedures...?	High importance	NA	25 (35%)	8 (13%)	NA	12 (29%)	6 (14%)
	Medium importance	NA	26 (36%)	37 (62%)	NA	19 (46%)	26 (63%)
	Low importance	NA	18 (25%)	15 (25%)	NA	9 (21%)	8 (21%)

*Items in Round 1 were written differently, e.g. “What is the optimal method for defining restrictions and search limits (e.g., years of inclusion, language, phase of study, study design)...?” was “What are the optimal restrictions and search limits (e.g., years of inclusion, language, phase of study, study design)...?”. NA: Not applicable.

Section VI: Additional analysis – results broke down by participant's profile

How data was analysed

Data was analysed by participants' profile considering the eligibility question: Please, select the category with which you most strongly identify (Researcher, including research-focus students; Healthcare practitioner, including trainees; Policymaker; Patient / community member / caregiver).

From the 60 participants that answered Round 3

- 47 self-identified themselves as researchers (79%)
- 7 healthcare professionals (11%)
- 6 policymakers (10%).
- No participant self-identified as a Patient / community member / caregiver.

Results

Results are presented in Table S12. The items currently in the final list are still highlighted in blue.

As it can be observed, for most of the included items (4 out of 6) all participant's profiles agree on their classification as high important (≥75%). This is not the case only for two items (highlighted in blue in table S12):

- Item n 14 "What is the optimal method for determining how to perform screening (e.g., independent screening, 1+1 approach/partial peer review, etc.) ...?". 1.4% of HCPs and 80.9% of researchers rated as high importance, while only 50% of policymakers did).
- Item n 21 "What is the optimal method for determining how to assess the quality of included studies and/or risk of bias...?". 85.7% of HCPs and 89.4% of researchers rated as high important while only 4 66.7% of policymakers also rated in the same way.

Table S12. Round 3 results of classification of the items in the final list broken down by participants' profile.

Items	Classification	n (%)		
		HCPS	Policymaker	Researcher
Search strategy				
6. What is the optimal method for developing search terms for the key elements in the rapid review...?	High importance	7 (100%)	5 (83.3%)	46 (97.9%)
	Medium importance	0 (0%)	1 (16.7%)	0 (0%)
	Low importance	0 (0%)	0 (0%)	1 (2.1%)
7. What is the optimal method for defining restrictions and search limits (e.g., years of inclusion, language, phase of study, study design)...?	High importance	6 (85.7%)	5 (83.3%)	42 (89.4%)
	Medium importance	1 (14.3%)	1 (16.7%)	4 (8.5%)
	Low importance	0 (0%)	0 (0%)	1 (2.1%)
Studies Selection				

12. What is the optimal method for determining the inclusion and exclusion criteria for studies...?	High importance	6 (85.7%)	5 (83.3%)	39 (83.0%)
	Medium importance	1 (14.3%)	1 (16.7%)	6 (12.8%)
	Low importance	0 (0%)	0 (0%)	2 (4.3%)
14. What is the optimal method for determining how to perform screening (e.g., independent screening, 1+1 approach/partial peer review, etc.)...?	**High importance**	**5 (71.4%)**	**3 (50%)**	**38 (80.9%)**
	Medium importance	**2 (28.6%)**	**2 (33.3%)**	**9 (19.1%)**
	Low importance	**0 (0%)**	**1 (16.7%)**	**0 (0%)**
Quality/Bias Assessment				
20. What is the optimal method for determining the dimensions of quality (e.g., trustworthiness, relevance) that should be considered when appraising studies...?	High importance	7 (100%)	5 (83.3%)	41 (87.2%)
	Medium importance	0 (0%)	1 (16.7%)	4 (8.5%)
	Low importance	0 (0%)	0 (0%)	2 (4.3%)
21. What is the optimal method for determining how to assess the quality of included studies and/or risk of bias...?	**High importance**	**6 (85.7%)**	**4 (66.7%)**	**42 (89.4%)**
	Medium importance	**1 (14.3%)**	**2 (33.3%)**	**4 (8.5%)**
	Low importance	**0 (0%)**	**0 (0%)**	**1 (2.1%)**

Table S13. Round 3 results of classification of the two-items close to cut-off point broke down by participant's profile.

Items	Classification	n (%)		
		HCPS	Policymaker	Researcher
Search Strategy				
4. What is the optimal method for defining the core set of databases that should be searched...?	High importance	6 (85.7%)	6 (100%)	32 (68.1%)
	Medium importance	1 (14.3%)	0 (0%)	14 (29.8%)
	Low importance	0 (0%)	0 (0%)	1 (2.1%)
Data Extraction				
17. What is the optimal method for defining the data extraction approach (e.g. peer review with independent extraction and comparison of discrepancies, one reviewer extracting and the other double checking)...?	High importance	6 (85.7%)	4 (66.7%)	29 (61.7%)
	Medium importance	1 (14.3%)	2 (33.3%)	17 (36.2%)
	Low importance	0 (0%)	0 (0%)	1 (2.1%)

Table S14. Round 3 results of classification of the remaining items – not included in the final list - broke down by participant's profile.

Items	Classification	n (%) HCPS	Policymaker	Researcher
Conceptual/high-level questions				
1. What is the optimal method for using guidelines (including checklists and templates)..?	High importance	1 (14.3%)	4 (66.7%)	21 (44.7%)
	Medium importance	5 (71.4%)	2 (33.3%)	22 (46.8%)
	Low importance	1 (14.3%)	0 (0%)	4 (8.5%)
2. What is the optimal method for balancing the skills and experience of the review team...?	High importance	1 (14.3%)	1 (16.7%)	15 (31.9%)
	Medium importance	4 (57.1%)	4 (66.7%)	22 (46.8%)
	Low importance	2 (28.6%)	1 (16.7%)	10 (21.3%)
Search strategy				
3. What is the optimal method for selecting the minimum number of databases for the search strategy...?	High importance	2 (28.6%)	2 (33.3%)	5 (10.6%)
	Medium importance	4 (57.1%)	3 (50%)	29 (61.7%)
	Low importance	1 (14.3%)	1 (16.7%)	13 (27.7%)
5. What is the optimal method for determining the non-peer-reviewed publication databases (i.e., pre-prints servers and high-volume producers) that should be included...?	High importance	2 (28.6%)	1 (16.7%)	3 (6.4%)
	Medium importance	1 (14.3%)	3 (50.0%)	24 (51.1%)
	Low importance	4 (57.1%)	2 (33.3%)	20 (42.6%)
8. What is the optimal method for gauging the sensitivity vs. specificity of your search strategy...?	High importance	3 (42.9%)	1 (16.7%)	9 (19.1%)
	Medium importance	4 (57.1%)	4 (66.7%)	27 (57.4%)
	Low importance	0 (0%)	1 (16.7%)	11 (23.4%)
9. What is the optimal method for evaluating the quality of your search (i.e., estimating coverage rate)	High importance	5(71.4%)	0 (0%)	13 (27.7%)
	Medium importance	2 (28.6%)	5 (83.3%)	26 (55.3%)
	Low importance	0 (0%)	1 (16.7%)	8 (17.0%)
10. What is the optimal method for using different types of search fields (e.g., controlled vocab vs. Title/abstract vs. Full text)	High importance	1 (14.3%)	0 (0%)	7 (14.9%)
	Medium importance	4 (57.1%)	4 (66.7%)	22 (46.8%)
	Low importance	2 (28.6%)	2 (33.3%)	18 (38.3%)
11. If you search in multiple languages, what is the optimal method for adapting your search for each language	High importance	2 (28.6%)	0 (0%)	2 (4.3%)
	Medium importance	2 (28.6%)	3 (50%)	14 (29.8%)
	Low importance	3 (42.9%)	3 (50%)	31 (66%)
Studies selection				
13. What is the optimal method for determining the use of automated and manual screening...?	High importance	3 (42.9%)	1 (16.7%)	16 (34.0%)
	Medium importance	4 (57.1%)	4 (66.7%)	25 (53.2%)
	Low importance	0 (0%)	1 (16.7%)	6 (12.8%)
15. When duplicating screening, what is the optimal method to	High importance	1 (14.3%)	2 (33.3%)	16 (34%)
	Medium importance	5 (71.4%)	4 (66.7%)	26 (55.3%)
	Low importance	1 (14.3%)	0 (0%)	5 (10.6%)

determine the screening approach (e.g. 10% duplication of title and abstract, full duplication of title and abstract with single reviewer for full-text review, and vice versa)...?				
16. What is the optimal method to determine the level of reliability between coders should we aim for during screening...?	High importance Medium importance Low importance	1 (14.3%) 6 (85.7%) 0 (0%)	0 (0%) 3 (50%) 3 (50%)	4 (8.5%) 20 (42.6%) 23 (48.9%)
Data extraction				
18. What is the optimal method to decide on the development of valid and reliable data extraction sheets...?	High importance Medium importance Low importance	4 (57.1%) 3 (42.9%) 0 (0%)	1 (16.7%) 3 (50.0%) 2 (33.3%)	10 (21.3%) 25 (53.2%) 12 (25.5%)
19. What is the optimal method to decide on the core sets of variables everyone should aim to extract (e.g., basic sample info from each sample)...?	High importance Medium importance Low importance	3 (42.9%) 4 (57.1%) 0 (0%)	3 (50.0%) 3 (50%) 0 (0%)	22 (46.8%) 22 (46.8%) 3 (6.4%)
Quality/bias assessment				
22. What is the optimal method for determining the usage of quality assessment results (e.g., to further exclude studies, to allow a more tailored synthesis)...?	High importance Medium importance Low importance	5 (71.4%) 2 (28.6%) 0 (0%)	4 (66.7%) 1 (16.7%) 1 (16.7%)	23 (48.9%) 20 (42.6%) 4 (8.5%)
23. What is the optimal method to decide on the standard set of quantitative analyses you can conduct to supplement reviews to examine various forms of bias (e.g., funnel plots, p-curves, Z-curves)...?	High importance Medium importance Low importance	2 (28.6%) 4 (57.1%) 1 (14.3%)	0 (0%) 4 (66.7%) 2 (33.3%)	2 (4.3%) 22 (46.8%) 23 (48.9%)
24. What is the optimal method to decide between having ratings of quality assessments, or having a set of RoB variables...?	High importance Medium importance Low importance	2 (28.6%) 3 (42.9%) 2 (28.6%)	0 (0%) 5 (83.3%) 1 (16.7%)	5 (10.6%) 23 (48.9%) 19 (40.4%)
Synthesis				
25. What is the optimal method for deciding on the synthesis methods to use (e.g., narrative, thematic, meta-analysis)...?	High importance Medium importance Low importance	4 (57.1%) 3 (42.9%) 0 (0%)	4 (66.7%) 1 (16.7%) 1 (16.7%)	12 (25.5%) 30 (63.8%) 5 (10.6%)
26. What is the optimal method for deciding the	High importance Medium importance	3 (42.9%) 4 (57.1%)	1 (16.7%) 4 (66.7%)	4 (8.5%) 29 (61.7%)

data synthesis quantitative procedures...?	Low importance	0 (0%)	1 (16.7%)	14 (29.8%)

APPENDIX III: Chapter 3

The impact of database choices: a methodological study using Cochrane cardiac rehabilitation reviews as a case study

Table S1A. Indexing rates by databases (considering studies)

Review	N of studies (n of trials)	MEDLINE	Embase	CINAHL	Web of Science	CENTRAL	PsycINFO	LILACS
	n of studies identified (%)							
CR1	145 (85)	105 (72.41)	67 (46.20)	45 (31.03)	60 (41.37)	127 (87.58)	NA	NA
CR2	104 (60)	87 (83.65)	72 (69.23)	5 (4.8)	NA	92 (88.46)	4 (3.84)	NA
CR3	23 (20)	19 (79.16)	21 (87.5)	13 (54.16)	13 (54.16)	20 (83.33)	1 (4.16)	0
CR4	10 (6)	1 (10)	3 (30)	NA	NA	NA	NA	NA
CR5	81 (35)	61 (75.30)	59 (72.83)	31 (38.27)	NA	74 (91.35)	10 (12.34)	NA
CR6	44 (21)	28 (63.63)	21 (47.72)	20 (45.45)	NA	38 (86.36)	8 (18.18)	NA
CR7	50 (24)	37 (74)	19 (38)	15 (30)	NA	47 (94)	1 (2)	NA
CR8	126 (54)	77 (61.11)	55 (43.65)	NA	39 (30.95)	111 (88)	NA	NA
	Mean (%)	***65.33***	***54.86***	***34.34***	***42.94***	***88.95***	***8.14***	***0***

Color code: ≥80% = dark green; 50 – 79 = yellow; 10 – 49 = light pink; 0 – 9 = grey. NA: not applicable.

Table S1B. Indexing rates by databases (considering studies and excluding the hand search studies- not indexed in any database)

Review	N of studies (n of trials)	MEDLINE	Embase	CINAHL	Web of Science	CENTRAL	PsycINFO	LILACS
				n of studies identified (%)				
CR1	135 (85)	105 (77.7)	67 (49.62)	45 (33.33)	60 (44.44)	127 (94.07)	NA	NA
CR2	97 (60)	87 (89.69)	72 (74.22)	5 (5.15)	NA	92 (94.84)	4 (4.12)	NA
CR3	23 (20)	19 (82.60)	21 (91.30)	13 (56.52)	13 (56.52)	20 (86.95)	1 (4.34)	0
CR4	3 (6)	1 (33.33)	3 (100)	NA	NA	NA	NA	NA
CR5	79 (35)	61 (77.21)	59 (74.68)	31 (39.24)	NA	74 (93.67)	10 (12.65)	NA
CR6	40 (21)	28 (70)	21 (51.5)	20 (50)	NA	38 (95)	8 (20)	NA
CR7	47 (24)	37 (78.72)	19 (40.42)	15 (31.91)	NA	47 (100)	1 (2.12)	NA
CR8	123 (54)	77 (62.6)	55 (44.71)	NA	39 (31.70)	111 (90.24)	NA	NA
	Mean (%)	***71.39***	***65.80***	***36.02***	***44.22***	***93.53***	***8.6***	***0***

Color code: ≥80% = dark green; 50 – 79 = yellow; 10 – 49 = light pink; 0 – 9 = white. NA: not applicable.

Table S2A. Included studies by the combination of databases pairs (considering Medline, Embase and CENTRAL as the main database)

Review	***N of studies (n of trials)***	Medline + Embase	Medline + CINAHL	Medline + WoS	Medline + PsycINFO	Medline + LILACS	Medline + CENTRAL	Embase + CINAHL	Embase + WoS	Embase + PsycINFO	Embase + LILACS	Embase + CENTRAL	CENTRAL + CINAHL	CENTRAL + WoS	CENTRAL + PsycINFO	CENTRAL + LILACS
		n of studies identified (%)														
CR1	145 (85)	112 (77.24)	113 (77.93)	114 (78.62)	NA	NA	129 (88.96)	86 (59.31)	99 (68.27)	NA	NA	130 (89.65)	130 (89.65)	132 (91.03)	NA	NA
CR2	104 (60)	94 (90.38)	88 (84.61)	NA	87 (83.65)	NA	97 (93.26)	75 (72.11)	NA	74 (71.15)	NA	93 (89.42)	92 (88.46)	NA	92 (88.46)	NA
CR3	23 (20)	23 (95.83)	20 (83.33)	20 (83.33)	19 (79.16)	19 (79.16)	20 (83.33)	23 (95.83)	23 (95.83)	22 (91.66)	21 (87.5)	23 (95.83)	21 (87.5)	20 (83.33)	20 (83.33)	20 (83.33)
CR4	10 (6)	3 (30)	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
CR5	81 (35)	75 (92.59)	62 (76.54)	NA	61 (75.30)	NA	77 (95.06)	65 (80.24)	NA	60 (74.07)	NA	77 (95.06)	77 (95.06)	NA	74 (91.35)	NA
CR6	44 (21)	29 (65.90)	31 (70.45)	NA	28 (63.63)	NA	38 (86.36)	38 (86.36)	NA	23 (52.27)	NA	39 (88.63)	39 (88.63)	NA	38 (86.36)	NA
CR7	50 (24)	38 (76)	38 (76)	NA	38 (76)	NA	47 (94)	26 (52)	NA	21 (42)	NA	47 (94)	47 (94)	NA	47 (94)	NA
CR8	126 (54)	82 (65.07)	NA	84 (66.66)	NA	NA	115 (91.26)	NA	66 (52.38)	NA	NA	115 (91.26)	NA	116 (92.06)	NA	NA
	Mean (%)	***74.12***	***78.14***	***76.20***	***75.54***	***79.16***	***90.31***	***74.30***	***72.16***	***66.23***	***87.5***	***91.97***	***90.55***	***88.80***	***88.7***	***83.33***

NA: not applicable.

Table S2B. Included studies (excluding the hand search studies) by the combination of databases pairs (considering Medline, Embase and CENTRAL as the main database)

Review	***N of studies (n of trials)***	Medline + Embase	Medline + CINAHL	Medline + WoS	Medline + PsycINFO	Medline + LILACS	Medline + CENTRAL	Embase + CINAHL	Embase + WoS	Embase + PsycINFO	Embase + LILACS	Embase + CENTRAL	CENTRAL + CINAHL	CENTRAL + WoS	CENTRAL + PsycINFO	CENTRAL + LILACS
		n of studies identified (%)														
CR1	135 (80)	112 (82.96)	113 (83.70)	114 (84.44)	NA	NA	129 (95.55)	86 (63.70)	99 (73.33)	NA	NA	130 (96.29)	130 (96.29)	132 (97.77)	NA	NA
CR2	97 (58)	94 (96.90)	88 (90.72)	NA	87 (89.69)	NA	97 (100)	75 (77.31)	NA	74 (76.28)	NA	93 (95.87)	92 (94.84)	NA	92 (94.84)	NA
CR3	23 (19)	23 (100)	20 (86.95)	20 (86.95)	19 (82.6)	19 (82.6)	20 (86.95)	23 (100)	23 (100)	22 (95.65)	21 (91.30)	23 (100)	21 (91.30)	20 (86.95)	20 (86.95)	20 (86.95)
CR4	3 (2)	3 (100)	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
CR5	79 (35)	75 (94.93)	62 (78.48)	NA	61 (77.21)	NA	77 (97.46)	65 (82.27)	NA	60 (75.94)	NA	77 (97.46)	77 (97.46)	NA	74 (93.67)	NA
CR6	40 (21)	29 (72.5)	31 (77.5)	NA	28 (70)	NA	38 (95)	38 (72.5)	NA	23 (57.5)	NA	39 (97.5)	39 (97.5)	NA	38 (95)	NA
CR7	47 (24)	38 (80.9)	38 (80.9)	NA	38 (80.9)	NA	47 (100)	26 (55.3)	NA	21 (44.7)	NA	47 (100)	47 (100)	NA	47 (100)	NA
CR8	123 (54)	82 (66.66)	NA	84 (68.29)	NA	NA	115 (93.49)	NA	66 (53.65)	NA	NA	115 (93.49)	NA	116 (94.30)	NA	NA
	Mean (%)	***86.85***	***83.04***	***79.89***	***80.08***	***82.6***	***95.49***	***75.18***	***75.66***	***70.04***	***91.3***	***97.23***	***96.23***	***93***	***94.09***	***86.95***

NA: not applicable.

Table S3A. Indexing rates by databases (considering trials)

Review	N of trials	MEDLINE	Embase	CINAHL	Web of Science	CENTRAL	PsycINFO	LILACS
	n of trials identified (%)							
CR1	85	68 (80)	45 (52.94)	35 (41.17)	45 (52.94)	77 (90.58)	NA	NA
CR2	60	54 (90)	43 (71.66)	5 (8.33)	NA	53 (88.33)	4 (6.66)	NA
CR3	20	16 (80)	17 (85)	13 (65)	12 (60)	16 (80)	1 (5)	0
CR4	6	1 (16.66)	2 (33.33)	NA	NA	NA	NA	NA
CR5	35	31 (88.57)	25 (71.42)	17 (48.57)	NA	35 (100)	9 (25.71)	NA
CR6	21	18 (85.71)	14 (66.66)	15 (71.42)	NA	19 (90.47)	8 (38.09)	NA
CR7	24	21 (87.5)	14 (58.33)	11 (45.83)	NA	24 (100)	2 (8.33)	NA
CR8	54	45 (83.33)	36 (66.66)	NA	27 (50)	47 (87.03)	NA	NA
	Mean (%)	***76.47***	***63.25***	***46.72***	***54.31***	***90.91***	***16.75***	***0***

Color code: ≥80% = dark green; 50 – 79 = yellow; 10 – 49 = light pink; 0 – 9 = white. NA: not applicable

Table S3B. Indexing rates by databases (considering trials and excluding hand search)

Review	N of trials	MEDLINE	Embase	CINAHL	Web of Science	CENTRAL	PsycINFO	LILACS
	n of trials identified (%)							
CR1	80	68 (85)	45 (56.25)	35 (43.75)	45 (56.25)	77 (96.25)	NA	NA
CR2	58	54 (93.10)	43 (74.13)	5 (8.62)	NA	53 (91.37)	4 (6.89)	NA
CR3	19	16 (84.21)	17 (89.47)	13 (68.42)	12 (63.15)	16 (84.21)	1 (5.26)	0
CR4	2	1 (50)	2 (100)	NA	NA	NA	NA	NA
CR5	35	31 (88.57)	25 (71.42)	17 (48.57)	NA	35 (100)	9 (25.71)	NA
CR6	21	18 (85.71)	14 (66.66)	15 (71.42)	NA	19 (90.47)	8 (38.09)	NA
CR7	24	21 (87.5)	14 (58.33)	11 (45.83)	NA	24 (100)	2 (8.33)	NA
CR8	54	45 (83.33)	36 (66.66)	NA	27 (50)	47 (87.03)	NA	NA
	Mean (%)	***82.17***	***72.86***	***47.76***	***56.46***	***92.76***	***16.85***	***0***

Color code: ≥80% = dark green; 50 – 79 = yellow; 10 – 49 = light pink; 0 – 9 = white. NA: not applicable

Table S4. Influential trials indexing rate

Review	N of trials (N of influential trials)	MEDLINE	Embase	CINAHL	Web of Science	CENTRAL	PsycINFO	LILACS
				n of trials identified (%)				
CR1	80 (9)	5 (55.55)	4 (44.44)	3 (33.33)	5 (55.55)	8 (88.88)	NA	NA
CR2	58 (5)	5 (100)	4 (80)	1 (20)	NA	5 (100)	0	NA
CR3	19 (1)	1 (100)	1 (100)	1 (100)	1 (100)	1 (100)	0	0
CR4	2 (0)	NA	NA	NA	NA	NA	NA	NA
CR5	35 (6)	6 (100)	3 (50)	4 (66.66)	NA	6 (100)	2 (33.33)	NA
CR6	21 (NA)*	NA	NA	NA	NA	NA	NA	NA
CR7	24 (1)	1 (100)	1 (100)	1 (100)	NA	1 (100)	0	NA
CR8	54 (NA)*	NA	NA	NA	NA	NA	NA	NA
	Mean (%)	***82.17***	***72.86***	***47.76***	***56.46***	***92.76***	***16.85***	***0***

NA: not applicable. *Information is not available for the reviews CR6 and CR8 because of the stage of the review update (no access to data extracted).

Table S5A. Included trials by the combination of databases pairs (considering Medline, Embase and CENTRAL as the main database)

Review	***N of studies (n of trials)***	Medline + Embase	Medline + CINAHL	Medline + WoS	Medline + PsycINFO	Medline + LILACS	Medline + CENTRAL	Embase + CINAHL	Embase + WoS	Embase + PsycINFO	Embase + LILACS	Embase + CENTRAL	CENTRAL + CINAHL	CENTRAL + WoS	CENTRAL + PsycINFO	CENTRAL + LILACS
		n of trials identified (%)														
CR1	145 (85)	73 (85.88)	70 (82.35)	70 (82.35)	NA	NA	79 (92.94)	56 (65.88)	66 (77.64)	NA	NA	79 (92.94)	78 (91.76)	78 (91.76)	NA	NA
CR2	104 (60)	56 (93.33)	55 (91.66)	NA	54 (90)	NA	58 (96.66)	46 (76.66)	NA	45 (75)	NA	54 (90)	53 (88.33)	NA	53 (88.33)	NA
CR3	24 (20)	19 (95)	17 (85)	16 (80)	16 (80)	16 (80)	16 (80)	19 (95)	19 (95)	18 (90)	17 (85)	19 (95)	17 (85)	16 (80)	16 (80)	16 (80)
CR4	10 (6)	2 (33.33)	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
CR5	81 (35)	33 (94.28)	31 (88.57)	NA	31 (88.57)	NA	35 (100)	27 (77.14)	NA	25 (71.42)	NA	35 (100)	35 (100)	NA	35 (100)	NA
CR6	44 (21)	19 (90.47)	20 (95.23)	NA	18 (85.71)	NA	19 (90.47)	18 (85.71)	NA	14 (66.66)	NA	20 (95.23)	20 (95.23)	NA	19 (90.47)	NA
CR7	50 (24)	21 (87.5)	21 (87.5)	NA	21 (87.5)	NA	24 (100)	17 (70.83)	NA	14 (58.33)	NA	24 (100)	24 (100)	NA	24 (100)	NA
CR8	126 (54)	49 (90.74)	NA	49 (90.74)	NA	NA	49 (90.74)	NA	42 (77.77)	NA	NA	49 (90.74)	NA	51 (94.44)	NA	NA
	Mean (%)	***83.81***	***88.38***	***84.36***	***86.35***	***80***	***92.97***	***78.53***	***83.47***	***72.28***	***85***	***94.84***	***93.38***	***88.73***	***91.76***	***80***

Table S5B. Included trials by the combination of databases pairs and excluding handsearch (considering Medline, Embase and CENTRAL as the main database)

Review	***N of studies (n of trials)***	Medline + Embase	Medline + CINAHL	Medline + WoS	Medline + PsycINFO	Medline + LILACS	Medline + CENTRAL	Embase + CINAHL	Embase + WoS	Embase + PsycINFO	Embase + LILACS	Embase + CENTRAL	CENTRAL + CINAHL	CENTRAL + WoS	CENTRAL + PsycINFO	CENTRAL + LILACS
		n of trials identified (%)														
CR1	135 (80)	73 (91.25)	70 (87.50)	70 (87.50)	NA	NA	79 (98.75)	56 (70)	66 (82.5)	NA	NA	79 (98.75)	78 (97.5)	78 (97.5)	NA	NA
CR2	97 (58)	56 (96.55)	55 (94.82)	NA	54 (93.10)	NA	58 (100)	46 (79.31)	NA	45 (77.58)	NA	54 (93.10)	53 (91.37)	NA	53 (91.37)	NA
CR3	23 (19)	19 (100)	17 (84.47)	16 (84.21)	16 (84.21)	16 (84.21)	16 (84.21)	19 (100)	19 (100)	18 (94.73)	17 (89.47)	19 (100)	17 (89.47)	16 (84.21)	16 (84.21)	16 (84.21)
CR4	3 (2)	2 (100)	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
CR5	79 (35)	33 (94.28)	31 (88.57)	NA	31 (88.57)	NA	35 (100)	27 (77.14)	NA	25 (71.42)	NA	35 (100)	35 (100)	NA	35 (100)	NA
CR6	40 (21)	19 (90.47)	20 (95.23)	NA	18 (85.71)	NA	19 (90.47)	18 (85.71)	NA	14 (66.66)	NA	20 (95.23)	20 (95.23)	NA	19 (90.47)	NA
CR7	47 (24)	21 (87.5)	21 (87.5)	NA	21 (87.5)	NA	24 (100)	17 (70.83)	NA	14 (58.33)	NA	24 (100)	24 (100)	NA	24 (100)	NA
CR8	123 (54)	49 (90.74)	NA	49 (90.74)	NA	NA	49 (90.74)	NA	42 (77.77)	NA	NA	49 (90.74)	NA	51 (94.44)	NA	NA
	Mean (%)	***93.84***	***89.68***	***87.48***	***87.81***	***84.21***	***94.88***	***80.49***	***86.75***	***73.74***	***89.47***	***96.83***	***95.59***	***92.05***	***93.21***	***84.21***

Table S6A. Included trials by the combination of databases excluding handsearch (considering Medline as the main database)

Review	***N studies (n trials)***	M + E + CE	M + E + CI	M + E + P	M + E + W	M + E + L	M + CE + CI	M + CE + P	M + CE + W	M + CE + L	M + CI + P	M + CI + W	M + CI + L	M + P + W	M + P + L	M + W + L
		n of trials identified (%)														
CR1	135 (80)	80 (100)	73 (91.25)	NA	75 (93.75)	NA	80 (100)	NA	79 (98.75)	NA	NA	72 (90)	NA	NA	NA	NA
CR2	97 (58)	58 (100)	57 (98.27)	56 (96.55)	NA	NA	58 (100)	58 (100)	NA	NA	55 (94.82)	NA	NA	NA	NA	NA
CR3	23 (19)	19 (100)	19 (100)	19 (100)	19 (100)	19 (100)	17 (89.47)	16 (84.21)	16 (84.21)	16 (84.21)	17 (89.47)	17 (89.47)	17 (89.47)	16 (84.21)	16 (84.21)	16 (84.21)
CR5	79 (35)	35 (100)	33 (94.28)	33 (94.28)	NA	NA	35 (100)	35 (100)	NA	NA	31 (88.57)	NA	NA	NA	NA	NA
CR6	40 (21)	20 (95.23)	21 (100)	19 (90.47)	NA	NA	20 (95.23)	19 (90.47)	NA	NA	20 (95.23)	NA	NA	NA	NA	NA
CR7	47 (24)	24 (100)	21 (87.5)	21 (87.5)	NA	NA	24 (100)	24 (100)	NA	NA	21 (87.5)	NA	NA	NA	NA	NA
CR8	123 (54)	51 (94.44)	NA	NA	52 (96.29)	NA	NA	NA	53 (98.14)	NA	NA	NA	NA	NA	NA	NA
	Mean (%)	***98.52***	***95.21***	***93.76***	***96.68***	***100***	***97.45***	***94.93***	***93.7***	***84.21***	***91.11***	***89.73***	***89.47***	***84.21***	***84.21***	***84.21***

M, MEDLINE; E, Embase; CE, CENTRAL; CI: CINAHL; W, Web of Science; P, PsycINFO ; L, LILACS. NA: not applicable.*

Table S6B. Included trials by the combination of databases excluding handsearch (considering Embase as the main database)

Review	***N studies (n trials)***	E + CE + CI	E + CE + P	E + CE + W	E + CE + L	E + CI + P	E + CI + W	E + CI + L	E + P + W	E + P + L	E + W + L
		n of trials identified (%)									
CR1	135 (80)	79 (98.75)	NA	79 (98.75)	NA	NA	70 (87.5)	NA	NA	NA	NA
CR2	97 (58)	54 (93.10)	54 (93.10)	NA	NA	47 (81.03)	NA	NA	NA	NA	NA
CR3	23 (19)	19 (100)	19 (100)	19 (100)	19 (100)	19 (100)	19 (100)	19 (100)	19 (100)	18 (94.73)	19 (100)
CR5	79 (35)	35 (100)	35 (100)	NA	NA	27 (77.14)	NA	NA	NA	NA	NA
CR6	40 (21)	20 (95.23)	20 (95.23)	NA	NA	20 (95.23)	NA	NA	NA	NA	NA
CR7	47 (24)	24 (100)	24 (100)	NA	NA	24 (100)	NA	NA	NA	NA	NA
CR8	123 (54)	NA	NA	52 (96.29)	NA	NA	NA	NA	NA	NA	NA
	Mean (%)	***97.84***	***97.66***	***98.34***	***100***	***90.68***	***93.75***	***100***	***100***	***94.73***	***100***

M, MEDLINE; E, Embase; CE, CENTRAL; CI: CINAHL; W, Web of Science; P, PsycINFO ; L, LILACS. NA: not applicable.*

Table S6C. Included trials by the combination of databases excluding handsearch (considering CENTRAL as the main database)

Review	***N studies (n trials)***	CE + CI + P	CE +CI + W	CE + CI + L	CE + P + W	CE + P + L	CE + W + L
		n of trials identified (%)					
CR1	135 (80)	NA	79 (98.75)	NA	NA	NA	NA
CR2	97 (58)	54 (93.10)	NA	NA	NA	NA	NA
CR3	23 (19)	17 (89.47)	17 (89.47)	17 (89.47)	16 (84.21)	16 (84.21)	16 (84.21)
CR5	79 (35)	35 (100)	NA	NA	NA	NA	NA
CR6	40 (21)	20 (95.23)	NA	NA	NA	NA	NA
CR7	47 (24)	24 (100)	NA	NA	NA	NA	NA
CR8	123 (54)	NA	NA	NA	NA	NA	NA
	Mean (%)	***95.56***	***94.11***	***89.47***	***84.21***	***84.21***	***84.21***

M, MEDLINE; E, Embase; CE, CENTRAL; CI: CINAHL; W, Web of Science; P, PsycINFO ; L, LILACS. NA: not applicable.*

*Not in the tables:

- CINAHL, PsycINFO, WoS
- CINAHL, PsycINFO, LILACS
- CINAHL, WoS, LILACS
- PsycINFO, WoS, LILACS
- CR4, as the number of databases is low, the combination of three is impossible.

There are 35 possible combinations of three databases chosen from Medline, Embase, CENTRAL, CINAHL, PsycINFO, WoS, and LILACS.
There are **21 combinations** of two databases chosen from Medline, Embase, CENTRAL, CINAHL, PsycINFO, WoS, and LILAC

Figure S1. Forest Plot CR1 – Mortality analysis by different databases combinations

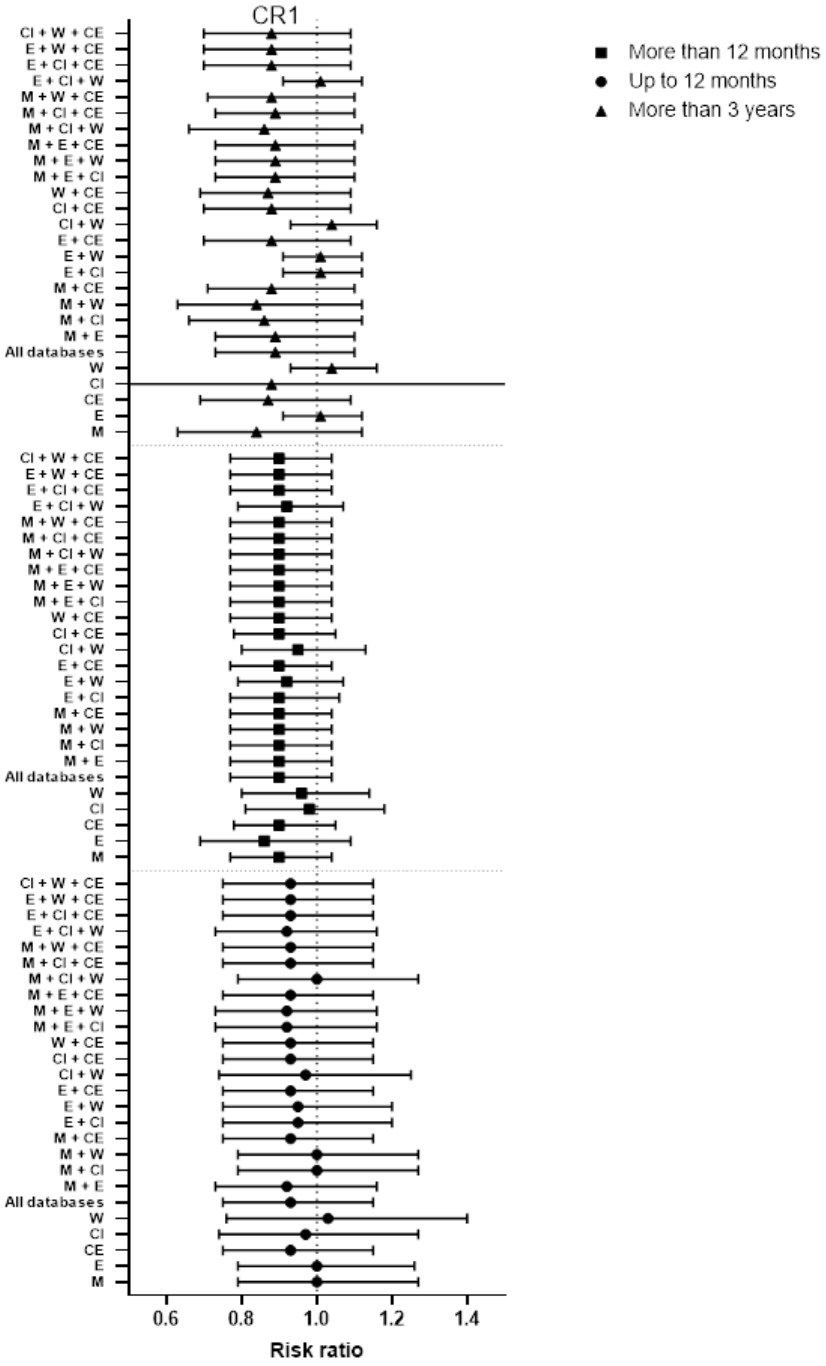


Figure S2. Forest Plot CR2 – Mortality analysis by different databases combinations

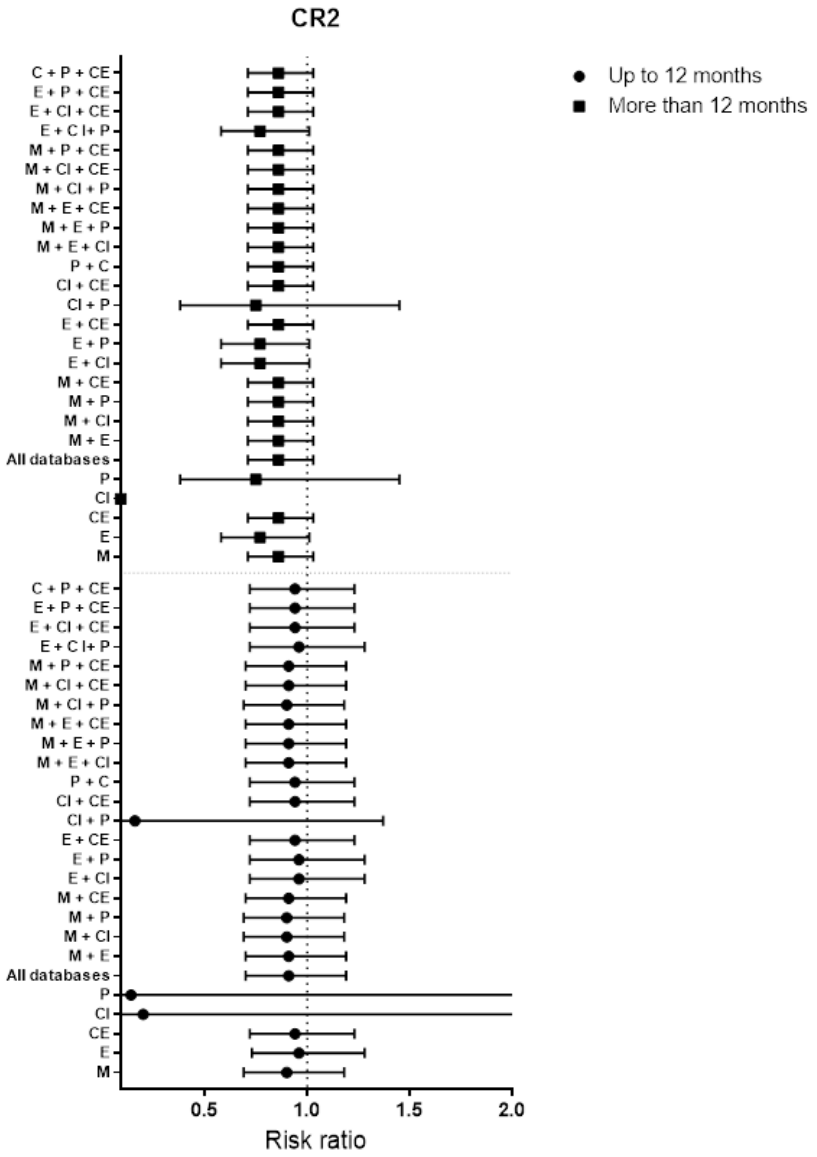
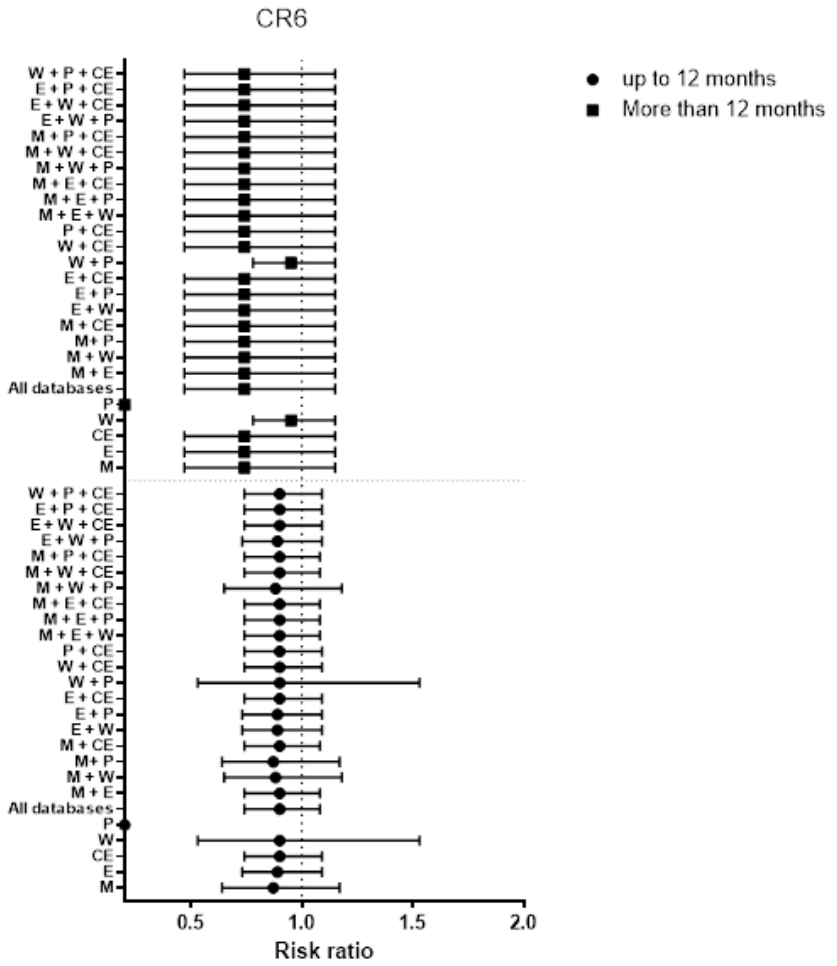


Figure S3. Forest Plot CR6 – Mortality analysis by different databases combinations



APPENDIX IV: Chapter 4

Methodological Insights: Comparing Single Review and Peer-Review Approaches in Study Selection for Evidence Synthesis

I. REVIEW QUESTIONS

The review previously developed by the META Group answered the following questions:

1. What is the effectiveness of isolation on reducing the transmission of respiratory infectious diseases (RIDs) (i.e., COVID-19, H1N1, severe acute respiratory syndrome (SARS), middle eastern respiratory syndrome (MERS), respiratory syncytial virus (RSV) and influenza?

2. What are the unintended health and social consequences/outcomes (e.g., mental health, financial circumstances) of isolation and quarantine (used for cases of COVID-19, H1N1, severe acute respiratory syndrome (SARS), middle eastern respiratory syndrome (MERS), respiratory syncytial virus (RSV) and influenza?

3. What is the effectiveness of quarantine on reducing the transmission of respiratory infectious diseases (RID), including COVID-19, H1N1, severe acute respiratory syndrome (SARS), middle eastern respiratory syndrome (MERS), respiratory syncytial virus (RSV) and influenza?

II. KEYWORDS - RAYYAN

Inclusion	Exclusion
Cohort, Case-control, Observational, Experimental, Quasi-experimental, Longitudinal, Modeling, Modelling, Mathematical, randomised control trial, RCT, randomized control trial	Review, this review, reviews, case report, case serie, opinion, protocol, editor
COVID-19, H1N1, SARS, MERS, influenza, isolation, quarantine, stay at home, stay home, Severe acute respiratory syndrome, Middle East Respiratory Syndrome	Lockdown, mass isolation, mass quarantine, feeling of isolation, travel, hospital